

# A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data\*

Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia

First Draft: July 2002.  
This version: October 20, 2003

## Abstract

It is a common practice in finance to estimate volatility from the sum of frequently-sampled squared returns. However market microstructure poses challenges to this estimation approach, as evidenced by recent empirical studies in finance. This work attempts to lay out theoretical grounds that reconcile continuous-time modeling and discrete-time samples. We propose an estimation approach that takes advantage of the rich sources in tick-by-tick data while preserving the continuous-time assumption on the underlying returns. Under our framework, it becomes clear why and where the “usual” volatility estimator fails when the returns are sampled at the highest frequency.

KEY WORDS: Measurement error; Subsampling; Market Microstructure; Martingale; Bias-correction; Realized volatility.

---

\*Lan Zhang is Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: lzhang@stat.cmu.edu. Per A. Mykland is Professor, Department of Statistics, The University of Chicago, Chicago, IL 60637. E-mail: mykland@galton.uchicago.edu. Yacine Aït-Sahalia is Professor, Bendheim Center for Finance, Princeton University and NBER, Princeton, NJ 08544-1021. E-mail: yacine@princeton.edu. We gratefully acknowledge the support of the National Science Foundation under grants DMS-0204639 (Zhang and Mykland) and SBR-0111140 (Aït-Sahalia).

## 1 INTRODUCTION

In the analysis of high frequency financial data, a major problem concerns the nonparametric determination of the volatility of an asset return process. A common practice is to estimate volatility from the sum of the frequently-sampled squared returns. Though this approach is justified under the assumption of a continuous stochastic model in an idealized world, it meets the challenge from market microstructure in real applications. We argue that this customary way of estimating volatility is flawed in that it overlooks observation error. The usual mechanism for dealing with the problem is to throw away some data, by sampling less frequently or constructing “time-aggregated” returns from the underlying high frequency asset prices. We propose here a statistically sounder device. Our device is model-free, it takes advantage of the rich sources in tick-by-tick data, and to a great extent it corrects the effect of the microstructure noise on volatility estimation. In the course of constructing our estimator, it becomes clear why and where the “usual” volatility estimator fails when the returns are sampled at high frequency.

Our interest lies in using high frequency intraday data to estimate the integrated volatility over some time periods. To fix the ideas, let  $\{S_t\}$  denote the price process of a security, and suppose the log-return process  $\{X_t\}$ , where  $X_t = \log S_t$ , follows an Itô process

$$X_t = \mu_t dt + \sigma_t dB_t \quad (1)$$

where  $B_t$  is a standard Brownian motion. Typically,  $\sigma_t^2$ , the instantaneous variance (or diffusion coefficient) of the return process  $\{X_t\}$ , will be stochastic. The parameter of interest is the integrated (cumulative) volatility over one or successive time periods,  $\int_0^{T_1} \sigma_t^2 dt$ ,  $\int_{T_1}^{T_2} \sigma_t^2 dt$ , .... A natural way to estimate the cumulative volatility over, say, a single time interval from 0 to  $T$ , is to use the sum of squared incremental returns,

$$\sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 \approx \int_0^T \sigma_t^2 dt, \quad (2)$$

where the  $X_{t_i}$ 's are all the observations of the return process in  $[0, T]$ . The estimator  $\sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2$  is commonly used and generally called “realized volatility” or “realized variance.” For a sample of the recent literature in integrated volatility, see Hull and White (1987), Jacod and Protter (1998), Gallant, Hsu, and Tauchen (1999), Chernov and Ghysels (2000), Gloter (2000), Andersen, Bollerslev, Diebold, and Labys (2001), Barndorff-Nielsen and Shephard (2001), Mykland and Zhang (2002) and others.

Under model (1), the approximation in (2) is justified by theoretical results in stochastic processes which state that

$$\text{plim} \sum_{t_i} (X_{t_{i+1}} - X_{t_i})^2 = \int_0^T \sigma_t^2 dt, \quad (3)$$

as the sampling frequency increases. In other words, the estimation error of the realized volatility diminishes. According to (3), realized volatility computed from the highest frequency data ought to provide the best possible estimate for  $\int_0^T \sigma_t^2 dt$  the integrated volatility.

However, this is not the general viewpoint from the finance literature. It is generally held there that the returns process  $X_t$  should not be sampled too often, regardless of the fact that the asset prices can often be observed with extremely high frequency, such as several times per second. It has been found empirically that the estimator is not robust when the sampling interval is quite small. Issues including bigger bias in the estimate and non-robustness to changes in sampling interval have been reported (see e.g., Brown (1990), Campbell, Lo, and MacKinlay (1997), Bai, Russell, and Tiao (2000)). The main explanation for this phenomenon is a vast array of issues collectively known as market microstructure, such as, but not limited to, the existence of the bid-ask spread: see Ait-Sahalia, Mykland, and Zhang (2003) for a description of these phenomena and their grounding in the vast theoretical literature describing the frictions inherent in the trading process. When prices are sampled at finer intervals, microstructure issues become more pronounced. It is then suggested that the bias induced by market microstructure effects makes the most finely sampled data unusable for the calculation, and many authors prefer to sample over longer time horizons to obtain reasonable estimates. The length of the typical choices in the literature is ad hoc and ranges from 5 to 30 minutes for exchange rate data, for instance.

This approach to handling the data poses a conundrum from the statistical point of view. We argue here that sampling over longer horizon merely reduces the impact of microstructure, rather than quantifying and correcting its effect for volatility estimation. And it goes against the grain to throw away data. On the other hand, market microstructure may pose so many problems that subsampling is the only way out.

In this paper we analyze the trade offs involved in the choice of sampling frequency and develop a method to estimate integrated volatility in such a way as to lessen this conflict. Our contention in the following is that the contamination due to market microstructure is, to first order, the same as what statisticians usually call “observation error”. We shall incorporate the observation error into the estimating procedure for integrated volatility. In other words, we shall suppose that the return process as observed at the sampling times is of the form

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \quad (4)$$

Here  $X_t$  is a latent true, or efficient, return process, and the  $\epsilon_{t_i}$ s are independent noise around the true return. A similar structure was used in the parametric context where  $\sigma_t = \sigma$  is constant by Ait-Sahalia, Mykland, and Zhang (2003). In that paper, due to the parametric nature of volatility, we proposed likelihood-based corrections for market microstructure.

We show in Section 2.2 that, if the data have a structure of the form (4), ignoring microstructure noise would have a devastating effect on the use of the realized volatility. Instead of (2), one gets

$$\sum_{t_i, t_{i+1} \in [0, T]} (Y_{t_{i+1}} - Y_{t_i})^2 = 2n \text{Var}(\epsilon) + O_p(n^{1/2}) \quad (5)$$

where the errors  $\epsilon_{t_i}$ 's are i.i.d. with mean 0, and  $n$  is the number of sampling intervals over  $[0, T]$ . As we will show, ignoring market microstructure noise in the context of stochastic volatility leads to

an even more dangerous situation than when  $\sigma$  is constant and  $T \rightarrow \infty$ . The results from equation (5) suggest that the realized volatility does not estimate the true integrated volatility, but rather the variance of the contamination noise. In fact, we will show that the true integrated volatility, which is  $O_p(1)$ , is even dwarfed by the magnitude of the asymptotically Gaussian  $O_p(n^{1/2})$  term in (5).

Of course, the model (4) may also not be correct. When made the basis of inference, it could still occur that one does not wish to sample as frequently as the data would permit. It may, however, make it possible to use substantially larger amounts of data than what would be possible under (2). Subsampling-based schemes can be made to incorporate all the data.

In seeking to create an inference procedure under measurement error, we have sought to draw some lessons from the empirical practice that one should not use all the data, while at the same time not violating basic statistical principles. Our procedure estimates parameters at two different frequencies of sampling, and then by cancellation removes the effect of the  $\epsilon$ 's to the required order. We show in Section 4 that this leads to a variance-variance trade-off between the effect in (5) and an effect due to the sampling frequencies.

The theory, including asymptotic distributions, is developed in Section 2-4 for the case of one time period  $[0, T]$ . The multi-period problem is treated in Section 5. Section 6 discusses how to estimate the asymptotic variance for equidistant sampling. And Section 7 discusses what to do if one really wants to use the customary realized volatility, sampled at high frequency despite the presence of microstructure noise.

## 2 ANALYSIS OF THE “REALIZED VOLATILITY”

### 2.1 Set-up

To spell out the model above, we let  $Y$  be the logarithm of the transaction price, which is observed at times  $0 = t_0, t_1, \dots, t_n = T$ . We assume that at these times,  $Y$  is related to a latent true price  $X$  (also in logarithm scale) through equation (4). The latent price  $X$  is given in (1). The noise  $\epsilon_{t_i}$  satisfies the following assumption,

$$\epsilon_{t_i} \text{ i.i.d. with } E\epsilon_{t_i} = 0, \text{ and } Var(\epsilon_{t_i}) = \nu. \text{ Also } \epsilon \perp\!\!\!\perp X \text{ process} \quad (6)$$

where  $\perp\!\!\!\perp$  denotes independence between two random quantities. Our modeling as in (4) does not require that  $\epsilon_t$  exists for every  $t$ , in other words, our interest in the noise is only at the observation times  $t_i$ 's.

For the moment, we focus on determining the integrated volatility of  $X$  for the entire time period  $[0, T]$ . This is also known as the continuous quadratic variation  $\langle X, X \rangle$  of  $X$ . In other

words,

$$\langle X, X \rangle_T = \int_0^T \sigma_t^2 dt \quad (7)$$

To describe succinctly the realized volatility, we use the notion of observed quadratic variation  $[\cdot, \cdot]$ . Given the grid  $\mathcal{G} = \{t_0, \dots, t_n\}$ , the observed quadratic variation for a generic process  $Z$  is

$$[Z, Z]_t = \sum_{t_{i+1} \leq t} (\Delta Z_{t_i})^2 \quad (8)$$

where  $\Delta Z_{t_i} = Z_{t_{i+1}} - Z_{t_i}$ . We shall later have occasion to vary the grid  $\mathcal{G}$ . Quadratic covariations are similarly defined (see e.g., Karatzas and Shreve (1991) for more details on quadratic variations.)

Our interest is to assess how well the realized volatility,  $[Y, Y]_T$ , approximates the integrated volatility  $\langle X, X \rangle_T$  of the true returns process. In our asymptotic considerations, we shall always assume that the number of observations in  $[0, T]$  goes to infinity, and also that the maximum distance in time between two observations goes to zero:

$$\max_i \Delta t_i \rightarrow 0 \text{ as } n \rightarrow \infty \quad (9)$$

## 2.2 The Realized Volatility: An Estimator of the Spread of the Noise?

Under the additive model  $Y_{t_i} = X_{t_i} + \epsilon_{t_i}$ , the realized volatility based on the observed returns  $Y_{t_i}$  now has the form

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

This gives the conditional mean and variance of  $[Y, Y]_T$ , given the process of latent true prices  $X$ . As derived in the Appendix,

$$E([Y, Y]_T \mid X \text{ process}) = [X, X]_T + 2n\nu, \quad (10)$$

under assumption (6) and the definition of  $[\cdot, \cdot]$  in (8). Similarly,

$$\text{Var}([Y, Y]_T \mid X \text{ process}) = 4nE\epsilon^4 + O_p(1), \quad (11)$$

subject to condition (6) and  $E\epsilon_{t_i}^4 < \infty$ , for all  $i$ . The exact expression for the variance is given in Appendix A.1.

Following the discussion in Appendix A.2, it is also the case that as  $n \rightarrow \infty$ , the distribution of  $n^{-1/2}([Y, Y]_T - 2n\nu)$  becomes normal conditionally on the  $X$  process, with mean 0 and variance  $4E\epsilon^4$ .

Equations (10) and (11) suggest that in the discrete world where microstructure effects are unfortunately present, realized volatility  $[Y, Y]_T$  is not a reliable estimator for the true variation

$[X, X]_T$  of the returns. *For large  $n$ , realized volatility could have little to do with the true returns.* As seen in (10),  $[Y, Y]_T$  has a positive bias whose magnitude increases linearly with the sample size  $n$ . If one really wants to live with this bias – which we do not recommend – and use the customary realized volatility as a measure of variation, the above provides theoretical evidence for not sampling too often (see Section 7 for a formal analysis).

Interestingly, apart from revealing the biased nature of  $[Y, Y]_T$  at high frequency, our analysis also delivers an estimator for the spread of the noise term. In other words, the realized volatility  $[Y, Y]_T$  yields a consistent and asymptotically normal estimator of noise spread  $\nu$ , namely  $\hat{\nu} = \frac{1}{2n}[Y, Y]_T$ . We have, for a fixed true return process  $X$ ,

$$n^{1/2}(\hat{\nu} - \nu) \rightarrow N(0, E\epsilon^4), \quad \text{as } n \rightarrow \infty, \quad (12)$$

see Theorem A.1 in the Appendix.

### 3 SAMPLING SPARSELY WHILE USING ALL THE DATA: ANALYSIS IN THE MULTIPLE GRID CASE

#### 3.1 Multiple Grids and Sufficiency

We have argued in the previous section that one can indeed benefit from using infrequently sampled data. And yet, one of the most basic lessons of statistics is that one should not do this. We present here two ways of tackling the problem. Both are based on selecting a number of subgrids of the original grid of observation times,  $\mathcal{G} = \{t_0, \dots, t_n\}$ , and then averaging the estimators derived from the subgrids. The principle is that to the extent that there is a benefit to subsampling, this benefit can now be retained, while the variation of the estimator can be lessened by the averaging. The benefit of the averaging is clear from sufficiency considerations, and many statisticians would say that subsampling without subsequent averaging is inferentially incorrect.

In the following, we first introduce a set of notations, and then turn to studying the realized volatility in the multi-grid context. In Section 4, we show how to explicitly estimate the model (4) by using a combination of the single grid  $\mathcal{G}$  and the multiple grids.

#### 3.2 Notation for the Multiple Grids

We specifically suppose that the total grid  $\mathcal{G}$ ,  $\mathcal{G} = \{t_0, \dots, t_n\}$  as before, is partitioned into  $K$  non-overlapping subgrids  $\mathcal{G}^{(k)}$ ,  $k = 1, \dots, K$ , in other words,

$$\mathcal{G} = \cup_{k=1}^K \mathcal{G}^{(k)} \quad \text{where } \mathcal{G}^{(k)} \cap \mathcal{G}^{(l)} = \emptyset \quad \text{when } k \neq l.$$

For most purposes, the natural way to select the  $k^{\text{th}}$  subgrid  $\mathcal{G}^{(k)}$  is to start with  $t_{k-1}$  and then pick every  $K$ th sample point after that, until  $T$ . That is to say that

$$\mathcal{G}^{(k)} = (t_{k-1}, t_{k-1+K}, t_{k-1+2K}, \dots, t_{k-1+n_k K})$$

for  $k = 1, \dots, K$ , and  $n_k$  is the integer making  $t_{k-1+n_k K}$  the last element in  $\mathcal{G}^{(k)}$ . We shall refer to this as *regular* allocation of sample points to subgrids.

Whether the allocation is regular or not, we let  $n_k$  be such that the subgrid  $\mathcal{G}^{(k)}$  has  $n_k + 1$  elements. As before, the number of elements in the total grid  $\mathcal{G}$  is  $n + 1$ . More general schemes for allocating sample points to grids can also be used, subject to the restrictions in Theorem A.1 in Appendix A.2. The realized volatility based on all observation points  $\mathcal{G}$ , so far denoted  $[Y, Y]_T$ , will now for clarity be written as  $[Y, Y]_T^{(all)}$ . Meanwhile, if one uses only the subsampled observations  $Y_t$ ,  $t \in \mathcal{G}^{(k)}$ , the realized volatility will be denoted as  $[Y, Y]_T^{(k)}$ . It has the form

$$[Y, Y]_T^{(k)} = \sum_{t_j, t_{j,+} \in \mathcal{G}^{(k)}} (Y_{t_{j,+}} - Y_{t_j})^2.$$

where, if  $t_i \in \mathcal{G}^{(k)}$ , then  $t_{i,-}$  and  $t_{i,+}$  denote respectively the preceding and following elements in  $\mathcal{G}^{(k)}$ .

A natural competitor to  $[Y, Y]_T^{(all)}$  is then given by

$$[Y, Y]_T^{(avg)} = \frac{1}{K} \sum_{k=1}^K [Y, Y]_T^{(k)}, \quad (13)$$

and this is the statistic we analyze in the following.

As before, we fix  $T$  and use only the observations within the time period  $[0, T]$ . Asymptotics will still be under (9) and under

$$\text{as } n \rightarrow \infty, \quad n/K \rightarrow \infty. \quad (14)$$

In general, the  $n_k$  need not be the same across  $k$ . We define

$$\bar{n} = \frac{1}{K} \sum_{k=1}^K n_k = \frac{n - K + 1}{K}. \quad (15)$$

### 3.3 Error Due to the Noise $\epsilon$

Recall that we are interested in determining the integrated volatility  $\langle X, X \rangle_T$ , or quadratic variation, of the true but unobservable returns. As an intermediate step, we study in this subsection how well the ‘‘pooled’’ realized volatility  $[Y, Y]_T^{(avg)}$  approximates  $[X, X]_T^{(avg)}$ , where the latter is the ‘‘pooled’’ true integrated volatility when  $X$  is considered only on the discrete time scale.  $[X, X]$  has the form given in equation (8).

From (10) and (13),

$$E([Y, Y]_T^{(avg)} \mid X \text{ process}) = [X, X]_T^{(avg)} + 2\bar{n}\nu. \quad (16)$$

Also, since  $\{\epsilon_t, t \in \mathcal{G}^{(k)}\}$  are independent for different  $k$ ,

$$\begin{aligned} \text{Var}([Y, Y]_T^{(avg)} \mid X \text{ process}) &= \frac{1}{K^2} \sum_{k=1}^K \text{Var}([Y, Y]_T^{(k)} \mid X \text{ process}) \\ &= \frac{1}{K^2} \sum_{k=1}^K n_k 4E\epsilon^4 + \text{Remainder}, \\ &= 4\frac{\bar{n}}{K} E\epsilon^4 + O_p\left(\frac{1}{K}\right) \end{aligned} \quad (17)$$

in the same way as in (11). The order of the remainder follows as in the single grid case, see (A.3) in the Appendix.

By Theorem A.1 in Appendix A.2, the conditional asymptotics for the estimator  $[Y, Y]_T^{(avg)}$  are as follows

**Theorem 1.** *Suppose  $X$  is an Itô process of form (1). Suppose  $Y$  is related to  $X$  through model (4), and that (6) is satisfied with  $E\epsilon^4 < \infty$ . Also suppose that  $t_i$  and  $t_{i+1}$  are not in the same subgrid for any  $i$ . Under assumption (14), as  $n \rightarrow \infty$*

$$\sqrt{\frac{K}{\bar{n}}}([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\nu\bar{n}) \xrightarrow{\mathcal{L}} 2\sqrt{E\epsilon^4}Z_\epsilon^{(avg)}, \quad (18)$$

conditional on the  $X$  process, where  $Z_\epsilon^{(avg)}$  is standard normal.

This can be compared with the result which is stated below in equation (48) in Section 7. Notice that  $Z_\epsilon^{(avg)}$  in (18) is almost never the same as  $Z_\epsilon$  in (48), in particular,  $\text{Cov}(Z_\epsilon, Z_\epsilon^{(avg)}) = \text{Var}(\epsilon^2)/E\epsilon^4$ , based on the proof in Theorem A.1 in the Appendix.

In comparison to the realized volatility using the full grid  $\mathcal{G}$ , the aggregated estimator  $[Y, Y]_T^{(avg)}$  provides an improvement in that both the asymptotic bias and variance are of smaller order of  $n$ . Cf. equations (10) and (11) in the preceding section. We shall use this in Section 4, and also in Section 7 below.

### 3.4 Error Due to the Discretization Effect: $[X, X]_T^{(avg)} - \langle X, X \rangle_T$

In this subsection, we study the impact of the time discretization. In other words, we investigate the deviation of  $[X, X]_T^{(avg)}$  from the integrated volatility  $\langle X, X \rangle_T$  of the true process. Denote



the discretization effect as  $D_T$ , where

$$\begin{aligned} D_t &= [X, X]_t^{(avg)} - \langle X, X \rangle_t \\ &= \frac{1}{K} \sum_{k=1}^K ([X, X]_t^{(k)} - \langle X, X \rangle_t) \end{aligned} \quad (19)$$

with

$$[X, X]_t^{(k)} = \sum_{t_i \in \mathcal{G}^{(k)}: t_{i,+} \leq t} (X_{t_{i,+}} - X_{t_i})^2 \quad (20)$$

We consider in the following the asymptotics of  $D_T$ . The problem is similar to that of finding the limit of  $[X, X]_T^{(all)} - \langle X, X \rangle_T$ , cf. equation (49) below. This present case, however, is more complicated due to the multiple grids.

We suppose in the following that the sampling points are allocated to subgrids as described by equation (A.20) in Appendix A.3. In particular, this covers the regular allocation, as defined in Section 3.2. We also assume that

$$\max_i |\Delta t_i| = O\left(\frac{1}{n}\right). \quad (21)$$

Define the weight function

$$h_i = \frac{8n}{TK^3} \sum_{l=1}^K [t_i - t_i^{(l)} + \frac{1}{2} \Delta t_i][\#\{k : t_i^{(k)} > t_i^{(l)}\} + \frac{1}{2}] \quad (22)$$

In the case where the  $t_i$  are equidistant, and under regular allocation of points to subgrids,  $\Delta t_i = \Delta t$ , and so all the  $h_i$  are equal, and

$$h_i = \frac{2}{\Delta t K^3} \Delta t (1 + 3^2 + \dots + (2K - 1)^2) = \frac{2K(2K - 1)(2K + 1)}{3K^3} = \frac{8}{3} + o(1). \quad (23)$$

More generally, assumptions (21) and (A.20) assure that

$$\sup_i h_i = O(1). \quad (24)$$

We take  $\langle D, D \rangle_T$  to be the quadratic variation of  $D_t$  when viewed as a continuous time process (19). This gives the best approximation to the variance of  $D_T$ .

We show the following results in Appendix A.3.

**Theorem 2.** *Suppose  $X$  is an Itô process of the form (1), with drift coefficient  $\mu_t$  and diffusion coefficient  $\sigma_t$ , both continuous almost surely. Assume (21) and (A.20). Then the quadratic variation of  $D_T$  is approximately*

$$\langle D, D \rangle_T = \frac{TK}{n} \eta_n^2 + o_p\left(\frac{K}{n}\right) \quad (25)$$

where

$$\eta_n^2 = \sum_i h_i (\langle X, X \rangle'_{t_i})^2 \Delta t_i. \quad (26)$$

In particular,  $D_T = O_p((K/n)^{1/2})$ . From this, we shall derive a variance-variance trade-off between the two effects that have been discussed – noise and discretization.

First, however, we discuss the asymptotic law of  $D_T$ . Stable convergence is discussed at the end of this section.

**Theorem 3.** *Assume the conditions of Theorem 2, and also that*

$$\eta_n^2 \xrightarrow{P} \eta^2 \tag{27}$$

Also assume Condition E in Appendix A.3. Then

$$D_T/(K/n)^{1/2} \xrightarrow{\mathcal{L}} \eta Z, \tag{28}$$

where  $Z$  is standard normal, and independent of the data. The convergence in law is stable.

In other words,  $D_T/(K/n)^{1/2}$  can be taken to be asymptotically mixed normal “ $N(0, \eta^2)$ .”

For most of our discussion, it is most convenient to suppose (27), and this is satisfied in many cases. For example, when the  $t_i$  are equidistant, and under regular allocation of points to subgrids,

$$\eta^2 = \frac{8}{3} \int_0^T \sigma_t^4 dt, \tag{29}$$

following (23). One does not need to rely on (27); we argue in Appendix A.3 that without this condition, one can take  $D_T/(K/n)^{1/2}$  to be approximately  $N(0, \eta_n^2)$ . For estimation of  $\eta^2$  or  $\eta_n^2$ , see Section 6 below.

Finally, *stable convergence* (Rényi (1963), Aldous and Eagleson (1978), Chapter 3 of Hall and Heyde (1980)) means for our purposes that the left hand side of (28) converges to the right hand side jointly with the  $X$  process, and that  $Z$  is independent of  $X$ . This is slightly weaker than convergence conditional on  $X$ , but serves the same function of permitting the incorporation of conditionality-type phenomena into arguments and conclusions, cf. the following sections.

### 3.5 Combining the Two Sources of Error

We can now combine the two error terms arising from discretization and from the observation noise, respectively. It follows from Theorems 1 and 3 that

$$[Y, Y]_T^{(avg)} - \langle X, X \rangle_T - 2\nu\bar{n} \approx \xi Z_{\text{total}}, \tag{30}$$

where  $Z_{\text{total}}$  is a standard normal random variable independent of the  $X$  process, and

$$\xi^2 = 4\frac{\bar{n}}{K} E\epsilon^4 + \frac{T}{\bar{n}}\eta^2. \tag{31}$$

Here, the symbol “ $\stackrel{\mathcal{L}}{\approx}$ ” means that when multiplied by a suitable factor, the convergence is in law (and stable, by the preceding results): see also the proof of Theorem 4 in the next section.

It is easily seen that if one takes  $K = cn^{2/3}$ , both components in  $\xi^2$  will be present in the limit, otherwise one of them will dominate. Based on (30),  $[Y, Y]_T^{(avg)}$  is yet a biased estimator of the quadratic variation  $\langle X, X \rangle_T$  of the true return process. In particular, the bias  $2\nu\bar{n}$  still increases with the number of the sub-samples. One can recognize that, as far as the asymptotic bias is concerned,  $[Y, Y]_T^{(avg)}$  is a better estimator than  $[Y, Y]_T^{(all)}$ , since  $\bar{n} \leq n$ , suggesting that the bias in the subsampled estimator  $[Y, Y]_T^{(avg)}$  increases in a slower pace than the full-sampled estimator. One can also construct a bias-adjusted estimator from (30), and this further development would involve the higher order analysis between the bias and the subsampled estimator. We show the methodology of bias correction in Section 4.

## 4 ESTIMATION FOR THE MODEL WITH MEASUREMENT ERROR: COMBINING TWO SAMPLING FREQUENCIES

### 4.1 The Estimator: Main Result

In previous sections, we have seen that the multigrid estimator  $[Y, Y]^{(avg)}$  is yet another biased estimator of the true integrated volatility  $\langle X, X \rangle$ . In this section we improve the multigrid estimator by adopting bias adjustment.

To access the bias, one utilizes the full grid. As mentioned from equation (12) in single-grid case (Section 2),  $\nu$  can be consistently approximated by

$$\hat{\nu} = \frac{1}{2n} [Y, Y]_T^{(all)}. \quad (32)$$

Hence the bias of  $[Y, Y]^{(avg)}$  can be consistently estimated by  $2\bar{n}\hat{\nu}$ . A bias-adjusted estimator for  $\langle X, X \rangle$  can thus be obtained by

$$\widehat{\langle X, X \rangle}_T = [Y, Y]_T^{(avg)} - 2\hat{\nu}\bar{n}. \quad (33)$$

To study the asymptotic behavior of  $\widehat{\langle X, X \rangle}_T$ , note first that under the conditions of Theorem A.1 in Appendix A.2

$$\begin{aligned} \left(\frac{K}{\bar{n}}\right)^{1/2} \left(\widehat{\langle X, X \rangle}_T - [X, X]_T^{(avg)}\right) &= \left(\frac{K}{\bar{n}}\right)^{1/2} \left([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\nu\bar{n}\right) \\ &\quad - 2(K\bar{n})^{1/2}(\hat{\nu} - \nu) \\ &\stackrel{\mathcal{L}}{\rightarrow} N(0, 8\nu^2), \end{aligned} \quad (34)$$

where the convergence in law is conditional on  $X$ .

We can now combine this with the results of Section 3.4 to determine the optimal choice of  $K$  as  $n \rightarrow \infty$ :

$$\begin{aligned} \langle \widehat{X}, \widehat{X} \rangle_T - \langle X, X \rangle_T &= (\langle \widehat{X}, \widehat{X} \rangle_T - [X, X]_T^{(avg)}) + ([X, X]_T^{(avg)} - \langle X, X \rangle_T) \\ &= O_p\left(\frac{\bar{n}^{1/2}}{K^{1/2}}\right) + O_p\left(\bar{n}^{-1/2}\right). \end{aligned} \quad (35)$$

The error is minimized by equating the two terms on the right hand side of (35), yielding that the optimal sampling step for  $[Y, Y]_T^{(avg)}$  is  $K = O(n^{2/3})$ . The right hand side of (35) then has order  $O_p(n^{-1/6})$ .

In particular, if we take

$$K = cn^{2/3}, \quad (36)$$

we find the limit in (35), as follows.

**Theorem 4.** *Suppose  $X$  is an Itô process of form (1), and assume the conditions of Theorem 3 in Section 3.4. Suppose  $Y$  is related to  $X$  through model (4), and that (6) is satisfied with  $E\epsilon^2 < \infty$ . Also suppose that  $t_i$  and  $t_{i+1}$  is not in the same subgrid for any  $i$ . Under assumption (36),*

$$\begin{aligned} n^{1/6} \left( \langle \widehat{X}, \widehat{X} \rangle_T - \langle X, X \rangle_T \right) &\xrightarrow{\mathcal{L}} N(0, 8c^{-2}\nu^2) + \eta\sqrt{T}N(0, c) \\ &= (8c^{-2}\nu^2 + c\eta^2T)^{1/2} N(0, 1), \end{aligned} \quad (37)$$

where the convergence is stable in law (see Section 3.4).

**Proof of Theorem 4.** Note that the first normal distribution comes from equation (34) and the second from Theorem 3 in Section 3.4. The two normal distributions are independent since the convergence of the first term in (35) is conditional of the  $X$  process, which is why they can be amalgamated as stated. The requirement that  $E\epsilon^4 < \infty$  (Theorem A.1 in the appendix) is not needed since only a law of large number is required for  $M_T^{(1)}$  (see the proof of that theorem) when considering the difference in (34) above. This finishes the proof.

The estimation of the asymptotic spread  $s^2 = 8c^{-2}\nu^2 + c\eta^2T$  of  $\langle \widehat{X}, \widehat{X} \rangle_T$  is deferred to Section 6 below. Also, note that, by Theorem A.1 and the same methods as in Appendix A.2, a consistent estimator of the asymptotic variance of  $\hat{\nu}$  is given by

$$\frac{1}{2n} \sum_i (\Delta Y_{t_i})^4 - 3\hat{\nu}^2. \quad (38)$$

## 4.2 Properties of $\langle \widehat{X}, \widehat{X} \rangle_T$ : Optimal Sampling, and Bias Adjustment

To further pin down the optimal sampling frequency  $K$  one can minimize the expected asymptotic variance in (37) to obtain

$$c = \left( \frac{16\nu^2}{T E\eta^2} \right)^{1/3} \quad (39)$$

which can be consistently estimated from data in past time periods (before time  $t_0 = 0$ ), using  $\hat{\nu}$  and an estimator of  $\eta^2$ , cf. Section 6. As mentioned in Section 3.4,  $\eta^2$  can be taken to be independent of  $K$  so long as one allocates sampling points to grids regularly, as defined in Section 3.2. Hence one can choose  $c$ , and so also  $K$ , based on past data.

**Example 1.** If  $\sigma_t^2$  is constant, and for equidistant sampling and regular allocation to grids,  $\eta^2 = \frac{8}{3}\sigma^4 T$ , and the asymptotic variance in equation (37) is

$$8c^{-2}\nu^2 + c\eta^2 T = 8c^{-2}\nu^2 + \frac{8}{3}c\sigma^4 T^2$$

and the optimal choice of  $c$  becomes

$$c_{opt} = \left( \frac{6\nu^2}{T^2\sigma^4} \right)^{1/3}. \quad (40)$$

In this case, the asymptotic variance is

$$4(6\nu^2)^{1/3}(\sigma^2 T)^{4/3}.$$

One can also, of course, estimate  $c$  to minimize the actual asymptotic variance in (37) from data in the current time period ( $0 \leq t \leq T$ ). It is beyond the scope of this paper to consider whether such a device for selecting the frequency has any impact on our asymptotic results.

In addition to large sample arguments, one can study  $\langle \widehat{X}, \widehat{X} \rangle_T$  from a “smallish” sample point of view. We argue in the following that one can apply a bias type adjustment to get

$$\langle \widehat{X}, \widehat{X} \rangle_T^{(adj)} = \left(1 - \frac{\bar{n}}{n}\right)^{-1} \langle \widehat{X}, \widehat{X} \rangle_T. \quad (41)$$

The difference from the estimator in (33) is of order  $O_p(\bar{n}/n) = O_p(K^{-1})$ , and thus the two estimators behave the same to the asymptotic order that we consider. The estimator (41), however, has the appeal of being, in a certain way, “unbiased”, as follows. Consider all estimators of the form

$$\langle \widehat{X}, \widehat{X} \rangle_T^{(adj)} = a[Y, Y]_T^{(avg)} - 2b\hat{\nu}\bar{n},$$

then, from (10) and (16),

$$\begin{aligned} E(\langle \widehat{X}, \widehat{X} \rangle_T^{(adj)} | X \text{ process}) &= a([X, X]_T^{(avg)} + 2\bar{n}\nu) - b\frac{\bar{n}}{n}([X, X]_T^{(all)} + 2n\nu) \\ &= a[X, X]_T^{(avg)} - b\frac{\bar{n}}{n}[X, X]_T^{(all)} + 2(a - b)\bar{n}\nu. \end{aligned}$$

It is natural to choose  $a = b$  to completely remove the effect of  $\nu$ . Also, following Section 3.4, both  $[X, X]_T^{(avg)}$  and  $[X, X]_T^{(all)}$  are asymptotically unbiased estimators of  $\langle X, X \rangle_T$ . Hence one can argue that one should take  $a(1 - \bar{n}/n) = 1$ , yielding (41).

Similarly, an adjusted estimator of  $\nu$  is given by

$$\hat{\nu}^{(adj)} = \frac{1}{2}(n - \bar{n})^{-1} \left( [Y, Y]_T^{(all)} - [Y, Y]_T^{(avg)} \right), \quad (42)$$

which satisfies that  $E(\hat{\nu}^{(adj)} | X \text{ process}) = \nu + \frac{1}{2}(n - \bar{n})^{-1} \left( [X, X]_T^{(all)} - [X, X]_T^{(avg)} \right)$ , and is therefore unbiased to high order. As for the asymptotic distribution, One can see from Theorem A.1 in the Appendix that

$$\begin{aligned} \hat{\nu}^{(adj)} - \nu &= (\hat{\nu} - \nu)(1 + O(K^{-1})) + O_p(Kn^{-3/2}) \\ &= \hat{\nu} - \nu + O_p(n^{-1/2}K^{-1}) + O_p(Kn^{-3/2}) \\ &= \hat{\nu} - \nu + O_p(n^{-5/6}) \end{aligned}$$

from (36). It follows that  $n^{1/2}(\hat{\nu} - \nu)$  and  $n^{1/2}(\hat{\nu}^{(adj)} - \nu)$  have the same asymptotic distribution.

## 5 MULTIPLE PERIOD INFERENCE

For a given family  $\mathcal{A} = \{\mathcal{G}^{(k)}, k = 1, \dots, K\}$ , we denote by

$$\langle \widehat{X}, \widehat{X} \rangle_t = [Y, Y]_t^{(avg)} - \frac{\bar{n}}{n}[Y, Y]_t \quad (43)$$

where, as usual,  $[Y, Y]_t = \sum_{t_{i+1} \leq t} \Delta Y_{t_i}^2$  and  $[Y, Y]_t^{(avg)} = \frac{1}{K} \sum_{k=1}^K [Y, Y]_t^{(k)}$ , with

$$[Y, Y]_t^{(k)} = \sum_{t_i \in \mathcal{G}^{(k)}: t_{i,+} \leq t} (Y_{t_{i,+}} - Y_{t_i})^2.$$

In order to estimate  $\langle X, X \rangle$  for several discrete time periods, say  $[0, T_1], [T_1, T_2], \dots, [T_{M-1}, T_M]$ , where  $M$  is fixed, this amounts to estimating  $\langle X, X \rangle_{T_m} - \langle X, X \rangle_{T_{m-1}} = \int_{T_{m-1}}^{T_m} \sigma_u^2 du$ , for  $m = 1, \dots, M$ , and the obvious estimator is  $\langle \widehat{X}, \widehat{X} \rangle_{T_m} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}}$ .

To carry out the asymptotics, let  $n_m$  be the number of points in the  $m^{\text{th}}$  time segment, and similarly let  $K_m = c_m n_m^{2/3}$ , where  $c_m$  is a constant. Then  $\{n_m^{1/6}(\langle \widehat{X}, \widehat{X} \rangle_{T_m} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}} - \int_{T_{m-1}}^{T_m} \sigma_u^2 du), m = 1, \dots, M\}$  converge stably to  $\{(8c_m^{-2}\nu^2 + c_m\eta_m^2(T_m - T_{m-1}))^{1/2}Z_m\}$ , where the  $Z_m$  are i.i.d. standard normals, independent of the underlying process, and  $\eta_m^2$  is the limit  $\eta^2$  (Theorem 3) for time period  $m$ . In the case of equidistant  $t_i$  and regular allocation of sample points to grids,  $\eta_m^2 = \frac{8}{3} \int_{T_{m-1}}^{T_m} \sigma_u^4 du$ .

In other words, the one period asymptotics generalizes straightforwardly to the multiperiod case. This is because  $\langle \widehat{X}, \widehat{X} \rangle_{T_m} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}} - \int_{T_{m-1}}^{T_m} \sigma_u^2 du$  has, to first order, a martingale structure. This can be seen from the Appendix.

An advantage of our proposed estimator is that if  $\epsilon_{t_i}$  has different variance in different time segments, say  $\text{Var}(\epsilon_{t_i}) = \nu_m$  for  $t_i \in (T_{m-1}, T_m]$ , then both consistency and asymptotic (mixed) normality continue to hold, provided that one replaces  $\nu$  by  $\nu_m$ . This adds a measure of robustness to the procedure. If one were convinced that  $\nu$  is the same across time segments, an alternative estimator has the form

$$\langle \widehat{X}, \widehat{X} \rangle_t = [Y, Y]_t^{(avg)} - \left( \frac{1}{K} \#\{t_{i+1} \leq t\} - 1 \right) \frac{1}{n} [Y, Y]_t^{(all)}. \quad (44)$$

The errors  $\langle \widehat{X}, \widehat{X} \rangle_{T_m} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}} - \int_{T_{m-1}}^{T_m} \sigma_u^2 du$ , however, are in this case not asymptotically independent. Note that for  $T = T_m$ , both candidates (43) and (44) for  $\langle \widehat{X}, \widehat{X} \rangle_t$  coincide with the quantity in (33).

## 6 DETERMINING THE ASYMPTOTIC VARIANCE

In the one period case, the main goal is to find the spread  $s^2 = 8c^{-2}\nu^2 + c\eta^2T$ , cf. (36)-(37). The multigrid case is a straightforward generalization, as indicated in Section 5.

Here, we shall only be concerned with the case where the points  $t_i$  are equally spaced ( $\Delta t_i = \Delta t$ ), and are regularly allocated to the grids  $\mathcal{A}_1 = \{\mathcal{G}^{(k)}, k = 1, \dots, K_1\}$ . The more general case where the sampling intervals are irregular, and proof of the method, is treated in Zhang and Mykland (2003). A richer set of ingredients are required to find the spread than just to estimate  $\langle \widehat{X}, \widehat{X} \rangle_T$ . To implement the estimator, create an additional family  $\mathcal{A}_2 = \{\mathcal{G}^{(k,i)}, k = 1, \dots, K_1, i = 1, \dots, I\}$  of grids where  $\mathcal{G}^{(k,i)}$  contains every  $i$ -th point of  $\mathcal{G}^{(k)}$ . We assume that  $K_1 \sim c_1 n^{2/3}$ . The new family then consists of  $K_2 \sim c_2 n^{2/3}$  grids, where  $c_2 = c_1 I$ .

In addition, we need a division of the time line into segments  $(T_{m-1}, T_m]$ , where  $T_m = \frac{m}{M}T$ . For the purposes of this discussion,  $M$  is large but finite.

We now get an initial estimator of spread as

$$\hat{s}_0^2 = n^{1/3} \sum_{m=1}^M \left( \langle \widehat{X}, \widehat{X} \rangle_{T_m}^{K_1} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}}^{K_1} - (\langle \widehat{X}, \widehat{X} \rangle_{T_m}^{K_2} - \langle \widehat{X}, \widehat{X} \rangle_{T_{m-1}}^{K_2}) \right)^2$$

where  $\langle \widehat{X}, \widehat{X} \rangle_t^{K_i}$  is the estimator (43) using grid family  $i$ ,  $i = 1, 2$ .

Using the discussion in Section 5, one can see that

$$\hat{s}_0^2 \approx s_0^2, \quad (45)$$

where, for  $c_1 \neq c_2$  ( $I \neq 1$ ),

$$\begin{aligned} s_0^2 &= 8\nu^2(c_1^{-2} + c_2^{-2} - c_1^{-1}c_2^{-1}) + (c_1^{1/2} - c_2^{1/2})^2 T\eta^2 \\ &= 8\nu^2 c_1^{-2} (1 + I^{-2} - I^{-1}) + c_1 (I^{1/2} - 1)^2 T\eta^2. \end{aligned} \quad (46)$$

In (45), the symbol  $\approx$  means first convergence in law as  $n \rightarrow \infty$ , and then a limit in probability as  $M \rightarrow \infty$ . Since  $\nu$  can be estimated by  $\hat{\nu} = [Y, Y]^{(all)}/2n$ , one can put hats on  $s_0^2$ ,  $\nu^2$ , and  $\eta^2$  in (46) to obtain an estimator of  $\eta^2$ . Similarly,

$$\begin{aligned} s^2 &= 8\nu^2 \left( c^{-2} - \frac{c(c_1^{-2} + c_2^{-2} - c_1^{-1}c_2^{-1})}{(c_1^{1/2} - c_2^{1/2})^2} \right) + \frac{c}{(c_1^{1/2} - c_2^{1/2})^2} s_0^2 \\ &= 8 \left( c^{-2} - cc_1^{-3} \frac{I^{-2} - I^{-1} + 1}{(I^{1/2} - 1)^2} \right) \nu^2 + \frac{c}{c_1 (I^{1/2} - 1)^2} s_0^2, \end{aligned} \quad (47)$$

where  $c \sim Kn^{-2/3}$  where  $K$  is the number of grids used originally to estimate  $\langle X, X \rangle_T$ .

Normally, one would take  $c_1 = c$ . Hence an estimator  $\hat{s}^2$  can be found from  $\hat{s}_0^2$  and  $\hat{\nu}$ . When  $c_1 = c$ , we argue that the optimal choice is  $I = 3$  or  $4$ , as follows. The coefficients in (47) become

$$\begin{aligned} \text{coeff}(s_0^2) &= (I^{1/2} - 1)^{-2} \\ \text{coeff}(\nu^2) &= 8c^{-2}(I^{1/2} - 1)^{-2} f(I) \end{aligned}$$

where  $f(I) = I - 2I^{1/2} - I^{-2} + I^{-1}$ . For  $I \geq 2$ ,  $f(I)$  is increasing, and  $f(I)$  crosses 0 for  $I$  between 3 and 4. These, therefore, are the two integer values of  $I$  which give the lowest ratio of  $\text{coeff}(\nu^2)/\text{coeff}(s_0^2)$ . Using  $I = 3$  or  $4$ , therefore, would maximally insulate against  $\hat{\nu}^2$  dominating over  $\hat{s}_0^2$ . This is desirable as  $\hat{s}_0^2$  is the estimator of carrying the information about  $\eta^2$ . Numerical values for the coefficients are given in Table 1. If  $c$  is such that  $\hat{\nu}^2$  still overwhelms  $\hat{s}_0^2$ , then a choice of  $c_1 \neq c$  should be considered.

Table 1. Coefficients of  $\hat{\nu}^2$  and  $\hat{s}^2$  when  $c_1 = c$

$I$	$\text{coeff}(s_0^2)$	$\text{coeff}(\nu^2)$
3	1.866	$-3.611c^{-2}$
4	1.000	$1.5000c^{-2}$

## 7 THE BENEFITS OF SAMPLING SPARSELY

In the above, we have constructed a method to directly estimate the integrated volatility of the process  $X$ , by combining different sampling frequencies. If one really insists, however, one can pretend that the noise term  $\epsilon$  is so negligible that one can ignore it. In the following, we discuss whether this approach can possibly have at least some merit.



## 7.1 The Single Grid Case

In Section 2, we argued that the realized volatility estimates the wrong quantity. This problem only gets worse when observations are sampled more frequently. Its financial interpretation boils down to market micro-structure, measured by  $\epsilon$  in (4). As the data record is sampled finely, the change in true returns gets smaller while the microstructure noise, such as bid-ask spread and transaction cost, remains at the same magnitude. In other words, when the sampling frequency is extremely high, the observed fluctuation in the returns process is more heavily contaminated by microstructure noise and becomes less representative of the true variation  $\langle X, X \rangle_T$  of the returns. Along this line of discussion, the broad opinion in financial applications is not to sample too often, at least when using realized volatility. We now discuss how this can be viewed in the context of the model (4) with stochastic volatility.

Intuitively, suppose that  $\nu$  is small. It could formally be taken to tend to zero as  $n \rightarrow \infty$ , along with  $E\epsilon^4$ . The asymptotic normality in Section 2.2 then takes the form

$$[Y, Y]_T \stackrel{\mathcal{L}}{\approx} [X, X]_T + 2\nu n + 2\sqrt{nE\epsilon^4}Z_\epsilon, \quad (48)$$

where the symbol “ $\stackrel{\mathcal{L}}{\approx}$ ” is used in a similar way to that of Section 3.5. Here  $Z_\epsilon$  is standard normal, the subscript  $\epsilon$  indicates that the randomness comes from the noise, or the deviation of the observables  $Y$  from the true process  $X$ . The convergence in law is conditional on the  $X$  process.

Following Rootzen (1980), Jacod and Protter (1998) and Mykland and Zhang (2002), and under the conditions stated in these papers, one can show that

$$\left(\frac{n}{T}\right)^{1/2} ([X, X]_T - \langle X, X \rangle_T) \xrightarrow{\mathcal{L}} \left( \int_0^T 2H'(t)\sigma_t^4 dt \right)^{1/2} \times Z_{\text{discr}}, \quad (49)$$

stably in law (see the end of Section 3.4).  $Z_{\text{discr}}$  is standard normal random variable, the subscript indicates that the randomness is due to the discretization effect in  $[X, X]_T$  when evaluating  $\langle X, X \rangle_T$ .  $H(t)$  is the asymptotic quadratic variation of time, as discussed in Mykland and Zhang (2002). In the case of equidistant observations which we consider here,  $\Delta t_0 = \dots = \Delta t_{n-1} = \Delta t$  and  $H'(t) = 1$ . Again the convergence is in law, and it is stable, cf. the end of Section 3.4. Since the  $\epsilon$ 's are independent of the  $X$  process,  $Z_\epsilon$  is independent of  $Z_{\text{discr}}$ .

For small  $\nu$ , one now has a chance at estimating  $\langle X, X \rangle_T$ . It follows from (48)-(49) that

$$[Y, Y]_T \stackrel{\mathcal{L}}{\approx} \langle X, X \rangle_T + 2\nu n + \Upsilon Z_{\text{total}}, \quad (50)$$

in the sense of stable convergence, where  $Z_{\text{total}}$  is standard normal, and where the variance has the form

$$\Upsilon^2 = 4nE\epsilon^4 + \frac{T}{n} \int_0^T 2H'(t)\sigma_t^4 dt \quad (51)$$

Seen from this angle, there is scope for using the realized volatility  $[Y, Y]$  to estimate  $\langle X, X \rangle$ . It is achieved with bias  $2\nu n$ , but the bias goes down if one uses fewer observations. This, then, is consistent with the practice in empirical finance.

As can be seen from (51), there is, however, a trade-off between sampling too often and sampling too rarely, as in the constant  $\sigma$  case. Consider again the situation where the observation times are equidistant, so that  $H'(t) = 1$  independently of the sampling frequency. It is then natural to minimize the mean squared error

$$MSE = (2\nu n)^2 + \Upsilon^2, \quad (52)$$

which means that one should choose  $n$  to satisfy  $\partial MSE/\partial n \approx 0$ , in other words,

$$8\nu^2 n + 4E\epsilon^4 - \frac{T}{n^2} \int_0^T 2H'(t)\sigma_t^4 dt \approx 0. \quad (53)$$

To solve for  $n$ , we suppose as mentioned above that  $\nu \rightarrow 0$  as  $n \rightarrow \infty$ , and we suppose that  $E\epsilon^4/(E\epsilon^2)^2$  is of order  $O(1)$ . Thus

$$n^3 + \frac{1}{2}n^2 \frac{E\epsilon^4}{(E\epsilon^2)^2} - \nu^{-2} \frac{T}{8} \int_0^T 2H'(t)\sigma_t^4 dt \approx 0. \quad (54)$$

Hence, finally,

$$n = \nu^{-2/3} \left( \frac{T}{8} \int_0^T 2H'(t)\sigma_t^4 dt \right)^{1/3} + o(\nu^{-2/3}) \quad \text{as } \nu \rightarrow 0. \quad (55)$$

The equation (55) is the formal statement saying that one can sample more frequently when the error spread is small. Note that to first order, the final trade-off is between the bias  $2\nu n$  and the variance due to discretization. The effect of the variance associated with  $Z_\epsilon$  is of lower order when comparing  $n$  and  $\nu$ . It should be emphasized that (55) is a feasible way of choosing  $n$ . One can estimate  $\nu$  using all the data following the procedure in Section 2.2. The integral  $\int_0^T 2H'(t)\sigma_t^4 dt$  can be estimated by the methods discussed in Section 6 below. For a general procedure, see Zhang and Mykland (2003).

We can do better, however, than using the ‘‘realized volatility’’, as we shall see in the following.

## 7.2 The Multiple Grid Case

Following the development in Section 3, one can go to the multi-grid case and search for an optimal frequency  $\bar{n}$  for subsampling to balance the coexistence of the bias and the variance in (30). To reduce the mean squared error of  $[Y, Y]_T^{(avg)}$ , we set  $\partial MSE/\partial \bar{n} = 0$ . From (31)-(30),  $bias = 2\nu\bar{n}$  and  $\xi^2 = 4\frac{\bar{n}}{K}E\epsilon^4 + \frac{T}{\bar{n}}\eta^2$ , then

$$MSE = bias^2 + \xi^2 = 4\nu^2\bar{n}^2 + 4\frac{\bar{n}}{K}E\epsilon^4 + \frac{T}{\bar{n}}\eta^2 = 4\nu^2\bar{n}^2 + \frac{T}{\bar{n}}\eta^2 \quad \text{to first order,}$$

thus the optimal  $\bar{n}^*$  satisfies that

$$\bar{n}^* = \left( \frac{T\eta^2}{8\nu^2} \right)^{1/3}. \quad (56)$$

Therefore, assuming the estimator  $[Y, Y]_T^{(avg)}$  is adopted, one could benefit from a minimum MSE if one subsamples  $\bar{n}^*$  data in an equidistant fashion. In other words, all  $n$  observations can be used if one uses  $K^*$ ,  $K^* \approx n/\bar{n}^*$ , subgrids. This is in contrast to the drawback of using all the data in the single grid case. The subsampling coupled with aggregation brings out the advantage of using the entire data. Of course, for the asymptotics to work, we need  $\nu^2 \rightarrow 0$ .

Our recommendation, however, is to use the methods in Sections 4 -6.

## 8 CONCLUSION

In this work, we have quantified and corrected the effect of noise on the nonparametric assessment of integrated volatility. In the setting of high frequency data, the usual financial practice is to use sparse sampling, in other words, throwing away most of the available data. We have argued that this is caused by not incorporating the noise in the model. While it is statistically unsound to throw away data, we have shown that it is possible to build on this practice to construct estimators that make statistical sense.

Specifically, we have found that the usual realized volatility mainly estimates the magnitude of the noise term rather than anything to do with volatility. An approach that is built on separating the observations into multiple “grids” lessens this problem. We found that the best results can be obtained by combining the usual (“single grid”) realized volatility with the multiple grid based device. This gives an estimator which is approximately unbiased, and we have also shown how to assess the (random) variance of this estimator. Most of the development is in the context of finding the integrated volatility over one time period; at the end, we extend this to multiple periods. Also, in the case where the noise can be taken to be almost negligible, we provide a way of optimizing the sampling frequency if one wishes to use the classical “realized volatility” or its multi-grid extension.

One important message of the paper: Any time one has an impulse to sample sparsely, one can always do better with a multi-grid method. No matter what the model is, no matter what quantity is being estimated.

## APPENDIX: PROOFS OF RESULTS

When the total grid  $\mathcal{G}$  is considered, we use  $\sum_{i=1}^{n-1}$ ,  $\sum_{t_{i+1} \leq T}$  and  $\sum_{t_i \in \mathcal{G}}$  interchangeably in the following proofs. And we write  $|X$  to indicate expressions that are conditional on the entire  $X$  process.

### A.1 Variance of $[Y, Y]_T$ Given the $X$ Process

We here calculate explicitly the variance in equation (11), from which the stated approximation follows. The explicit remainder term is also used for equation (17).

Let a partition of  $[0, T]$  be  $0 = t_0 \leq t_1 \leq \dots \leq t_n = T$ . Under assumption (6),

$$\begin{aligned} \text{Var}([Y, Y]_T|X) &= \text{Var}\left[\sum_{t_{i+1} \leq T} (\Delta Y_{t_i})^2 | X\right] \\ &= \underbrace{\sum_{t_{i+1} \leq T} \text{Var}[(\Delta Y_{t_i})^2 | X]}_{I_T} + 2 \underbrace{\sum_{t_{i+1} \leq T} \text{Cov}[(\Delta Y_{t_{i-1}})^2, (\Delta Y_{t_i})^2 | X]}_{II_T} \end{aligned}$$

since  $\Delta Y_{t_i} = \Delta X_{t_i} + \Delta \epsilon_{t_i}$  is 1-dependent given  $X$  process.

$$\begin{aligned} \text{Var}[(\Delta Y_{t_i})^2 | X] &= \kappa_4(\Delta Y_{t_i} | X) + 2[\text{Var}(\Delta Y_{t_i} | X)]^2 \\ &\quad + 4[E(\Delta Y_{t_i} | X)]^2 \text{Var}(\Delta Y_{t_i} | X) \\ &\quad + 4E(\Delta Y_{t_i} | X) \kappa_3(\Delta Y_{t_i} | X) \\ &= \kappa_4(\Delta \epsilon_{t_i}) + 2[\text{Var}(\Delta \epsilon_{t_i})]^2 + 4(\Delta X_{t_i})^2 \text{Var}(\Delta \epsilon_{t_i}) \\ &\quad + 4(\Delta X_{t_i}) \kappa_3(\Delta \epsilon_{t_i}), \text{ under Assumption 6} \\ &= 2\kappa_4(\epsilon) + 8\nu^2 + 8(\Delta X_{t_i})^2 \nu \end{aligned}$$

since  $\kappa_3(\Delta \epsilon_{t_i}) = 0$ . The  $\kappa$ 's are the cumulants of the relevant order. So,  $I_T = n(2\kappa_4(\epsilon) + 8\nu^2) + 8\nu[X, X]_T$ .

Similarly, for the covariance,

$$\begin{aligned} \text{Cov}[(\Delta Y_{t_{i-1}})^2, (\Delta Y_{t_i})^2 | X] &= \text{Cov}[(\Delta \epsilon_{t_{i-1}})^2, (\Delta \epsilon_{t_i})^2] + 4(\Delta X_{t_{i-1}})(\Delta X_{t_i}) \text{Cov}(\Delta \epsilon_{t_{i-1}}, \Delta \epsilon_{t_i}) \\ &\quad + 2(\Delta X_{t_{i-1}}) \text{Cov}[\Delta \epsilon_{t_{i-1}}, (\Delta \epsilon_{t_i})^2] + 2(\Delta X_{t_i}) \text{Cov}[(\Delta \epsilon_{t_{i-1}})^2, \Delta \epsilon_{t_i}] \\ &= \kappa_4(\epsilon) + 2\nu^2 - 4(\Delta X_{t_{i-1}})(\Delta X_{t_i}) \kappa_2(\epsilon) \\ &\quad - 2(\Delta X_{t_i}) \kappa_3(\epsilon) + 2(\Delta X_{t_{i-1}}) \kappa_3(\epsilon) \end{aligned} \tag{A.1}$$

because  $\kappa_1(\epsilon) = 0$ ,  $\kappa_2(\epsilon) = \text{Var}(\epsilon) = E(\epsilon^2)$ ,  $\kappa_3(\epsilon) = E\epsilon^3$ , and  $\kappa_4(\epsilon) = E(\epsilon^4) - 3\nu^2$ .

Thus, assuming the coefficients in (A.1)

$$\begin{aligned} II_T &= 2(n-1)(\kappa_4(\epsilon) + 2\nu^2) \\ &\quad - 8\nu \sum_{t_{i+1} \leq T} (\Delta X_{t_{i-1}})(\Delta X_{t_i}) - 4\kappa_3(\epsilon)(\Delta X_{t_{n-1}} - \Delta X_{t_0}) \end{aligned}$$

Amalgamating the two expressions one obtains

$$\begin{aligned} \text{Var}([Y, Y]_T|X) &= n(2\kappa_4(\epsilon) + 8\nu^2) + 8\nu[X, X]_T + 2(n-1)(\kappa_4(\epsilon) + 2\nu^2) \\ &\quad - 8\nu \sum (\Delta X_{t_{i-1}})(\Delta X_{t_i}) - 4\kappa_3(\epsilon)(\Delta X_{t_{n-1}} - \Delta X_{t_0}) \\ &= 4nE\epsilon^4 + R_n, \end{aligned} \tag{A.2}$$

where the remainder term  $R_n$  satisfies

$$\begin{aligned} |R_n| &\leq 8\nu[X, X]_T + 2(\kappa_4(\epsilon) + 2\nu^2) \\ &\quad + 8\nu \left| \sum (\Delta X_{t_{i-1}}) |(\Delta X_{t_i})| + 4|\kappa_3(\epsilon)| (|\Delta X_{t_{n-1}}| + |\Delta X_{t_0}|) \right| \\ &\leq 16\nu[X, X]_T + 2(\kappa_4(\epsilon) + 2\nu^2) + 2|\kappa_3(\epsilon)|(2 + [X, X]_T) \end{aligned} \quad (\text{A.3})$$

by the Cauchy-Schwarz inequality and since  $|x| \leq (1 + x^2)/2$ .

Since  $[X, X]_T = O_p(1)$ , (11) follows. Similarly, (17) follows since  $[X, X]_T^{(avg)} = O_p(1)$ .

## A.2 Relevant Central Limit Theorem

**Lemma A.1.** *Suppose  $X$  is an Itô process. Suppose  $Y$  is related to  $X$  through model (4). Then under assumption (6) and definitions (8) and (13),*

$$[Y, Y]_T^{(all)} = [\epsilon, \epsilon]_T^{(all)} + O_p(1), \text{ and } [Y, Y]_T^{(avg)} = [\epsilon, \epsilon]_T^{(avg)} + [X, X]_T^{(avg)} + O_p\left(\frac{1}{\sqrt{K}}\right)$$

### Proof of Lemma A.1

(a) The one grid case:

$$[Y, Y]_T^{(all)} = [X, X]_T^{(all)} + [\epsilon, \epsilon]_T^{(all)} + 2[X, \epsilon]_T^{(all)} \quad (\text{A.4})$$

We show:

$$E \left( ([X, \epsilon]_T^{(all)})^2 | X \right) = O_p(1) \quad (\text{A.5})$$

and in particular

$$[X, \epsilon]_T^{(all)} = O_p(1) \quad (\text{A.6})$$

To see (A.5):

$$\begin{aligned} [X, \epsilon]_T^{(all)} &= \sum_{i=0}^{n-1} (\Delta X_{t_i}) (\Delta \epsilon_{t_i}) \\ &= \sum_{i=0}^{n-1} (\Delta X_{t_i}) \epsilon_{t_{i+1}} - \sum_{i=0}^{n-1} (\Delta X_{t_i}) \epsilon_{t_i} \\ &= \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}} - \Delta X_{t_i}) \epsilon_{t_i} + \Delta X_{t_{n-1}} \epsilon_{t_n} - \Delta X_{t_0} \epsilon_{t_0} \end{aligned}$$

Since  $E([X, \epsilon]_T^{(all)} | X) = 0$  and  $\epsilon_{t_i}$  i.i.d. for different  $t_i$ , we get

$$\begin{aligned}
E\left(\left([X, \epsilon]_T^{(all)}\right)^2 | X\right) &= \text{Var}([X, \epsilon]_T^{(all)} | X) \\
&= \nu \left[ \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}} - \Delta X_{t_i})^2 + \Delta X_{t_{n-1}}^2 + \Delta X_{t_0}^2 \right] \\
&= 2\nu [X, X]_T - 2\nu \sum_{i=1}^{n-1} (\Delta X_{t_{i-1}})(\Delta X_{t_i}) \\
&\leq 4\nu [X, X]_T
\end{aligned} \tag{A.7}$$

by the Cauchy-Schwarz Inequality, from which and from  $[X, X]^{(all)}$  being of order  $O_p(1)$ , (A.5) follows. Hence (A.6) follows by the Markov Inequality.

(b) The multiple grid case:

Notice that

$$[Y, Y]^{(avg)} = [X, X]_T^{(avg)} + [\epsilon, \epsilon]_T^{(avg)} + 2[X, \epsilon]_T^{(avg)} \tag{A.8}$$

(A.8) strictly follows from model (4) and the definitions of grids and  $[\cdot, \cdot]_t^{(avg)}$ , see Section 3.2.

We need to show:

$$E([X, \epsilon]_T^{(avg)} | X) = O_p\left(\frac{1}{K}\right) \tag{A.9}$$

in particular

$$[X, \epsilon]_T^{(avg)} = O_p\left(\frac{1}{K^{1/2}}\right). \tag{A.10}$$

and  $\text{Var}([X, \epsilon]_T^{(avg)} | X) = E\left(\left([X, \epsilon]_T^{(avg)}\right)^2 | X\right)$ .

To show (A.9), note that  $E([X, \epsilon]_T^{(avg)} | X) = 0$ ,

$$\begin{aligned}
E\left(\left([X, \epsilon]_T^{(avg)}\right)^2 | X\right) &= \text{Var}([X, \epsilon]_T^{(avg)} | X) \\
&= \frac{1}{K^2} \sum_{k=1}^K \text{Var}([X, \epsilon]_T^{(k)} | X) \\
&\leq \frac{4\nu}{K} [X, X]_T^{(avg)} = O_p\left(\frac{1}{K}\right)
\end{aligned}$$

where the second equality follows from the disjointness of different grids as well as  $\epsilon \perp\!\!\!\perp X$ . The inequality follows from the same argument as in (A.7). Then the order follows since  $[X, X]_T^{(avg)} = O_p(1)$ : see the method in Mykland and Zhang (2002) if one wants a rigorous development for the order of  $[X, X]_T^{(avg)}$ .

**Theorem A.1.** *Suppose  $X$  is an Itô process of form (1). Suppose  $Y$  is related to  $X$  through model (4), and that (6) is satisfied with  $E\epsilon^4 < \infty$ . Also suppose that  $t_i$  and  $t_{i+1}$  is not in the same subgrid*

for any  $i$ . Under assumption (14), as  $n \rightarrow \infty$ ,  $(\sqrt{n}(\hat{\nu} - \nu), \sqrt{\frac{K}{n}}([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\nu\bar{n}))$  converges in law to a bivariate normal, with mean 0 and covariance matrix

$$\begin{pmatrix} E\epsilon^4 & 2\text{Var}(\epsilon^2) \\ 2\text{Var}(\epsilon^2) & 4E\epsilon^4 \end{pmatrix} \quad (\text{A.11})$$

conditional on  $X$  process. where the limiting random variable is independent of the  $X$  process.

### Proof of Theorem A.1:

By Lemma A.1, need the distribution of  $[\epsilon, \epsilon]^{(avg)}$  and  $[\epsilon, \epsilon]^{(all)}$ .

First, we explore the convergence of

$$\frac{1}{\sqrt{n}} \left( [\epsilon, \epsilon]_T^{(all)} - 2\nu n, [\epsilon, \epsilon]_T^{(avg)} K - 2\nu\bar{n}K \right) \quad (\text{A.12})$$

Recall that all the sampling points  $t_0, t_1, \dots, t_n$  are within  $[0, T]$ . We use  $\mathcal{G}$  to denote the time points in the full sampling, as in the single grid.  $\mathcal{G}^{(k)}$  denotes the subsamplings from  $k^{\text{th}}$  grid.

As before, if  $t_i \in \mathcal{G}^{(k)}$ , then  $t_{i,-}$  and  $t_{i,+}$  are, respectively, the previous and next element in  $\mathcal{G}^{(k)}$ .  $\epsilon_{t_i,-} = 0$  for  $t_i = \min \mathcal{G}^{(k)}$  and  $\epsilon_{t_i,+} = 0$  for  $t_i = \max \mathcal{G}^{(k)}$ .

Set

$$\begin{aligned} M_T^{(1)} &= \frac{1}{\sqrt{n}} \sum_{t_i \in \mathcal{G}} (\epsilon_{t_i}^2 - \nu) \\ M_T^{(2)} &= \frac{1}{\sqrt{n}} \sum_{t_i \in \mathcal{G}} \epsilon_{t_i} \epsilon_{t_{i-1}} \\ M_T^{(3)} &= \frac{1}{\sqrt{n}} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_i} \epsilon_{t_{i,-}} \end{aligned} \quad (\text{A.13})$$

We first find the asymptotic distribution of  $(M_T^{(1)}, M_T^{(2)}, M_T^{(3)})$  using the martingale central limit theorem, and then we use the result to find the limit of (A.12).

Note that  $(M_T^{(1)}, M_T^{(2)}, M_T^{(3)})$  are the end points of martingales with respect to filtration  $\mathcal{F}_i = \sigma(\epsilon_{t_j}, j \leq i, X_t, \text{all } t)$ . We now derive its (discrete-time) predictable quadratic variation  $\langle M^{(l)}, M^{(k)} \rangle$ ,  $l, k = 1, 2, 3$ . (Discrete time predictable quadratic variations are only used in this proof, and are different from the continuous time quadratic variations in (7)).

$$\begin{aligned} \langle M^{(1)}, M^{(1)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \text{Var}(\epsilon_{t_i}^2 - \nu \mid \mathcal{F}_{t_{i-1}}) = \text{Var}(\epsilon^2) \\ \langle M^{(2)}, M^{(2)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \text{Var}(\epsilon_{t_i} \epsilon_{t_{i-1}} \mid \mathcal{F}_{t_{i-1}}) = \frac{\nu}{n} \sum_{t_i \in \mathcal{G}} \epsilon_{t_{i-1}}^2 = \nu^2 + o_p(1) \\ \langle M^{(3)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \text{Var}(\epsilon_{t_i} \epsilon_{t_{i,-}} \mid \mathcal{F}_{t_{i-1}}) = \frac{\nu}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i,-}}^2 = \nu^2 + o_p(1) \end{aligned}$$

by the law of large numbers.

Similarly, for the predictable quadratic covariations,

$$\begin{aligned}
\langle M^{(1)}, M^{(2)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \text{Cov}(\epsilon_{t_i}^2 - \nu, \epsilon_{t_i} \epsilon_{t_{i-1}} \mid \mathcal{F}_{t_{i-1}}) = E \epsilon^3 \frac{1}{n} \sum_{t_i \in \mathcal{G}} \epsilon_{t_{i-1}} = o_p(1) \\
\langle M^{(1)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \text{Cov}(\epsilon_{t_i}^2 - \nu, \epsilon_{t_i} \epsilon_{t_{i,-}} \mid \mathcal{F}_{t_{i-1}}) = E \epsilon^3 \frac{1}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i,-}} = o_p(1) \\
\langle M^{(2)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \text{Cov}(\epsilon_{t_i} \epsilon_{t_{i-1}}, \epsilon_{t_i} \epsilon_{t_{i,-}} \mid \mathcal{F}_{t_{i-1}}) \\
&= \frac{\nu}{n} \sum_{k=1}^K \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i-1}} \epsilon_{t_{i,-}} = o_p(1)
\end{aligned}$$

since  $t_{i+1}$  is not in the same grid as  $t_i$ .

Since the  $\epsilon_{t_i}$ 's are i.i.d. and  $E \epsilon_{t_i}^4 < \infty$ , the conditional Lindeberg conditions are satisfied. Hence by the martingale CLT (see condition 3.1, p. 58 of Hall and Heyde (1980)),  $(M^{(1)}, M^{(2)}, M^{(3)})$  are asymptotically normal, with covariance matrix as the asymptotic value of  $\langle M^{(l)}, M^{(k)} \rangle$ . In other words, asymptotically,  $(M^{(1)}, M^{(2)}, M^{(3)})$  are independent normal with respective variances  $\text{Var}(\epsilon)$ ,  $\nu^2$ , and  $\nu^2$ .

Returning to (A.12),

$$\begin{aligned}
[\epsilon, \epsilon]^{(all)} - 2n\nu &= 2 \sum_{i \neq 0, n} (\epsilon_{t_i}^2 - \nu) + (\epsilon_{t_0}^2 - \nu) + (\epsilon_{t_n}^2 - \nu) - 2 \sum_{t_i > 0} \epsilon_{t_i} \epsilon_{t_{i-1}} \\
&= 2\sqrt{n}(M^{(1)} - M^{(2)}) + O_p(1)
\end{aligned} \tag{A.14}$$

Meanwhile:

$$\begin{aligned}
[\epsilon, \epsilon]^{(k)} - 2n_k\nu &= \sum_{t_i \in \mathcal{G}^{(k)}, t_i \neq \max \mathcal{G}^{(k)}} (\epsilon_{t_{i,+}} - \epsilon_{t_i})^2 - 2n_k\nu \\
&= 2 \sum_{t_i \in \mathcal{G}^{(k)}} (\epsilon_{t_i}^2 - \nu) - (\epsilon_{\min \mathcal{G}^{(k)}}^2 - \nu) - (\epsilon_{\max \mathcal{G}^{(k)}}^2 - \nu) \\
&\quad - 2 \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_i} \epsilon_{t_{i,-}}
\end{aligned} \tag{A.15}$$

where  $n_k + 1$  is the total number of sampling points in  $\mathcal{G}^{(k)}$ .

Hence,

$$[\epsilon, \epsilon]_T^{(avg)} K - 2\bar{n}\nu K = \sqrt{n}(2M^{(1)} - 2M^{(3)}) - R = 2\sqrt{n}(M^{(1)} - M^{(3)}) + O_p(K^{1/2}), \tag{A.16}$$

since  $R = \sum_{k=1}^K \left[ (\epsilon_{\min \mathcal{G}^{(k)}}^2 - \nu) + (\epsilon_{\max \mathcal{G}^{(k)}}^2 - \nu) \right]$  satisfying  $ER^2 = \text{Var}(R) \leq 4KV \text{Var}(\epsilon^2)$ .



Since  $n^{-1}K \rightarrow 0$ , and since the error terms in (A.14) and (A.15) are uniformly integrable, it follows that

$$(A.12) = 2(M^{(1)} - M^{(2)}, M^{(1)} - M^{(3)}) + o_p(1) \quad (A.17)$$

Hence, (A.12) is also asymptotically normal with covariance matrix

$$\begin{pmatrix} 4E\epsilon^4 & 4\text{Var}(\epsilon^2) \\ 4\text{Var}(\epsilon^2) & 4E\epsilon^4 \end{pmatrix}.$$

By Lemma A.1, and as  $n^{-1}K \rightarrow 0$ ,

$$\frac{1}{\sqrt{n}} \left( [Y, Y]_T^{(all)} - 2\nu n, K([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\nu\bar{n}) \mid X \right)$$

is asymptotically normal,

$$\begin{aligned} \frac{1}{\sqrt{n}} \left( \begin{array}{c} [Y, Y]_T^{(all)} - 2\nu n \\ [Y, Y]_T^{(avg)} K - 2\nu\bar{n}K - [X, X]_T^{(avg)} K \end{array} \mid X \right) &= 2 \begin{pmatrix} M^{(1)} - M^{(2)} \\ M^{(1)} - M^{(3)} \end{pmatrix} + o_p(1) \\ &\xrightarrow{\mathcal{L}} 2N \left( 0, \begin{pmatrix} E\epsilon^4 & \text{Var}(\epsilon^2) \\ \text{Var}(\epsilon^2) & E\epsilon^4 \end{pmatrix} \right) \end{aligned} \quad (A.18)$$

Since

$$\hat{\nu} = \frac{1}{2n} [Y, Y]_T^{(all)} \quad \text{and} \quad \frac{K}{\sqrt{n}} = \sqrt{\frac{K}{n}} (1 + o(1)), \quad (A.19)$$

Theorem A.1 follows.

### A.3 Asymptotics of $D_T$

For transparency of notation, we take  $\bar{\Delta}t = T/n$ , in other words, the average of the  $\Delta t_i$ . For given  $s \in [0, T]$ , let  $s_-^{(k)}$  be the closest point on grid  $\mathcal{G}^{(k)}$  smaller than  $s$ , i.e.,  $s_-^{(k)} = \max\{u \leq s : u \in \mathcal{G}^{(k)}\}$ . In particular, for grid points  $t_i$ , let  $t_i^{(k)}$  be the closest point on the grid  $\mathcal{G}^{(k)}$  smaller than  $t_i$ , i.e.,  $t_i^{(k)} = \max\{u \leq t_i : u \in \mathcal{G}^{(k)}\}$ . Observe that  $t_i^{(k)} = (t_i)_-^{(k)}$ .

We do not assume regular allocation of sample points to subgrids, but instead that

$$\max_i \sum_{l=1}^K \{\#k : t_i^{(k)} > t_i^{(l)}\}^2 = O(K^3). \quad (A.20)$$

Note that  $\{\#k : t_i^{(k)} > t_i^{(l)}\}$  is the number of points in the total grid  $\mathcal{G}$  between  $t_i^{(l)}$  and  $t_i$ . The requirement (A.20) is satisfied under regular allocation of sample points to subgrids, as defined in Section 3.2, in other words,  $\mathcal{G}^{(l)} = \{t_{l-1}, t_{K+l-1}, \dots\}$ .

**Proof of Theorem 2.** Rewrite

$$\begin{aligned}
D_T &= [X, X]_T^{(avg)} - \langle X, X \rangle_T \\
&= \frac{1}{K} \sum_{k=1}^K ([X, X]_T^{(k)} - \langle X, X \rangle_T) \\
&= \frac{1}{K} \sum_{k=1}^K \sum_{\{t_i, t_{i,+} \in \mathcal{G}^{(k)}\}} 2 \int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) dX_s \text{ by It\^o's formula} \\
&= 2 \int_0^T \frac{1}{K} \sum_{k=1}^K (X_s - X_{s_-^{(k)}}) dX_s
\end{aligned}$$

Denote the integrand as  $Z_s$ . We can write  $Z_s = X_s - \frac{1}{K} \sum_{k=1}^K X_{s_-^{(k)}}$

Following the arguments in Mykland and Zhang (2002), the quadratic variation of  $D_T$  is

$$\begin{aligned}
\langle D, D \rangle_T &= 4 \int_0^T Z_s^2 d\langle X, X \rangle_s \\
&= 4 \int_0^T \langle Z, Z \rangle_s d\langle X, X \rangle_s + o_p\left(\frac{K}{n}\right) \\
&= 4 \int_0^T \langle Z, Z \rangle_s \langle X, X \rangle'_s ds + o_p\left(\frac{K}{n}\right) \\
&= 4 \sum_i \int_{t_i}^{t_{i+1}} \langle Z, Z \rangle_s \langle X, X \rangle'_s ds + o_p\left(\frac{K}{n}\right),
\end{aligned}$$

where the sum is over all (except the last) observation points  $t_i$ . To calculate the integrand, note that for  $t_i \leq s < t_{i+1}$ ,

$$\begin{aligned}
\langle Z, Z \rangle_s \langle X, X \rangle'_s &= \frac{1}{K^2} \sum_{k=1}^K \sum_{l=1}^K (\langle X, X \rangle_s - \langle X, X \rangle_{s_-^{(k)} \wedge s_-^{(l)}}) \langle X, X \rangle'_s \\
&= \frac{1}{K^2} \sum_{l=1}^K (\langle X, X \rangle_s - \langle X, X \rangle_{s_-^{(l)}}) (2\{\#k : s_-^{(k)} > s_-^{(l)}\} + 1) \langle X, X \rangle'_s \\
&= \frac{1}{K^2} \sum_{l=1}^K (s - t_i^{(l)}) (2\{\#k : t_i^{(k)} > t_i^{(l)}\} + 1) (\langle X, X \rangle'_{t_i})^2 + o_p\left(\frac{K}{n}\right),
\end{aligned}$$

since the  $s_-^{(l)} = t_i^{(l)}$  as  $s$  varies over the relevant time interval  $[t_i, t_{i+1})$ . Hence

$$\begin{aligned}
 & \int_{t_i}^{t_{i+1}} \langle Z, Z \rangle_s \langle X, X \rangle'_s ds \\
 &= \frac{1}{K^2} \sum_{l=1}^K \left( \int_{t_i}^{t_{i+1}} (s - t_i^{(l)}) ds \right) (2\{\#k : t_i^{(k)} > t_i^{(l)}\} + 1) (\langle X, X \rangle'_{t_i})^2 + o_p\left(\frac{K}{n^2}\right) \\
 &= \frac{1}{K^2} \sum_{l=1}^K \left( \frac{1}{2} \Delta t_i^2 + (t_i - t_i^{(l)}) \Delta t_i \right) (2\{\#k : t_i^{(k)} > t_i^{(l)}\} + 1) (\langle X, X \rangle'_{t_i})^2 + o_p\left(\frac{K}{n^2}\right) \\
 &= \frac{1}{4} \bar{\Delta} t K h_i (\langle X, X \rangle'_{t_i})^2 \Delta t_i + o_p\left(\frac{K}{n^2}\right)
 \end{aligned}$$

where the  $h_i$  are defined by (22). Hence, since the error term above is uniform in  $i$  (in probability),

$$\langle D, D \rangle_T = \bar{\Delta} t K \sum_i h_i (\langle X, X \rangle'_{t_i})^2 \Delta t_i + o_p\left(\frac{K}{n}\right) \quad (\text{A.21})$$

thus showing Theorem 2.

We now proceed to the asymptotic distribution of  $D_T$ . We first state a technical condition on the filtration  $(\mathcal{F}_t)_{0 \leq t \leq T}$  to which  $X_t$  and  $\mu_t$  (but not the  $\epsilon'_s$ ) are assumed to be adapted.

**Condition E** (*Description of the filtration*): *There is a continuous multidimensional  $P$ -local martingale  $\mathcal{X} = (\mathcal{X}^{(1)}, \dots, \mathcal{X}^{(p)})$ , any  $p$ , so that  $\mathcal{F}_t$  is the smallest sigma-field containing  $\sigma(\mathcal{X}_s, s \leq t)$  and  $\mathcal{N}$ , where  $\mathcal{N}$  contains all the null sets in  $\sigma(\mathcal{X}_s, s \leq T)$ .*

For example,  $\mathcal{X}$  can be a collection of Brownian motions.

**Proof of Theorem 3.** One shows by methods similar to those in the proof of Theorem 2 that if  $L$  is any martingale adapted to the filtration generated by  $\mathcal{X}$ , then

$$\sup_t \left| \frac{1}{\Delta t K} \langle D, L \rangle_t \right| \xrightarrow{p} 0, \quad (\text{A.22})$$

The stable convergence with respect to the filtration  $(\mathcal{F}_t)_{0 \leq t \leq T}$  then follows in view of Rootzen (1980) or Jacod and Protter (1998). This ends the proof of Theorem 3.

Finally, in the case where  $\eta^2$  does not converge, one can still use the mixed normal with variance  $\eta_n^2$ . This is because every subsequence of  $\eta_n^2$  has a further subsequence which does converge in probability to some  $\eta^2$  in probability, and hence for which the assumption (27) in Theorem 3 would be satisfied.

The reason for this is that one can define the distribution function of a finite measure by

$$G_n(t) = \sum_{t_{i+1} \leq t} h_i \Delta t_i \quad (\text{A.23})$$

Since  $G_n(t) \leq T \sup_i h_i$ , it follows from (24) that the sequence  $G_n$  is weakly compact in the sense of weak convergence (see Helly's Theorem, e.g. Billingsley (1995) p. 336). For any convergent subsequence  $G_n \rightarrow G$ , we then get that

$$\eta_n^2 = \int_0^T (\langle X, X \rangle'_t)^2 dG_n(t) \rightarrow \int_0^T (\langle X, X \rangle'_t)^2 dG(t). \quad (\text{A.24})$$

almost surely, since we have assumed  $\langle X, X \rangle'_t$  to be a continuous function of  $t$ . One then defines  $\eta^2$  to be the (subsequence dependent) right hand side of (A.24).

To proceed further with the asymptotics, continue the subsequence from above, and note that

$$\begin{aligned} \frac{1}{\Delta t K} \langle D, D \rangle_t &\approx \int_0^t (\langle X, X \rangle'_s)^2 dG_n(s) \\ &\rightarrow \int_0^t (\langle X, X \rangle'_s)^2 dG(s) \end{aligned}$$

to conclude.

## REFERENCES

- Ait-Sahalia, Y., Mykland, P. A., and Zhang, L. (2003), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," Tech. rep., Princeton University.
- Aldous, D. J. and Eagleson, G. K. (1978), "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001), "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55.
- Bai, X., Russell, J. R., and Tiao, G. C. (2000), "Beyond Merton's Utopia I: Effects of Non-Normality and Dependence on the Precision of Variance Estimates Using High Frequency Financial Data," Tech. rep., University of Chicago.
- Barndorff-Nielsen, O. E. and Shephard, N. (2001), "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of Royal Statistical Society, B*, 63, 167–241.
- Billingsley, P. (1995), *Probability and Measure*, New York: Wiley, 3rd ed.
- Brown, S. J. (1990), "Estimating Volatility," in *Financial Options: From Theory to Practice*, eds. Figlewski, S., Silber, W., and Subrahmanyam, M., Homewood, IL: Business One-Irwin, pp. 516–537.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997), *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

- Chernov, M. and Ghysels, E. (2000), “A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation,” *Journal of Financial Economics*, 57, 407–458.
- Gallant, A. R., Hsu, C.-T., and Tauchen, G. T. (1999), “Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance,” *The Review of Economics and Statistics*, 81, 617–631.
- Gloter, A. (2000), “Estimation des Paramètres d’une Diffusion Cachée,” Ph.D. thesis, Université de Marne-la-Vallée.
- Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Application*, Boston: Academic Press.
- Hull, J. and White, A. (1987), “The Pricing of Options on Assets with Stochastic Volatilities,” *Journal of Finance*, 42, 281–300.
- Jacod, J. and Protter, P. (1998), “Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations,” *Annals of Probability*, 26, 267–307.
- Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.
- Mykland, P. A. and Zhang, L. (2002), “ANOVA for Diffusions,” Tech. rep., The University of Chicago, Department of Statistics.
- Rényi, A. (1963), “On Stable Sequences of Events,” *Sankyā Series A*, 25, 293–302.
- Rootzen, H. (1980), “Limit Distributions for the Error in Approximations of Stochastic Integrals,” *Annals of Probability*, 8, 241–251.
- Zhang, L. and Mykland, P. A. (2003), “Interval Estimation for the Variability of a Contaminated Ito Process,” Tech. rep., Carnegie-Mellon University, Department of Statistics.