Estimating Covariation: Epps Effect, Microstructure Noise *

Lan Zhang
Department of Finance
University of Illinois at Chicago
Chicago, IL 60607
E-mail: lanzhang@uic.edu

Original version: June 21, 2005
Previous version: October 24, 2008
This version: July 2, 2009

Abstract

This paper is about how to estimate the integrated covariance $\langle X, Y\rangle_T$ of two assets over a fixed time horizon $[0, T]$, when the observations of X and Y are "contaminated" and when such noisy observations are at discrete, but not synchronized, times. We show that the usual previous-tick covariance estimator is biased, and the size of the bias is more pronounced for less liquid assets. This is an analytic characterization of the Epps effect. We also provide optimal sampling frequency which balances the tradeoff between the bias and various sources of stochastic error terms, including nonsynchronous trading, microstructure noise, and time discretization. Finally, a two-scales covariance estimator is provided which simultaneously cancels (to first order) the Epps effect and the effect of microstructure noise. The gain is demonstrated in data.

KEYWORDS: Bias-variance tradeoff; Epps effect; High frequency data; Measurement error; Market Microstructure; Martingale; Nonsynchronous trading; Realized covariance; Realized variance; Two scales estimation.

JEL CODES: C01; C13; C14; C46; G11

*The author would like to thank the editors and referees for helpful and constructive comments.

1. Introduction

This paper is about how to estimate the integrated covariance $\langle X,Y\rangle_T$ over a fixed time horizon $[0,T]$, when the observations of X and Y are "contaminated" and when such noisy observations of X and of Y are at discrete, but not synchronized, times.

Consider the price processes of two assets, $\{X_t\}$ and $\{Y_t\}$, both in logarithmic scale. Suppose both $\{X_t\}$ and $\{Y_t\}$ follow an Itô process, namely,

$$dX_t = \mu_t^X dt + \sigma_t^X dB_t^X, \tag{1}$$

$$dY_t = \mu_t^Y dt + \sigma_t^Y dB_t^Y, \tag{2}$$

where B^X and B^Y are standard Brownian motions, with correlation $corr(B_t^X, B_t^Y) = \rho_t$. The drift coefficient μ_t, and the instantaneous variance σ_t^2 of the returns process X_t will be stochastic processes, which are assumed to be locally bounded.

Our interest is to estimate the integrated covariation $\langle X,Y\rangle_T$,

$$\langle X,Y\rangle_T = \int_0^T \sigma_t^X \sigma_t^Y d\langle B^X, B^Y\rangle_t, \tag{3}$$

using the ultra-high frequency observations of X and Y within the fixed time horizon $[0,T]$. Inference for (3) is a well-understood problem if X and Y are observed *simultaneously* and *without contamination* (say, in the form of microstructure noise). A limit theorem in stochastic processes states that $\sum_{i:\tau_i\in[0,T]}(X_{\tau_i} - X_{\tau_{i-1}})(Y_{\tau_i} - Y_{\tau_{i-1}})$, commonly called *realized covariance*, is a consistent estimator for $\langle X,Y\rangle_T$ as the observation intervals get closer; furthermore its estimation error follows a mixed normal distribution, see, for example, Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002a), Zhang (2001), and Mykland and Zhang (2006). For a glimpse of the econometric literature on this inference problem when $X = Y$, one can read Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2002b), and Gençay, Ballocchi, Dacorogna, Olsen, and Pictet (2002), among others.

In ultra-high frequency data, the exact observation times of X and Y are rarely simultaneous, and estimating $\langle X,Y\rangle_T$ in this *asynchronous* case becomes a relevant and pressing problem. This lack of synchronicity often causes some undesirable features in the inference. In particular, as documented by Epps (1979), correlation estimates tend to decrease when sampling is done at high frequencies. Even in daily data, asynchronicity can cause difficulties (Scholes and Williams (1977)). Lo and MacKinlay (1990) propose a solution based on a stochastic model of censoring. In practice, most nonparametric estimation procedures for $\langle X,Y\rangle_T$ start with creating an approximately synchronized pair (X,Y) by either *previous-tick interpolation* or *linear interpolation*, then construct the estimator on the basis of the synchronized approximations. These interpolation-based estimators are often biased, as witnessed in empirical studies (Dacorogna, Gençay, Müller, Olsen, and Pictet (2001)).

A different issue when one deals with high frequency data is the existence of microstructure noise. In the early papers (Aït-Sahalia, Mykland, and Zhang (2005), Zhang, Mykland, and Aït-Sahalia (2005)), they found that when the microstructure noise is present in the observed prices, then the *realized variance* estimator for $\langle X,X\rangle_T$ – a special case of realized covariance – is biased and this bias can get progressively worse as more high frequency data is employed.[1] However, it is not well understood how an estimator for the covariation $\langle X,Y\rangle_T$ behaves, when the estimation uses ultra-high frequency noisy data.

[1]Recent developments on volatility estimation include multi-scale estimation (Zhang (2006), Aït-Sahalia, Mykland, and Zhang

In this paper, we are concerned with the behavior of the previous-tick approach to estimation of $\langle X, Y\rangle_T$ when the observation times of X and Y are not synchronized and when the microstructure noise is present in the observed price processes. We show that asynchronicity leads to a bias in the previous-tick estimator for $\langle X, Y\rangle_T$, thus giving an analytic form of the Epps effect. The variance of the estimator, meanwhile, comes from three sources – discrete observation/transaction times, nonsynchronization, and the microstructure noise. We provide the optimal sampling frequency to balance the tradeoff among different error sources, and present the explicit expression for the asymptotic bias and variance when the observations times of X and Y follow Poisson process.

A further advantage of the previous tick estimator is that it permits easy analysis of microstructure noise. It is here shown that in the presence of noise, one can create two and multi scale versions of the previous tick estimator. As we shall see in Section 8, the bias due to asynchronicity cancels in the same way as the bias due to microstructure noise, while the variance asymptotically behaves as if there is no asynchronicity (in the subsample of previous ticks). Thus, while the previous tick approach does throw away data, it can retain rate efficiency.

In terms of microstructure noise, this paper provides a two- and multiscale alternative to the multivariate autocovariance-based estimator of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b). Other work investigating the combination of asynchronicity and microstructure noise includes Lunde and Voev (2007) and Griffin and Oomen (2007).

The paper is organized as follows: we introduce the concepts and notations in Section 2.1, and give a preview of the main findings in Section 2.2. Section 3 and 4 provide the asymptotic stochastic bias and variance of the previous-tick estimator, assuming the absence of microstructure noise in the price processes. Section 5 deals with the case when the trading times are random. An application when the transaction times follow Poisson processes is provided in Section 6. Section 7 focuses on the inference when the microstructure noise is present. Two scales estimation is presented in Section 8. Finally, Section 9 concludes.

2. Setting, and some Main Findings

2.1. Setup and Notations

Our interest is to estimate the covariation $\langle X, Y\rangle_T$ between two returns in a fixed time period $[0, T]$, when X and Y are observed asynchronously.

Let the observation/transaction times of X be recorded in $\mathcal{T}_n$, and those of Y in $\mathcal{S}_m$. At the moment we assume X and Y are free of microstructure noise (in short, *noise*). Later in Section 7 we study the cases when these two price processes are observed with noise ϵ^X and ϵ^Y respectively. We denote the elements in $\mathcal{T}_n$ by $\tau_{n,i}$, and the elements in $\mathcal{S}_m$ by $\theta_{m,i}$. Specifically, $0 = \tau_{n,0} \leq \tau_{n,1} \leq \ldots \leq \tau_{n,n} = T$, $0 = \theta_{m,0} \leq \theta_{m,1} \leq \ldots \leq \theta_{m,m} = T$. For the ease of the notation, we often suppress the subscript n and m from the τs and θs unless the context is misleading. The τ and θ sequences may be irregular and random but independent of the price process, so long as the spacings are not allowed to be too large. An extension to more general random times is considered in Section 5.

We focus on a particular type of covariance estimator called *previous-tick* estimator. Intuitively, it is

(2009)), kernel methods (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a, 2009)), and pre-averaging (Podolskij and Vetter (2009) , Jacod, Li, Mykland, Podolskij, and Vetter (2009)).

a sample covariance estimator based on the prices that immediately precede (or are at) the pre-specified sampling points. One can view this previous-tick approach as a special way to subsample the raw data.

To formulate the previous-tick covariance estimator, we introduce the concepts related to sampling points. Let

$$N = n + m,$$

write n and m as n_N and m_N from here on.

We consider a subset of $[0, T]$ which satisfies the following.

$$\mathcal{V}_N \subset [0, T]; \quad 0, T \in \mathcal{V}_N, \text{ also } \mathcal{V}_N \text{ is finite for each } N. \tag{4}$$

We use v_i to denote the elements in $\mathcal{V}_N$, $\mathcal{V}_N = \{v_0, v_1, \ldots, v_{M_N}\}$, with $v_0 = 0$ and $v_{M_N} = T$, where M_N is the sampling frequency. A simple case of $\mathcal{V}$ would be a *regular grid*, where the elements are equally spaced out in time, that is, $v_i - v_{i-1} = \Delta v, \forall i$. This sampling scheme is the most common one in analyzing time-dependent data, for example, typical sampling interval in high-frequency financial application includes every 5 minutes, 15 minutes, 30 minutes and hourly.

An alternative way of setting the grid $\mathcal{V}_N$ is to let the v_i's depend on the observation times, for example by setting v_i to be the maximum of $\min\{\tau \in \mathcal{T}_n : \tau > v_{i-1}\}$ and $\min\{\theta \in \mathcal{S}_m : \theta > v_{i-1}\}$. This is the concept of refresh time, as introduced by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b). One can also implement this for more than two stocks.

We assume the following regarding the relation between v_is, τ_is, and θ_is:

CONDITION C1: There is at least one pair of (τ, θ) in between the consecutive v_is.

Under Condition C1, the previous ticks are then defined as:

$$t_i = \max\{\tau \in \mathcal{T}_n : \tau \le v_i\}, \quad \text{and} \quad s_i = \max\{\theta \in \mathcal{S}_m : \theta \le v_i\}, \tag{5}$$

so that the t_is and the s_is are the sampling points in X and Y, respectively, according to the previous-tick sampling scheme. We note that condition C1 holds so long as there are sufficiently many data in both X and Y within the time window $[0, T]$. A sufficient criterion for C1 is provided by Conditions C2-C3 below. C1 is also valid (without C3) when the v's are the refresh times mentioned above.

We need more assumptions to pursue the analysis for the covariance estimator. We assume that the transaction times of X and Y satisfy:

CONDITION C2: $\sup_i |\theta_{m,i} - \theta_{m,i-1}| = O(\frac{1}{N})$, and $\sup_i |\tau_{n,i} - \tau_{n,i-1}| = O(\frac{1}{N})$.

Note that Condition C2 implies that on the one hand $\liminf_{N\to\infty} \frac{m_N}{N} > 0$ and $\liminf_{N\to\infty} \frac{n_N}{N} > 0$; on the other hand, it is obvious that $\frac{m_N}{N} \le 1$ and $\frac{n_N}{N} \le 1$. In particular, $n_N = O(m_N)$ and $m_N = O(n_N)$. We sometimes assume that the sampling frequency M_N satisfies:

CONDITION C3: $\sup_i |v_{N,i} - v_{N,i-1}| = O(\frac{1}{M_N})$, and $M_N = o(N)$.

Conditions C2-C3 imply Condition C1 when N is large enough.

There are two reasons for imposing C3. One is technical, it arises naturally in connection with both two scales estimation (Section 8) and bias-variance tradeoffs (Sections 4.3, 6.2 and 7.1). The other is more conceptual: the observation times are often not known exactly or incorrectly recorded. If one assumes that the times are known up to, say, order $O(N^{-\alpha})$, having the distance between consecutive grid points in $\mathcal{V}$, $v_i - v_{i-1}$, bigger than $O(N^{-\alpha})$ ensures the previous-tick estimator to be consistent.

Definition 1. *At last, the previous-tick estimator for the covariation is defined as*

$$[X, Y]_T = \sum_{i=1}^{M_N} (X_{t_i} - X_{t_{i-1}})(Y_{s_i} - Y_{s_{i-1}}), \tag{6}$$

where the t_is and s_is are the previous ticks in (5).

2.2. Some Main Findings

We here summarize the most important results from the practitioner point of view. First of all, the bias in the estimator (6) is given by

$$-\int_0^T \langle X, Y\rangle'_u dF_N(u) + O_p(\frac{1}{N}), \tag{7}$$

where $F_N(t) = \sum_{i:\max(t_i,s_i)\leq t} |t_i - s_i|$. Typically, $F_N(t)$ and the bias are of order $O_p(M_N/N)$. See Theorem 1 for precise statements. Second, when Condition C3 is in place, then

$$[X, Y]_T = \widetilde{[X, Y]}_T - \int_0^T \langle X, Y\rangle'_u dF_N(u) + O_p(N^{-1/2}) \tag{8}$$

where $\widetilde{[X, Y]}_T$ is the unobserved value of the synchronized estimator

$$\widetilde{[X, Y]}_T = \sum_{i=1}^{M_N} (X_{v_i} - X_{v_{i-1}})(Y_{v_i} - Y_{v_{i-1}}). \tag{9}$$

See Theorem 4 (in Section 4.1) for details. Under condition C3, therefore, one can behave as if observations were synchronously obtained at times v_i, provided that one can deal with the bias. This has important consequences. On the one hand, it provides an analytic characterization of the Epps (1979) effect. As described further in Section 3.2, the Epps effect is essentially the bias (7), and it is typically negative for positively associated processes (X, Y). Also, from (8), the Epps effect is only a matter of bias; except at the highest sampling frequencies, it does not substantially affect the variance of the estimator. On the other hand, (8) suggests that when suitably adapted, existing theory for the synchronized case can be applied to the asynchronous case.

We shall show two types of applications. In Sections 4.3, 6.2, and 7.1, we carry out a bias-variance trade-off to remove the effect of asynchronicity. In Section 8 we show that both asynchronicity and microstructure noise can be removed with the help of two-scales estimation, along the lines of Zhang, Mykland, and Aït-Sahalia (2005).

3. Previous-tick covariance estimator under zero noise

We start with an idealized world, where the mechanics of the trading process is perfect so that there is no microstructure noise in both X and Y. We shall see that $[X, Y]_T$ can be decomposed based on the impact of different data structure.

3.1. Decomposition for the estimator $[X, Y]_T$

Let X and Y be Itô processes satisfying (1)-(2). Let $[X, Y]_T$ be the previous-tick covariance estimator in (6). From the Kunita-Watanabe inequality, $\langle X, Y\rangle_t$ is absolutely continuous in t. Assuming Condition C1, we can therefore decompose $[X, Y]_T$ into:

$$
\begin{aligned}
[X,Y]_T \;=\; & \underbrace{\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} \langle X,Y\rangle'_u du}_{\text{drift term}} \\
& + \underbrace{\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2]}_{L_N,\ \text{discretization error}} + \underbrace{R_N}_{\text{asynchronicity error}},
\end{aligned} \tag{10}
$$

where the t_is and s_is are the previous ticks defined in (5), and see Lemma 1 in Appendix for the exact form of R_N. We have used the following symbol (cf. McCullagh (1987)):

Notation 1. *The symbol "[2]" is used as follows: if* (a, b) *is an expression in* a *and* b*, then* $(a, b)[2]$ *means* $(a, b) + (b, a)$*, so that* $(X_u - X_{\max(t_{i-1},s_{i-1})})dY_u[2]$ *means* $(X_u - X_{\max(t_{i-1},s_{i-1})})dY_u + (Y_u - Y_{\max(t_{i-1},s_{i-1})})dX_u$*.*

Each component in (10) plays different role in the distribution of $[X, Y]_T$. To proceed the discussion, we need first to define *stochastic bias* and *stochastic variance* of an estimator.

Definition 2. *Consider a semimartingale* Z*. Let* $\hat{Z}$ *be an estimator for* Z*. Suppose that* $Z_t - \hat{Z}_t$ *has the following Doob-Meyer decomposition, for* $t \in [0, T]$*,*

$$\hat{Z}_t - Z_t = A_t + M_t,$$

where $\{M_t\}$ *is a martingale and* $\{A_t\}$ *is a predictable process. Then for fixed* t*,* $t \in [0, T]$*, we call* A_t *the stochastic bias of* $\hat{Z}_t$*, denoted as* $SBias(\hat{Z}_t)$*; we call the quadratic variation* $\langle M, M\rangle_t$ *of* M_t *the stochastic variance of* $\hat{Z}_t$*.*

Note that if A_t is nonrandom, it is also the exact bias; if $\langle M, M\rangle_t$ is nonrandom, it gives the exact variance.

In light of Definition 2 and the decomposition equation (10), the stochastic bias of $[X, Y]_T$ is

$$\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} \langle X,Y\rangle'_u du - \langle X,Y\rangle_T;$$

meanwhile, both the discretization error L_N and the asynchronicity error R_N contribute to the stochastic variance of $[X, Y]_T$. It is apparent that, in the situation when X and Y are traded simultaneously, the asynchronicity error R_N becomes zero. When the trading is not synchronous, however, it is not obvious to see the relative impact and the trade-off between L_N and R_N. We pursue these next. First, we need the following concept.

Definition 3. *A sequence of càdlàg processes* $G_N(t), 0 \le t \le T$ *is said to be relatively compact in probability (RCP) if every subsequence has a further subsequence* G_{N_k} *so that there is a process* $G(t)$*, where* $G_{N_k}(t)$ *converges in probability to* $G(t)$ *at every continuity point* $t \in [0, T]$ *of* $G(t)$*.*

For applied purposes, if the sequence G_N is RCP, one can act as if the limit exists, cf. the discussion on p. 1411 of Zhang, Mykland, and Aït-Sahalia (2005).

3.2. Stochastic bias: The Epps Effect

Theorem 1. *Let X and Y be Itô processes satisfying (1)-(2), with μ_t and σ_t locally bounded. Let $[X, Y]_T$ be the previous-tick covariance estimator. Let $\mathcal{V}_N = \{0 = v_0, v_1, \ldots, v_{M_N}\}$ be a collection of sampling points which span across $[0, T]$, and let t_i and s_i be the transaction times of X and Y, respectively, that immediately precede v_i. Then, under Conditions* C1-C2*, the stochastic bias of $[X, Y]_T$ is*

$$-\int_0^T \langle X, Y\rangle'_u dF_N(u) + O_p(\frac{1}{N}), \tag{11}$$

where

$$F_N(t) \quad = \sum_{i:\max(t_i,s_i)\leq t} |t_i - s_i|.$$

Furthermore, the sequences $\frac{N}{M_N}F_N(t)$ and $\frac{N}{M_N}\int_0^T \langle X, Y\rangle'_t dF_N(t)$ are RCP in the sense of Definition 3. ∎

The function F_N takes non-negative value and it will play a central rôle in our narrative. To see an example of a limit of $\frac{N}{M_N}F_N(t)$, we refer to the Poisson example in Corollary 4 (Section 6).

From Theorem 1, one should note that $F_N(t) = 0$ – thus the previous-tick estimator is unbiased – when the two processes X and Y are traded simultaneously, or more generally if the selected subsample has synchronized observation times. If the two assets X and Y are not traded simultaneously, the stochastic bias typically has order M_N/N, the previous-tick estimator $[X, Y]$ is then asymptotically unbiased under Condition C3.

However, there is a finite sample effect in (11), and (11) is an analytic representation of the Epps effect in cases where the subsampling is moderate (see the discussion in Section 2.2). Also Theorem 1 implies the magnitude of the bias $-\int_0^T \langle X, Y\rangle'_u dF_N(u)$ is greater for less liquid assets (large $|t_i - s_i|$ on average).

Remark 1. *When the previous tick estimator is used for all of $[X, Y]_T$, $[X, X]_T$, and $[Y, Y]_T$, the correlation estimator is no larger than one in absolute value. If one uses a different type of estimator for $[X, X]_T$, and $[Y, Y]_T$, the estimated correlation should just be truncated at 1 or -1 as appropriate. Similar comments apply when a covariation matrix $\langle X, X\rangle_T$ is estimated for a vector process X. If a different estimator is used to compute the diagonal elements, one can take the estimated matrix and write $\widetilde{\langle X, X\rangle}_T = \Gamma\Lambda\Gamma^*$, where Γ is orthogonal and Λ is a diagonal matrix. If one sets Λ^+ as the matrix Λ with negative elements replaced by zero, a nonnegative definite estimator of $\langle X, X\rangle_T$ is given by $\widehat{\langle X, X\rangle}_T = \Gamma\Lambda^+\Gamma^*$. Asymptotically all these procedures have the same properties when the true $\langle X, X\rangle_T$ is positive definite, since then Λ is eventually positive with probability one as $N \to \infty$.*

3.3. Stochastic variance

Theorem 2. *Under the same conditions and setup as in Theorem 1, the following processes*

$$U_{N,u}^{(dis)} = \sum_{i:t_i,s_i\le u} \frac{(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^2}{T/M_N},$$

$$U_{N,u}^{(nonsyn)} = \sum_{i:s_i,t_i\le u} \frac{(s_i-s_{i-1})(t_i-t_{i-1})-(\max(t_{i-1},s_{i-1})-\min(t_i,s_i))^2}{T/N},$$

are RCP in the sense of Definition 3, and the process

$$M_N^{3/2}Q_{N,u}[2] = 2\sum_{i:t_i,s_i\le u}\langle X,X\rangle'_{t_i}\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(\min(t_i,s_i)-u)(Y_u-Y_{\max(t_{i-1},s_{i-1})})dY_u[2]$$
$$+2\sum_{i:t_i,s_i\le u}\langle X,Y\rangle'_{t_i}\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(\min(t_i,s_i)-u)(X_u-X_{\max(t_{i-1},s_{i-1})})dY_u[2].$$

is tight. Also, the leading terms in the stochastic variance of $[X,Y]_T-\langle X,Y\rangle_T$ *are*

$$\frac{T}{M_N}\int_0^T\left[\langle X,X\rangle'_u\langle Y,Y\rangle'_u+(\langle X,Y\rangle'_u)^2\right]dU_{N,u}^{(dis)}+Q_{N,T}[2]+\frac{T}{N}\int_0^T\langle X,X\rangle'_u\langle Y,Y\rangle'_u dU_{N,u}^{(nonsyn)}. \quad (12)$$

Finally,

$$U_{N,u}^{(dis)} = u-2F_N(u)+O(1/M_N)+O((M_N/N)^2). \quad (13)$$

∎

As we shall see from the proof of Theorem 2 (in Appendix), the $1/M_N$ term and $1/M_N^{3/2}$ term (i.e. $Q_{N,T}$) in (12) correspond to the first- and the second-order effect, from the quadratic variation of the discretization error in (10), whereas the $1/N$ term comes from the quadratic variation of the asynchronous error. We can call $U^{(dis)}$ *quadratic covariation of time due to discretization*, and call $U^{(nonsyn)}$ *quadratic covariation of time due to nonsynchronization*.

Remark 2. *In the special case where X and Y are traded simultaneously, $U^{(nonsyn)}$ becomes zero, and the total asymptotic variance in (12) reduces to*

$$\frac{T}{M_N}\int_0^T\left[\langle X,X\rangle'_u\langle Y,Y\rangle'_u+(\langle X,Y\rangle'_u)^2\right]dH(u), \quad (14)$$

where $H(u) = U(u) = \lim\sum_{\tau_{M_N,i}\le u}\frac{(\tau_{M_N,i}-\tau_{M_N,i-1})^2}{T/M_N}$ is the quadratic variation of time. A further specialization is when $X=Y$, (12) becomes

$$\frac{T}{M_N}\int_0^T\left[2(\langle X,X\rangle'_u)^2\right]dH(u), \quad (15)$$

both (14) and (15) are consistent with the results in Mykland and Zhang (2006).

Note that in (12), the relevant component in $Q_T[2]$ is the end-point of a martingale with quadratic variations as follows:

$$\langle Q_N[2], Q_N[2]\rangle = \frac{2}{3}\lim\sum_{i=1}^{M_N}(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^4 \tag{16}$$
$$\times\{(\langle X,X\rangle'_{t_{i-1}})^2(\langle Y,Y\rangle'_{t_{i-1}})^2+6(\langle X,X\rangle'_{t_{i-1}})(\langle Y,Y\rangle'_{t_{i-1}})(\langle X,Y\rangle'_{t_{i-1}})^2+(\langle X,Y\rangle'_{t_{i-1}})^4\}\times(1+o_p(1)).$$

When taking expectation (which is relevant when the trading times are random), the $Q_T[2]$ term yields zero, thus it disappears in the final expression for the variance.

4. The case when $M_N = o(N)$

We shall see in this section that under C3, the $1/M_N$ term (i.e. the discretization effect) in (12) is the sole leading term in the asymptotic variance of the previous-tick estimator. The source of the second-order term in the asymptotic variance depends on the exact order of M_N. An interesting case is when $M_N = O(N^{2/3})$. This choice is optimal in the sense of minimizing mean squared error of $[X,Y]_T$, when the stochastic bias of $[X,Y]_T$ is $O_p(M_N/N)$ (Theorem 1) and the stochastic variance is $O_p(1/M_N)$ (Theorem 2). We can see that in this scenario the $1/N$ term and the $1/M_N^{3/2}$ term in (12) share the second-order effects. We shall elaborate on the higher-order behaviors in this Section.

We emphasize that regardless of the order of M_N, the interaction between the discretization and the asynchronous effect is at most a third-order effect, with order $1/(M_N\sqrt{N})$ (see the proof of Theorem 2).

4.1. *First order behavior*

This is an immediate conclusion from Theorem 2:

Corollary 1. *Assume* C2-C3. *Then $U_u^{(dis)}$ exists and equals the scaled quadratic variation of the grid points $\mathcal{V}$. In the case of equispaced grid points, $U_u^{(dis)} = u$. The total variance term of the previous tick estimator is, to first order,*

$$\frac{T}{M_N}\int_0^T \left[\langle X,X\rangle'_u\langle Y,Y\rangle'_u + (\langle X,Y\rangle'_u)^2\right] dU_u^{(dis)} \tag{17}$$

Corollary 1 says that when the data (X,Y) arrive faster than the sampling frequency, the asynchronization effect disappear and only the discretization effect $U^{(dis)}$ remains in the variance term.

We can also assert something about the asymptotic distribution of the estimator. Let L_N be the discretization term in (10). Then, in view of Theorem 1 and Lemma 1 (in the Appendix),

$$[X,Y]_T - \langle X,Y\rangle_T = -\int_0^T \langle X,Y\rangle'_u dF_N(u) + L_N + O_p(N^{-1/2}) \tag{18}$$

The quantity in (17) is simply the asymptotic version of $\langle L_N, L_N\rangle_T$. By extending the arguments above to all time points $t\in[0,T]$ and using the theory in Chapters VI and IX or Jacod and Shiryaev (2003), we thus obtain

Theorem 3. *Assume the conditions of Theorem 1. Under conditions C2-C3,* $M_N^{1/2} L_N$ *converges in law to*

$$Z\left(T\int_0^T \left[\langle X,X\rangle'_u \langle Y,Y\rangle'_u + (\langle X,Y\rangle'_u)^2\right] dU_u^{(dis)}\right)^{1/2}, \tag{19}$$

where Z *is standard normal, and independent of* X *and* Y.

Note that the convergence is stable, in the sense of Rényi (1963), Aldous and Eagleson (1978), Chapter 3 (p. 56) of Hall and Heyde (1980), Rootzén (1980) and Section 2 (p. 169-170) of Jacod and Protter (1998). For the connection to this type of high frequency data problem, see Zhang, Mykland, and Aït-Sahalia (2005) and Zhang (2006).

By using the same methods, we relate the estimator to the hypothetical unobserved "gold standard" (9):

Theorem 4. *Assume the conditions of Theorem 1. Under conditions C2-C3,*

$$[X,Y]_T = \widetilde{[X,Y]}_T - \int_0^T \langle X,Y\rangle'_u dF_N(u) + O_p(N^{-1/2}) \tag{20}$$

The result also holds if $\widetilde{[X,Y]}_T$ *is defined with* $w_i = \max(t_i, s_i)$ *replacing* v_i.

One can, in fact, deduce Theorem 3 from this result using the standard theorems for synchronous observation in Barndorff-Nielsen and Shephard (2002a), Jacod and Protter (1998), Mykland and Zhang (2006) and Zhang (2001).

4.2. Higher order behavior

We can also say something about the higher order terms in the variance. First the non-martingale part.

Corollary 2. *Assume* C2-C3. *In addition to the conclusions of Corollary 1, we also have that*

$$U_{N,u}^{(nonsyn)} = 2\frac{N}{M_N}F_N(u) + o(1) \tag{21}$$

If we for the moment ignore the martingale $Q_N[2]$, we can therefore assert that the effect of nonsynchronization is to high order fully characterized by the function $F_N(u)$, since this is the quantity one encounters in both the bias, the $U_{N,u}^{(dis)}$ and $U_{N,u}^{(nonsyn)}$ terms.

Theorem 2 put together with Corollary 2 then yields

$$\begin{aligned}\text{stochastic variance } &= \frac{T}{M_N}\int_0^T \left[\langle X,X\rangle'_u \langle Y,Y\rangle'_u + (\langle X,Y\rangle'_u)^2\right] du - 2\frac{T}{M_N}\int_0^T (\langle X,Y\rangle'_u)^2 dF_N(u) \\ &+ Q_N[2] + o_p(N^{-1}) + o_p(M_N^{-3/2}) + o_p(M_N/N^2)\end{aligned} \tag{22}$$

Putting this in turn together with Theorem 1, one obtains that the stochastic MSE (bias^2 + variance) of $[X,Y]_T - \langle X,Y\rangle_T$ is

$$\begin{aligned}\text{stoch MSE } &= \left(\int_0^T \langle X,Y\rangle'_u dF_N(u)\right)^2 + \frac{T}{M_N}\int_0^T \left[\langle X,X\rangle'_u \langle Y,Y\rangle'_u + (\langle X,Y\rangle'_u)^2\right] du \\ &- 2\frac{T}{M_N}\int_0^T (\langle X,Y\rangle'_u)^2 dF_N(u) + Q_N[2] + o_p(N^{-1}) + o_p(M_N^{-3/2}) + o_p(M_N/N^2)\end{aligned} \tag{23}$$

Recall that $Q_N[2] = O_p(M_N^{-3/2})$. What about the term due to $Q_N[2]$? First order answers can be provided by considering the case when the quadratic variations $\langle X, X\rangle$, $\langle Y, Y\rangle$ and $\langle X, Y\rangle$ are nonrandom. In this case, by taking the expected MSE, the martingale term Q_N disappears. One can behave as if the MSE is the first three elements of (23). We return to the question of the meaning of $Q_N[2]$ in later Section 4.4.

4.3. Bias-Variance Tradeoff

In view of Theorem 1, "typical" behavior is that

$$\frac{N}{M_N} F_N(u) \to F(u) \text{ as } N \to \infty, \tag{24}$$

where F is a nondecreasing function. (In particular, every subsequence will have a further subsequence displaying this behavior). In this case, we obtain

$$\begin{aligned} \text{stoch MSE } &= \left(\frac{M_N}{N}\right)^2 \left(\int_0^T \langle X, Y\rangle'_u dF(u)\right)^2 + \frac{T}{M_N}\int_0^T \left[\langle X, X\rangle'_u \langle Y, Y\rangle'_u + (\langle X, Y\rangle'_u)^2\right] du \\ &- 2\frac{T}{N}\int_0^T (\langle X, Y\rangle'_u)^2 dF(u) + Q_N[2] + o_p(N^{-1}) + o_p(M_N^{-3/2}) + o_p(M_N^2/N^2) \end{aligned} \tag{25}$$

A tradeoff between bias^2 (the first term) and variance (the later terms) is therefore obtained by setting $(M_N/N)^2 = O(M_N^{-1})$, yielding $M_N = O(N^{2/3})$. Thus the first order terms in the MSE is given by the two first terms in (23) or (25).

4.4. The meaning of the martingale Q_N

Let K_N be the martingale (non-drift) term in (10). In other words, $[X, Y]_T - \langle X, Y\rangle_T = -\int_0^T \langle X, Y\rangle u' dF_N(u) + K_N + O_p(N^{-1})$. By the same methods as before, we obtain.

Corollary 3. *Assume conditions C2-C3. Then, in probability,*

$$\begin{aligned} &M_N^3 \langle Q_N[2], Q_N[2]\rangle \to \\ &\frac{2}{3}T^3 \int_0^T \{(\langle X, X\rangle'_u)^2 (\langle Y, Y\rangle'_u)^2 + 6(\langle X, X\rangle'_u)(\langle Y, Y\rangle'_u)(\langle X, Y\rangle'_u)^2 + (\langle X, Y\rangle'_u)^4\} du \end{aligned} \tag{26}$$

and

$$\langle M_N^{1/2} K_N, M_N^{3/2} Q_N[2]\rangle\rangle \to \frac{1}{3}T^2 \int_0^T \{5\langle X, X\rangle'_u \langle Y, Y\rangle'_u \langle X, Y\rangle'_u + (\langle X, Y\rangle'_u)^3\} du \tag{27}$$

In fact, the corollary asserts that $[X, Y]_T - \langle X, Y\rangle_T$ is correlated with its own stochastic variance! What could this possibly imply?

Again, to get a first order answer, by considering what would happen if the quadratic variations $\langle X, X\rangle$, $\langle Y, Y\rangle$ and $\langle X, Y\rangle$ are nonrandom. We then obtain that the third cumulant of K_N is given by

$$\begin{aligned} \text{cum}_3(K_N) &= 3\text{cov}(K_N, \langle K_N, K_N\rangle) \\ &= 3M_N^{-2}\text{cov}(M_N^{1/2} K_N, Q_N[2]) + o(M_N^{-2}) \\ &= 3M_N^{-2} E\langle M_N^{1/2} K_N, Q_N[2]\rangle + o(M_N^{-2}) \\ &= M_N^{-2} T^2 \int_0^T \{5\langle X, X\rangle'_u \langle Y, Y\rangle'_u \langle X, Y\rangle'_u + (\langle X, Y\rangle'_u)^3\} du \end{aligned} \tag{28}$$

(for the first transition, cf. equation (2.14) (p. 23) of Mykland (1994)). Similar methods can be used to compute the fourth cumulant.

Thus, the Q_N is more of a contribution to the Edgeworth expansions of our estimator, rather than an adjustment to variance.

5. When trading times are random

It is often natural to assume that the trading times τ and θ can be described as the event times of a counting process. Let the arrival times τs have intensity $\lambda^X(t)$ and the θs have intensity $\lambda^Y(t)$. For the moment we assume that both these intensities can be random (but predictable) processes.

This type of model requires some modification on the earlier development. For one thing, the counts m and n are random, so is $N = m + n$. Also, and more seriously, Conditions C1 and/or C2 may not be satisfied. We consider these issues in turn.

First of all, to get an asymptotic framework, we assume the following.

CONDITION C4: There is a sequence of experiments indexed by nonrandom α, $\alpha > 0$, so that $\lambda^X = \lambda^X_\alpha$ and $\lambda^Y = \lambda^Y_\alpha$. In general, the intensities can be any function of α, but we suppose that there are constants $\bar{c}$ and $\underline{c}$, independent of α, $0 < \underline{c} \leq 1 \leq \bar{c} < \infty$, so that for all $t \in [0, T]$,

$$\alpha\underline{c} \leq \lambda^X_\alpha(t) \leq \alpha\bar{c} \quad \text{and} \quad \alpha\underline{c} \leq \lambda^Y_\alpha(t) \leq \alpha\bar{c} \tag{29}$$

Remark 3. *(Asymptotic framework.) We do asymptotics as $\alpha \to \infty$. Note that since $N/\alpha = O_p(1)$ but not $o_p(1)$, this is the same as supposing that $N \to \infty$. The same argument yields the same orders for m and n.*

The assumption that will run into trouble is Condition C2. This is a natural assumption for developing analytical results when the trading/sampling times are nonrandom, but Condition C2 is neither true nor necessary if the sampling times are random. In fact, if the intensities λ^X_α and λ^Y_α are independent of time t, then conditionally on m and n, the sampling times for the X and Y processes are like the order statistics from a uniform distribution on $[0, T]$ (see, for example, Theorem 2.3.1 (p. 67) of Ross (1996)). Thus $\sup_i |\theta_{m,i} - \theta_{m,i-1}| = O(\log N/N)$, but not $O(1/N)$, and similarly for the τ's (see Devroye (1981, 1982), Aldous (1989), Shorack and Wellner (1986), for example). By the subsampling argument used in the proof of Theorem 5 below, this extends to all sampling schemes covered by condition C4.

Fortunately, this problem does not affect us in view of the upcoming Theorem 5. A restriction that ensures Condition C1 to be satisfied (eventually as $N \to \infty$) is sufficient. For this, we require that the size of the regular grid satisfies

$$M_\alpha = o_p(\alpha/\log\alpha). \tag{30}$$

Theorem 5. *Let X and Y be Itô processes satisfying (1)-(2), and let μ_t and σ_t be locally bounded. Let $[X, Y]_T$ be the previous-tick covariance estimator. Let $\mathcal{V}_\alpha = \{0 = v_0, v_1, \ldots, v_{M_\alpha}\}$ be a collection of nonrandom time points which span across $[0, T]$. Let t_i and s_i be the transaction times of X and Y, respectively, that immediately precede v_i. Assume Conditions C3, C4 and (30). Then the conclusions of Theorems 1, 2 and 4 remain valid (with M_α replacing M_N).*

6. Application: trading times follow a Poisson process

We now consider an application where the transaction times for assets X and Y follow two *independent* Poisson processes with (constant) intensities λ_α^X and λ_α^Y, respectively. The meaning of condition C4 is now simply that λ_α^X and λ_α^Y have the same order as index $\alpha \to \infty$.

6.1. Stochastic bias and variance in the case of Poisson arrivals

Corollary 4. *In the setting of Theorem 5, suppose that the consecutive sampling points v_i's are evenly spaced. Also, suppose that the transaction times for assets X and Y follow two independent Poisson processes with intensities λ_α^X and λ_α^Y (constant for each α), respectively. Also suppose that $\lambda_\alpha^X/\alpha \to \ell^X$ and $\lambda_\alpha^Y/\alpha \to \ell^Y$, as $\alpha \to \infty$. Then*

$$\frac{N}{M_\alpha} F_N(t) \to t \left(\frac{\ell^Y}{\ell^X} + \frac{\ell^X}{\ell^Y} \right) \tag{31}$$

in probability.

Since the stochastic bias is given by $SBias([X,Y]_T) = -\int_0^T \langle X,Y\rangle_t' dF_N(t) + O_p(N^{-1})$, we obtain that

$$\frac{N}{M_\alpha} SBias([X,Y]_T) \to -\langle X,Y\rangle_T \left(\frac{\ell^Y}{\ell^X} + \frac{\ell^X}{\ell^Y} \right). \tag{32}$$

It is obvious that the bias has opposite sign with the covariation between X and Y, and its magnitude reaches its minimum when $\ell^X = \ell^Y$ (for given value of $\ell = \ell^X + \ell^Y$).

We now move on to the asymptotics of stochastic variance in the case of Poisson processes. In analogy with the result in Section 4.2, we obtain:

Corollary 5. *In the setting of Corollary 4, then the asymptotic stochastic variance of the previous tick estimator becomes (leaving out the term that is due to Q_N in Theorem 2)*

$$\frac{T}{M_\alpha} \int_0^T (\langle X,Y\rangle_u')^2 + \langle X,X\rangle_u' \langle Y,Y\rangle_u' du - 2\frac{T}{N} \left(\frac{\ell^Y}{\ell^X} + \frac{\ell^X}{\ell^Y} \right) \int_0^T (\langle X,Y\rangle_u')^2 du \tag{33}$$

∎

The Q_N term is excluded for the reasons discussed in Sections 4.2 and 4.4.

6.2. Bias-variance tradeoff

Assuming that the observed (X,Y) are true (efficient) logarithmic prices, we have demonstrated that the previous-tick estimator has an asymptotically bounded bias. This bias is induced by asynchronous trading of two assets. Naturally the variance estimator in this case is unbiased as the price series is inherently synchronized with itself.

In analogy with the development in Section 4.3, we can now find an optimal sampling frequency. In this Poisson application, we can obtain very straightforward expressions.

Recall from Corollary 5 that the variance of the previous tick estimator consists of two terms, the $1/M_\alpha$ term and $1/N$ term. Under Condition (30), the latter is of smaller order. So the main terms in the mean squared error (MSE) are:

$$\left(\frac{M_\alpha}{N}\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})\right)^2+\frac{T}{M_\alpha}\int_0^T\left[(\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'\right]du. \tag{34}$$

Setting $\partial MSE/\partial M_\alpha=0$ gives $M_\alpha=O(N^{2/3})$. In particular,

$$M_\alpha^*=\left[\frac{T\int_0^T(\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'du}{2(\langle X,Y\rangle_T)^2(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})^2}\right]^{1/3}N^{2/3}.$$

With this choice in the sampling frequency, the MSE becomes

$$3(2^{-2/3})\left[\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})T\int_0^T\left((\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'\right)du\right]^{2/3}N^{-2/3}.$$

Note that our assumption is that M_α is nonrandom while N is random. We are using N for simplicity of notation only, to stand in for $(\lambda_\alpha^X+\lambda_\alpha^Y)T$. We can do this since $N/(\lambda_\alpha^X+\lambda_\alpha^Y)T\to 1$ in probability as $\alpha\to\infty$.

7. Effect of microstructure noise

Like many other applications, financial data usually are noisy. In the finance literature, this noise is commonly referred to as microstructure noise. One can also view microstructure noise as observation or measurement error caused by "imperfect trading".

A simple yet natural way to view high frequency transaction data is to use a hidden semimartingale argument. One model is to write the logarithmic price process of the observables as the sum of a latent process (say, efficient price), which follows a semimartingale model, and a microstructure noise process. That is,

$$X^o_{\tau_{n,i}}=X_{\tau_{n,i}}+\epsilon^X_{\tau_{n,i}}\text{ and }Y^o_{\theta_{m,i}}=Y_{\theta_{m,i}}+\epsilon^Y_{\theta_{m,i}} \tag{35}$$

where X^o and Y^o are the observed transaction prices in logarithmic scale, X and Y are the latent efficient (log) prices which satisfy the Itô-process models (1) and (2), respectively. Following the same notations as in Section 2.1, we suppose X^o is observed at grid $\mathcal{T}_n$, $\mathcal{T}_n=\{0=\tau_{n,0}\le\tau_{n,1}\le\ldots\le\tau_{n,n}=T\}$, suppose Y^o is sampled at grid $\mathcal{S}_m$, $\mathcal{S}_m=\{0=\theta_{m,0}\le\theta_{m,1}\le\ldots\le\theta_{m,m}=T\}$.

In the following, we present two approaches to handling microstructure noise. One is the classical bias-variance tradeoff. We then turn to two scales estimation in the next section.

It should be emphasized that the main recommendation is to use two- or multiscale estimation. The purpose of carrying out the trade-off below is mainly to show that the effect of microstructure can be integrated into the same scheme as the Epps effect, also for the purpose of sampling frequency.

7.1. Tradeoff between discretization, asynchronization, and microstructure noise

To demonstrate the idea without delving into the mathematical details, we let the noise be independent of the latent processes, that is, $\epsilon^X\perp\!\!\!\perp X$, $\epsilon^Y\perp\!\!\!\perp Y$, also ϵ^X and ϵ^Y are independent. A simple structure for the

ϵ's is white noise. We note that this model structure can be extended to incorporate the correlation structure between the latent prices and the noises, as well as that of the noises from two securities, but we shall not consider this here. (We shall use more relaxed assumptions in Section 8 below).

As was argued in Section 2 of Zhang, Mykland, and Aït-Sahalia (2005), to rigorously implement a bias-variance trade-off in the presence of microstructure noise, one needs to work with a shrinking noise asymptotics: $E(\epsilon^X)^2$ and $E(\epsilon^Y)^2$ will be taken to be of order $o(1)$ as $N \to \infty$. See also Zhang, Mykland, and Aït-Sahalia (2009). A similar approach was used in Delattre and Jacod (1997).

Similar to the definition and notations in (6), the previous-tick estimator for covariation now becomes the cross product of X^o and Y^o:

$$[X^o, Y^o]_T = \sum_{i=1}^{M_N} (X^o_{t_i} - X^o_{t_{i-1}})(Y^o_{s_i} - Y^o_{s_{i-1}}), \tag{36}$$

where the t_is and the s_is are the corresponding time ticks immediately preceding the sampling point v_i. (Because of the law of large numbers, we shall in this section identify M_α and M_N.)

Our question in this section is, given the observations X^o and Y^o at the nonsynchronized discrete grids and assuming the model (35), how close is $[X^o, Y^o]_T$ to the latent quantity $[X, Y]_T$? How well can $[X^o, Y^o]_T$ estimate the target $\langle X, Y\rangle_T$? We next study $[X^o, Y^o]_T - [X, Y]_T$, termed as the *error due to noise* .

7.1.1 Signal-noise decomposition

When the microstructure noise is present in the observed price processes, we can decompose the covariation estimator into those induced by the latent prices and those related to the noise. From (36), we get

$$[X^o, Y^o]_T = [X, Y]_T + [X, \epsilon^Y]_T[2] + [\epsilon^X, \epsilon^Y]_T,$$

where $[X, Y]_T$ is the same as (6), $[\epsilon^X, \epsilon^Y]_T = \sum_{i=1}^{M_N} (\epsilon^X_{t_i} - \epsilon^X_{t_{i-1}})(\epsilon^Y_{s_i} - \epsilon^Y_{s_{i-1}})$, and

$$[X, \epsilon^Y]_T[2] \quad = \quad \sum_{i=1}^{M_N} (X_{t_i} - X_{t_{i-1}})(\epsilon^Y_{s_i} - \epsilon^Y_{s_{i-1}}) + \sum_{i=1}^{M_N} (Y_{s_i} - Y_{s_{i-1}})(\epsilon^X_{t_i} - \epsilon^X_{t_{i-1}}).$$

We shall see that the main-order term in the above decomposition is from the noise covariation $[\epsilon^X, \epsilon^Y]$ and from the signal-noise interaction $[X, \epsilon^Y]_T[2]$. To see this, write

$$[X, \epsilon^Y]_T = \sum_{i=1}^{M_N} (X_{t_i} - X_{t_{i-1}})\epsilon^Y_{s_i} - \sum_{i=1}^{M_N} (X_{t_i} - X_{t_{i-1}})\epsilon^Y_{s_{i-1}}.$$

Because of the white-noise property in ϵ^Y, we obtain $E[([X, \epsilon^Y]_T)^2|X] = 2[X, X]_T E(\epsilon^Y)^2 + O_p(E(\epsilon^Y)^2 M_N^{-1})$ where the order $o_p(E(\epsilon^Y)^2)$ is from the cross term. To find the exact formula for the cross term, we refer to the method in Zhang, Mykland, and Aït-Sahalia (2005). So far we have $[X, \epsilon^Y]_T = O_p([E(\epsilon^Y)^2]^{1/2})$, similarly, $[Y, \epsilon^X]_T = O_p([E(\epsilon^X)^2]^{1/2})$.

For the noise variation, notice that

$$[\epsilon^X, \epsilon^Y]_T \quad = \quad 2\sum_{i=1}^{M_N} \epsilon^X_{t_i}\epsilon^Y_{s_i} - \sum_{i=1}^{M_N} \epsilon^X_{t_i}\epsilon^Y_{s_{i-1}}[2] + \epsilon^X_{t_0}\epsilon^Y_{s_0} - \epsilon^X_{t_M}\epsilon^Y_{s_M}, \tag{37}$$

where we recall that $\epsilon^X_{t_i}\epsilon^Y_{s_{i-1}}[2] = \epsilon^X_{t_i}\epsilon^Y_{s_{i-1}} + \epsilon^X_{t_{i-1}}\epsilon^Y_{s_i}$. Because ϵ^X and ϵ^Y are both white noise with mean zero, and uncorrelated with each other, we have

$$\mathrm{var}(\epsilon^X_{t_i}\epsilon^Y_{s_i}) = \mathrm{var}(\epsilon^X_{t_i}\epsilon^Y_{s_{i-1}}) = \mathrm{var}(\epsilon^X_{t_{i-1}}\epsilon^Y_{s_i}) = E(\epsilon^X)^2 E(\epsilon^Y)^2.$$

Hence, $\mathrm{var}([\epsilon^X,\epsilon^Y]_T) = 6M_N E(\epsilon^X)^2 E(\epsilon^Y)^2$. Therefore, the pure noise variation $[\epsilon^X,\epsilon^Y]_T$ has order $[M_N E(\epsilon^X)^2 E(\epsilon^Y)^2]^{1/2}$. In summary,

$$[X^o,Y^o]_T = \underbrace{[X,Y]_T}_{O_p(1)} + \underbrace{[X,\epsilon^Y]_T[2]}_{O_p([E(\epsilon^X)^2]^{1/2}) + O_p([E(\epsilon^Y)^2]^{1/2})} + \underbrace{[\epsilon^X,\epsilon^Y]_T}_{O_p([M_N E(\epsilon^X)^2 E(\epsilon^Y)^2]^{1/2})}.$$

7.1.2 The Tradeoff

We study the tradeoff when the observation times of X and of Y follow Poisson processes with intensities λ^X_α and λ^Y_α, respectively. As in section 6.2, we shall for ease of exposition identify N and $(\lambda^X_\alpha + \lambda^Y_\alpha)T$ (by the law of large numbers, these two quantities are interchangeable in our formulas in this section).

We note that the previous tick estimator is asymptotically unbiased, with order M_N/N. As far as its variance is concerned, the part due to asynchronicity and discretization decreases with sampling frequency (Theorem 2) whereas the part due to microstructure noise increases (Section 7.1.1). It would be desirable to balance the terms between the bias and the variance terms, in the sense that minimizes the MSE. We do so in the following. We let ν_N^2 represent the order of the variance of ϵ^X and ϵ^Y, so that $O(E(\epsilon^X)^2) = O(E(\epsilon^Y)^2) = O(\nu_N^2)$. For $\mu^X = \mu^Y = 0$, $[X,\epsilon^Y]_T[2]$ is the end point of a martingale and it has zero covariation with the pure noise. Then, the leading terms in MSE of $[X^o,Y^o]$ is,

$$\left(\frac{M_N}{N}\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})\right)^2 + \frac{T}{M_N}\int_0^T \left[(\langle X,Y\rangle'_u)^2 + \langle X,X\rangle'_u\langle Y,Y\rangle'_u\right] du$$
$$+2\langle X,X\rangle_T E(\epsilon^Y)^2 + 2\langle Y,Y\rangle_T E(\epsilon^X)^2 + 6M_N E(\epsilon^X)^2 E(\epsilon^Y)^2.$$

In order to capture all three effects consisting of microstructure noise, discretization variance and nonsynchronization bias, we have elected to let the size of the microstructure go to zero as $N\to\infty$ in such a way that the variance due to noise will have the same size as the discretization and nonsynchronization MSE. To this effect, we can select $M_N = O(N^{2/3})$, and $\nu_N^2 = O(N^{-2/3})$. To be specific, suppose that r, r_X and r_Y are nonnegative constants to that

$$M_N = rN^{2/3}, \quad E(\epsilon^X)^2 = r_X N^{-2/3}, \quad \text{and} \quad E(\epsilon^Y)^2 = r_Y N^{-2/3}.$$

Note that r, r_X and r_Y regulate the triangular array type of asymptotics described just before Section 7.1.1. Only r is assumed to be controllable by the econometrician; r_X and r_Y are given by nature.

Setting $\partial MSE/\partial M_N = 0$, we get

$$\frac{2}{N^2}\left[\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})\right]^2 M_N - \frac{T\int_0^T\left[(\langle X,Y\rangle'_u)^2 + \langle X,X\rangle'_u\langle Y,Y\rangle'_u\right] du}{M_N^2} + 6E(\epsilon^X)^2E(\epsilon^Y)^2 = 0,$$

yielding an optimal choice of r satisfying

$$r_X r_Y = \frac{1}{6}\left\{r^{-2}T\int_0^T\left[(\langle X,Y\rangle'_u)^2 + \langle X,X\rangle'_u\langle Y,Y\rangle'_u\right] du - 2r\left[\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})\right]^2\right\}.$$

(This equation uniquely defines r). Hence, the optimal mean squared error is

$$MSE = N^{-2/3}\left\{2r^{-1}T\int_0^T\left[(\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'\right]du - r^2\left[\langle X,Y\rangle_T(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y})\right]^2+2\langle X,X\rangle_T r_Y+2\langle Y,Y\rangle_T r_X\right\}.$$

8. Two Scales Estimation with Previous Tick Covariances

We here continue to look at the case (35) where there is microstructure noise on top of X and Y. In this section, we *do not* make the assumptions from Section 7.1 above. (In particular, there is no need to take noise to be "shrinking").

8.1. Definition and analysis of two scales estimation

First, the average lag K previous tick realized covariance is defined by

$$[X^o,Y^o]_T^{(K)}=\frac{1}{K}\sum_{i=K}^{M_N}(X^o_{t_i}-X^o_{t_{i-K}})(Y^o_{s_i}-Y^o_{s_{i-K}}).$$

In analogy with development in Zhang, Mykland, and Aït-Sahalia (2005), we can now define a previous tick two scales realized covariance (TSCV) by

$$\widehat{\langle X,Y\rangle}_T=c_N\left([X^o,Y^o]_T^{(K)}-\frac{\bar n_K}{\bar n_J}[X^o,Y^o]_T^{(J)}\right)$$

where c_N is a constant that can be tuned for small sample precision and for our purposes must satisfy that $c_N=1+o_p(M_N^{-1/6})$ (see, in particular, Section 4.2 of Zhang, Mykland, and Aït-Sahalia (2005)), and $\bar n_K=(M_N-K+1)/K$, and similarly for $\bar n_J$.

The two scales are chosen such that $1\le J << K$. Specifically, for the asymptotics, we require $K=K_N=O(N^{2/3})$. J can be fixed or go to infinity with N. In the classical two scales setting, $J=1$.

If we assume that the sequence $(\epsilon^X_{t_i},\epsilon^Y_{s_i})$ is independent of the latent X and Y processes, has bounded fourth moments, and is exponentially α-mixing, it follows from the proof of Lemma A.2 in Zhang, Mykland, and Aït-Sahalia (2005) that

$$\widehat{\langle X,Y\rangle}_T=\left([X,Y]_T^{(K)}-\frac{\bar n_K}{\bar n_J}[X,Y]_T^{(J)}\right)+\frac{1}{K}\left(\sum_{i=J}^{M_N}\epsilon^X_{t_i}\epsilon^Y_{s_{i-J}}[2]-\sum_{i=K}^{M_N}\epsilon^X_{t_i}\epsilon^Y_{s_{i-K}}[2]\right)+o_p(M_N^{-1/6}). \quad (38)$$

where "$\epsilon^X_{t_i}\epsilon^Y_{s_{i-J}}[2]$" means $\epsilon^X_{t_i}\epsilon^Y_{s_{i-J}}+\epsilon^Y_{s_i}\epsilon^X_{t_{i-J}}$ (see Notation 1 in Section 3).

We now use the results in this paper to analyze the first term in (38). If we assume conditions C1 and C2, and if also $J\to\infty$, it follows that condition C3 is satisfied for the subsamples, and so from (8),

$$[X,Y]_T^{(K)}-\frac{\bar n_K}{\bar n_J}[X,Y]_T^{(J)}=\widetilde{[X,Y]}_T^{(K)}-\frac{\bar n_K}{\bar n_J}\widetilde{[X,Y]}_T^{(J)}+o_p(M_N^{-1/6}).$$

where $\widetilde{[X,Y]}_T^{(K)}$ is the averaged subsampled version of the (unobserved) synchronized estimator $\widetilde{[X,Y]}_T$. This is because the bias cancels to the relevant order! Specifically, the bias in $[X,Y]_T^{(K)}$ is the expression in (7), multiplied by $1/K$, and similarly for $[X,Y]_T^{(J)}$. Hence, in the end,

$$\widehat{\langle X,Y\rangle}_T = \left(\widetilde{[X,Y]}_T^{(K)} - \frac{\bar{n}_K}{\bar{n}_J}\widetilde{[X,Y]}_T^{(J)}\right) + \frac{1}{K}\left(\sum_{i=J}^{M_N}\epsilon_{t_i}^X\epsilon_{s_{i-J}}^Y[2] - \sum_{i=K}^{M_N}\epsilon_{t_i}^X\epsilon_{s_{i-K}}^Y[2]\right) + o_p(M_N^{-1/6}). \quad (39)$$

It can be verified directly that this is also true when J is fixed as $n \to \infty$, using the more complicated Theorems 1-2.

Two things have ben achieved in this development. On the one hand, we have shown that in this case, previous tick estimation reduces, for purposes of analysis, to using synchronized observations. The other is that we do not need to be overly concerned with the precise dependence structure between $\epsilon_{t_i}^X$ and $\epsilon_{s_i}^Y$.

From (39), we can now obtain the asymptotic mixed normality of $M_N^{1/6}(\widehat{\langle X,Y\rangle}_T - \langle X,Y\rangle_T)$ by just recycling the results in Zhang, Mykland, and Aït-Sahalia (2005). We again stress that Condition C3 is not required on the original grid $\mathcal{V}$, so that one can take $M_N = O(N)$.

A concrete limit theorem would be as follows:

Theorem 6. *Assume that (X_t) and (Y_t) are Itô processes given by (1)-(2), with σ_t^X and σ_t^Y continuous, and μ_t^X and μ_t^Y locally bounded. Observables $X_{\tau_{n,i}}^o$ and $Y_{\theta_{m,i}}^o$ are given by (35), and the grid $\mathcal{V}$ satisfies (C1)-(C2), with $M_N/N \to c_1 > 0$ as $N \to \infty$. The scales $J = J_N$ and $K = K_N$ satisfy that $K_N/N^{2/3} \to c_2$ and $J_N/N^{2/3} \to 0$ as $N \to \infty$. Assume that $J = \lim_{N\to\infty} J_N$ is either infinity, or exists and is finite. Also assume that the noise processes are independent of (X_t) and (Y_t), and that the process $(\epsilon_{t_i}^X, \epsilon_{s_i}^Y)$ is stationary and exponentially α-mixing, with $E\epsilon^X = E\epsilon^Y = 0$. Also suppose that $\epsilon_{t_i}^X$ and $\epsilon_{s_i}^Y$ have finite $(4+\delta)^{\text{th}}$ moment for some $\delta > 0$. Finally, define h_i as in equation (43) in Zhang, Mykland, and Aït-Sahalia (2005), with v_i replacing t_i, and set $G_n(t) = \sum_{v_{i+1}\le t} h_i \Delta v_i$. Assume that G_n converges pointwise to G. Then: $N^{1/6}(\widehat{\langle X,Y\rangle}_T - \langle X,Y\rangle_T)$ converges stably in law to ωZ, where Z is standard normal (independent of X and Y), and*

$$\omega^2 = \frac{1}{2}c_1^{-1}c_2 T\int_0^T \left[\langle X,X\rangle_u'\langle Y,Y\rangle_u' + (\langle X,Y\rangle_u')^2\right] dG(t) + c_2^{-2}c_1\left[\gamma_{0,J} + \gamma_{0,\infty} + 2\sum_{i=1}^{\infty}(\gamma_{i,J} + \gamma_{i,\infty})\right],$$

where $\gamma_{i,j} = Cov(\epsilon_{t_0}^X\epsilon_{s_{-j}}^Y[2], \epsilon_{t_i}^X\epsilon_{s_{i-j}}^Y[2])$ (the precise form is rather tedious and is given in (52)) and $\gamma_{i,\infty} = \lim_{j\to\infty}\gamma_{i,j}$, given by

$$\gamma_{i,\infty} = 2Cov(\epsilon_{t_0}^X, \epsilon_{t_i}^X)Cov(\epsilon_{s_0}^Y, \epsilon_{s_i}^Y) + 2Cov(\epsilon_{t_0}^X, \epsilon_{s_i}^Y)Cov(\epsilon_{s_0}^Y, \epsilon_{t_i}^X). \quad (40)$$

Note that the functions $G_n(t)$ and $G(t)$ are exactly as in Zhang, Mykland, and Aït-Sahalia (2005), and as argued on p. 1411 in that paper, $G(t)$ always exists under Condition (C2), if necessary by using subsequences. If the v_is are equidistant, we have $G'(t) \equiv 4/3$.

The assumption of stationarity of subsequences is one of convenience, and is not required, at the cost of more complicated expressions for the asymptotic variance. The deeper result is (39), which is not dependent on stationarity.

8.2. An illustration of behavior in data

We here provide an instance of how the estimators behave in data. The daily covariance of Microsoft (MSFT) and Google (GOOG) were estimated based on the previous tick method from the transactions reported in the TAQ database. The grid points $\mathcal{V}$ were based on the refresh time method (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008b); see Section 2.1 above for a description). Figure 1-2 give the average of the daily estimates from the trading days of October 2005. In Figure 1, subsampling and averaging was used; Figure 2 is based on the two scales estimator.

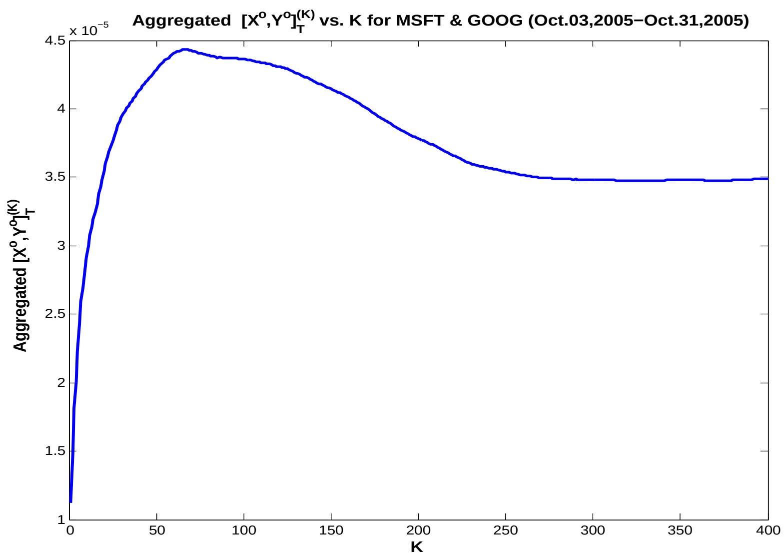

Figure 1: Average estimator $[X^o, Y^o]_T^{(K)}$ as a function of K.

From Figure 1, one can see that the Epps effect does kick in at the highest frequencies (at very small K, the estimator sharply drops from 4×10^{-5} to 1.1×10^{-5}), while at more moderately small K, there is an upward bias which is presumably due to microstructure. The Epps effect is substantially removed in the two scale covariance estimator. In Figure 2, the TSCV is stable around 4×10^{-5} for large enough K. Regardless of the choice in K and J, it looks like that TSCV fluctuates in a much narrower range (between 4×10^{-5} to 5.4×10^{-5}).

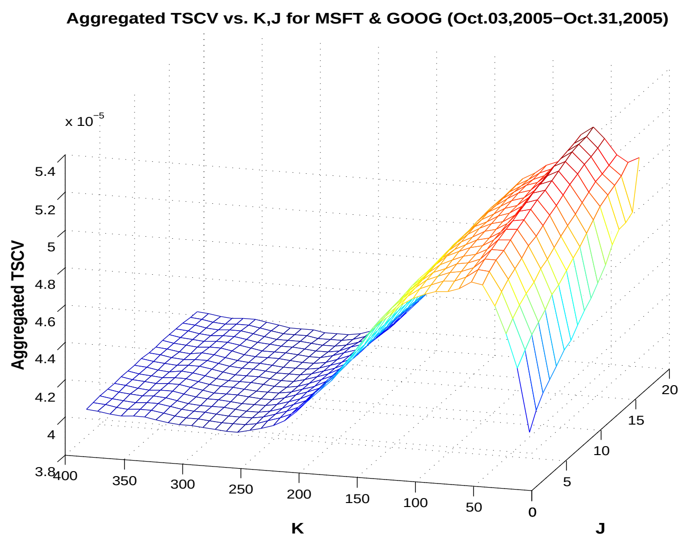

Figure 2: Two scales estimator $\widehat{\langle X, Y \rangle}_T$ as a function of J and K.

9. Conclusion: the Epps effect and its remedies

This paper is about how to estimate $\langle X, Y \rangle_T$ when the observation times of X and Y are not synchronized and when the microstructure noise is present in the observed price processes. Using the previous-tick estimator for $\langle X, Y \rangle_T$, we show in Theorem 1 that for positively associated assets X and Y, nonsynchronization induces a negative bias in the estimator. The magnitude of this bias increases in sampling frequency, up to a point; On the other hand, it decreases for more liquid assets. This is an analytic characterization of the Epps effect (Epps (1979)).

To cope with this effect, the paper offers two approaches. On the one hand, the effect can be controlled through a bias-variance trade-off. This trade-off provides a optimal scheme for subsampling observations. The scheme can incorporate microstructure noise.

A more satisfying approach is two- or multiscale estimation. Section 8 shows that this approach eliminates, *at the same time*, the biases due to asynchronicity and microstructure noise. The rate of convergence is the same as that achieved in the scalar process case, where there is no asynchronicity.

The principles outlined can be applied similarly to multi scale estimation (Zhang (2006)), thus achieving rate efficiency. A full development of this approach is deferred to later work.

10. APPENDIX: PROOFS

PROOF OF THEOREM 1:

Assume first that $\mu^X = 0$ and $\mu^Y = 0$. We know that the stochastic bias of $[X,Y]_T$ is

$$\sum_{i=1}^{M_N} \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} \langle X,Y\rangle'_u du - \langle X,Y\rangle_T \quad = \quad \int_0^T \langle X,Y\rangle'_u d[G_N(u)-u],$$

where

$$\begin{aligned} G_N(t) &= \int_0^t \sum_{i=1}^{M_N} I(\max(t_{i-1},s_{i-1}) < v < \min(t_i,s_i))dv \\ &= \sum_{i:\min(t_i,s_i)\le t} (\min(t_i,s_i) - \max(t_{i-1},s_{i-1})) + (t-\underline{t}) \\ &= \sum_{i:\min(t_i,s_i)\le t} (\min(t_i,s_i) - \max(t_{i-1},s_{i-1})) + O_p(\frac{1}{N}) \end{aligned}$$

where $\underline{t} = \max\{l \in \mathcal{T}\cup\mathcal{S} : l \le t\}$. Hence (11) follows since

$$\begin{aligned} G_N(t) - t &= -\max(t_1,s_1) - \sum_{i:\max(t_i,s_i)\le t} (\max(t_i,s_i) - \min(t_i,s_i)), \\ &= -\sum_{i:\max(t_i,s_i)\le t} (\max(t_i,s_i) - \min(t_i,s_i)) + O_p(\frac{1}{N}). \end{aligned}$$

To see why $\frac{N}{M_N}\int_0^T \langle X,Y\rangle'_t dF_N(t)$ is RCP, note that under C2, $0 \le v_i - t_i \le \inf\{\tau > v_i : \tau \in \mathcal{T}_n\} - t_i \le \frac{c_1}{N}$ for some positive constant c_1. Similarly, $v_i - s_i \le \frac{c_2}{N}$ for some positive constant c_2.

Set $0 \le \delta_i^t = N(v_i - t_i) \le c$ and set $0 \le \delta_i^s = N(v_i - s_i) \le c$ for some c. Then,

$$\begin{aligned} \frac{N}{M_N}F_N(t) &= \frac{1}{M_N} \sum_{i:\max(t_i,s_i)\le t} [\max(\delta_i^t,\delta_i^s) - \min(\delta_i^t,\delta_i^s)] \\ &= \frac{1}{M_N}\sum_{i:v_i\le t}[\max(\delta_i^t,\delta_i^s) - \min(\delta_i^t,\delta_i^s)] + o_p(1) = O_p(1). \end{aligned}$$

Hence, $\frac{N}{M_N}F_N(t)$ is RCP by Helly's Theorem (Ash (1972), p 329). The same result for the stochastic bias follows since $\langle X,Y\rangle'_t$ is continuous.

If we do not assume that $\mu^X = 0$ and $\mu^Y = 0$, it is easy to see that the contribution to the bias from such terms is asymptotically negligible. To see this, we refer to Girsanov's Theorem and the device used at the beginning of the proof of Theorem 2 in Zhang, Mykland, and Aït-Sahalia (2005) (Section A.3, p. 1410). This works unless the instantaneous correlation between X and Y is one. In this latter case, one should use the methods of Mykland and Zhang (2006). ∎

Lemma 1. *Let X and Y be Itô processes satisfying (1)-(2). Let v_i, t_i, and s_i be the i-th sampling point, and the previous ticks in X and in Y, respectively, as defined in Section 2.1. Let $R_N = \sum_i (R_{1,i}, R_{2,i}$ and $R_{3,i})$,* *where*

$$\begin{aligned} R_{1.i} &= (X_{t_i} - X_{\min(t_i,s_i)})(Y_{s_i} - Y_{s_{i-1}}) \\ R_{2,i} &= (X_{\max(t_{i-1},s_{i-1})} - X_{t_{i-1}})(Y_{s_i} - Y_{s_{i-1}}) \\ R_{3,i} &= (X_{\min(t_i,s_i)} - X_{\max(t_{i-1},s_{i-1})})[(Y_{s_i} - Y_{\min(t_i,s_i)}) + (Y_{\max(t_{i-1},s_{i-1})} - Y_{s_{i-1}})] \end{aligned}$$

Then, under Conditions C1 *and* C2, $U^{(nonsyn)}_{N,u}$ *is RCP in the sense of Definition 3, and*

$$R_N = O_p(\frac{1}{\sqrt{N}}), \tag{41}$$

in particular, its quadratic variation

$$\langle R_N, R_N\rangle = \frac{T}{N}\int_0^T \langle Y,Y\rangle'_u\langle X,X\rangle'_u dU^{(nonsyn)}_u + o_p(\frac{1}{N}), \tag{42}$$

through any subsequence for which $U^{(nonsyn)}_u$ *(the limit of* $U^{(nonsyn)}_{N,u}$*) exists.* □

PROOF OF LEMMA 1:
$U^{(nonsyn)}_{N,u}$ is RCP by the same methods as in the proof of Theorem 1.
Note that the leading terms in $R_{1,i}$ -$R_{3,i}$ are martingale increments, with order root N. This is because

$$\begin{aligned} \langle\sum_i R_{1,i}, \sum_i R_{1,i}\rangle &= \sum_i (Y_{s_i} - Y_{s_{i-1}})^2 \int_{\min(s_i,t_i)}^{t_i} d\langle X,X\rangle_u \\ &= \sum_i (\langle Y,Y\rangle_{s_i} - \langle Y,Y\rangle_{s_{i-1}}) \int_{\min(s_i,t_i)}^{t_i} d\langle X,X\rangle_u + o_p(\frac{1}{N}) \\ &= \sum_i \langle Y,Y\rangle'_{s_i}\langle X,X\rangle'_{s_i}(s_i - s_{i-1})(t_i - \min(s_i,t_i)) + o_p(\frac{1}{N}). \end{aligned}$$

Similarly,

$$\langle\sum_i R_{2,i}, \sum_i R_{2,i}\rangle = \sum_i \langle Y,Y\rangle'_{s_{i-1}}\langle X,X\rangle'_{s_{i-1}}(\max(t_{i-1},s_{i-1}) - t_{i-1})(s_i - s_{i-1}) + o_p(\frac{1}{N}).$$

At last, for $\sum_i R_{3,i}$:

$$\begin{aligned} \langle R_{3,i}, R_{3,i}\rangle &= \sum_i \int_{\min(t_i,s_i)}^{s_i} (X_{\min(t_i,s_i)} - X_{\max(t_{i-1},s_{i-1})})^2 d\langle Y,Y\rangle_u \\ &+ \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (Y_{\max(t_{i-1},s_{i-1})} - Y_{s_{i-1}})^2 d\langle X,X\rangle_u \\ &= \sum_i \langle X,X\rangle'_{t_i}\langle Y,Y\rangle'_{t_i}(\min(t_i,s_i) - \max(t_{i-1},s_{i-1}))[(s_i - \min(s_i,t_i)) + (\max(t_{i-1},s_{i-1}) - s_{i-1})] \\ &+ o_p(\frac{1}{N}) \end{aligned}$$

The first transition in above is because

$$\langle\sum_i\int_{\min(t_i,s_i)}^{s_i}(X_{\min(t_i,s_i)}-X_{\max(t_{i-1},s_{i-1})})dY_u,\sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(Y_{\max(t_{i-1},s_{i-1})}-Y_{s_{i-1}})dX_u\rangle=0.$$

Now, left to compute that covariations between the different terms. As it turns out, the covariations are either zero or of order $o_p(\frac{1}{N})$. In particular, it follows directly from the definitions of $R_{1,i}$, $R_{2,i}$ and $R_{3,i}$ that on the one hand

$$\langle\sum_i R_{1,i},\sum_i R_{2,i}\rangle=0 \text{ and } \langle\sum_i R_{1,i},\sum_i R_{3,i}\rangle=0,$$

while on the other hand,

$$\begin{aligned}\langle\sum_i R_{2,i},\sum_i R_{3,i}\rangle &= \sum_i\int_{\min(t_i,s_i)}^{s_i}(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}})(X_{\min(t_i,s_i)}-X_{\max(t_{i-1},s_{i-1})})d\langle Y,Y\rangle_u\\ &= \sum_i\langle Y,Y\rangle'_{t_i}(s_i-\min(t_i,s_i))\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}})dX_u+o_p(\frac{1}{N}),\end{aligned}$$

whose main order has a quadratic variation

$$\begin{aligned}&\sum_i(\langle Y,Y\rangle'_{t_i})^2(s_i-\min(t_i,s_i))^2\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}})^2d\langle X,X\rangle_u\\ =&\sum_i(s_i-\min(t_i,s_i))^2(\langle Y,Y\rangle'_{t_i})^2(\langle X,X\rangle'_{t_i})^2(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))(\max(t_{i-1},s_{i-1})-t_{i-1})+o_p(\frac{1}{N})\\ =&\ o_p(\frac{1}{N})\end{aligned}$$

the order in the final step is because: under Conditions C1 and C2, $0\le v_i-s_i\le\inf\{\tau>v_i:\tau\in\mathcal{T}_n\}-s_i\le c_1/N$ for some c_1, and $v_i-t_i\le c_2/N$ for some c_2. Hence, $|s_i-\min(t_i,s_i)|=O(1/N)$, $|\max(t_{i-1},s_{i-1})-t_{i-1}|=O(1/N)$, and $|\min(t_i,s_i)-\max(t_{i-1},s_{i-1})|\le|\min(t_i,s_i)-v_i|+|v_i-v_{i-1}|+|v_{i-1}-\max(t_{i-1},s_{i-1})|=O(1/M)$. Therefore, $\langle\sum_i R_{2,i},\sum_i R_{3,i}\rangle=O_p(N^{-3/2})=o_p(N^{-1})$.

Put together the above results, we have

$$\begin{aligned}&\langle\sum_i(R_{1,i}+R_{2,i}+R_{3,i}),\sum_i(R_{1,i}+R_{2,i}+R_{3,i})\rangle=\sum_i(\langle R_{1,i},R_{1,i}\rangle+\langle R_{2,i},R_{2,i}\rangle+\langle R_{3,i},R_{3,i}\rangle)+o_p(\frac{1}{N})\\ =&\sum_i\langle X,X\rangle'_{t_i}\langle Y,Y\rangle'_{t_i}\left[(s_i-s_{i-1})(t_i-t_{i-1})-(\max(t_{i-1},s_{i-1})-\min(t_i,s_i))^2\right]+o_p(\frac{1}{N}),\end{aligned}$$

taking the limit (for convergent subsequences), Lemma 1 follows from the theory in Chapter VI in Jacod and Shiryaev (2003). ∎

PROOF OF THEOREM 2:

Note that

$$[X,Y]_T = \sum_i(X_{\min(t_i,s_i)}-X_{\max(t_{i-1},s_{i-1})})(Y_{\min(t_i,s_i)}-Y_{\max(t_{i-1},s_{i-1})})+\sum_i(R_{1,i}+R_{2,i}+R_{3,i}).$$

Invoking Itô's formula on the first term, we obtain,

$$\begin{aligned}
& \sum_i (X_{\min(t_i,s_i)} - X_{\max(t_{i-1},s_{i-1})})(Y_{\min(t_i,s_i)} - Y_{\max(t_{i-1},s_{i-1})}) \\
= & \sum_i (\langle X,Y\rangle_{\min(t_i,s_i)} - \langle X,Y\rangle_{\max(t_{i-1},s_{i-1})}) + \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2] \\
= & \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} \langle X,Y\rangle'_u du + \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2].
\end{aligned}$$

The asymptotic variance of $[X,Y]_T$ has two components, one – $\sum_i (R_{1,i} + R_{2,i} + R_{3,i})$ – comes from the nonsynchronization, while the other – $\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2]$ – is because of the discrete trading (or recording) time. The former is analyzed in Lemma 1. Now we are left to show the result in the latter term and the interaction between the two terms.

We start with the quadratic variation of $\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2]$.

$$\begin{aligned}
& \langle \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2], \quad \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2] \rangle \\
= & \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})^2 d\langle Y,Y\rangle_u \\
& + \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (Y_u - Y_{\max(t_{i-1},s_{i-1})})^2 d\langle X,X\rangle_u \\
& + 2 \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})(Y_u - Y_{\max(t_{i-1},s_{i-1})}) d\langle X,Y\rangle_u \\
= & \sum_i \langle X,X\rangle'_{t_i} \langle Y,Y\rangle'_{t_i} (\min(t_i,s_i) - \max(t_{i-1},s_{i-1}))^2 \\
& + \sum_i (\langle X,Y\rangle'_{t_i})^2 (\min(t_i,s_i) - \max(t_{i-1},s_{i-1}))^2 + Q_{N,T}[2] + o_p(\frac{1}{M_N^{3/2}}),
\end{aligned}$$

where

$$\begin{aligned}
Q_{N,T}[2] = & \; 2 \sum_i \langle X,X\rangle'_{t_i} \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (\min(t_i,s_i) - u)(Y_u - Y_{\max(t_{i-1},s_{i-1})}) dY_u[2] \\
& + 2 \sum_i \langle X,Y\rangle'_{t_i} \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (\min(t_i,s_i) - u)(X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u[2].
\end{aligned}$$

By the results and the methods in Lemma 2 (with $\alpha_i = \max(t_{i-1},s_{i-1})$ and $\beta_i = \min(t_i,s_i)$), we obtain that the quadratic variation of $Q_N[2]$ is as follows:

$$\begin{aligned}
& \langle Q_N[2], Q_N[2] \rangle \\
= & \frac{2}{3} \sum_{i=1}^{M_N} (\min(t_i,s_i) - \max(t_{i-1},s_{i-1}))^4 \\
& \times \left\{ (\langle X,X\rangle'_{t_{i-1}})^2 (\langle Y,Y\rangle'_{t_{i-1}})^2 + 6(\langle X,X\rangle'_{t_{i-1}})(\langle Y,Y\rangle'_{t_{i-1}})(\langle X,Y\rangle'_{t_{i-1}})^2 + (\langle X,Y\rangle'_{t_{i-1}})^4 \right\} \times (1 + o_p(1)).
\end{aligned}$$

Under Condition C2, we know that $Q_{N,T}[2] = O_p(\frac{1}{M_N^{3/2}})$.

By continuity, and using (13)

$$\begin{aligned} &\sum_i \langle X,X\rangle'_{t_i}\langle Y,Y\rangle'_{t_i}(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^2+\sum_i\left(\langle X,Y\rangle'_{t_i}\right)^2(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^2 \\ \sim\ & \frac{T}{M}\int_0^T\left[\langle X,X\rangle'_u\langle Y,Y\rangle'_u+(\langle X,Y\rangle'_u)^2\right]dU^{(dis)}_{N,u}. \end{aligned}$$

For a rigorous proof of similar statements under lesser regularity conditions, see Proposition 1 and Proposition 3 in Mykland and Zhang (2006).

To see equation (13), let δ_i^s and δ_i^t be as defined in the proof of Theorem 1, and set $\Delta v = T/M_N$. We then get that

$$\min(t_i,s_i)-\max(t_{i-1},s_{i-1}) = \Delta v - N^{-1}(\max(\delta_i^t,\delta_i^s)-\min(\delta_{i-1}^t,\delta_{i-1}^s)).$$

Hence,

$$\begin{aligned} U^{(dis)}_{N,u} &= \frac{M_N}{T}\sum_{i:t_i,s_i\le u}\left((\Delta v)^2-2\frac{\Delta v}{N}(\max(\delta_i^t,\delta_i^s)-\min(\delta_{i-1}^t,\delta_{i-1}^s))+O(N^{-2})\right) \\ &= u-2F_N(u)+O(M_N^{-1})+O((M_N/N)^2), \end{aligned}$$

proving (13).

Next we study the interaction term. Since $\sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u-X_{\max(t_{i-1},s_{i-1})})dY_u[2]$ is symmetric, we only need to show the interaction between $\sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u-X_{\max(t_{i-1},s_{i-1})})dY_u$ and $\sum_i(R_{1,i}+R_{2,i}+R_{3,i})$. First, it is obvious that

$$\langle\sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u-X_{\max(t_{i-1},s_{i-1})})dY_u,\sum_i R_{1,i}\rangle = 0.$$

For the rest, we have

$$\begin{aligned} &\langle\sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u-X_{\max(t_{i-1},s_{i-1})})dY_u,\sum_i R_{2,i}\rangle \\ =\ & \sum_i\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u-X_{\max(t_{i-1},s_{i-1})})(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}})d\langle Y,Y\rangle_u \\ =\ & \sum_i(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}})\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}\langle Y,Y\rangle'_{t_{i-1}}(\min(t_i,s_i)-u)dX_u\times(1+o_p(1)), \end{aligned}$$

which has a quadratic variation

$$\begin{aligned} &\sum_i\left(X_{\max(t_{i-1},s_{i-1})}-X_{t_{i-1}}\right)^2\left(\langle Y,Y\rangle'_{t_{i-1}}\right)^2\int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(\min(t_i,s_i)-u)^2d\langle X,X\rangle_u\times(1+o_p(1)) \\ \le\ & \frac{1}{3}\sup_t\left(\langle Y,Y\rangle'_t\right)^2\sup_t\left(\langle X,X\rangle'_t\right)^2\sum_i(\max(t_{i-1},s_{i-1})-t_{i-1})(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^3\times(1+o_p(1)) \\ =\ & O_p(\frac{1}{NM^2}). \end{aligned}$$

Hence

$$\langle \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u, \sum_i R_{2,i} \rangle = O_p(\frac{1}{\sqrt{N}M}).$$

By same method,

$$\langle \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})}) dY_u, \sum_i R_{3,i} \rangle = O_p(\frac{1}{\sqrt{N}M}).$$

In sum, the interaction term is negligible. Theorem 2 is proved. ∎

For next Lemma, we use the notation $\int X dY[2] = \int X dY + \int Y dX$.

Lemma 2. *Let X and Y be Itô processes satisfying (1)-(2). Let N and M_N be as defined in Section 2.1. Denote*

$$Q_N = 2 \sum_{i=1}^{M_N} \langle Y, Y \rangle'_{\alpha_i} \int_{\alpha_i}^{\beta_i} (\beta_i - u)(X_u - X_{\alpha_i}) dX_u,$$

and

$$R_N = \sum_{i=1}^{M_N} \langle X, Y \rangle'_{\alpha_i} \int_{\alpha_i}^{\beta_i} (\beta_i - u)(X_u - X_{\alpha_i}) dY_u[2].$$

Assuming Condition C1*, then*

$$\sum_{i=1}^{M_N} \int_{\alpha_i}^{\beta_i} (X_u - X_{\alpha_i})^2 d\langle Y, Y \rangle_u = \frac{1}{2} \sum_{i=1}^{M_N} \langle X, X \rangle'_{\alpha_i} \langle Y, Y \rangle'_{\alpha_i} (\beta_i - \alpha_i)^2 + Q_N + o_p(\frac{1}{M_N^{3/2}}),$$

where Q_N has quadratic variation

$$\frac{1}{3} \sum_{i=1}^{M_N} (\langle X, X \rangle'_{\alpha_i})^2 (\langle Y, Y \rangle'_{\alpha_i})^2 (\beta_i - \alpha_i)^4 \times (1 + o_p(1)).$$

Similarly,

$$\sum_{i=1}^{M_N} \int_{\alpha_i}^{\beta_i} (X_u - X_{\alpha_i})(Y_u - Y_{\alpha_i}) d\langle X, Y \rangle_u = \frac{1}{2} \sum_{i=1}^{M_N} (\langle X, Y \rangle'_{\alpha_i})^2 (\beta_i - \alpha_i)^2 + R_N + o_p(\frac{1}{M_N^{3/2}}),$$

where R_N has quadratic variation

$$\frac{1}{6} \sum_{i=1}^{M_N} (\langle X, Y \rangle'_{\alpha_i})^2 (\langle X, X \rangle'_{\alpha_i})(\langle Y, Y \rangle'_{\alpha_i})(\beta_i - \alpha_i)^4 \times (1 + o_p(1)).$$

∎

PROOF OF LEMMA 2:

We first show that

$$\langle \int_{\alpha_i}^{\beta_i} \int_{\alpha_i}^{u} (X_v - X_{\alpha_i}) dX_v du, \int_{\alpha_i}^{\beta_i} \int_{\alpha_i}^{u} (X_v - X_{\alpha_i}) dX_v du \rangle = \frac{1}{12} (\langle X, X \rangle'_{\alpha_i})^2 (\beta_i - \alpha_i)^4 \times (1 + o_p(1)). \quad (43)$$

Use integration by parts on the outer integration, we get

$$\begin{aligned}\int_{\alpha_i}^{\beta_i}\int_{\alpha_i}^{u}(X_v - X_{\alpha_i})dX_v du &= (\beta_i-\alpha_i)\int_{\alpha_i}^{\beta_i}(X_u - X_{\alpha_i})dX_u - \int_{\alpha_i}^{\beta_i}(u-\alpha_i)(X_u - X_{\alpha_i})dX_u \\ &= \int_{\alpha_i}^{\beta_i}(\beta_i - u)(X_u - X_{\alpha_i})dX_u,\end{aligned}$$

which has quadratic variation

$$\begin{aligned}\int_{\alpha_i}^{\beta_i}(\beta_i-u)^2(X_u - X_{\alpha_i})^2 d\langle X,X\rangle_u &= \int_{\alpha_i}^{\beta_i}(\beta_i-u)^2(\langle X,X\rangle_u - \langle X,X\rangle_{\alpha_i})d\langle X,X\rangle_u \times (1+o_p(1)) \\ &= (\langle X,X\rangle'_{\alpha_i})^2\int_{\alpha_i}^{\beta_i}(\beta_i-u)^2(u-\alpha_i)du \times (1+o_p(1)) \\ &= \frac{1}{12}(\langle X,X\rangle'_{\alpha_i})^2(\beta_i-\alpha_i)^4 \times (1+o_p(1))\end{aligned}$$

Next, invoke Itô's formula,

$$\begin{aligned}&\sum_{i=1}^{M_N}\int_{\alpha_i}^{\beta_i}(X_u - X_{\alpha_i})^2 d\langle Y,Y\rangle_u \\ =\ &\sum_{i=1}^{M_N}\langle Y,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}(X_u - X_{\alpha_i})^2 du + o_p(M_N^{-3/2}) \\ =\ &\sum_{i=1}^{M_N}\langle Y,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}(\langle X,X\rangle_u - \langle X,X\rangle_{\alpha_i})du + 2\sum_{i=1}^{M_N}\langle Y,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}\int_{\alpha_i}^{u}(X_v - X_{\alpha_i})dX_v du + o_p(M_N^{-3/2}) \\ =\ &\frac{1}{2}\sum_{i=1}^{M_N}\langle X,X\rangle'_{\alpha_i}\langle Y,Y\rangle'_{\alpha_i}(\beta_i-\alpha_i)^2 + Q_N + o_p(M_N^{-3/2}) \qquad \text{by (43)}\end{aligned}$$

Similarly,

$$\begin{aligned}&\sum_{i=1}^{M_N}\int_{\alpha_i}^{\beta_i}(X_u - X_{\alpha_i})(Y_u - Y_{\alpha_i})d\langle X,Y\rangle_u \\ =\ &\sum_{i=1}^{M_N}\langle X,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}(X_u - X_{\alpha_i})(Y_u - Y_{\alpha_i})du + o_p(M_N^{-3/2}) \\ =\ &\sum_{i=1}^{M_N}\langle X,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}(\langle X,Y\rangle_u - \langle X,Y\rangle_{\alpha_i})du + \sum_{i=1}^{M_N}\langle X,Y\rangle'_{\alpha_i}\int_{\alpha_i}^{\beta_i}\left(\int_{\alpha_i}^{u}(X_v - X_{\alpha_i})dY_v[2]\right)du + o_p(M_N^{-3/2}) \\ =\ &\frac{1}{2}\sum_{i=1}^{M_N}(\langle X,Y\rangle'_{\alpha_i})^2(\beta_i-\alpha_i)^2 + R_N + o_p(M_N^{-3/2})\end{aligned}$$

∎

PROOF OF THEOREM 4. Since in this case, $U_{N,t}^{(nonsyn)} \to 0$, it follows from the proof of Theorem 2 that we can take L_N in Theorem 3 to be $\sum_i \int_{\max(t_{i-1},s_{i-1})}^{\min(t_i,s_i)}(X_u - X_{\max(t_{i-1},s_{i-1})})dY_u[2]$. Now set

$\tilde{L}_N = \sum_i \int_{\max(t_{i-1},s_{i-1})}^{\max(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})dY_u[2]$. To assess $\tilde{L}_N - L_N = \sum_i \int_{\min(t_i,s_{i}))}^{\max(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})dY_u[2]$, note that

$$\begin{aligned}
&\langle \sum_i \int_{\min(t_i,s_i)}^{\max(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})dY_u, \sum_i \int_{\min(t_i,s_i)}^{\max(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})dY_u\rangle \\
&= \sum_i \int_{\min(t_i,s_i)}^{\max(t_i,s_i)} (X_u - X_{\max(t_{i-1},s_{i-1})})^2 d\langle Y,Y\rangle_u \\
&= \sum_i \langle X,X\rangle'_{\max(t_{i-1},s_{i-1})} \langle Y,Y\rangle'_{\min(t_i,s_i)} \int_{\min(t_i,s_i)}^{\max(t_i,s_i)} (u - \max(t_{i-1},s_{i-1}))du(1+o_p(1)) \\
&\le \frac{1}{N}\int_0^T \langle X,X\rangle'_t \langle Y,Y\rangle'_t dt \\
&= O_p(1/N),
\end{aligned}$$

and similarly for the second term. This shows the result for $w_i = \max(t_i, s_i)$. The result for v_i follows similarly. ∎

PROOF OF COROLLARY 2: To show equation (21), define δ_i^t and δ_i^s as in the proof of Theorem 1. We then obtain that (with $\Delta v = T/M_N$)

$$\begin{aligned}
&\frac{N}{T}\sum_{i:s_i,t_i\le u}(s_i - s_{i-1})(t_i - t_{i-1}) \\
&= \frac{N}{T}\sum_{i:s_i,t_i\le u}((\Delta v)^2 - N^{-1}\Delta v(\delta_i^s - \delta_{i-1}^s) - N^{-1}\Delta v(\delta_i^t - \delta_{i-1}^t) + N^{-2}(\delta_i^s - \delta_{i-1}^s)(\delta_i^t - \delta_{i-1}^t)) \\
&= \frac{N}{T}\sum_{i:s_i,t_i\le u}(\Delta v)^2 + o(1) \qquad (44)
\end{aligned}$$

(by telescope sum) and

$$\begin{aligned}
&\frac{N}{T}\sum_{i:s_i,t_i\le u}(\max(t_{i-1},s_{i-1}) - \min(t_i,s_i))^2 \\
&= \frac{N}{T}\sum_{i:s_i,t_i\le u}((\Delta v)^2 - 2N^{-1}\Delta v(\max(\delta_i^s,\delta_i^t) - \min(\delta_{i-1}^s,\delta_{i-1}^t))) + o(1) \\
&= \frac{N}{T}\sum_{i:s_i,t_i\le u}(\Delta v)^2 - 2\frac{N}{M_N}F_N(u) + o(1) \qquad (45)
\end{aligned}$$

Since $U_{N,u}^{(nonsyn)}$ is defined as the difference between the two above terms, the result (21) follows. ∎

PROOF OF THEOREM 5. The proof is similar to that of the earlier results; the main difference lies in verifying that the relevant sequences are RCP in the sense of Definition 3. We here provide the argument in the case of the bias; entirely similar considerations apply to the variance.

Consider first the term

$$\frac{\alpha}{M_\alpha}\sum_i(v_i - \max(s_i,t_i)) \le \frac{\alpha}{M_\alpha}\sum_i(v_i - s_i) + \frac{\alpha}{M_\alpha}\sum_i(v_i - t_i) \qquad (46)$$

Suppose one considers the following subsampling scheme: every time τ_i occurs, it is sampled with probability $\underline{c}\alpha/\lambda_\alpha^X(\tau_i)$. By standard considerations, the subsampled times τ_i' are derived from a Poisson process with intensity $\underline{c}\alpha$. Suppose that the number of such τ_i' is n'. If $t_i' = \max\{\tau_j' \le v_i\}$ one obtains that

$$\frac{\alpha}{M_\alpha}\sum_i (v_i - t_i) \le \frac{\alpha}{M_\alpha}\sum_i (v_i - t_i') \tag{47}$$

Note that by the Poisson property of the τ_i', the expectation of the right hand side of (47) is bounded by $1/\underline{c}$, hence (47) is $O_p(1)$. By using the same argument on the s_is, one thus obtains that (46) is $O_p(1)$. Finally, if $N' = m' + n'$, by the law of large numbers, $N'/\alpha \to 2\underline{c}T$ as $\alpha \to \infty$. Hence $\frac{N}{M_N}F_N(T)$ is $O_p(1)$. Again, by Helly's Theorem (Ash (1972), p. 329), $\frac{N}{M_N}F_N(t)$ is RCP. The rest of the proof follows similarly. ∎

Proof of Corollary 4:

Let v_i, s_i, and t_i be the same as in Section 2.1. Since X and Y are Poisson with intensities λ_α^X and λ_α^Y, respectively, we get $v_i - t_i \sim \exp(\lambda_\alpha^X)$[2], and $v_i - s_i \sim \exp(\lambda_\alpha^Y)$, thus

$$v_i - \max(t_i, s_i) = \min\left((v_i - t_i), (v_i - s_i)\right) \sim \exp(\lambda_\alpha^X + \lambda_\alpha^Y).$$

Also, since

$$\begin{aligned} v_i - \min(t_i, s_i) &= \max\left(v_i - t_i, v_i - s_i\right) \\ &= (v_i - t_i) + (v_i - s_i) - \min\left(v_i - t_i, v_i - s_i\right) \\ &\sim \exp(\lambda_\alpha^X) + \exp(\lambda_\alpha^Y) - \exp(\lambda_\alpha^X + \lambda_\alpha^Y) \end{aligned} \tag{48}$$

then,

$$F_N(v_k) = -\sum_{i=1}^{k}\left[-(v_i - \min(t_i, s_i)) + (v_i - \max(t_i, s_i))\right] + O_p(\frac{1}{N}). \tag{49}$$

Under our assumptions,

$$E[F_N(v_k)] = -k(-\frac{1}{\lambda_\alpha^X} - \frac{1}{\lambda_\alpha^Y} + \frac{2}{\lambda_\alpha^X + \lambda_\alpha^Y}) + O(\frac{1}{\alpha}).$$

Also note that $N/(\lambda_\alpha^X T + \lambda_\alpha^Y T) \to 1$ in probability. By appropriate normalization, it follows that (31) holds in expectation. By observing that (49) is an independent sum, it also follows that (31) holds in probability. Thus, (32) yields accordingly. ∎

Proof of Corollary 5:

This is a direct consequence of Theorem 2 and Corollaries 2 and 4. We here provide an independent proof as an addition. Again use the relation (48), we obtain

$$E[v_i - \min(t_i, s_i)|v_i] = \frac{1}{\lambda_\alpha^X} + \frac{1}{\lambda_\alpha^Y} - \frac{1}{\lambda_\alpha^X + \lambda_\alpha^Y}, \tag{50}$$

[2] $X \sim \exp(\lambda)$ means X follow exponential distribution with intensity λ.

and

$$\begin{aligned}
& E\left[((\min(t_i,s_i)-v_i)^2|v_i)\right] \\
= & \frac{2}{(\lambda_\alpha^X)^2}+\frac{2}{(\lambda_\alpha^Y)^2}+\frac{2}{(\lambda_\alpha^X+\lambda_\alpha^Y)^2}+\frac{2}{\lambda_\alpha^X\lambda_\alpha^Y}-2E[(v_i-t_i)\min(v_i-t_i,v_i-s_i)|v_i][2] \\
= & \frac{2}{(\lambda_\alpha^X)^2}+\frac{2}{(\lambda_\alpha^Y)^2}+\frac{2}{(\lambda_\alpha^X+\lambda_\alpha^Y)^2}-\frac{2}{\lambda_\alpha^X\lambda_\alpha^Y}+\frac{2}{(\lambda_\alpha^X+\lambda_\alpha^Y)^2}(\frac{\lambda_\alpha^X}{\lambda_\alpha^Y}+\frac{\lambda_\alpha^Y}{\lambda_\alpha^X})
\end{aligned} \tag{51}$$

where $E[(v_i-t_i)\min(v_i-t_i,v_i-s_i)|v_i][2]=E[((v_i-t_i)+(v_i-s_i))\min(v_i-t_i,v_i-s_i)|v_i]$. The first step is due to the independence between X and Y, and the second step is because

$$E[(v_i-s_i)\min(v_i-t_i,v_i-s_i)|v_i] \quad = \quad -\frac{\lambda_\alpha^Y}{\lambda_\alpha^X}\frac{1}{(\lambda_\alpha^X+\lambda_\alpha^Y)^2}+\frac{1}{\lambda_\alpha^X\lambda_\alpha^Y}.$$

Thus, by independent increment and by (50)-(51), we get

$$\begin{aligned}
& E[(\min(t_i,s_i)-\max(t_{i-1},s_{i-1}))^2|v_i] \\
= & E[((\min(t_i,s_i)-v_i)+(v_i-v_{i-1})+(v_{i-1}-\max(t_{i-1},s_{i-1}))^2|v_i] \\
= & (\frac{T}{M_\alpha})^2\left\{1+2\frac{M_\alpha}{T}\left(\frac{2}{\lambda_\alpha^X+\lambda_\alpha^Y}-\frac{1}{\lambda_\alpha^X}-\frac{1}{\lambda_\alpha^Y}\right)\right\}+O(\frac{1}{\alpha^2}).
\end{aligned}$$

Therefore,

$$EU_{N,v_k}^{(dis)} \quad = \quad k\frac{T}{M_\alpha}\left[1+2\frac{M_\alpha}{T}(\frac{2}{\lambda_\alpha^X+\lambda_\alpha^Y}-\frac{1}{\lambda_\alpha^X}-\frac{1}{\lambda_\alpha^Y})+O(\frac{1}{\alpha^2})\right].$$

Similarly, $E[(s_i-s_{i-1})]=E[(s_i-v_i)+(v_i-v_{i-1})+(v_{i-1}-s_{i-1})|v_i]=\frac{T}{M_\alpha}=E[(t_i-t_{i-1})|v_i]$, thus,

$$E(U_{N,v_k}^{(nonsyn)}) \quad = \quad 2(\lambda_\alpha^X+\lambda_\alpha^Y)k[\frac{T}{M_\alpha}(-\frac{2}{\lambda_\alpha^X+\lambda_\alpha^Y}+\frac{1}{\lambda_\alpha^X}+\frac{1}{\lambda_\alpha^Y})+O(\frac{1}{\alpha^2})].$$

By the same argument as in the proof of Corollary 4 and the results in Theorem 2, the asymptotic stochastic variance due to discretization is

$$\frac{T}{M_\alpha}\int_0^T(\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'du-2\frac{T}{N}\left(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y}\right)\int_0^T(\langle X,Y\rangle_u')^2+\langle X,X\rangle_u'\langle Y,Y\rangle_u'du,$$

whereas the the asymptotic stochastic variance due to nonsynchonization is

$$2\frac{T}{N}\int_0^T\langle X,X\rangle_u'\langle Y,Y\rangle_u'du\left(\frac{\ell^Y}{\ell^X}+\frac{\ell^X}{\ell^Y}\right)$$

Adding up, (33) follows by the law of large numbers. ∎

PROOF OF THEOREM 6: Consider separately the signal and noise terms in (39). This is legitimate since the two terms are independent. It is easy to see that the term involving the semimartingales X and Y is handled exactly in analogy with the similar development (Theorem 3) in Zhang, Mykland, and Aït-Sahalia (2005), integrating the methodology from Theorem 2 in the current paper. The constant appears as follows: the constant c from the earlier paper is here $c\sim K_N/M_N^{2/3}\sim c_2c_1^{-2/3}$.

For the noise term, replace normalization by $N^{1/6}/K_N$ by $M_N^{-1/2}$ (thus creating a constant of $c_1^{1/2}c_2^{-1}$, which is squared in the variance). We now have to deal with two suitably normalized mixing sums. The asymptotic normality follows as in Chapter 5 of Hall and Heyde (1980). It is easy to verify that the two sums are asymptotically uncorrelated. If one sets $\gamma_{i,j} = \text{Cov}(\epsilon^X_{t_0}\epsilon^Y_{s_{-j}}[2], \epsilon^X_{t_i}\epsilon^Y_{s_{i-j}}[2])$, the asymptotic variance of the "J" term thus gets the form $\gamma_{0,J} + 2\sum_{i=1}^{\infty}\gamma_{i,J}$, and similarly for the "K" term (let $J \to \infty$ is it isn't already there). To see the expression for $\gamma_{i,j}$, note that

$$\begin{aligned}
\gamma_{i,j} &= \text{Cov}(\epsilon^X_{t_0}\epsilon^Y_{s_{-j}} + \epsilon^Y_{s_0}\epsilon^X_{t_{-j}}, \epsilon^X_{t_i}\epsilon^Y_{s_{i-j}} + \epsilon^Y_{s_i}\epsilon^X_{t_{i-j}}) \\
&= 2\text{Cov}(\epsilon^X_{t_0}, \epsilon^X_{t_i})\text{Cov}(\epsilon^Y_{s_0}, \epsilon^Y_{s_i}) + 2\text{Cov}(\epsilon^X_{t_0}, \epsilon^Y_{s_i})\text{Cov}(\epsilon^Y_{s_0}, \epsilon^X_{t_i}) \\
&+ \text{Cov}(\epsilon^X_{t_0}, \epsilon^Y_{s_{i-j}})\text{Cov}(\epsilon^Y_{s_{-j}}, \epsilon^X_{t_i}) + \text{Cov}(\epsilon^Y_{s_0}, \epsilon^X_{t_{i-j}})\text{Cov}(\epsilon^X_{t_{-j}}, \epsilon^Y_{s_i}) \\
&+ \text{Cov}(\epsilon^X_{t_0}, \epsilon^X_{t_{i-j}})\text{Cov}(\epsilon^Y_{s_{-j}}, \epsilon^Y_{s_i}) + \text{Cov}(\epsilon^Y_{s_0}, \epsilon^Y_{s_{i-j}})\text{Cov}(\epsilon^X_{t_i}, \epsilon^X_{t_{-j}}) \\
&+ \text{cum}(\epsilon^X_{t_0}, \epsilon^X_{t_i}, \epsilon^Y_{s_{-j}}, \epsilon^Y_{s_{i-j}}) + \text{cum}(\epsilon^Y_{s_0}, \epsilon^Y_{s_i}, \epsilon^X_{t_{-j}}, \epsilon^X_{t_{i-j}}) \\
&+ \text{cum}(\epsilon^X_{t_0}, \epsilon^X_{t_{i-j}}, \epsilon^Y_{s_{-j}}, \epsilon^Y_{s_i}) + \text{cum}(\epsilon^X_{t_i}, \epsilon^X_{t_{-j}}, \epsilon^Y_{s_0}, \epsilon^Y_{s_{i-j}}) \qquad (52)
\end{aligned}$$

Obviously, as $j \to \infty$, $\gamma_{i,j}$ tends to the expression in (40). ∎

REFERENCES

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Mykland, P. A., Zhang, L., 2009. Ultra high frequency volatility estimation with dependent microstructure noise. Tech. rep., in revision for *JOE*.

Aldous, D., 1989. Probability Approximations via the Poisson Clumping Heuristic. Springer-Verlag.

Aldous, D. J., Eagleson, G. K., 1978. On mixing and stability of limit theorems. Annals of Probability 6, 325–331.

Andersen, T. G., Bollerslev, T., 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review 39, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 96, 42–55.

Ash, R. B., 1972. Real Analysis and Probability. Academic Press.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008a. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008b. Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Tech. rep., University of Oxford.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2009. Subsampling realised kernels. Journal of Econometrics, forthcoming .

Barndorff-Nielsen, O. E., Shephard, N., 2002a. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, B 64, 253–280.

Barndorff-Nielsen, O. E., Shephard, N., 2002b. Estimating quadratic variation using realised variance. Journal of Applied Econometrics 17, 457–477.

Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., Pictet, O. V., 2001. An Introduction to High-Frequency Finance. Academic Press, San Diego.

Delattre, S., Jacod, J., 1997. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. Bernoulli 3, 1–28.

Devroye, L., 1981. Law of the iterated logarithm for order statistics for uniform spacing. Annals of Probability 9, 860–867.

Devroye, L., 1982. A log log law for maximal uniform spacings. Annals of Probability 10, 863–868.

Epps, T. W., 1979. Comovements in stock prices in the very short run. Journal of the American Statistical Association 74, 291–298.

Gençay, R., Ballocchi, G., Dacorogna, M., Olsen, R., Pictet, O., 2002. Real-time trading models and the statistical properties of foreign exchange rates. International Economic Review 43, 463–491.

Griffin, J., Oomen, R., 2007. Covariance measurement in the presence of non-synchronous trading and market microstructure noise, preprint.

Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Academic Press, Boston.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M., 2009. Microstructure noise in the continuous case: The pre-averaging approach. Stochastic Processes and Their Applications 119, 2249–2276.

Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. Annals of Probability 26, 267–307.

Jacod, J., Shiryaev, A. N., 2003. Limit Theorems for Stochastic Processes, 2nd Edition. Springer-Verlag, New York.

Lo, A. W., MacKinlay, A. C., 1990. An econometric analysis of nonsynchronous trading. Journal of Econometrics 45, 181–211.

Lunde, A., Voev, V., 2007. Integrated covariance estimation using high-frequency data in the presence of noise. Journal of Financial Econometrics 5, 68–104.

McCullagh, P., 1987. Tensor Methods in Statistics. Chapman and Hall, London, U.K.

Mykland, P. A., 1994. Bartlett type identities for martingales. Annals of Statistics 22, 21–38.

Mykland, P. A., Zhang, L., 2006. ANOVA for diffusions and Itô processes. Annals of Statistics 34, 1931–1963.

Podolskij, M., Vetter, M., 2009. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. Bernoulli, forthcoming .

Rényi, A., 1963. On stable sequences of events. Sankyā Series A 25, 293–302.

Rootzén, H., 1980. Limit distributions for the error in approximations of stochastic integrals. Annals of Probability 8, 241–251.

Ross, S., 1996. Stochastic Processes, 2nd Edition. Wiley, New York, N.Y.

Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.

Shorack, G. R., Wellner, J. A., 1986. Empirical Processes with Application to Statistics. Wiley.

Zhang, L., 2001. From martingales to ANOVA: Implied and realized volatility. Ph.D. thesis, The University of Chicago, Department of Statistics.

Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. Bernoulli 12, 1019–1043.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. Journal of the American Statistical Association 100, 1394–1411.

Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2009. Edgeworth expansions for realized volatility and related estimators. Tech. rep., University of Illinois at Chicago.