Bernoulli **11**(2), 2005, 359–379

On covariance estimation of non-synchronously observed diffusion processes

TAKAKI HAYASHI[1] and NAKAHIRO YOSHIDA[2]

[1]*Department of Statistics, Columbia University, 1255 Amsterdam Avenue, New York NY 10027, USA. E-mail: hayashi@stat.columbia.edu*
[2]*Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan. E-mail: nakahiro@ms.u-tokyo.ac.jp*

We consider the problem of estimating the covariance of two diffusion processes when they are observed only at discrete times in a non-synchronous manner. The modern, popular approach in the literature, the realized covariance estimator, which is based on (regularly spaced) synchronous data, is problematic because the choice of regular interval size and data interpolation scheme may lead to unreliable estimation. We propose a new estimator which is free of any 'synchronization' processing of the original data, hence free of bias or other problems caused by it.

Keywords: diffusions; discrete-time observations; high-frequency data; mathematical finance; non-synchronous trading; quadratic variation; realized volatility

1. Introduction

The modelling and analysis of high-frequency financial data are growing in importance, both in theory and in practice (see, for instance, Goodhart and O'Hara 1997 and Dacorogna *et al*. 2001 for an overiew). In this paper, the problem of estimating the covariance of two diffusion processes is studied when they are observed only at discrete times in a non-synchronous manner.

In the literature Andersen *et al*. (2001) propose the use of prelimits of quadratic variations as 'estimators' of variances and covariances of multivariate security price processes. Although the theory underlying their methodology is not new, its practical implications may be significant today. By the defining property of quadratic variations, we are sure that the greater the sampling frequency the more accurately (realizations of) 'true' variances and covariances are obtained; these were long considered and treated as 'latent' variables when intraday data were hard to come by.

Suppose we have discrete observations of two security 'prices' – or 'logarithmic prices', depending on the context – $(P^1_{t_i}, P^2_{t_i})_{i=0,1,\ldots,m}$ of size $m+1$, which are taken from continuous-time Itô semimartingales, $0 \leqslant t \leqslant T$. We are interested in the covariation, $V := \langle P^1, P^2 \rangle_T$, of two processes. Based on the sample, consider a statistic

$$V_{\pi(m)} := \sum_{i=1}^{m} (P^1_{t_i} - P^1_{t_{i-1}})(P_{t_i}{}^2 - P^2_{t_{i-1}}), \tag{1.1}$$

which is usually called the *realized covariance* estimator. We know that as $\pi(m) := \max_{1 \leqslant i \leqslant m} |t_i - t_{i-1}| \to 0$, $V_{\pi(m)} \to V$ in probability. Although it is not essential to this scheme, it is often the case that equal spacing is chosen, that is, $t_i - t_{i-1} = T/m$ $(=: h)$, for each i, for practical reasons. For ease of treatment we will follow that convention in this paper; however, the basic argument remains intact even if we replaced equal spacing by arbitrary, non-random spacing.

There are two crucial points pertaining to practical implementation of (1.1). First, actual transaction data are recorded at random times. Thus, two transaction prices are usually not observed (or recorded) at the same time. Secondly, due to such randomness of spacing, a significant portion of the original data sets should be missing at prespecified grid points. Consequently, in order to make the scheme work, we must choose the common interval length h first, and impute or interpolate the missing observations in some way. The cleaned data sets are then utilized for the computation of (1.1) as if they were regularly and concurrently observed, even if the two original processes may have very different observation frequencies. (We will refer to this preprocessing of data sets as 'synchronization'.) As is clear from the construction, (1.1) must depend heavily on the choice of h. On top of that, the methods of imputation may be another potential source of bias. Indeed, some authors (for instance Barucci and Renò 2002) point out a potential bias caused by the linear interpolation adopted by the aforementioned paper by Andersen *et al.* (2001) on estimating volatilities. Not to mention that the choice of an imputation method must be crucial for covariance estimation, too.

The impact of non-synchronous data on covariance measurement (daily returns or longer) had been long studied in the finance and econometrics literature even before the age of high-frequency data; see, for instance, Scholes and Williams (1977) and Lo and MacKinlay (1990). In particular, empirical evidence for the dependence of correlation estimators of high-frequency stock returns on sampling frequency is reported by Epps (1979). Using intraday returns data (up to 10 minutes) on four US stocks, he finds that correlations have a tendency to diminish as sampling frequency increases, known as the *Epps effect*. This phenomenon has been observed across different markets. See, for instance, Renò (2003) and the references therein. Regarding covariance/correlation estimation problems by use of non-synchronous high-frequency data, the reader is suggested to refer to Lundin *et al.* (1999), Muthuswamy *et al.* (2001), Brandt and Diebold (2003), and Tsay and Yeh (2003), for instance. They independently proposed an estimation procedure to cope with non-synchronicity in a various model assumption. Most of the existing approaches (not to mention the realized covariance and the sample covariance), however, need intrinsically to rely on synchronization of the original data – either explicitly or implicitly – due to their construction. See Martens (2004) for a surveillance of some of the approaches mentioned here and others. A few exceptions found in de Jong and Nijman (1997) and Malliavin and Mancino (2002). We will comment on them in the remark in Section 2.2. Our model assumption (that is, continuous-time diffusions) has commonality with Malliavin and Mancino (2002); however, our estimation procedure is similar to none of the existing ones.

Our first objective is to investigate the bias of the realized covariance estimator (1.1) when processes are not synchronously observed. We will show in a simple model that the realized covariance estimator has potentially a serious bias, especially when the regular interval size h is small relative to the frequency of actual trades. The result may provide a partial account for the Epps-type effects. Next, we will propose a new estimator which is based only on original data, that is to say, which requires no prior synchronization of transaction-based data. Since it is independent of the choice of h and of imputation of missing values, it is free of extraneous biases or any other problems caused by them. Under general conditions we will show consistency of the estimator as the observation time intensity (which represents the liquidity of the market) increases to infinity. As a by-product of the proof, unbiasedness will also be obtained when the diffusion processes have no drift.

Estimation problems of the diffusion parameter for diffusion processes based on discrete-time samples have been well studied in statistics; see Prakasa Rao (1983; 1988), Yoshida (1992), Genon-Catalot and Jacod (1993; 1994) and Kessler (1997), for instance. In particular, Genon-Catalot and Jacod (1994) have discussed (synchronous) random sampling. To the best of our knowledge, however, non-synchronous cases have seldom been investigated.

2. Non-synchronous observations and downward biases

Throughout the paper, we will suppose that processes evolve continuously, in continuous time, especially as diffusion-type processes. Detailed specifications will be given below.[2]

In this section, we aim to show why the *de facto* standard approach – the realized covariance estimator – is inadequate. We consider a market with insufficient liquidity – but not too excessive illiquidity – so that price observations are less and less frequent as the sampling window shrinks. For simplicity of analysis, when imputing missing data, we adopt the *previous-tick* (or piecewise constant) interpolation scheme (Dacorogna *et al.* 2001), that is, we take $P_t^k = P_{t_i}^k$, $k = 1, 2$, where t_i is the largest observation time before and including time t. This should be reasonable, for it will produce no obvious extraneous bias when estimating quadratic variations of *univariate* processes via (1.1) with $P^1 = P^2$.

2.1. Continuous martingales with random-time sampling

To begin with, suppose P^1 and P^2 are L^2 continuous martingales. Let $(T^{1,i})_{i=0,1,2,\ldots}$ and $(T^{2,j})_{j=0,1,2,\ldots}$ be random times independent of P^1 and P^2, representing the ith and jth observation times of P^1 and P^2, respectively, with $T^{1,0} := 0$ and $T^{2,0} := 0$. $T^{1,i}$ and $T^{2,j}$ may be mutually dependent. Similarly, dependence may be allowed among $(T^{1,i})$ or $(T^{2,j})$. Let $\Pi := ((T^{1,i})_{i=0,1,2,\ldots}, (T^{2,j})_{j=0,1,2,\ldots})$. Also, let N^k be the counting process associated with $(T^{k,i})$, starting at $N_0^k := 0$, $k = 1, 2$. Let $T \in]0, \infty[$ be a terminal time for observing P^1 and P^2 (for instance $T = 1$ day).

[1] An alternative approach, which is often adopted in this growing area, is to model prices as pure jump processes; see, for instance, Rydberg and Shephard (1999).

Actual observation processes, denoted by $\overline{P}^1$ and $\overline{P}^2$, are defined as piecewise constant processes by

$$\overline{P}^k_t := P^k_{T^{k,i}}, \qquad t \in [T^{k,i}, T^{k,i+1}[, \ k = 1, 2.$$

Suppose the observations are to be 'synchronized' on grid points $(ih)_{i=0,\dots,m}$, where h is the given length of each interval and $m = T/h$ an integer, for simplicity. That is, 'synchronized observations' are defined as discrete-time processes by

$$\overline{P}^{k,h}_i := \overline{P}^k_{ih}, \qquad i = 0, 1, \dots, m, \ k = 1, 2,$$

which may be treated in continuous time as regularly spaced, piecewise constant processes.

In this set-up, the quantity of interest is the covariation of the two processes, $\langle P^1, P^2 \rangle_T$ (or its expectation $\mathrm{E}[\langle P^1, P^2 \rangle_T]$, depending on the situation). To estimate this, we consider the realized covariance estimator (1.1) with previous-tick interpolation, that is,

$$V_h := \sum_{i=1}^m \left(\overline{P}^{1,h}_i - \overline{P}^{1,h}_{i-1}\right)\left(\overline{P}^{2,h}_i - \overline{P}^{2,h}_{i-1}\right) = \sum_{i=1}^m \left(\overline{P}^1_{ih} - \overline{P}^1_{(i-1)h}\right)\left(\overline{P}^2_{ih} - \overline{P}^2_{(i-1)h}\right)$$

$$= \sum_{i=1}^m \left(P^1_{\tau^1(ih)} - P^1_{\tau^1((i-1)h)}\right)\left(P^2_{\tau^2(ih)} - P^2_{\tau^2((i-1)h)}\right),$$

where $\tau^k(t) := \max_{0 \leqslant j < \infty}\{T^{k,j} \colon T^{k,j} \leqslant t\}$, the largest observation time of the kth security, $k = 1, 2$, up to and including time t. Note that if there is no jump of N^k by time t, $\tau^k(t) = 0$, $k = 1, 2$.

Observe that we always have $\tau^k(t) \leqslant t$ for any t, and $\tau^k(s) \leqslant \tau^k(t)$ for any $s < t$. Moreover,

$$\begin{aligned} N^k \text{ has no jumps in }](i-1)h, ih] &\Leftrightarrow \tau^k(ih) \leqslant (i-1)h \\ &\Leftrightarrow \tau^k(ih) = \tau^k((i-1)h) \\ &\Rightarrow P^k_{\tau^k(ih)} - P^k_{\tau^k((i-1)h)} = 0. \end{aligned} \tag{2.1}$$

Let $G^k_i := \{N^k \text{ jumps at least once in }](i-1)h, ih]\}$, $k = 1, 2$.

Throughout the paper, for a continuous-time process X, the increment of X over an interval $I :=]a, b]$, $0 \leqslant a < b < \infty$, may be written as $\Delta X(I) := X(b) - X(a)$.

Proposition 2.1. *V_h has the expected value*

$$\mathrm{E}[V_h] = \mathrm{E}\left[\sum_{i=1}^m \{\langle P^1, P^2 \rangle_{\tau^1(ih)\wedge\tau^2(ih)} - \langle P^1, P^2 \rangle_{\tau^1((i-1)h)\vee\tau^2((i-1)h)}\} 1_{G^1_i \cap G^2_i}\right].$$

If $\langle P^1, P^2 \rangle$ is increasing (decreasing) almost surely, then

$$E[V_h] \leqslant (\geqslant) E[\langle P^1, P^2 \rangle_{\tau^1(T)\wedge\tau^2(T)}] \qquad \textit{for every } h > 0.$$

Furthermore, if $\langle P^1, P^2 \rangle$ is strictly monotone with positive probability and if

$P[\tau^1(T) \vee \tau^2(T) > 0] > 0$, *then strict inequality holds if and only if* $P[\tau^1(ih) \neq \tau^2(ih)] > 0$ *for some* $i \in \{1, \ldots, m-1\}$.

Proof. Let $I^{k,i} :=]\tau^k((i-1)h), \tau^k(ih)]$, $i = 0, 1, 2, \ldots$, $k = 1, 2$. Observe first that $\tau^1(ih) \wedge \tau^2(ih) \geqslant \tau^1((i-1)h) \vee \tau^2((i-1)h)$ on $G_i^1 \cap G_i^2$. Due to (2.1),

$$E[V_h] = \mathrm{E}\left[\sum_{i=1}^m \Delta P^1(I^{1,i})\Delta P^2(I^{2,i})\right] = \mathrm{E}\left[\sum_{i=1}^m \mathrm{E}\left\{\Delta P^1(I^{1,i})\Delta P^2(I^{2,i})1_{G_i^1\cap G_i^2}\middle|\Pi\right\}\right].$$

$$= \mathrm{E}\left[\sum_{i=1}^m \mathrm{E}\left\{\left(\Delta P^1(I_1^{1,i}) + \Delta P^1(I_2^i) + \Delta P^1(I_3^{1,i})\right)\left(\Delta P^2(I_1^{2,i}) + \Delta P^2(I_2^i) + \Delta P^2(I_3^{2,i})\right)1_{G_i^1\cap G_i^2}\middle|\Pi\right\}\right]$$

$$= \mathrm{E}\left[\sum_{i=1}^m \mathrm{E}\left\{\Delta P^1(I_2^i)\Delta P^2(I_2^i)\middle|\Pi\right\}1_{G_i^1\cap G_i^2}\right] = \mathrm{E}\left[\sum_{i=1}^m \Delta\langle P^1, P^2\rangle(I_2^i)1_{G_i^1\cap G_i^2}\right],$$

where, for $k = 1, 2$,

$$I_1^{k,i} :=]\tau^k((i-1)h), \tau^1((i-1)h) \vee \tau^2((i-1)h)],$$

$$I_2^i :=]\tau^1((i-1)h) \vee \tau^2((i-1)h), \tau^1(ih) \wedge \tau^2(ih)],$$

$$I_3^{k,i} :=]\tau^1(ih) \wedge \tau^2(ih), \tau^k(ih)].$$

We have used the orthogonality of increments of P^k in the fourth equality. Therefore, the first assertion is obtained.

In particular, if $\langle P^1, P^2\rangle$ is increasing, then

$$\begin{aligned}\sum_{i=1}^m \Delta\langle P^1, P^2\rangle(I_2^i)1_{G_i^1\cap G_i^2} &\leqslant \sum_{i=1}^m \left(\langle P^1, P^2\rangle_{\tau^1(ih)\wedge\tau^2(ih)} - \langle P^1, P^2\rangle_{\tau^1((i-1)h)\wedge\tau^2((i-1)h)}\right)\\ &= \langle P^1, P^2\rangle_{\tau^1(mh)\wedge\tau^2(mh)} - \langle P^1, P^2\rangle_{\tau^1(0)\wedge\tau^2(0)}\\ &= \langle P^1, P^2\rangle_{\tau^1(T)\wedge\tau^2(T)}.\end{aligned}$$

The second assertion follows. The last assertion is immediate. □

Case of perfect synchronicity. If prices are observed synchronously (with probability one), regardless of the monotonicity of $\langle P^1, P^2\rangle$,

$$\mathrm{E}[V_h] = \mathrm{E}\left[\langle P^1, P^2\rangle_{\tau^1(T)}\right] \qquad \text{for every } h > 0.$$

The difference $\mathrm{E}\left[\langle P^1, P^2\rangle_{\tau^1(T)} - \langle P^1, P^2\rangle_T\right]$ is the bias of V_h, which is no more than the cost of 'right-censoring' by time T, that is due to the random sampling, the true values at T may not in general be observable.

2.2. Brownian motion with Poisson sampling

We now consider a special case, which would enable us to evaluate the downward (or upward) bias more explicitly.

Let $P^1 := W^1$ and $P^2 := W^2$, where W^1 and W^2 are standard but correlated Brownian motions with $\langle W^1, W^2\rangle_t = \rho t$ and $\rho \in\,]-1, 1[$ constant. Let N^1 and N^2 be two independent Poisson processes with intensity λ^1 (> 0) and λ^2 (> 0), respectively. These processes are also independent of W^1 and W^2. Let $T^{1,i}$ and $T^{2,i}$ be the respective arrival times of the ith Poisson jumps, with $T^{1,0} := 0$ and $T^{2,0} := 0$.

Note that, in light of Proposition 2.1, if $\rho > (<)\, 0$, then V_h is 'downward' ('upward') biased, that is, $\mathrm{E}[V_h] < (>)\, \rho\mathrm{E}[\tau^1(T) \wedge \tau^2(T)]$ for every $h > 0$. Moreover:

Proposition 2.2. *As* $h \downarrow 0$,

$$\mathrm{E}[V_h] = \rho \frac{\lambda^1\lambda^2}{\lambda^1+\lambda^2}\left(T - \frac{1-\mathrm{e}^{-(\lambda^1+\lambda^2)T}}{\lambda^1+\lambda^2}\right)h + O(h^2).$$

That is, approximately linearly, $\mathrm{E}[V_h]$ *vanishes as h decreases for any fixed intensities* λ^1 *and* λ^2.

Therefore, the downward bias (in magnitude) of the realized *correlation*, $(1/T)\,V_h$, is measured explicitly – here we assume the variances are known to be the unity – in the case when the regular interval size h is small relative to the average trading intervals, $1/\lambda^1$ and $1/\lambda^2$.

Proof. Let $\tau^k(h) := \max_{0 \leqslant j < \infty}\{T^{k,j}\colon T^{k,j} \leqslant h\}$, the largest arrival time up to time h (> 0) of N^k. Again, we set $G_i^k := \{N^k \text{ jumps at least once in }](i-1)h, ih]\}$, $k = 1, 2$. From Proposition 2.1,

$$\mathrm{E}[V_h] = \mathrm{E}\left[\sum_{i=1}^m \Delta\langle P^1, P^2\rangle(I_2^i) 1_{G_i^1\cap G_i^2}\right]$$

$$= \rho\sum_{i=1}^m \mathrm{E}\Big[\big\{\tau^1(ih)\wedge\tau^2(ih) - \tau^1((i-1)h)\vee\tau^2((i-1)h)\big\}1_{G_i^1\cap G_i^2}\Big].$$

Put

$$I(i) := \mathrm{E}\Big[\big\{\tau^1(ih)\wedge\tau^2(ih)\big\}1_{G_i^1\cap G_i^2}\Big], \qquad II(i) := \mathrm{E}\Big[\big\{\tau^1((i-1)h)\vee\tau^2((i-1)h)\big\}1_{G_i^1\cap G_i^2}\Big].$$

For notational convenience, put $X_{2,i} := \tau^1(ih)$, $Y_{2,i} := \tau^2(ih)$, $X_{1,i} := \tau^1((i-1)h)$, $Y_{1,i} := \tau^2((i-1)h)$ in what follows. We need the distributions of $X_{2,i}\wedge Y_{2,i}$ and $X_{1,i}\vee Y_{1,i}$ on $G_i^1\cap G_i^2$ to compute $I(i)$ and $II(i)$.

For this purpose we require the following fact which will be invoked repeatedly. Suppose that N is a Poisson process with intensity λ (> 0) and that $(T^j)_{j=0,1,2,\dots}$ are its jump arrival times with $T^0 := 0$. Let $\tau(h) := \max_{0 \leqslant j < \infty}\{T^j\colon T^j \leqslant h\}$, the largest arrival time up to a given reference time h (> 0). Then, we claim,

$$P[\tau(h) \leqslant s,\ N_h \geqslant 1] = \mathrm{e}^{-\lambda h}(\mathrm{e}^{\lambda s} - 1), \qquad 0 < s \leqslant h. \tag{2.2}$$

Indeed, when there are k $(\geqslant 1)$ jumps in $]0, h]$, the largest jump time has the distribution of the kth order statistic of k independent uniform random variables,

$$P[\tau(h) \leqslant s | N_h = k] = \left(\frac{s}{h}\right)^k, \qquad 0 < s \leqslant h.$$

Hence,

$$P[\tau(h) \leqslant s,\ N_h \geqslant 1] = \sum_{k=1}^{\infty}\left(\frac{s}{h}\right)^k \frac{\mathrm{e}^{-\lambda h}(\lambda h)^k}{k!} = \mathrm{e}^{-\lambda h}(\mathrm{e}^{\lambda s} - 1), \qquad 0 < s \leqslant h,$$

as claimed.

To calculate $I(i)$, put $\tilde{X}_2 := X_{2,i} - (i-1)h$ and $\tilde{Y}_2 := Y_{2,i} - (i-1)h$ on $G_i^1 \cap G_i^2$. Due to the weak Markov property of N^1 and N^2 at $t = (i-1)h$, if we put $\tilde{N}_u^k := N_{u+(i-1)h}^k - N_{(i-1)h}^k$, $0 \leqslant u \leqslant h$, $k = 1, 2$, which are new Poisson processes starting from 0 at $t = (i-1)h$, then

$$\begin{aligned} \mathrm{E}[X_{2,i} \wedge Y_{2,i} 1_{G_i^1 \cap G_i^2} | N_u^1, N_u^2, 0 \leqslant u \leqslant (i-1)h] &= \mathrm{E}[X_{2,i} \wedge Y_{2,i} 1_{G_i^1 \cap G_i^2} | N_{(i-1)h}^1, N_{(i-1)h}^2] \\ &= \mathrm{E}[\{\tilde{X}_2 \wedge \tilde{Y}_2 + (i-1)h\} 1_{\tilde{G}^1 \cap \tilde{G}^2}], \end{aligned}$$

where $\tilde{G}^k := \{\tilde{N}^k$ jumps at least once in $]0, h]\}$, $k = 1, 2$. Therefore,

$$I(i) = \mathrm{E}\Big[X_{2,i} \wedge Y_{2,i} 1_{G_i^1 \cap G_i^2}\Big] = \mathrm{E}\big[\tilde{X}_2 \wedge \tilde{Y}_2 1_{\tilde{G}^1 \cap \tilde{G}^2}\big] + (i-1)h \cdot P\big[\tilde{G}^1 \cap \tilde{G}^2\big].$$

By (2.2) and the independence between $(T^{1,i})$ and $(T^{2,j})$, the first term can be evaluated as

$$\begin{aligned} \mathrm{E}\big[\tilde{X}_2 \wedge \tilde{Y}_2 1_{\tilde{G}^1 \cap \tilde{G}^2}\big] &= \int_{\tilde{x}_2=0}^{h}\int_{\tilde{y}_2=0}^{h} \tilde{x}_2 \wedge \tilde{y}_2 \mathrm{P}\big[\tau^1(h) \in \mathrm{d}\tilde{x}_2,\ \tilde{N}_{\mathrm{h}}^1 \geqslant 1\big] P\big[\tau^2(h) \in \mathrm{d}\tilde{y}_2,\ \tilde{N}_h^2 \geqslant 1\big] \\ &= \int_{\tilde{x}_2=0}^{h}\int_{\tilde{y}_2=0}^{\tilde{x}_2} \tilde{y}_2 \Big(\mathrm{e}^{-\lambda^1 h} \cdot \lambda^1 \mathrm{e}^{\lambda^1 \tilde{x}_2}\Big)\Big(\mathrm{e}^{-\lambda^2 h} \cdot \lambda^2 \mathrm{e}^{\lambda^2 \tilde{y}_2}\Big) \mathrm{d}\tilde{y}_2 \mathrm{d}\tilde{x}_2 \\ &\quad + \int_{\tilde{y}_2=0}^{h}\int_{\tilde{x}_2=0}^{\tilde{y}_2} \tilde{x}_2 \Big(\mathrm{e}^{-\lambda^1 h} \cdot \lambda^1 \mathrm{e}^{\lambda^1 \tilde{x}_2}\Big)\Big(\mathrm{e}^{-\lambda^2 h} \cdot \lambda^2 \mathrm{e}^{\lambda^2 \tilde{y}_2}\Big) \mathrm{d}\tilde{x}_2 \mathrm{d}\tilde{y}_2. \end{aligned}$$

The second term is given by

$$(i-1)h \cdot P\big[\tilde{G}^1\big] P\big[\tilde{G}^2\big] = (i-1)h \cdot \Big(1 - \mathrm{e}^{-\lambda^1 h}\Big)\Big(1 - \mathrm{e}^{-\lambda^2 h}\Big).$$

Turning to $II(i)$, let $H_i := \{$either N^1 or N^2 jumps at least once in $]0, (i-1)h]\}$. Then, since $X_{1,i} \vee Y_{1,i} = 0$ on $H_i^c = \{$neither N^1 nor N^2 jump in $]0, (i-1)h]\}$,

$$\begin{aligned}
II(i) &= \mathrm{E}\Big[X_{1,i} \vee Y_{1,i} 1_{H_i} 1_{G_i^1 \cap G_i^2}\Big] \\
&= \mathrm{E}\Big[\mathrm{E}\Big\{X_{1,i} \vee Y_{1,i} 1_{H_i} 1_{G_i^1 \cap G_i^2} \Big| N_u^1,\, N_u^2,\, 0 \leqslant u \leqslant (i-1)h\Big\}\Big]. \\
&= \mathrm{E}\Big[X_{1,i} \vee Y_{1,i} 1_{H_i} \mathrm{E}\Big\{1_{G_i^1 \cap G_i^2} \Big| N_{(i-1)h}^1,\, N_{(i-1)h}^2\Big\}\Big] \\
&= \mathrm{E}[X_{1,i} \vee Y_{1,i} 1_{H_i}] P\big[\tilde{G}^1 \cap \tilde{G}^2\big],
\end{aligned}$$

again due to the weak Markov property, as well as the stationarity of increments of N^k, $k = 1, 2$.

Notice, on H_i, that $S := X_{1,i} \vee Y_{1,i}$ is the last jump time, up to time $(i-1)h$, of the aggregate Poisson process $N^* := N^1 + N^2$ with intensity $\lambda^* := \lambda^1 + \lambda^2$. Using (2.2), we have

$$P\Big[S \leqslant s,\, N^*_{(i-1)h} \geqslant 1\Big] = \mathrm{e}^{-\lambda^*(i-1)h}\Big(\mathrm{e}^{\lambda^* s} - 1\Big), \qquad 0 < s \leqslant (i-1)h,$$

so that

$$\mathrm{E}[X_{1,i} \vee Y_{1,i} 1_{H_i}] = \int_{s=0}^{(i-1)h} s e^{-\lambda^*(i-1)h} \cdot \lambda^* \mathrm{e}^{\lambda^* s} \mathrm{d}s = -\frac{1}{\lambda^*} + (i-1)h + \mathrm{e}^{-\lambda^*(i-1)h} \frac{1}{\lambda^*}.$$

Therefore,

$$II(i) = \Big(1 - \mathrm{e}^{-\lambda^1 h}\Big)\Big(1 - \mathrm{e}^{-\lambda^2 h}\Big)\Big\{-\frac{1}{\lambda^*} + (i-1)h + \mathrm{e}^{-\lambda^*(i-1)h}\frac{1}{\lambda^*}\Big\}.$$

Now, after tedious asymptotic computation, we have

$$\sum_{i=1}^{m} I(i) = \frac{1}{2}\lambda^1\lambda^2 T^2 h + \Big(-\frac{1}{6}\lambda^1\lambda^2 T - \frac{1}{4}\lambda^1\lambda^2\lambda^* T^2\Big) h^2 + O(h^3)$$

and

$$\sum_{i=1}^{m} II(i) = \Big\{-\frac{\lambda^1\lambda^2}{\lambda^*} T + \frac{\lambda^1\lambda^2}{2} T^2 + \frac{\lambda^1\lambda^2}{\lambda^{*2}}(1 - \mathrm{e}^{-\lambda^* T})\Big\} h - \frac{1}{4}\lambda^1\lambda^2\lambda^* T^2 h^2 + O(h^3).$$

Taking the difference, $\sum_{i=1}^m I(i) - \sum_{i=1}^m II(i)$, and multiplying by ρ, we obtain the result. □

Remark 2.1. As far as we see in the literature, most of the existing approaches for covariance estimation using non-synchronous data need to rely on synchronization of the original data in one way or another due to their construction. A few exceptions include de Jong and Nijman (1997) and Malliavin and Mancino (2002), both of which utilize the original and hence circumvent problems caused by synchronization. De Jong and Nijman (1997) have proposed a regression based approach, regressing observed cross products of returns on the number of common time units to these returns under a discrete-time model with stationary increments. Although it is not clear how to extend their approach to a continuous-time setting, it is worthy of mention. Malliavin and Mancino (2002) have proposed a *Fourier* transform based estimator for the variance-covariance matrix of multivariate diffusion processes, which can apply to data that are not synchronous in continuous time. Its statistical properties are yet to be explored. Renò (2003) utilizes it to investigate numerically biases of Epps type for a bivariate continuous-time version of the GARCH(1,1) process. Martens (2004) provides an overview of some of the existing approaches and compares numerically their performance assuming the (discrete-time) non-trading model of Lo and MacKinlay (1990).

The asymptotic analysis breaks down for liquid securities, that is, when h is small relative to $1/\lambda$, and $1/\lambda_2$. To deal with such situations, more complicated models may need to be adopted. Also, when the variances are unknown and to be estimated, incorporation of 'microstructure noise' with the same spirit as Bandi and Russell (2003) and Zhang *et al.* (2003) – they study 'optimal' use of high-frequency data (contaminated with such noise) for volatility estimation – may provide a reasonable explanation for the Epps-type effects.

As mentioned in the paragraph of 'Case of perfect synchronicity' at the end of Section 2.1, the bias due to missing observations at T is inevitable in so far as we adopt random sampling schemes. To simplify the discussion, we will therefore assume that processes are observable at T (that is, $T^{k,i}$ will be replaced by $T^{k,i} \wedge T$, $i \geqslant 1$, $k = 1, 2$, in terms of the notation difined above) throughout the rest of the paper. Note that when the observation frequencies explode this treatment will be asymptotically negligible.

3. Consistent estimation

3.1. The covariance estimator: main result

We will propose an estimator of the (cumulative) covariance of two diffusion processes when they are recorded at random times, thus not necessarily regularly spaced. Estimators such as (1.1) require that we transform such data sets into regularly spaced in advance; first the length of each interval, h, will be chosen, then an appropriate imputation method is conducted. As we have seen above, the choice of h and the method of imputation may cause serious biases. Our proposed estimator is free of such arbitrariness.

Throughout the rest of the paper, we suppose that P^l follows the one-dimensional Itô process

$$\mathrm{d}P^l_t = \mu^l_t \mathrm{d}t + \sigma^l_t \mathrm{d}W^l_t, \qquad P^l_0 = p^l,\ l = 1, 2,$$

with $\mathrm{d}\langle W^1, W^2\rangle_t = \rho_t \mathrm{d}t$, where $\rho_\cdot \in\,]-1, 1[$ is an unknown, deterministic function, $p^l > 0$ is a constant, $\mu^l_\cdot$ is a progressively measurable (possibly unknown) function, and $\sigma^l_\cdot > 0$ is a deterministic and bounded (possibly unknown) function.

Let $T \in]0, \infty[$ be an arbitrary terminal time for observing P^ls. Let $\Pi^1 := (I^i)_{i=1,2,\ldots}$ and $\Pi^2 := (J^i)_{i=1,2,\ldots}$ be random intervals reading from left to right, each of which partitions $]0, T]$. Let $T^{1,i} := \inf\{t \in I^{i+1}\}$ represent the ith observation time of P^1, and $T^{2,i} := \inf\{t \in J^{i+1}\}$ that of P^2. Let n be an index representing the size of Π^1 and Π^2.

The length of an interval I is denoted by $|I|$. We assume that the sampling intervals $\Pi := (\Pi^1, \Pi^2)$ satisfy the following:

Condition C.

(i) (I^i) *and* (J^i) *are independent of* P^1 *and* P^2.
(ii) As $n \to \infty$, $\mathrm{E}[\max_i|I^i| \vee \max_j|J^j|] = o(1)$.

Remark 3.1. Apparently, (ii) is equivalent to the condition $\max_i|I^i| \vee \max_j|J^j| \to 0$ in probability as $n \to \infty$. Since

$$\left(\max_i|I^i| \vee \max_j|J^j|\right)^2 \leqslant \sum_i |I^i|^2 + \sum_j |J^j|^2 \leqslant T\max_i|I^i| + T\max_j|J^j|,$$

(ii) is also equivalent to the condition
(iii) $\sum_i|I^i|^2 + \sum_j|J^j|^2 \to 0$ in probability as $n \to \infty$.
Moreover, for (ii) to hold it is sufficient that
(iv) $P[\max_i|I^i| \vee \max_j|J^j| > n^{-q}] = \mathrm{o}(1)$ for some $q > 0$.

Example 3.1 *Poisson* random sampling scheme. Let N^1 and N^2 be two independent Poisson processes with intensity $\lambda^1 = np^1$ and $\lambda^2 = np^2$ for $p^1 > 0$, $p^2 > 0$ and $n \in \mathbb{N}$, which are also independent of P^1 and P^2. If $\tilde{T}^{1,i}$ and $\tilde{T}^{2,i}$ are the respective arrival times of the ith Poisson jumps with $\tilde{T}^{1,0} := 0$ and $\tilde{T}^{2,0} := 0$, we construct $\Pi^1 := (I^i)_{i=1,2,\dots}$ and $\Pi^2 := (J^i)_{i=1,2,\dots}$, by setting $I^i :=]\tilde{T}^{1,i-1}, \tilde{T}^{1,i}] \cap]0, T]$ and $J^i :=]\tilde{T}^{2,i-1}, \tilde{T}^{2,i}] \cap]0, T]$. This Poisson random sampling scheme satisfies the Condition C (see Corollary 3.1).

Example 3.2. Synchronous sampling scheme. Any sampling scheme (deterministic or random) with $I^i = J^i$, for every i, is covered so far as Condition C is met.

We will now estimate the (deterministic) covariation of P^1 and P^2,

$$\langle P^1, P^2\rangle_T = \int_0^T \sigma_t^1\sigma_t^2\rho_t \mathrm{d}t =: \theta.$$

We define the following estimator for θ constructed only from the observations of P^1 and P^2.

Definition 3.1. *Cumulative covariance estimator.*

$$U_n := \sum_{i,j} \Delta P^1(I^i)\Delta P^2(J^j)1_{\{I^i\cap J^j\neq\varnothing\}}. \tag{3.1}$$

That is, the product of any pair of increments $\Delta P^1(I^i)$ and $\Delta P^2(J^j)$ will make a contribution to the summation only when the respective observation intervals I^i and J^j are overlapping.

Notice that the proposed estimator utilizes the information regarding not only the price changes *per se* but also the transaction times at which those changes took place, in the form of the indicator function in (3.1), in contrast to the realized covariance estimator (1.1).

Remark 3.2. As we have seen in the previous section, the 'downward' bias (in magnitude) of the realized covariance estimator derives from the fact that each increment in (1.1), $(P^1_{t_i} - P^1_{t_{i-1}})(P^2_{t_i} - P^2_{t_{i-1}})$, contributes to the sum when and only when both P^1 and P^2 'jump' together during the interval $]t_{i-1}, t_i]$ of length h, thus all the other occasions – when at most one of the two prices jumps – are ignored. Such occasions of zero increment will become dominant if h becomes *finer*. On the other hand, *coarser* h may not be able to capture rapid movements of processes – multiple jumps that may have occurred during h – so that the realized covariance estimator may fail to reflect such microscopic movements (which are crucial for variance–covariance estimation). In other words, large h leads to inefficient use of data. The proposed estimator (3.1) circumvents this dilemma by avoiding the introduction of the parameter h, which in its origination has nothing to do with the original record of processes. It allows for *all* the jumps (even for multiplicity) of P^1 and P^2 without reference to their *synchronicity*. Notice that each increment $\Delta P^1(I^i)$ contributes to the sum possibly *multiple* times so long as the corresponding interval I^i intersects J^j (which may prevent the downward bias).

Theorem 3.1. *Suppose Condition C holds.*

(a) *If* $\sup_{0\leqslant t\leqslant T}|\mu^k_t| \in L^4$, $k = 1, 2$, *then* $U_n \to \theta$ *in* L^2 *as* $n \to \infty$.
(b) *If* $\sup_{0\leqslant t\leqslant T}|\mu^k_t| < \infty$ *almost surely,* $k = 1, 2$, *then* U_n *is consistent for* θ, *that is,* $U_n \to \theta$ *in probability as* $n \to \infty$.

Remark 3.3. As is seen in the proof below, if $\mu^l_t \equiv 0$, $0 \leqslant t \leqslant T$, then U_n is unbiased for each $n \geqslant 1$.

Proof of Theorem 3.1. (i) First we assume $\mu^l_t \equiv 0$, $0 \leqslant t \leqslant T$ for the time being. We will show that $U_n \to \theta$ in L^2 as $n \to \infty$.

We need to introduce some auxiliary symbols. Set $K_{ij} := 1_{\{I^i \cap J^j \neq \varnothing\}}$. For each measurable set I on $[0, \infty[$, we define (signed) measures by

$$v(I) := v^0(I) := \int_I \sigma^1_t \sigma^2_t \rho_t \mathrm{d}t,$$

$$v^k(I) := \int_I \left(\sigma^k_t\right)^2 \mathrm{d}t, \qquad k = 1, 2.$$

We will repeatedly use identities such as, for $k = 0, 1, 2$,

$$\sum_i v^k(I^i)1_{\{I^i\neq\emptyset\}} = \sum_i v^k(I^i) = v^k(]0,\, T]),$$

$$\sum_{i,j} v^k(I^i \cap J^j)K_{ij} = v^k(]0,\, T]),$$

$$\sum_j v^k(I^i \cap J^j)K_{ij} = v^k(I^i),$$

which hold due to the fact that the I^is and J^js are the partitions of $]0,\, T]$. Moreover, for each measurable set I on $[0,\, \infty[$, define

$$\Delta P^k(I) := \int_0^T 1_I(t)\sigma_t^k \mathrm{d}W_t^k, \qquad k = 1,\, 2.$$

Observe that $U_n$is unbiased, because, for each n,

$$\mathrm{E}[U_n] = \mathrm{E}\left[\sum_{i,j} \mathrm{E}\{\Delta P^1(I^i)\Delta P^2(J^j)|\Pi\}K_{ij}\right] = \mathrm{E}\left[\sum_{i,j} v(I^i \cap J^j)K_{ij}\right] = \theta.$$

We claim that $\mathrm{E}[U_n^2] = \theta^2 + o(1)$. If so, then $\mathrm{var}[U_n] = o(1)$ so that $U_n \to \theta$ in L^2 as $n \to \infty$.

To this end, first note that

$$\mathrm{E}[U_n^2] = \mathrm{E}\left[\sum_{i,j,i',j'} \mathrm{E}\{\Delta P^1(I^i)\Delta P^2(J^j)\Delta P^1(I^{i'})\Delta P^2(J^{j'})|\Pi\}K_{ij}K_{i'j'}\right].$$

We decompose the inside summation into

$$\sum_{i,j,i',j'} = \sum_{\substack{i,j,i',j':\\ i'=i,j'=j}} + \sum_{\substack{i,j,i',j':\\ i'=i,j'\neq j}} + \sum_{\substack{i,j,i',j':\\ i'\neq i,j'=j}} + \sum_{\substack{i,j,i',j':\\ i'\neq i,j'\neq j}} =: D_1 + D_2 + D_3 + D_4,$$

the respective expectation of which will be computed in the following four cases.

Case 1: $i = i'$, $j = j'$. Let $L_1 := I^i \cap J^j$, $L_2 := I^i \backslash L_1$ and $L_3 := J^j \backslash L_1$. Note that

$$D_1 = \sum_{i,j} \mathrm{E}\{\Delta P^1(I^i)^2 \Delta P^2(J^j)^2|\Pi\}K_{ij}.$$

Using the independence of the increments and identities such as $v^k(L_2) = v^k(I^i) - v^k(I^i \cap J^j)$ and $v^k(L_3) = v^k(J^j) - v^k(I^i \cap J^j)$, we have

$$\begin{aligned}
\mathrm{E}\{\Delta P^1(I^i)^2\Delta P^2(J^j)^2|\Pi\} &= \mathrm{E}\{(\Delta P^1(L_2)+\Delta P^1(L_1))^2(\Delta P^2(L_3)+\Delta P^2(L_1))^2|\Pi\}\\
&= \mathrm{E}\{\Delta P^1(L_2)^2\Delta P^2(L_1)^2|\Pi\}+\mathrm{E}\{\Delta P^1(L_1)^2\Delta P^2(L_1)^2|\Pi\}\\
&\quad+\mathrm{E}\{\Delta P^1(L_1)^2\Delta P^2(L_3)^2|\Pi\}+\mathrm{E}\{\Delta P^1(L_2)^2\Delta P^2(L_3)^2|\Pi\}\\
&= v^1(L_2)v^2(L_1)+2v(L_1)^2+v^1(L_1)v^2(L_1)\\
&\quad+v^1(L_1)v^2(L_3)+v^1(L_2)v^2(L_3)\\
&= v^1(I^i)v^2(J^j)+2v(I^i\cap J^j)^2.
\end{aligned}$$

In the third identity, we have used the fact that, for any (deterministic) interval I, $\Delta P^1(I)$ and $\Delta P^2(I)$ are jointly normal with respective mean and variance, 0 and $v^k(I)$, $k=1, 2$, and with covariance $v(I)$, so that $\mathrm{E}[\Delta P^1(I)^2\Delta P^2(I)^2]=2v(I)^2+v^1(I)v^2(I)$. Therefore,

$$D_1=\sum_{i,j}v^1\big(I^i\big)v^2\big(J^j\big)K_{ij}+2\sum_{i,j}v\big(I^i\cap J^j\big)^2K_{ij}. \tag{3.2}$$

Consider the first term on the right-hand side of (3.2). Noting that the σ^k are bounded, we have

$$\begin{aligned}
\sum_{i,j}v^1\big(I^i\big)v^2\big(J^j\big)K_{ij} &= \sum_{i,j}^{k}\bigg(\int_{I^i}\big(\sigma_t^1\big)^2\mathrm{d}t\bigg)\bigg(\int_{J^j}\big(\sigma_t^2\big)^2\mathrm{d}t\bigg)K_{ij}\\
&\leqslant \sup_{0\leqslant t\leqslant T}\big(\sigma_t^1\big)^2\sup_{0\leqslant t\leqslant T}\big(\sigma_t^2\big)^2\sum_{i,j}|I^i||J^j|K_{ij}.
\end{aligned}$$

We claim that

$$\mathrm{E}\sum_{i,j}|I^i||J^j|K_{ij}=o(1).$$

To this end, decompose

$$\sum_{i,j}|I^i||J^j|K_{ij}=\sum_{i,j}|I^i||J^j|K_{ij}1_{\{|I^i|\geqslant|J^j|\}}+\sum_{i,j}|I^i||J^j|K_{ij}1_{\{|I^i|<|J^j|\}}.$$

Noting that $\sum_j|J^j|K_{ij}1_{\{|I^i|\geqslant|J^j|\}}\leqslant 3|I^i|$ (for fixed i), we have

$$\sum_{i,j}|I^i||J^j|K_{ij}1_{\{|I^i|\geqslant|J^j|\}}=\sum_i|I^i|\sum_j|J^j|K_{ij}1_{\{|I^i|\geqslant|J^j|\}}\leqslant 3\sum_i|I^i|^2,$$

hence,

$$\mathrm{E}\sum_{i,j}|I^i||J^j|K_{ij}1_{\{|I^i|\geqslant|J^j|\}}\leqslant 3\mathrm{E}\sum_i|I^i|^2.$$

By symmetry, we have

$$\mathrm{E}\sum_{i,j}|I^i||J^j|K_{ij} \leqslant 3\mathrm{E}\sum_i |I^i|^2 + 3\mathrm{E}\sum_j |J^j|^2.$$

Now, the last expression is $o(1)$ under the Condition C(ii) (see Remark 3.1).

Similarly, we can ascertain that, for any random partition $(\tilde{I}^i)$ of $]0, T]$ satisfying Condition C(ii),

$$\mathrm{E}\sum_i v\big(\tilde{I}^i\big)^2 = o(1). \tag{3.3}$$

The second term on the right of (3.2) is shown to be of $o_P(1)$ by choosing $(I^i \cap J^j)$ as such a partition. It follows that $\mathrm{E}[D_1] = o(1)$.

Case 2: $i = i'$, $j \neq j'$. Note that

$$D_2 = \sum_{i,j,j':j\neq j'} \mathrm{E}\{\Delta P^1(I^i)^2 \Delta P^2(J^j)\Delta P^2(J^{j'})|\Pi\}K_{ij}K_{ij'}.$$

Let $L_1 := I^i \cap J^j$, $L_2 := I^i \cap J^{j'}$, and $L_3 := I^i \backslash (L_1 \cup L_2)$. Then, using the independence of increments,

$$\begin{aligned}
\mathrm{E}\{\Delta P^1(I^i)^2\Delta P^2(J^j)\Delta P^2(J^{j'})|\Pi\} &= \mathrm{E}\{\Delta P^1(I^i)^2\Delta P^2(L_1)\Delta P^2(L_2)|\Pi\}\\
&= \mathrm{E}\{(\Delta P^1(L_1) + \Delta P^1(L_3) + \Delta P^1(L_2))^2\Delta P^2(L_1)\Delta P^2(L_2)|\Pi\}\\
&= 2\mathrm{E}\{\Delta P^1(L_1)\Delta P^2(L_1)|\Pi\}\mathrm{E}\{\Delta P^1(L_2)\Delta P^2(L_2)|\Pi\}\\
&= 2v(L_1)v(L_2) = 2v(I^i \cap J^j)v(I^i \cap J^{j'}).
\end{aligned}$$

Hence,

$$\begin{aligned}
D_2 &= 2\sum_{i,j,j':j'\neq j} v\big(I^i \cap J^j\big)v\big(I^i \cap J^{j'}\big)K_{ij}K_{ij'}\\
&= 2\sum_i\left\{\sum_j v\big(I^i \cap J^j\big)K_{ij}\left(\sum_{j'} v\big(I^i \cap J^{j'}\big)K_{ij'} - v\big(I^i \cap J^j\big)\right)\right\}\\
&= 2\sum_i v\big(I^i\big)^2 - 2\sum_i\sum_j v\big(I^i \cap J^j\big)^2K_{ij},
\end{aligned}$$

so that $\mathrm{E}[D_2] = o(1)$ by use of (3.3), and by the fact that $(I^i \cap J^j)$ partitions $]0, T]$.

Case 3: $i \neq i'$, $j = j'$. The same argument as in case 2 applies by symmetry to obtain $\mathrm{E}[D_3] = o(1)$.

Case 4: $i \neq i'$, $j \neq j'$. By analogy with case 2 we first let $L_1 := I^i \cap J^j$, $L_2 := I^{i'} \cap J^{j'}$. Observe, for i, j, i', j' such that $i' \neq i$, $j' \neq j$ and $K_{ij}K_{i'j'} = 1$, we must always have $K_{i'j}K_{ij'} = 0$. Hence, due to the identity

$$(1 - K_{i'j})(1 - K_{ij'}) + K_{i'j} + K_{ij'} \equiv 1,$$

we can decompose further the event $\{K_{ij}K_{i'j'} = 1\}$ into three subcases, $\{I^{i'} \cap J^j = \varnothing,$

$I^i \cap J^{j'} = \varnothing\}$, $\{I^{i'} \cap J^j \neq \varnothing, I^i \cap J^{j'} = \varnothing\}$, and $\{I^{i'} \cap J^j = \varnothing, I^i \cap J^{j'} \neq \varnothing\}$, each of which respectively corresponds to $\{(1 - K_{i'j})(1 - K_{ij'}) = 1\}$, $\{K_{i'j} = 1\}$, and $\{K_{ij'} = 1\}$.

Case 4a: $\{I^{i'} \cap J^j = \varnothing, I^i \cap J^{j'} = \varnothing\}$. We have

$$\begin{aligned} &\sum_{i,j,i',j':i\neq i',j\neq j'} \mathrm{E}\{\Delta P^1(I^i)\Delta P^2(J^j)\Delta P^1(I^{i'})\Delta P^2(J^{j'})|\Pi\} K_{ij}K_{i'j'}(1 - K_{i'j})(1 - K_{ij'}) \\ &\quad = \sum_{i,j,i',j':i\neq i',j\neq j'} \mathrm{E}\{\Delta P^1(L_1)\Delta P^2(L_1)\Delta P^1(L_2)\Delta P^2(L_2)|\Pi\} K_{ij}K_{i'j'}(1 - K_{i'j})(1 - K_{ij'}) \\ &\quad = \sum_{i,j,i',j':i\neq i',j\neq j'} v(L_1)v(L_2)K_{ij}K_{i'j'}(1 - K_{i'j})(1 - K_{ij'}). \end{aligned}$$

Case 4b: $\{I^{i'} \cap J^j \neq \varnothing, I^i \cap J^{j'} = \varnothing\}$. Letting $L_3 := I^{i'} \cap J^j$ and $L_4 := J^j \backslash (L_1 \cup L_3)$, $L_5 := I^{i'} \backslash (L_2 \cup L_3)$,

$$\begin{aligned} &\mathrm{E}\{\Delta P^1(L_1)(\Delta P^2(L_1) + \Delta P^2(L_4) + \Delta P^2(L_3))(\Delta P^1(L_3) + \Delta P^1(L_5) + \Delta P^1(L_2))\Delta P^2(L_2)|\Pi\} \\ &\quad = \mathrm{E}\{\Delta P^1(L_1)\Delta P^2(L_1)\Delta P^1(L_2)\Delta P^2(L_2)|\Pi\} = v(L_1)v(L_2). \end{aligned}$$

Therefore,

$$\begin{aligned} &\sum_{i,j,i',j':i\neq i',j\neq j'} \mathrm{E}\{\Delta P^1(I^i)\Delta P^2(J^j)\Delta P^1(I^{i'})\Delta P^2(J^{j'})|\Pi\} K_{ij}K_{i'j'}K_{i'j} \\ &\quad = \sum_{i,j,i',j':i\neq i',j\neq j'} v(L_1)v(L_2)K_{ij}K_{i'j'}K_{i'j}. \end{aligned}$$

Case 4c: $\{I^{i'} \cap J^j = \varnothing, I^i \cap J^{j'} \neq \varnothing\}$. By symmetry, we can obtain the equivalent result to case 4b, but with $K_{ij'}$ in place of $K_{i'j}$.

Putting the three subcases together, we have

$$\begin{aligned} D_4 &= \sum_{i,j,i',j':i\neq i',j\neq j'} v\big(I^i \cap J^j\big)v\big(I^{i'} \cap J^{j'}\big)K_{ij}K_{i'j'} \\ &= \sum_{i,j} v\big(I^i \cap J^j\big)K_{ij}\left(\sum_{i',j':i'\neq i,j'\neq j} v\big(I^{i'} \cap J^{j'}\big)K_{i'j'}\right). \end{aligned}$$

Since, for fixed i and j,

$$\begin{aligned}
&\sum_{i',j':i'\neq i,j'\neq j} v\big(I^{i'}\cap J^{j'}\big)K_{i'j'}\\
&\quad=\sum_{i',j'} v\big(I^{i'}\cap J^{j'}\big)K_{i'j'}-v\big(I^{i}\cap J^{j}\big)-\sum_{j':j'\neq j} v\big(I^{i}\cap J^{j'}\big)K_{ij'}-\sum_{i':i'\neq i} v\big(I^{i'}\cap J^{j}\big)K_{i'j}\\
&\quad=\sum_{i',j'} v\big(I^{i'}\cap J^{j'}\big)K_{i'j'}+v\big(I^{i}\cap J^{j}\big)-v\big(I^{i}\big)-v\big(J^{j}\big),
\end{aligned}$$

we have

$$\begin{aligned}
D_4 =&\ \sum_{i,j} v\big(I^{i}\cap J^{j}\big)K_{ij}\sum_{i',j'} v\big(I^{i'}\cap J^{j'}\big)K_{i'j'}+\sum_{i,j} v\big(I^{i}\cap J^{j}\big)^2 K_{ij}\\
&-\sum_{i,j} v\big(I^{i}\cap J^{j}\big)K_{ij}v\big(I^{i}\big)-\sum_{i,j} v\big(I^{i}\cap J^{j}\big)K_{ij}v\big(J^{j}\big)\\
=&\ v(]0,\,T])^2+\sum_{i,j} v\big(I^{i}\cap J^{j}\big)^2 K_{ij}-\sum_{i} v\big(I^{i}\big)^2-\sum_{j} v\big(J^{j}\big)^2.
\end{aligned}$$

Thus $\mathrm{E}[D_4]=\theta^2+o(1)$ by (3.3).

Therefore, $\mathrm{E}[U_n^2]=E[D_1+D_2+D_3+D_4]=\theta^2+o(1)$, as desired.

(ii) Now we consider the case with non-zero drift such that $\sup_{0\leqslant t\leqslant T}|\mu_t^k|\in L^4$, $k=1,2$ Let $A_\cdot^k:=\int_0^\cdot \mu_t^k\mathrm{d}t$, $M_\cdot^k:=\int_0^\cdot \sigma_t^k\mathrm{d}W_t^k$, $k=1,2$, and

$$B_0:=\sum_{i,j}\Delta M^1(I^i)\Delta M^2(J^j)K_{ij},\qquad B_1:=\sum_{i,j}\Delta A^1(I^i)\Delta M^2(J^j)K_{ij};$$

$$B_2:=\sum_{i,j}\Delta M^1(I^i)\Delta A^2(J^j)K_{ij},\qquad B_3:=\sum_{i,j}\Delta A^1(I^i)\Delta A^2(J^j)K_{ij}.$$

Note that

$$\begin{aligned}
|B_1|&=\left|\sum_i\int_{I^i}\mu_t^1\mathrm{d}t\left(\sum_j\int_{J^j}\sigma_t^2\mathrm{d}W_t^2K_{ij}\right)\right|\leqslant\sum_i\int_{I^i}|\mu_t^1|\mathrm{d}t\left|\sum_j\int_{J^j}\sigma_t^2\mathrm{d}W_t^2K_{ij}\right|\\
&\leqslant T\sup_{0\leqslant t\leqslant T}|\mu_t^1|\cdot\max_i\sup\left\{\left|\int_s^t\sigma_t^2\mathrm{d}W_t^2\right|\,|t-s|\leqslant|I^i|+2\max_j|J^j|,\ s,\ t\in[0,\ T]\right\}\\
&\leqslant T\sup_{0\leqslant t\leqslant T}|\mu_t^1|\cdot\sup\left\{\left|\int_s^t\sigma_t^2\mathrm{d}W_t^2\right|\,|t-s|\leqslant\max_i|I^i|+2\max_j|J^j|,\ s,\ t\in[0,\ T]\right\},\qquad(3.4)
\end{aligned}$$

from which we can ascertain that B_1 is in L^2 because σ^2 is bounded and (the supremum of) μ^1 is in L^4. Moreover, under Condition C(ii), $\mathrm{E}\big[B_1^2\big]=o(1)$ as $n\to\infty$ by the dominated convergence theorem. $\mathrm{E}\big[B_2^2\big]=o(1)$ and $\mathrm{E}\big[B_3^2\big]=o(1)$ can be shown similarly.

Because

$$\mathrm{E}[(U_n-\theta)^2] \leqslant 2\mathrm{E}[(B_0-\theta)^2] + 8\mathrm{E}[B_1^2+B_2^2+B_3^2],$$

together with result (i), assertion (a) has been shown.

Assertion (b) is obvious from (3.4), noting that what we need is merely $B_k \to 0$ in probability as $n \to \infty$, $k = 1, 2, 3$. □

We have the following corollary as an immediate application of Theorem 3.1.

Corollary 3.1. *If Π is created as per the Poisson sampling stated in Example 3.1, then the same conclusion as the Theorem 3.1 holds.*

For the proof, one only needs to show that the Poisson sampling satisfies the Condition C. See the Appendix.

Remark 3.4. The realized covariance estimator (1.1) is not generally consistent as $n \to \infty$, once the regular interval size h (> 0) is fixed. For instance, consider the case of Brownian motions with arbitrary sampling scheme satisfying Condition C, such as, the Poisson sampling of Example 3.1. For simplicity, let $h = T$, hence only two pairs, $(W_0^1, W_0^2) \equiv (0, 0)$ and (W_T^1, W_T^2), are utilized to construct the estimator $V_h = V_T = W_T^1 W_T^2$. Then, regardless of n, $\mathrm{E}[V_h] = \rho T$, but $\mathrm{var}[V_h] = (\rho^2+1)T^2$, which never vanishes as $n \to \infty$.

3.2. Correlation estimators

Suppose further that $\rho_t \equiv \rho$ and $\sigma_t^l \equiv \sigma^l$ for some constants, $\rho \in (-1, 1)$ and $\sigma^l > 0$, $l = 1, 2$. We are now interested in estimating the correlation ρ. Obviously, we are able to form statistics that estimate ρ consistently by 'standardizing' U_n as follows.

Definition 3.2. *Correlation estimators.*

$$R_n^{(1)} := \frac{1}{T}\sum_{i,j}\frac{\Delta P^1(I^i)\Delta P^2(J^j)}{\sigma^1\sigma^2}1_{\{I^i\cap J^j\neq\varnothing\}} \qquad (\sigma^l \textit{ known});$$

$$R_n^{(2)} := \frac{\displaystyle\sum_{i,j}\Delta P^1(I^i)\Delta P^2(J^j)1_{\{I^i\cap J^j\neq\varnothing\}}}{\left\{\displaystyle\sum_i \Delta P^1(I^i)^2\right\}^{1/2}\left\{\displaystyle\sum_j \Delta P^2(J^j)^2\right\}^{1/2}} \qquad (\sigma^1 \textit{ known/unknown}).$$

Corollary 3.2. *Under Condition C, $R_n^{(1)}$ and $R_n^{(2)}$ are consistent for ρ as $n \to \infty$.*

An obvious drawback of the estimators $R_n^{(1)}$ and $R_n^{(2)}$ is that they are not bounded by 1 in magnitude; hence, it may make sense to apply to them a bounded transformation. In particular, let $\varphi(x) := (x \wedge 1) \vee (-1)$ be such a transformation. Then we may define modified correlation estimators by

$$\tilde{R}_n^{(k)} := \varphi(R_n^{(k)}), \qquad k = 1, 2.$$

Although such modification may induce extraneous biases, due to the continuous mapping theorem $\tilde{R}_n^{(k)}$ is still expected to estimate ρ well for large n. Besides, preliminary numerical experiments (with small or medium-size n) have indicated some potential improvements in terms of mean square error, especially when ρ is large in magnitude. Further investigation is needed in this direction, including development of alternative estimators.

3.3. Application to finance: multi-dimensional Black–Scholes model

Consider a market with d securities, $(P^1, \ldots, P^{\mathrm{d}})$, with P_t^k the price of the kth stock at $t \in [0, T]$, $k = 1, \ldots, d$. We suppose each P^k follows a one-dimensional geometric Brownian motion,

$$\mathrm{d}P_t^k = \mu_t^k P_t^k \,\mathrm{d}t + \sigma_t^k P_t^k \,\mathrm{d}W_t^k, \qquad P_0^k = p^k,\ k = 1, \ldots, d,$$

where W^ks are Brownian motions with $\mathrm{d}\langle W^k, W^l \rangle_t = \rho_t^{k,l}\mathrm{d}t$, $\rho_t^{k,l} \in]-1, 1[$, $\mu^k(t)$, and $\sigma_t^k > 0$ are all (unknown) deterministic and bounded functions, k (or l) $= 1, \ldots, d$. Let $I^{k,i}$ denote the ith observation interval for the kth security, satisfying Condition C and being independent of $W^1, \ldots, W^{\mathrm{d}}$.

Put $\mathbf{X} := (X^1, \ldots, X^d)^{\mathrm{T}}$, where $X_t^k := \ln P_t^k$. Then, the $d \times d$ matrix $\mathbf{U}_n := \left(U_n^{k,l}\right)_{1 \leqslant k,l \leqslant d}$, the (k, l)th element of which is defined by

$$U_n^{k,l} := \sum_{i,j} \Delta X^k(I^{k,i}) \Delta X^l(I^{l,j}) 1_{\{I^{k,i} \cap I^{l,j} \neq \emptyset\}},$$

is a consistent estimator for the (cumulative) covariance matrix of returns $\langle \mathbf{X}, \mathbf{X} \rangle_T := (\langle X^k, X^l \rangle_T)_{1 \leqslant k,l \leqslant d}$ with $\langle X^k, X^l \rangle_T = \int_0^T \sigma_t^k \sigma_t^l \rho_t^{k,l} \mathrm{d}t$. □

4. Concluding remarks

In this paper, a new procedure for estimating the covariation of two diffusion processes is proposed when they are observed only at discrete times in a non-synchronous manner. The proposed estimator is free of any synchronization of the original data, hence free from any problems caused by it, especially, biases of Epps type. The estimator is shown to have consistency as the observation frequency tends to infinity.

The theory may be extended in several directions. One may wonder, for instance, about *rates of convergence* for the estimator, for rates could be a useful way to understand the effect of non-synchronous data for the accuracy of estimation. In fact, in the simplest case – Brownian motions with the Poisson sampling scheme discussed in Example 3.1 – the rate has turned out to be $\sqrt{n}$. Moreover, we have established *asymptotic normality* under general conditions covering this as a special case. The result will be presented in a separate paper (Hayashi and Yoshida 2004).

Among interesting questions as to the convergence result is whether some *uniformity* in the parameter holds for the convergence in probability. It would be worthwhile to investigate this.

Diffusion processes considered in the paper have deterministic and bounded diffusion coefficients with no feedback from the processes, which would certainly limit broad applications. Extension of the methodology to *general* diffusion processes is under way.

The independence assumption of observation times from underlying diffusion processes may be restrictive in financial modelling. For instance, it is more natural to assume that trading takes place dependent on movements of security prices within intraday. The relaxation of the assumption to allow observations to be made at arbitrary *stopping times* is currently under investigation.

Similar arguments can hold even by replacing Brownian motions with *Lévy processes*, for the key properties used to prove the main theorem are independence of increments and finiteness of moments. This assertion needs to be verified.

Appendix: Proof of Corollary 3.1

To prove Corollary 3.1, one only needs to check whether the Poisson sampling scheme in Example (3.1) satisfies Condition C(ii). Without loss of generality we assume $T = 1$. Also, to simplify the notation $\tilde{T}^{1,i}$ will be written as $T^{1,i}$.

We will show that $P\big[\max_i |I^i| > n^{-3/4}\big] = o(1)$, which is sufficient for Conditon C(ii) to hold as in Remark 3.1. Put $i^* := \min\{i : T^{1,i} \geqslant 1\}$. Notice that $|I^i| = 0$ for all i such that $T^{1,i-1} > 1$, hence $|I^i| = 0$ for all $i > i^*$. Moreover, $|I^{i^*}| = 1 - T^{1,i^*-1} \leqslant T^{1,i^*} - T^{1,i^*-1}$. These facts imply

$$\left\{\max_i |I^i| > n^{-3/4}\right\} = \left\{\max_{i\leqslant i^*} |I^i| > n^{-3/4}\right\} \subset \left\{\max_{i\leqslant i^*} (T^{1,i} - T^{1,i-1}) > n^{-3/4}\right\},$$

furthermore, since $i^* \leqslant [3np^1]$ on $\{T^{1,[3np^1]} > 1\} =: H$,

$$\begin{aligned}\left\{\max_i |I^i| > n^{-3/4}\right\} \cap H &\subset \left\{\max_{i\leqslant i^*} (T^{1,i} - T^{1,i-1}) > n^{-3/4}\right\} \cap H \\ &\subset \left\{\max_{i\leqslant [3np^1]} (T^{1,i} - T^{1,i-1}) > n^{-3/4}\right\}.\end{aligned}$$

So,

$$\begin{aligned}P\left[\max_i |I^i| > n^{-3/4}\right] &= P\left[\max_i |I^i| > n^{-3/4},\, H\right] + P\left[\max_i |I^i| > n^{-3/4},\, H^c\right] \\ &\leqslant P\left[\max_{i\leqslant [3np^1]} (T^{1,i} - T^{1,i-1}) > n^{-3/4}\right] + P[H^c]. \qquad \text{(A.1)}\end{aligned}$$

By assumption, $(T^{1,i} - T^{1,i-1})$ are i.i.d. exponentially distributed with mean $1/np^1$, therefore, the first term on the right-hand side of (A.1) is evaluated as

$$P\left[\max_{i\leqslant[3np^1]}(T^{1,i}-T^{1,i-1})>n^{-3/4}\right]=1-P\left[\max_{i\leqslant[3np^1]}(T^{1,i}-T^{1,i-1})\leqslant n^{-3/4}\right]$$
$$=1-(1-\exp\{-(np^1)\cdot n^{-3/4}\})^{[3np^1]}$$
$$=[3np^1]\exp\{-n^{1/4}p^1\}+o(1).$$

On the other hand, due to the Stirling formula

$$k!\sim k^{k+\frac{1}{2}}\mathrm{e}^{-k}\sqrt{2\pi}>k^k\mathrm{e}^{-k}\sqrt{2\pi}$$

for all large k, as well as the fact that $(np^1/[3np^1])\mathrm{e}<1$ for all large n, the second term on the right of (A.1) is bounded by

$$P[H^c]=P[T^{1,[3np^1]}\leqslant 1]\leqslant\sum_{k=[3np^1]}^{\infty}\mathrm{e}^{-np^1}\frac{(np^1)^k}{k!}\leqslant C_1\sum_{k=[3np^1]}^{\infty}\mathrm{e}^{-np^1}\frac{(np^1)^k}{k^k\mathrm{e}^{-k}}$$
$$\leqslant C_1\mathrm{e}^{-np^1}\sum_{k=[3np^1]}^{\infty}\left(\frac{np^1}{[3np^1]}\mathrm{e}\right)^k\leqslant C_2\mathrm{e}^{-np^1}$$

for all large n, for some constants C_1 and C_2. Therefore, Condition C(ii) is satisfied, hence the corollary is proved.

Acknowledgement

The authors are grateful to the editor and the anonymous referees for helpful comments. This study was initiated during the first author's stay at the Graduate School of Economics, Kyoto University, in the summer of 2003, and completed at the Graduate School of Mathematical Sciences, University of Tokyo, in the spring and summer of 2004. He is grateful to Professors Masaaki Kijima and Shigeo Kusuoka for their hospitality. Financial support from Daiwa Securities Chairship at Kyoto University, the 21st Century COE Program at the University of Tokyo (Hayashi), the Research Fund for Scientists of the Ministry of Science, Education and Culture, Japan, and the Cooperative Research Program of the Institute of Statistical Mathematics (Yoshida) is also gratefully acknowledged.

References

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001) The distribution of realized exchange rate volatility. *J. Amer. Statist. Assoc.*, **96**, 42–55.

Bandi, F.M. and Russell, J.R. (2003) Microstructure noise, realized volatility, and optimal sampling. Preprint, Graduate School of Business, University of Chicago.

Barucci, E. and Renò, R. (2002) On measuring volatility and the GARCH forecasting performance. *J. Internat. Financial Markets Institutions Money*, **12**, 183–200.

Brandt, M.W. and Diebold, F.X. (2003) A no-arbitrage approach to range-based estimation of covariances and correlations. Penn Institute for Economic Research Working Paper 03-013.

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

de Jong, F. and Nijman, T. (1997) High frequency analysis of lead-lag relationships between financial Markets. *J. Empir. Finance*, **4**, 259–277.

Epps, T.W. (1979) Comovements in stock prices in the very short run. *J. Amer. Statist. Assoc.*, **74**, 291–298.

Genon-Catalot, V. and Jacod, J. (1993) On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. H. Poincaré Probab. Statist.*, **29**(1), 119–151.

Genon-Catalot, V. and Jacod, J. (1994) Estimation of the diffusion coefficient for diffusion processes: Random sampling. *Scand. J. Statist.*, **21**(3), 193–221.

Goodhart, C.A. and O'Hara, M. (1997) High frequency data in financial markets: issues and applications. *J. Empir. Finance*, **4**, 73–114.

Hayashi, T. and Nakahiro, N. (2004) Asymptotic normality of nonsynchronous covariance estimators for diffusion processes. In preparation.

Kessler, M. (1997) Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.*, **24**, 211–229.

Lo, A.W. and MacKinlay, A.C. (1990) An econometric analysis of nonsynchronous trading. *J. Econometrics*, **45**, 181–211.

Lundin, M.C., Dacorogna, M.M. and Müller, U.A. (1999) Correlation of high frequency financial time series. In P. Lequex (eds), *The Financial Markets Tick by Tick*, pp. 91–126. New York: Wiley.

Malliavin, P., and Mancino, M.E. (2002) Fourier series method for measurement of multivariate volatilities. *Finance Stochastics*, **6**, 49–61.

Martens, M. (2004) Estimating unbiased and precise realized covariances. European Finance Association 2004 Maastricht Meetings Paper no. 4299. SSRN Electronic Library.

Muthuswamy, J., Sarkar, S., Low, A. and Terry, E. (2001) Time variation in the correlation structure of exchange rates: High-frequency analyses. *J. Futures Markets*, **21**, 127–144.

Prakasa Rao, B. (1983) Asymptotic theory for non-linear least square estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.*, **14**, 195–209.

Prakasa Rao, B. (1988) Statistical inference from sampled data for stochastic processes. In N.U. Prabhu (ed.), *Statistical Inference from Stochastic Processes*, Contemp. Math, **80**, pp. 249–287. Providence, RI: American Mathematical Society.

Renò, R. (2003) A closer look at the Epps effect. *Int. J. Theor. Appl. Finance*, **6**(1), 87–102.

Rydberg, T.H. and Shephard, N. (1999) A modelling framework for the prices and times of trades made on the NYSE. Technical report, Nuffield College, Oxford University.

Scholes, M. and Williams, J. (1977) Estimating beta from nonsynchronous data. *J. Financial Economics*, **5**, 309–327.

Tsay, R.S. and Yeh, J.-H. (2003) Non-synchronous trading and high-frequency beta. Preprint.

Yoshida, N. (1992) Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, **41**, 220–242.

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2003) A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Technical Report 542, Department of Statistics, University of Chicago (in revision). To appear in *J. Amer. Statist. Assoc*.

Received November 2003 and revised April 2004