A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data

Lan ZHANG, Per A. MYKLAND, and Yacine AÏT-SAHALIA

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. The present 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 frequencies. If the noise is asymptotically small, our work provides a way of finding the optimal sampling frequency. A better approach, the "two-scales estimator," works for any size of the noise.

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

1. INTRODUCTION

1.1 High-Frequency Financial Data With Noise

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. Although this approach is justified under the assumption of a continuous stochastic model in an idealized world, it runs into the challenge from market microstructure in real applications. We argue that this customary way of estimating volatility is flawed, in that it overlooks the observation error. The usual mechanism for dealing with the problem is to throw away a large fraction of the available data by sampling less frequently or constructing "time-aggregated" returns from the underlying high-frequency asset prices. Here we propose a statistically sounder device. Our device is model-free, takes advantage of the rich sources of tick-by-tick data, and to a great extent corrects for the adverse effects of microstructure noise on volatility estimation. In the course of constructing our estimator, it becomes clear why and where the "usual" volatility estimator fails when returns are sampled at high frequencies.

To fix ideas, let S_t denote the price process of a security, and suppose that the process $X_t = \log S_t$ follows an Itô process,

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

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

$$[X,X]_T \stackrel{\triangle}{=} \sum_{t_i} \bigl(X_{t_{i+1}} - X_{t_i}\bigr)^2, \tag{2}$$

where the X_{t_i}'s are all of 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." A sample of the recent literature on realized and integrated volatilities includes works by 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), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2002).

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

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

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

However, this is not the general viewpoint adopted in the empirical 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 in some instances. It has been found empirically that the realized volatility estimator is not robust when the sampling interval is small. Such issues as larger bias in the estimate and nonrobustness to changes in the sampling interval have been reported (see, e.g., Brown 1990). In pure probability terms, the observed log return process is not in fact a semimartingale. Compelling visual evidence of this can be found by comparing a plot of the realized volatility as a function of the sampling frequency with theorem I.4.47 of Jacod and Shiryaev (2003); the realized volatility does not converge as the sampling frequency increases. In particular this phenomenon can be observed for any of the 30 stocks composing the Dow Jones Industrial Average.

The main explanation for this phenomenon is a vast array of issues collectively known as *market microstructure*, including, but not limited to, the existence of the bid–ask spread. When

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, Department of Economics and Bendheim Center for Finance, Princeton University and NBER, Princeton, NJ 08544 (E-mail: *yacine@princeton.edu*). This research was supported by National Science Foundation grants DMS-02-04639 (Zhang and Mykland) and SBR-0111140 (Aït-Sahalia). The authors are grateful for comments received at the IMS New Researchers' Conference at UC Davis (July 2003), the Joint Statistical Meetings in San Francisco (August 2003) and the CIRANO Conference on Realized Volatility in Montréal (November 2003). They also thank Peter Hansen for interesting discussions and the referees for their feedback and suggestions.

Journal of the American Statistical Association
December 2005, Vol. 100, No. 472, Theory and Methods
DOI 10.1198/016214505000000169

prices are sampled at finer intervals, microstructure issues become more pronounced. The empirical finance literature then suggests that the bias induced by market microstructure effects makes the most finely sampled data unusable, and many authors prefer to sample over longer time horizons to obtain more reasonable estimates. The sampling length of the typical choices in the literature is ad hoc and usually ranges from 5 to 30 minutes for exchange rate data, for instance. If the original data are sampled once every second, say, then retaining an observation every 5 minutes amounts to discarding 299 out of every 300 data points.

This approach to handling the data poses a conundrum from the statistical standpoint. It is difficult to accept that throwing away data, especially in such quantities, can be an optimal solution. We argue here that sampling over longer horizons merely reduces the impact of microstructure, rather than quantifying and correcting its effect for volatility estimation. Ultimately, we seek a solution that makes use of all of the data, yet results in an unbiased and consistent estimator of the integrated volatility. Our contention is that the contamination due to market microstructure is, to first order, the same as what statisticians usually call "observation error." In other words, the transaction will be seen as an observation device for the underlying price process. We incorporate the observation error into the estimating procedure for integrated volatility. That is, we suppose that the return process as observed at the sampling times is of the form

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{4}$$

Here X_t is a latent true, or efficient, return process that follows (4). The ϵ'_{t_i}s are independent noise around the true return.

In the process of constructing our final estimator for the integrated volatility, which we call the "first best" approach, we are led to develop and analyze a number of intermediary estimators, starting with a "fifth-best" approach that computes realized volatility using all of the data available. This fifth-best approach has some clearly undesirable properties. In fact, we show in Section 2.2 that ignoring microstructure noise would have a devastating effect on the use of the realized volatility. Instead of (2), one gets

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

where n is the number of sampling intervals over $[0, T]$. Thus the realized volatility estimates not the true integrated volatility, but rather the variance of the contamination noise. In fact, we 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. This section also discusses why it is natural to want the quadratic variation of the latent process X, as opposed to trying to construct, say, prices with the help of some form of quadratic variation for Y.

Faced with such dire consequences, the usual empirical practice where one does not use all of the data is likely to produce an improvement. We show that this is indeed the case by analyzing the estimation approach where one selects an arbitrary sparse sampling frequency (such as one every 5 minutes, for instance). We call this the "fourth-best" approach. A natural question to ask, then, is how that sparse sampling frequency should be determined. We show how to determine it optimally by minimizing the mean squared error (MSE) of the sparse realized volatility estimator, yielding what we term the "third-best" estimator. Similar results to our third-best approach have been discussed independently (when σ_t is conditionally nonrandom) by Bandi and Russell (2003) in a paper presented at the 2003 CIRANO conference.

The third-best answer may be to sample, say, every 6 minutes and 15 seconds instead of an arbitrary 5 minutes. Thus, even if one determines the sampling frequency optimally, it remains the case that one is not using a large fraction of the available data. Driven by one of statistics' first principles—"thou shall not throw data away"—we go further by proposing two estimators that make use of the full data. Our next estimator, the "second-best," involves sampling sparsely over subgrids of observations. For example, one could subsample every 5 minutes, starting with the first observation, then starting with the second, and so on. We then average the results obtained across those subgrids. We show that subsampling and averaging result in a substantial decrease in the bias of the estimator.

Finally, we show how to bias-correct this estimator, resulting in a centered estimator for the integrated volatility despite the presence of market microstructure noise. Our final estimator, the "first-best" estimator, effectively combines the second-best estimator (an average of realized volatilities estimated over subgrids on a slow time scale of, say, 5 minutes each) with the fifth-best estimator (realized volatility estimated using all of the data) providing the bias-correction device. This combination of slow and fast time scales gives our article its title.

Models akin to (4) have been studied in the constant σ case by Zhou (1996), who proposed a bias-correcting approach based on autocovariances. The behavior of this estimator has been studied by Zumbach, Corsi, and Trapletti (2002). Other contributions in this direction have been made by Hansen and Lunde (2004a,b), who considered time-varying σ in the conditionally nonrandom case, and by Oomen (2004). Efficient likelihood estimation of σ has been considered by Aït-Sahalia, Mykland, and Zhang (2005).

The article is organized as follows. Section 1.2 provides a summary of the five estimators. The main theory for these estimators, including asymptotic distributions, is developed successively in Sections 2–4 for the case of one time period $[0, T]$. The multiperiod problem is treated in Section 5. Section 6 discusses how to estimate the asymptotic variance for equidistant sampling. Section 7 presents the results of Monte Carlo simulations for all five estimators. Section 8 concludes and provides a discussion for the case where the process contains jumps. The Appendix provides proofs.

1.2 The Five Estimators: A User's Guide

Here we provide a preview of our results in Sections 2–4. To describe the results, we proceed through five estimators, from the worst to the (so far) best candidate. In this section, all limit results and optimal choices are stated in the equidistantly sampled case; the more general results are given in the subsequent sections.

1.2.1 The Fifth-Best Approach: Completely Ignoring the Noise. The naïve estimator of quadratic variation is $[Y, Y]_T^{(all)}$, which is the realized volatility based on the entire sample. We show in Section 2 that (5) holds. We therefore conclude that realized volatility $[Y, Y]_T^{(all)}$ is not a reliable estimator for the true variation $\langle X, X\rangle_T$ of the returns. For large n, the realized volatility diverges to infinity linearly in n. Scaled by $(2n)^{-1}$, it estimates consistently the variance of microstructure noise, $E\epsilon^2$, rather than the object of interest $\langle X, X\rangle_T$. Said differently, market microstructure noise totally swamps the variance of the price signal at the level of the realized volatility. Our simulations in Section 7 also document this effect.

1.2.2 The Fourth-Best Approach: Sampling Sparsely at Some Lower Frequency. Of course, completely ignoring the noise and sampling as prescribed by $[Y, Y]_T^{(all)}$ is not what empirical researchers do in practice. Much, if not all, of the existing literature seems to have settled on sampling sparsely, that is, constructing lower-frequency returns from the available data. For example, using essentially the same exchange rate series, these somewhat ad hoc choices range from 5-minute intervals to as long as 30 minutes. That is, the literature typically uses the estimator $[Y, Y]_T^{(sparse)}$ described in Section 2.3. This involves taking a subsample of n_{sparse} observations. For example, with $T = 1$ day, or 6.5 hours of open trading for stocks, and we start with data sampled on average every $\overline{\Delta t} = 1$ second, then, for the full dataset, $n = T/\overline{\Delta t} = 23{,}400$; but once we sample sparsely every 5 minutes, then we sample every 300th observation, and $n_{\text{sparse}} = 78$.

The distribution of $[Y, Y]_T^{(sparse)}$ is described by Lemma 1 and Proposition 1, both in Section 2.3. To first approximation,

$$[Y, Y]_T^{(sparse)} \overset{\mathcal{L}}{\approx} \langle X, X\rangle_T + \underbrace{2n_{\text{sparse}}E\epsilon^2}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{4n_{\text{sparse}}E\epsilon^4}_{\text{due to noise}} + \underbrace{\frac{2T}{n_{\text{sparse}}}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}\right]^{1/2} Z_{\text{total}}}_{\text{total variance}}. \quad (6)$$

Here Z_{total} is a standard normal random variable. The symbol "$\approx^{\mathcal{L}}$" means that when multiplied by a suitable factor, the convergence is in law. For precise statements, consult Sections 2–4.

1.2.3 The Third-Best Approach: Sampling Sparsely at an Optimally Determined Frequency. Our results in Section 2.3 show how, if one insists on sampling sparsely, it is possible to determine an optimal sampling frequency instead of selecting the frequency in a somewhat ad hoc manner as in the fourth-best approach. In effect, this is similar to the fourth-best approach, except that the arbitrary n_{sparse} is replaced by an optimally determined n_{sparse}^*. Section 2.3 shows how to minimize over n_{sparse} the MSE and calculate it; for equidistant observations,

$$n_{\text{sparse}}^* = \left(\frac{T}{4(E\epsilon^2)^2}\int_0^T \sigma_t^4\,dt\right)^{1/3}.$$

The more general version is given by (31).

We note that the third- and the fourth-best estimators, which rely on comparatively small sample sizes, could benefit from the higher-order adjustments. A suitable Edgeworth adjustment is currently under investigation.

1.2.4 The Second-Best Approach: Subsampling and Averaging. As discussed earlier, even if one determines the optimal sampling frequency as we have just described, it is still the case that the data are not used to their full extent. We therefore propose in Section 3 the estimator $[Y, Y]_T^{(avg)}$ constructed by averaging the estimators $[Y, Y]_T^{(k)}$ across K grids of average size $\bar{n}$. We show in Section 3.5 that

$$[Y, Y]_T^{(avg)} \overset{\mathcal{L}}{\approx} \langle X, X\rangle_T + \underbrace{2\bar{n}E\epsilon^2}_{\text{bias due to noise}} + \underbrace{\left[\underbrace{4\frac{\bar{n}}{K}E\epsilon^4}_{\text{due to noise}} + \underbrace{\frac{4T}{3\bar{n}}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}\right]^{1/2} Z_{\text{total}}}_{\text{total variance}}, \quad (7)$$

where Z_{total} is a standard normal term. The above expression is in the equidistantly sampled case; the general expression is given by (51)–(52) in Section 3.5.

As can be seen from the foregoing equation, $[Y, Y]_T^{(avg)}$ remains a biased estimator of the quadratic variation $\langle X, X\rangle_T$ of the true return process. But the bias $2\bar{n}E\epsilon^2$ now increases with the average size of the subsamples, and $\bar{n} \le n$. Thus, as far as the bias is concerned, $[Y, Y]_T^{(avg)}$ is a better estimator than $[Y, Y]_T^{(all)}$. The optimal trade-off between the bias and variance for the estimator $[Y, Y]_T^{(avg)}$ is described in Section 3.6; we set $K^* \approx n/\bar{n}^*$ with $\bar{n}^*$ determined in (53), namely (in the equidistantly sampled case)

$$\bar{n}^* = \left(\frac{T}{6(E\epsilon^2)^2}\int_0^T \sigma_t^4\,dt\right)^{1/3}. \quad (8)$$

1.2.5 The First-Best Approach: Subsampling and Averaging, and Bias-Correction. Our final estimator is based on bias-correcting the estimator $[Y, Y]_T^{(avg)}$. This is the subject of Section 4. We show that a bias-adjusted estimator for $\langle X, X\rangle_T$ can be constructed as

$$\widehat{\langle X, X\rangle}_T = [Y, Y]_T^{(avg)} - \frac{\bar{n}}{n}[Y, Y]_T^{(all)},$$

that is, by combining estimators obtained over the two time scales "all" and "avg." A small-sample adjustment,

$$\widehat{\langle X, X\rangle}_T^{(adj)} = \left(1 - \frac{\bar{n}}{n}\right)^{-1}\widehat{\langle X, X\rangle}_T,$$

is given in (64) in Section 4.2, which shares the same asymptotic distribution as $\widehat{\langle X, X\rangle}_T$ to the order considered.

We show in Theorem 4 in Section 4.1 that if the number of subgrids is selected as $K = cn^{2/3}$, then

$$\widehat{\langle X, X\rangle}_T \overset{\mathcal{L}}{\approx} \langle X, X\rangle_T + \frac{1}{n^{1/6}}\underbrace{\left[\underbrace{\frac{8}{c^2}(E\epsilon^2)^2}_{\text{due to noise}} + \underbrace{c\frac{4T}{3}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}\right]^{1/2} Z_{\text{total}}}_{\text{total variance}}. \quad (9)$$

Unlike all of the previously considered estimators, this estimator is now correctly centered at $\langle X, X\rangle_T$. It converges only at rate $n^{-1/6}$, but from a MSE perspective, this is better than being (badly) biased. In particular, the prescription is to use all the available data, so there is no limit to how often one should sample: every few seconds or more often would be typical in tick-by-tick financial data, so n can be quite large. Based on Section 4.2, we use an estimate of the optimal value

$$c^* = \left(\frac{T}{12(E\epsilon^2)^2}\int_0^T \sigma_t^4\,dt\right)^{-1/3}. \tag{10}$$

2. ANALYSIS OF REALIZED VOLATILITY UNDER MARKET MICROSTRUCTURE NOISE

2.1 Setup

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

$$\epsilon_{t_i} \text{ iid with } E\epsilon_{t_i} = 0, \text{ and } \operatorname{var}(\epsilon_{t_i}) = E\epsilon^2; \quad \text{also, } \epsilon \perp\!\!\!\perp X \text{ process}, \tag{11}$$

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

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

$$\langle X, X\rangle_T = \int_0^T \sigma_t^2\,dt. \tag{12}$$

To succinctly describe the realized volatility, we use the notions of *grid* and *observed quadratic variation*, as follows.

Definition 1. The full grid containing all of the observation points is given by

$$\mathcal{G} = \{t_0, \ldots, t_n\}. \tag{13}$$

Definition 2. We also consider arbitrary grids, $\mathcal{H} \subseteq \mathcal{G}$. To denote successive elements in such grids, we proceed as follows. If $t_i \in \mathcal{H}$, then $t_{i,-}$ and $t_{i,+}$ denote the preceding and following elements in $\mathcal{H}$. t_i will always denote the ith point in the full grid $\mathcal{G}$. Hence, when $\mathcal{H} = \mathcal{G}$, $t_{i,-} = t_{i-1}$ and $t_{i,+} = t_{i+1}$. When $\mathcal{H}$ is a strict subgrid of $\mathcal{G}$, it will normally be the case that $t_{i,-} < t_{i-1}$, and $t_{i,+} > t_{i+1}$. Finally, we take

$$|\mathcal{H}| = (\#\text{ points in grid } \mathcal{H}) - 1. \tag{14}$$

This is to say that $|\mathcal{H}|$ is the number of time increments $(t_i, t_{i,+}]$, so that both endpoints are contained in $\mathcal{H}$. In particular, $|\mathcal{G}| = n$.

Definition 3. The observed quadratic variation $[\cdot, \cdot]$ for a generic process Z (such as Y or X) on an arbitrary grid $\mathcal{H} \subseteq \mathcal{G}$ is given by

$$[Z, Z]_t^{\mathcal{H}} = \sum_{t_j, t_{j,+} \in \mathcal{H}, t_{j,+} \le t} (Z_{t_{j,+}} - Z_{t_j})^2. \tag{15}$$

When the choice of grid follows from the context, $[Z, Z]_t^{\mathcal{H}}$ may be denoted as just $[Z, Z]_t$. On the full grid $\mathcal{G}$, the quadratic variation is given by

$$[Z, Z]_t^{(all)} = [Z, Z]_t^{\mathcal{G}} = \sum_{t_i, t_{i+1} \in \mathcal{G}, t_{i+1} \le t} (\Delta Z_{t_i})^2, \tag{16}$$

where $\Delta Z_{t_i} = Z_{t_{i+1}} - Z_{t_i}$. Quadratic covariations are defined similarly (see, e.g., Karatzas and Shreve 1991; Jacod and Shiryaev 2003 for more details on quadratic variations).

Our first objective is to assess how well the realized volatility, $[Y, Y]_T^{(all)}$, approximates the integrated volatility $\langle X, X\rangle_T$ of the true, latent process. In our asymptotic considerations, we always assume that the number of observations in $[0, T]$ goes to infinity, and also that the maximum distance in time between two consecutive observations goes to 0,

$$\max_i \Delta t_i \to 0 \quad \text{as } n \to \infty. \tag{17}$$

For the sake of preciseness, it should be noted that when $n \to \infty$, we are dealing with a sequence $\mathcal{G}_n$ of grids, $\mathcal{G}_n = \{t_{0,n}, \ldots, t_{n,n}\}$, and similarly for subgrids. We have avoided using double subscripts so as to not complicate the notation, but this is how all our conditions and results should be interpreted.

Note finally that we have adhered to the time series convention that $\Delta Z_{t_i} = Z_{t_{i+1}} - Z_{t_i}$. This is in contrast to the stochastic calculus convention $\Delta Z_{t_i} = Z_{t_i} - Z_{t_{i-1}}$.

2.2 The Realized Volatility: An Estimator of the Variance 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

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

This gives the conditional mean and variance of $[Y, Y]_T^{(all)}$,

$$E\big([Y, Y]_T^{(all)} \big| X \text{ process}\big) = [X, X]_T^{(all)} + 2nE\epsilon^2, \tag{18}$$

under assumption (11). Similarly,

$$\operatorname{var}\big([Y, Y]_T^{(all)} \big| X \text{ process}\big) = 4nE\epsilon^4 + O_p(1), \tag{19}$$

subject to condition (11) and $E\epsilon_{t_i}^4 = E\epsilon^4 < \infty$, for all i. Subject to slightly stronger conditions,

$$\begin{aligned}\operatorname{var}\big([Y, Y]_T^{(all)} \big| X \text{ process}\big) &= 4nE\epsilon^4 + \big(8[X, X]_T^{(all)} E\epsilon^2 - 2\operatorname{var}(\epsilon^2)\big) \\ &\quad + O_p(n^{-1/2}).\end{aligned} \tag{20}$$

It is also the case that as $n \to \infty$, conditionally on the X process, we have asymptotic normality,

$$n^{-1/2}\big([Y, Y]_T^{(all)} - 2nE\epsilon^2\big) \xrightarrow{\mathcal{L}} 2(E\epsilon^4)^{1/2} Z_{\text{noise}}. \tag{21}$$

Here Z_{noise} is standard normal, with the subscript "noise" indicating that the randomness comes from the noise ϵ, that is, the deviation of the observables Y from the true process X.

The derivations of (19)–(20), along with the conditions for the latter, are provided in Section A.1. The result (21) is derived in Section A.2, where it is part of Theorem A.1.

Equations (18) and (19) suggest that in the discrete world where microstructure effects are unfortunately present, realized volatility $[Y,Y]_T^{(all)}$ is not a reliable estimator for the true variation $[X,X]_T^{(all)}$ of the returns. *For large n, realized volatility could have little to do with the true returns*. Instead, it relates to the noise term, $E\epsilon^2$ in the first order and $E\epsilon^4$ in the second order. Also, one can see from (18) that $[Y,Y]_T^{(all)}$ has a positive bias whose magnitude increases linearly with the sample size.

Interestingly, apart from revealing the biased nature of $[Y,Y]_T^{(all)}$ at high frequency, our analysis also delivers a consistent estimator for the variance of the noise term. In other words, let

$$\widehat{E\epsilon^2} = \frac{1}{2n}[Y,Y]_T^{(all)}.$$

We have, for a fixed true return process X,

$$n^{1/2}(\widehat{E\epsilon^2} - E\epsilon^2) \to \mathrm{N}(0, E\epsilon^4), \quad \text{as } n \to \infty; \tag{22}$$

see Theorem A.1 in the Appendix.

By the same methods as in Section A.2, a consistent estimator of the asymptotic variance of $\widehat{E\epsilon^2}$ is then given by

$$\widehat{E\epsilon^4} = \frac{1}{2n}\sum_i (\Delta Y_{t_i})^4 - 3(\widehat{E\epsilon^2})^2. \tag{23}$$

A question that naturally arises is why one is interested in the quadratic variation of X (either $\langle X,X\rangle$ or $[X,X]$), as opposed to the quadratic variation of Y ($[Y,Y]$), because "$\langle Y,Y\rangle$" would have to be taken to be infinite in view of the foregoing. For example, in options pricing or hedging, one could take the opposite view that $[Y,Y]$ is the volatility that one actually faces.

A main reason why we focus on the quadratic variation of X is that the variation caused by the ϵ's is tied to each transaction, as opposed to the price process of the underlying security. From the standpoint of trading, the ϵ's represent trading costs, which are different from the costs created by the volatility of the underlying process. Different market participants may even face different types of trading costs depending on institutional arrangements. In any case, for the purpose of integrating trading costs into options prices, it seems more natural to approach the matter via the extensive literature on market microstructure (see O'Hara 1995 for a survey).

A related matter is that continuous finance would be difficult to implement if one were to use the quadratic variation of Y. If one were to use the quadratic variation $[Y,Y]$, then one would also be using a quantity that would depend on the data frequency.

Finally, apart from the specific application, it is interesting to be able to say something about the underlying log return process and to be able to separate this from the effects introduced by the mechanics of the trading process, where the idiosyncratics of the trading arrangement play a role.

2.3 The Optimal Sampling Frequency

We have just 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 microstructure, summarized by ϵ 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.

Formally, sparse sampling is implemented by sampling on a subgrid $\mathcal{H}$ of $\mathcal{G}$, with, as before,

$$[Y,Y]^{(sparse)} = [Y,Y]^{(\mathcal{H})} = \sum_{t_i,t_{i,+}\in\mathcal{H}} (Y_{t_{i,+}} - Y_{t_i})^2.$$

For the moment, we take the subgrid as given, but later we consider how to optimize the sampling. We call $n_{\text{sparse}} = |\mathcal{H}|$.

To give a rationale for sparse sampling, we propose an asymptotic where the law of ϵ, although still iid, is allowed to vary with n and n_{sparse}. Formally, we suppose that the distribution of ϵ, $\mathcal{L}(\epsilon)$, is an element of the set $\mathcal{D}$ of all distributions for which $E(\epsilon) = 0$ and where $E(\epsilon^2)$ and $E(\epsilon^4)/E(\epsilon^2)^2$ are bounded by arbitrary constants. The asymptotic normality in Section 2.2 then takes the following, more nuanced form.

Lemma 1. Suppose that X is an Itô process of form (1), where $|\mu_t|$ and σ_t are bounded above by a constant. Suppose that for given n, the grid $\mathcal{H}_n$ is given, with $n_{\text{sparse}} \to \infty$ as $n \to \infty$, and that for each n, Y is related to X through model (4). Assume (11), where $\mathcal{L}(\epsilon) \in \mathcal{D}$. Like the grid $\mathcal{H}_n$, the law $\mathcal{L}(\epsilon)$ can depend on n (whereas the process X is fixed). Let (17) be satisfied for the sequence of grids $\mathcal{H}_n$. Then

$$\begin{aligned}[Y,Y]_T^{\mathcal{H}} &= [X,X]_T^{\mathcal{H}} + 2n_{\text{sparse}}E\epsilon^2 \\ &\quad + \big(4n_{\text{sparse}}E\epsilon^4 \\ &\qquad + \big(8[X,X]_T^{\mathcal{H}}E\epsilon^2 - 2\operatorname{var}(\epsilon^2)\big)\big)^{1/2} Z_{\text{noise}} \\ &\quad + O_p\big(n_{\text{sparse}}^{-1/4}(E\epsilon^2)^{1/2}\big),\end{aligned} \tag{24}$$

where Z_{noise} is a quantity that is asymptotically standard normal.

The lemma is proved at the end of Section A.2. Note now that the relative order of the terms in (24) depends on the quantities n_{sparse} and $E\epsilon^2$. Therefore, if $E\epsilon^2$ is small relative to n_{sparse}, then $[Y,Y]_T^{\mathcal{H}}$ is, after all, not an entirely absurd substitute for $[X,X]_T^{\mathcal{H}}$.

One would be tempted to conclude that the optimal choice of n_{sparse} is to make it as small as possible. But that would overlook the fact that the bigger the n_{sparse}, the closer the $[X,X]_T^{\mathcal{H}}$ to the target integrated volatility $\langle X,X\rangle_T$ from (12).

To quantify the overall error, we now combine Lemma 1 with the results on discretization error to study the total error $[Y,Y]_T^{\mathcal{H}} - \langle X,X\rangle_T$. Following Rootzen (1980), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2002), and under the conditions that these authors stated, we can show that

$$\begin{aligned}&\left(\frac{n_{\text{sparse}}}{T}\right)^{1/2}([X,X]_T^{\mathcal{H}} - \langle X,X\rangle_T) \\ &\qquad \xrightarrow{\mathcal{L}} \left(\int_0^T 2H'(t)\sigma_t^4\,dt\right)^{1/2} \times Z_{\text{discrete}},\end{aligned} \tag{25}$$

stably in law. (We discuss this concept at the end of Sec. 3.4.) Z_{discrete} is a standard normal random variable, with the subscript "discrete" indicating that the randomness is due to the discretization effect in $[X,X]_T^{\mathcal{H}}$ when evaluating $\langle X,X\rangle_T$. $H(t)$ is the asymptotic quadratic variation of time, as discussed by Mykland and Zhang (2002),

$$H(t)=\lim_{n\to\infty}\frac{n_{\text{sparse}}}{T}\sum_{t_i,t_{i,+}\in\mathcal{H},t_{i,+}\le t}(t_{i,+}-t_i)^2.$$

In the case of equidistant observations, $\Delta t_0=\cdots=\Delta t_{n-1}=\Delta t=T/n_{\text{sparse}}$ and $H'(t)=1$. Because the ϵ's are independent of the X process, Z_{noise} is independent of Z_{discrete}.

For small $E\epsilon^2$, one now has an opportunity to estimate $\langle X,X\rangle_T$. It follows from our Lemma 1 and proposition 1 of Mykland and Zhang (2002) that

Proposition 1. Assume the conditions of Lemma 1, and also that $\max_{t_i,t_{i,+}\in\mathcal{H}}(t_{i,+}-t_i)=O(1/n_{\text{sparse}})$. Also assume Condition E in Section A.3. Then H is well defined, and

$$[Y,Y]_T^{\mathcal{H}}=\langle X,X\rangle_T+2E\epsilon^2 n_{\text{sparse}}+\Upsilon Z_{\text{total}}+O_p\bigl(n_{\text{sparse}}^{-1/4}(E\epsilon^2)^{1/2}\bigr)+o_p\bigl(n_{\text{sparse}}^{-1/2}\bigr),\quad(26)$$

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

$$\Upsilon^2=\underbrace{4n_{\text{sparse}}E\epsilon^4+\bigl(8[X,X]_T^{\mathcal{H}}E\epsilon^2-2\operatorname{var}(\epsilon^2)\bigr)}_{\text{due to noise}}+\underbrace{\frac{T}{n_{\text{sparse}}}\int_0^T 2H'(t)\sigma_t^4\,dt}_{\text{due to discretization}}.\quad(27)$$

Seen from this angle, there is scope for using the realized volatility $[Y,Y]^{\mathcal{H}}$ to estimate $\langle X,X\rangle$. There is a bias, $2E\epsilon^2 n_{\text{sparse}}$, but the bias goes down if one uses fewer observations. This, then, is consistent with the practice in empirical finance, where the estimator used is not $[Y,Y]^{(all)}$, but instead $[Y,Y]^{(sparse)}$, by sampling sparsely. For instance, faced with data sampled every few seconds, empirical researchers would typically use square returns (i.e., differences of log-prices Y) over, say, 5-, 15-, or 30-minute time intervals. The intuitive rationale for using this estimator is to attempt to control the bias of the estimator; this can be assessed based on our formula (26), replacing the original sample size n by the lower number reflecting the sparse sampling. But one should avoid sampling too sparsely, because formula (27) shows that decreasing n_{sparse} has the effect of increasing the variance of the estimator via the discretization effect which is proportional to n_{sparse}^{-1}. Based on our formulas, this trade-off between sampling too often and sampling too rarely can be formalized, and an optimal frequency at which to sample sparsely can be determined.

It is natural to minimize the MSE of $[Y,Y]^{(sparse)}$,

$$MSE=(2n_{\text{sparse}}E\epsilon^2)^2+\left\{4n_{\text{sparse}}E\epsilon^4+\bigl(8[X,X]_T^{\mathcal{H}}E\epsilon^2-2\operatorname{var}(\epsilon^2)\bigr)+\frac{T}{n_{\text{sparse}}}\int_0^T 2H'(t)\sigma_t^4\,dt\right\}.\quad(28)$$

One has to imagine that the original sample size n is quite large, so that $n_{\text{sparse}}<n$. In this case minimizing the MSE (28) means that one should choose n_{sparse} to satisfy $\partial MSE/\partial n_{\text{sparse}}\approx 0$; in other words,

$$8n_{\text{sparse}}(E\epsilon^2)^2+4E\epsilon^4-\frac{T}{n_{\text{sparse}}^2}\int_0^T 2H'(t)\sigma_t^4\,dt\approx 0\quad(29)$$

or

$$n_{\text{sparse}}^3+\frac{1}{2}n_1^2\frac{E\epsilon^4}{(E\epsilon^2)^2}-(E\epsilon^2)^{-2}\frac{T}{8}\int_0^T 2H'(t)\sigma_t^4\,dt\approx 0.\quad(30)$$

Finally, still under the conditions on Proposition 1, the optimum n_{sparse}^* becomes

$$n_{\text{sparse}}^*=(E\epsilon^2)^{-2/3}\left(\frac{T}{8}\int_0^T 2H'(t)\sigma_t^4\,dt\right)^{1/3}(1+o_p(1))\quad\text{as } E\epsilon^2\to 0.\quad(31)$$

Of course, if the actual sample size n were smaller than n_{sparse}^*, then one would simply take $n_{\text{sparse}}^*=n$, but this is unlikely to occur for heavily traded stock.

Equation (31) is the formal statement saying that one can sample more frequently when the error spread is small. Note from (28) that to first order, the final trade-off is between the bias $2n_{\text{sparse}}E\epsilon^2$ and the variance due to discretization. The effect of the variance associated with Z_{noise} is of lower order when comparing n_{sparse} and $E\epsilon^2$. It should be emphasized that (31) is a feasible way of choosing n_{sparse}. One can estimate $E\epsilon^2$ using all of 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.

Hence, if one decides to address the problem by selecting a lower frequency of observation, then one can do so by subsampling the full grid $\mathcal{G}$ at an arbitrary sparse frequency and use $[Y,Y]^{(sparse)}$. Alternatively, one can use n_{sparse} as optimally determined by (31). We denote the corresponding estimator as $[Y,Y]^{(sparse,opt)}$. These are our fourth- and third-best estimators.

3. SAMPLING SPARSELY WHILE USING ALL OF THE DATA: SUBSAMPLING AND AVERAGING OVER MULTIPLE GRIDS

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. Instead of selecting (arbitrarily or optimally) a subsample, our methods are based on selecting a number of subgrids of the original grid of observation times, $\mathcal{G}=\{t_0,\ldots,t_n\}$, 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, whereas the variation of the estimator can be lessened by the averaging. The benefit of averaging is clear from sufficiency considerations, and many statisticians would say that subsampling without subsequent averaging is inferentially incorrect.

In what follows we first introduce a set of notations, then turn to studying the realized volatility in the multigrid context. In Section 4 we show how to eliminate the bias of the estimator by using two time scales, a combination of single grid and multiple grids.

3.2 Notation for the Multiple Grids

We specifically suppose that the full grid $\mathcal{G}$, $\mathcal{G} = \{t_0, \ldots, t_n\}$ as in (13), is partitioned into K nonoverlapping subgrids $\mathcal{G}^{(k)}$, $k = 1, \ldots, K$; in other words,

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

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

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

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

Whether the allocation is regular or not, we let $n_k = |\mathcal{G}^{(k)}|$. As in Definition 2, $n = |\mathcal{G}|$. Recall that the realized volatility based on all observation points $\mathcal{G}$ is written as $[Y, Y]_T^{(all)}$. Meanwhile, if one uses only the subsampled observations Y_t, $t \in \mathcal{G}^{(k)}$, then the realized volatility is denoted by $[Y, Y]_T^{(k)}$. It has the form

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

where if $t_i \in \mathcal{G}^{(k)}$, then $t_{i,+}$ denotes the following elements in $\mathcal{G}^{(k)}$.

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

$$[Y, Y]_T^{(avg)} = \frac{1}{K} \sum_{k=1}^{K} [Y, Y]_T^{(k)}, \tag{32}$$

and this is the statistic we analyze in the what following. As before, we fix T and use only the observations within the time period $[0, T]$. Asymptotics are still under (17) and under

$$\text{as } n \to \infty \qquad \text{and} \qquad n/K \to \infty. \tag{33}$$

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

$$\bar{n} = \frac{1}{K} \sum_{k=1}^{K} n_k = \frac{n - K + 1}{K}. \tag{34}$$

3.3 Error Due to the Noise: $[Y, Y]_T^{(avg)} - [X, X]_T^{(avg)}$

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, here we study 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.

From (18) and (32),

$$E\left([Y, Y]_T^{(avg)} \middle| X \text{ process}\right) = [X, X]_T^{(avg)} + 2\bar{n} E\epsilon^2. \tag{35}$$

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

$$\begin{aligned} \operatorname{var}\left([Y, Y]_T^{(avg)} \middle| X \text{ process}\right) &= \frac{1}{K^2} \sum_{k=1}^{K} \operatorname{var}\left([Y, Y]_T^{(k)} \middle| X \text{ process}\right) \\ &= 4\frac{\bar{n}}{K} E\epsilon^4 + O_p\left(\frac{1}{K}\right), \end{aligned} \tag{36}$$

in the same way as in (19). Incorporating the next-order term in the variance yields that

$$\begin{aligned} &\operatorname{var}\left([Y, Y]_T^{(avg)} \middle| X\right) \\ &\quad = 4\frac{\bar{n}}{K} E\epsilon^4 + \frac{1}{K}\left[8[X, X]_T^{(avg)} E\epsilon^2 - 2\operatorname{var}(\epsilon^2)\right] \\ &\qquad + o_p\left(\frac{1}{K}\right), \end{aligned} \tag{37}$$

as in (20).

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

Theorem 1. Suppose that X is an Itô process of form (1). Suppose that Y is related to X through model (4), and that (11) 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 (33), as $n \to \infty$,

$$\sqrt{\frac{K}{\bar{n}}}\left([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\bar{n} E\epsilon^2\right) \xrightarrow{\mathcal{L}} 2\sqrt{E\epsilon^4} Z_{\text{noise}}^{(avg)}, \tag{38}$$

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

This can be compared with the result stated in (24). Notice that $Z_{\text{noise}}^{(avg)}$ in (38) is almost never the same as Z_{noise} in (24); in particular, $\operatorname{cov}(Z_{\text{noise}}, Z_{\text{noise}}^{(avg)}) = \operatorname{var}(\epsilon^2)/E\epsilon^4$, based on the proof of Theorem A.1 in the Appendix.

In comparison with 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, compare (18) and (19). We use this in Section 4.

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

In this section 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 by D_T, where

$$\begin{aligned} D_t &= [X, X]_t^{(avg)} - \langle X, X \rangle_t \\ &= \frac{1}{K} \sum_{k=1}^{K} \left([X, X]_t^{(k)} - \langle X, X \rangle_t\right), \end{aligned} \tag{39}$$

with

$$[X, X]_t^{(k)} = \sum_{t_i, t_{i,+} \in \mathcal{G}^{(k)}, t_{i,+} \le t} \left(X_{t_{i,+}} - X_{t_i}\right)^2. \tag{40}$$

In what follows we consider 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. (25)]. However, this present case is more complicated due to the multiple grids.

We suppose in the following that the sampling points are regularly allocated to subgrids; in other words, $\mathcal{G}^{(l)} = \{t_{l-1}, t_{K+l-1}, \ldots\}$. We also assume that

$$\max_i |\Delta t_i| = O\left(\frac{1}{n}\right) \tag{41}$$

and

$$K/n \to 0. \tag{42}$$

Define the weight function

$$h_i = \frac{4}{K\overline{\Delta t}} \sum_{j=1}^{(K-1)\wedge i} \left(1 - \frac{j}{K}\right)^2 \Delta t_{i-j}. \tag{43}$$

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 of the h_i's (except the first $K-1$) are equal, and

$$\begin{aligned} h_i &= 4\frac{1}{K} \sum_{j=1}^{(K-1)\wedge i} \left(1 - \frac{j}{K}\right)^2 \\ &= \begin{cases} 4\dfrac{2K^2 - 4K + 3}{6K^2} = \dfrac{4}{3} + o(1), & i \geq K-1 \\ \dfrac{4}{3}\dfrac{i}{K^2}(3K^2 - 3Ki + i^2) + o(1), & i < K-1. \end{cases} \end{aligned} \tag{44}$$

More generally, assumptions (41) and (42) ensure that

$$\sup_i h_i = O(1). \tag{45}$$

We take $\langle D, D\rangle_T$ to be the quadratic variation of D_t when viewed as a continuous-time process (39). This gives the best approximation to the variance of D_T. We show the following results in Section A.3.

Theorem 2. Suppose that X is an Itô process of the form (1), with drift coefficient μ_t and diffusion coefficient σ_t, both continuous almost surely. Also suppose that σ_t is bounded away from 0. Assume (41) and (42), and that sampling points are regularly allocated to grids. Then the quadratic variation of D_T is approximately

$$\langle D, D\rangle_T = \frac{TK}{n}\eta_n^2 + o_p\left(\frac{K}{n}\right), \tag{46}$$

where

$$\eta_n^2 = \sum_i h_i \sigma_{t_i}^4 \Delta t_i. \tag{47}$$

In particular, $D_T = O_p((K/n)^{1/2})$. From this, we 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. We discuss stable convergence at the end of this section.

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

$$\eta_n^2 \xrightarrow{P} \eta^2 \tag{48}$$

where η is random. Also assume condition E in Section A.3. Then

$$D_T/(K/n)^{1/2} \xrightarrow{\mathcal{L}} \eta\sqrt{T}Z_{\text{discrete}}, \tag{49}$$

where Z_{discrete} is standard normal and independent of the process X. The convergence in law is stable.

In other words, $D_T/(K/n)^{1/2}$ can be taken to be asymptotically mixed normal "N$(0, \eta^2 T)$." For most of our discussion, it is most convenient to suppose (48), and this is satisfied in many cases. For example, when the t_i are equidistant, and under regular allocation of points to subgrids,

$$\eta^2 = \frac{4}{3}\int_0^T \sigma_t^4\, dt, \tag{50}$$

following (44). One does not need to rely on (48); we argue in Section A.3 that without this condition, one can take $D_T/(K/n)^{1/2}$ to be approximately N$(0, \eta_n^2 T)$. For estimation of η^2 or η_n^2, see Section 6.

Finally, *stable convergence* (Rényi 1963; Aldous and Eagleson 1978; Hall and Heyde 1980, chap. 3) means for our purposes that the left side of (49) converges to the right 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.

3.5 Combining the Two Sources of Error

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

$$[Y, Y]_T^{(avg)} - \langle X, X\rangle_T - 2\bar{n}E\epsilon^2 = \xi Z_{\text{total}} + o_p(1), \tag{51}$$

where Z_{total} is an asymptotically standard normal random variable independent of the X process and

$$\xi^2 = \underbrace{4\frac{\bar{n}}{K}E\epsilon^4}_{\text{due to noise}} + \underbrace{T\frac{1}{\bar{n}}\eta^2}_{\text{due to discretization}}. \tag{52}$$

It is easily seen that if one takes $K = cn^{2/3}$, then both components in ξ^2 will be present in the limit; otherwise, one of them will dominate. Based on (51), $[Y, Y]_T^{(avg)}$ is still a biased estimator of the quadratic variation $\langle X, X\rangle_T$ of the true return process. One can recognize that as far as the asymptotic bias is concerned, $[Y, Y]_T^{(avg)}$ is a better estimator than $[Y, Y]_T^{(all)}$, because $\bar{n} \leq n$, demonstrating that the bias in the subsampled estimator $[Y, Y]_T^{(avg)}$ increases in a slower pace than the full-grid estimator. One can also construct a bias-adjusted estimator from (51); this further development would involve the higher-order analysis between the bias and the subsampled estimator. We describe the methodology of bias correction in Section 4.

3.6 The Benefits of Sampling Sparsely: Optimal Sampling Frequency in the Multiple-Grid Case

As in Section 2.3, when the noise is negligible asymptotically, we can search for an optimal $\bar{n}$ for subsampling to balance the coexistence of the bias and the variance in (51). To reduce the MSE of $[Y,Y]_T^{(avg)}$, we set $\partial MSE/\partial\bar{n}=0$. From (52)–(51), $bias = 2\bar{n}E\epsilon^2$ and $\xi^2 = 4\frac{\bar{n}}{K}E\epsilon^4 + \frac{T}{\bar{n}}\eta^2$. Then

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

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

$$\bar{n}^* = \left(\frac{T\eta^2}{8(E\epsilon^2)^2}\right)^{1/3}. \tag{53}$$

Therefore, assuming that 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 of the data in the single-grid case. The subsampling coupled with aggregation brings out the advantage of using all of data. Of course, for the asymptotics to work, we need $E\epsilon^2 \to 0$. Our recommendation, however, is to use the bias-corrected estimator given in Section 4.

4. THE FINAL ESTIMATOR: SUBSAMPLING, AVERAGING, AND BIAS CORRECTION OVER TWO TIME SCALES

4.1 The Final Estimator $\widehat{\langle X,X\rangle}_T$: Main Result

In the preceding 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 uses the full grid. As mentioned before, from (22) in the single-grid case (Sec. 2), $E\epsilon^2$ can be consistently approximated by

$$\widehat{E\epsilon^2} = \frac{1}{2n}[Y,Y]_T^{(all)}. \tag{54}$$

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

$$\widehat{\langle X,X\rangle}_T = [Y,Y]_T^{(avg)} - \frac{\bar{n}}{n}[Y,Y]_T^{(all)}, \tag{55}$$

thereby combining the two time scales (*all*) and (*avg*).

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

$$\left(\frac{K}{\bar{n}}\right)^{1/2}\left(\widehat{\langle X,X\rangle}_T - [X,X]_T^{(avg)}\right)$$
$$= \left(\frac{K}{\bar{n}}\right)^{1/2}\left([Y,Y]_T^{(avg)} - [X,X]_T^{(avg)} - 2\bar{n}E\epsilon^2\right)$$
$$- 2(K\bar{n})^{1/2}(\widehat{E\epsilon^2} - E\epsilon^2)$$
$$\xrightarrow{\mathcal{L}} \mathrm{N}\big(0, 8(E\epsilon^2)^2\big), \tag{56}$$

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 $n\to\infty$,

$$\widehat{\langle X,X\rangle}_T - \langle X,X\rangle_T$$
$$= \left(\widehat{\langle X,X\rangle}_T - [X,X]_T^{(avg)}\right) + \left([X,X]_T^{(avg)} - \langle X,X\rangle_T\right)$$
$$= O_p\left(\frac{\bar{n}^{1/2}}{K^{1/2}}\right) + O_p\big(\bar{n}^{-1/2}\big). \tag{57}$$

The error is minimized by equating the two terms on the right side of (57), yielding that the optimal sampling step for $[Y,Y]_T^{(avg)}$ is $K = O(n^{2/3})$. The right side of (57) then has order $O_p(n^{-1/6})$.

In particular, if we take

$$K = cn^{2/3}, \tag{58}$$

then we find the limit in (57), as follows.

Theorem 4. Suppose that X is an Itô process of form (1), and assume the conditions of Theorem 3 in Section 3.4. Suppose that Y is related to X through model (4), and that (11) is satisfied with $E\epsilon^2 < \infty$. Under assumption (58),

$$n^{1/6}\left(\widehat{\langle X,X\rangle}_T - \langle X,X\rangle_T\right)$$
$$\xrightarrow{\mathcal{L}} \mathrm{N}\big(0, 8c^{-2}(E\epsilon^2)^2\big) + \eta\sqrt{T}\mathrm{N}(0,c)$$
$$= \left(8c^{-2}(E\epsilon^2)^2 + c\eta^2 T\right)^{1/2}\mathrm{N}(0,1), \tag{59}$$

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

Proof. Note that the first normal distribution comes from (56) and that the second comes from Theorem 3 in Section 3.4. The two normal distributions are independent because the convergence of the first term in (57) is conditional on the X process, which is why they can be amalgamated as stated. The requirement that $E\epsilon^4 < \infty$ (Thm. A.1 in the App.) is not needed, because only a law of large numbers is required for $M_T^{(1)}$ (see the proof of Thm. A.1) when considering the difference in (56). This finishes the proof.

The estimation of the asymptotic spread $s^2 = 8c^{-2}(E\epsilon^2)^2 + c\eta^2 T$ of $\widehat{\langle X,X\rangle}_T$ is deferred to Section 6. It is seen in Section A.1 that for $K = cn^{2/3}$, the second-order conditional variance of $\widehat{\langle X,X\rangle}_T$, given the X process, is given by

$$\mathrm{var}\left(\widehat{\langle X,X\rangle}_T \middle| X\right) = n^{-1/3}c^{-2}8(E\epsilon^2)^2$$
$$+ n^{-2/3}c^{-1}\left[8[X,X]_T^{(avg)}E\epsilon^2 - 2\,\mathrm{var}(\epsilon^2)\right]$$
$$+ o_p\big(n^{-2/3}\big). \tag{60}$$

The evidence from our Monte Carlo simulations in Section 7 suggests that this correction may matter in small samples, and that the (random) variance of $\widehat{\langle X,X\rangle}_T$ is best estimated by

$$s^2 + n^{-2/3}c^{-1}\left[8[X,X]_T^{(avg)}E\epsilon^2 - 2\,\mathrm{var}(\epsilon^2)\right]. \tag{61}$$

For the correction term, $[X,X]_T^{(avg)}$ can be estimated by $\widehat{\langle X,X\rangle}_T$ itself. Meanwhile, by (23), a consistent estimator of $\mathrm{var}(\epsilon^2)$ is given by

$$\widehat{\mathrm{var}(\epsilon^2)} = \frac{1}{2n}\sum_i \big(\Delta Y_{t_i}\big)^4 - 4(\widehat{E\epsilon^2})^2.$$

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

To further pin down the optimal sampling frequency K, one can minimize the expected asymptotic variance in (59) to obtain

$$c^* = \left(\frac{16(E\epsilon^2)^2}{T\,E\eta^2}\right)^{1/3}, \tag{62}$$

which can be consistently estimated from data in past time periods (before time $t_0 = 0$), using $\widehat{E\epsilon^2}$ and an estimator of η^2 (cf. Sec. 6). As mentioned in Section 3.4, η^2 can be taken to be independent of K as 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 σ_t^2 is constant, and for equidistant sampling and regular allocation to grids, $\eta^2 = \frac{4}{3}\sigma^4 T$, then the asymptotic variance in (59) is

$$8c^{-2}(E\epsilon^2)^2 + c\eta^2 T = 8c^{-2}(E\epsilon^2)^2 + \tfrac{4}{3}c\sigma^4 T^2,$$

and the optimal choice of c becomes

$$c_{opt} = \left(\frac{12(E\epsilon^2)^2}{T^2\sigma^4}\right)^{1/3}. \tag{63}$$

In this case, the asymptotic variance in (59) is

$$2\left(12(E\epsilon^2)^2\right)^{1/3}\ (\sigma^2 T)^{4/3}.$$

One can also, of course, estimate c to minimize the actual asymptotic variance in (59) from data in the current time period ($0 \le t \le T$). It is beyond the scope of this article 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 $\widehat{\langle X,X\rangle}_T$ from a "smallish" sample standpoint. We argue in what follows that one can apply a bias-type adjustment to get

$$\widehat{\langle X,X\rangle}_T^{(adj)} = \left(1 - \frac{\bar n}{n}\right)^{-1}\widehat{\langle X,X\rangle}_T. \tag{64}$$

The difference from the estimator in (55) is of order $O_p(\bar n/n) = O_p(K^{-1})$, and thus the two estimators behave the same as the asymptotic order that we consider. The estimator (64), however, has the appeal of being, in a certain way, "unbiased," as follows. For arbitrary (a, b), consider all estimators of the form

$$\widehat{\langle X,X\rangle}_T^{(adj)} = a[Y,Y]_T^{(avg)} - b\frac{\bar n}{n}[Y,Y]_T^{(all)}.$$

Then, from (18) and (35),

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

It is natural to choose $a = b$ to completely remove the effect of $E\epsilon^2$. 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 (64).

Similarly, an adjusted estimator of $E\epsilon^2$ is given by

$$\widehat{E\epsilon^2}^{(adj)} = \tfrac{1}{2}(n-\bar n)^{-1}\big([Y,Y]_T^{(all)} - [Y,Y]_T^{(avg)}\big), \tag{65}$$

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

$$\begin{aligned}
&\widehat{E\epsilon^2}^{(adj)} - E\epsilon^2\\
&\quad= (\widehat{E\epsilon^2} - E\epsilon^2)(1 + O(K^{-1})) + O_p\big(Kn^{-3/2}\big)\\
&\quad= \widehat{E\epsilon^2} - E\epsilon^2 + O_p\big(n^{-1/2}K^{-1}\big) + O_p\big(Kn^{-3/2}\big)\\
&\quad= \widehat{E\epsilon^2} - E\epsilon^2 + O_p\big(n^{-5/6}\big),
\end{aligned}$$

from (58). It follows that $n^{1/2}(\widehat{E\epsilon^2} - E\epsilon^2)$ and $n^{1/2}(\widehat{E\epsilon^2}^{(adj)} - E\epsilon^2)$ have the same asymptotic distribution.

5. EXTENSION TO MULTIPLE PERIOD INFERENCE

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

$$\widehat{\langle X,X\rangle}_t = [Y,Y]_t^{(avg)} - \frac{\bar n}{n}[Y,Y]_t^{(all)}, \tag{66}$$

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

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

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 $\widehat{\langle X,X\rangle}_{T_m} - \widehat{\langle X,X\rangle}_{T_{m-1}}$.

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

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

An advantage of our proposed estimator is that if ϵ_{t_i} has different variance in different time segments, say $\mathrm{var}(\epsilon_{t_i}) = (E\epsilon^2)_m$ for $t_i \in (T_{m-1}, T_m]$, then both consistency and asymptotic (mixed) normality continue to hold, provided that one replaces $E\epsilon^2$ by $(E\epsilon^2)_m$. This adds a measure of robustness to the procedure. If one were convinced that $E\epsilon^2$ were the same across time segments, then an alternative estimator would have the form

$$\widehat{\langle X,X\rangle}_t = [Y,Y]_t^{(avg)} - \left(\frac{1}{K}\#\{t_{i+1}\le t\} - 1\right)\frac{1}{n}[Y,Y]_T^{(all)}$$
$$\text{for } t = T_1,\dots,T_M. \tag{67}$$

However, the errors $\widehat{\langle X,X\rangle}_{T_m} - \widehat{\langle X,X\rangle}_{T_{m-1}} - \int_{T_{m-1}}^{T_m} \sigma_u^2\,du$ in this case are not asymptotically independent. Note that for $T = T_m$, both candidates (66) and (67) for $\widehat{\langle X,X\rangle}_t$ coincide with the quantity in (55).

6. ESTIMATING THE ASYMPTOTIC VARIANCE OF $\widehat{\langle X,X\rangle}_T$

In the one-period case, the main goal is to estimate the asymptotic variance $s^2 = 8c^{-2}(E\epsilon^2)^2 + c\eta^2 T$ [cf. (58) and (59)]. The multigrid case is a straightforward generalization, as indicated in Section 5.

Here, we are concerned only with the case where the points t_i are equally spaced ($\Delta t_i = \Delta t$) and are regularly allocated to the grids $\mathcal{A}_1 = \{\mathcal{G}^{(k)}, k = 1,\ldots,K_1\}$. A richer set of ingredients are required to find the spread than to just estimate $\widehat{\langle X,X\rangle}_T$. To implement the estimator, we create an additional family $\mathcal{A}_2 = \{\mathcal{G}^{(k,i)}, k = 1,\ldots,K_1, i = 1,\ldots,I\}$ of grids where $\mathcal{G}^{(k,i)}$ contains every Ith point of $\mathcal{G}^{(k)}$, starting with the ith point. 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 to divide 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

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

where $\langle X,X\rangle_t^{K_i}$ is the estimator (66) using the grid family i, $i = 1,2$.

Using the discussion in Section 5, we can show that

$$\hat{s}_0^2 \approx s_0^2, \tag{68}$$

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

$$\begin{aligned} s_0^2 &= 8(E\epsilon^2)^2(c_1^{-2} + c_2^{-2} - c_1^{-1}c_2^{-1}) + \big(c_1^{1/2} - c_2^{1/2}\big)^2 T\eta^2 \\ &= 8(E\epsilon^2)^2 c_1^{-2}(1 + I^{-2} - I^{-1}) + c_1\big(I^{1/2} - 1\big)^2 T\eta^2. \end{aligned} \tag{69}$$

In (68), the symbol "$\approx$" denotes first convergence in law as $n \to \infty$, and then a limit in probability as $M \to \infty$. Because $E\epsilon^2$ can be estimated by $\widehat{E\epsilon^2} = [Y,Y]^{(all)}/2n$, we can put hats on s_0^2, $(E\epsilon^2)^2$, and η^2 in (69) to obtain an estimator of η^2. Similarly,

$$\begin{aligned} s^2 &= 8(E\epsilon^2)^2\left(c^{-2} - \frac{c(c_1^{-2} + c_2^{-2} - c_1^{-1}c_2^{-1})}{\big(c_1^{1/2} - c_2^{1/2}\big)^2}\right) + \frac{c}{\big(c_1^{1/2} - c_2^{1/2}\big)^2}s_0^2 \\ &= 8\left(c^{-2} - cc_1^{-3}\frac{I^{-2} - I^{-1} + 1}{\big(I^{1/2} - 1\big)^2}\right)(E\epsilon^2)^2 + \frac{c}{c_1}\frac{1}{\big(I^{1/2} - 1\big)^2}s_0^2, \end{aligned} \tag{70}$$

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

Table 1. Coefficients of $(\widehat{E\epsilon^2})^2$ and $\hat{s}^2$ When $c_1 = c$

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

Normally, one would take $c_1 = c$. Hence an estimator $\hat{s}^2$ can be found from $\hat{s}_0^2$ and $\widehat{E\epsilon^2}$. When $c_1 = c$, we argue that the optimal choice is $I = 3$ or 4, as follows. The coefficients in (70) become

$$\text{coeff}(s_0^2) = \big(I^{1/2} - 1\big)^{-2}$$

and

$$\text{coeff}((E\epsilon^2)^2) = 8c^{-2}\big(I^{1/2} - 1\big)^{-2} f(I),$$

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 that give the lowest ratio of coeff$((E\epsilon^2)^2)$/coeff(s_0^2). Using $I = 3$ or 4, therefore, would maximally insulate against $(\widehat{E\epsilon^2})^2$ dominating over $\hat{s}_0^2$. This is desirable, because $\hat{s}_0^2$ is the estimator of carrying the information about η^2. Numerical values for the coefficients are given in Table 1. If c were such that $(\widehat{E\epsilon^2})^2$ still overwhelms $\hat{s}_0^2$, then a choice of $c_1 \neq c$ should be considered.

7. MONTE CARLO EVIDENCE

In this article we have discussed five approaches to dealing with the impact of market microstructure noise on realized volatility. In this section we examine the performance of each approach in simulations, and compare the results with those predicted by the aforementioned asymptotic theory.

7.1 Simulations Design

We use the stochastic volatility model of Heston (1993) as our data-generating process,

$$dX_t = (\mu - v_t/2)\,dt + \sigma_t\,dB_t \tag{71}$$

and

$$dv_t = \kappa(\alpha - v_t)\,dt + \gamma v_t^{1/2}\,dW_t, \tag{72}$$

where $v_t = \sigma_t^2$. We assume that the parameters $(\mu,\kappa,\alpha,\gamma)$ and ρ, the correlation coefficient between the two Brownian motions B and W, are constant in this model. We also assume Feller's condition $2\kappa\alpha \geq \gamma^2$, to make the zero boundary unattainable by the volatility process.

We simulate $M = 25{,}000$ sample paths of the process using the Euler scheme at a time interval $\Delta t = 1$ second, and set the parameter values at values reasonable for a stock: $\mu = .05$, $\kappa = 5$, $\alpha = .04$, $\gamma = .5$, and $\rho = -.5$. As for the market microstructure noise ϵ, we assume that it is Gaussian and small. Specifically, we set $(E\epsilon^2)^{1/2} = .0005$ (i.e., the standard deviation of the noise is .05% of the value of the asset price). We purposefully set this value to be smaller than the values calibrated using the empirical market microstructure literature.

On each simulated sample path, we estimate $\langle X,X\rangle_T$ over $T = 1$ day (i.e., $T = 1/252$, because the parameter values are all annualized) using the five estimation strategies described earlier: $[Y,Y]_T^{(all)}$, $[Y,Y]_T^{(sparse)}$, $[Y,Y]_T^{(sparse,opt)}$, $[Y,Y]_T^{(avg)}$, and

finally, $\widehat{\langle X,X\rangle}_T^{(adj)}$. We assume that a day consists of 6.5 hours of open trading, as is the case on the NYSE and NASDAQ. For the fourth-best estimator, we represent this approach using sparse sampling at a frequency of once every 5 minutes.

7.2 Simulations Results

Table 2 reports the simulation results for the five estimation strategies. For each estimator, the bias is calculated as $E[\text{estimator} - \langle X,X\rangle_T]$. The variance row reports the quantity

$$E\left[\operatorname{var}\left(\text{estimator} - \langle X,X\rangle_T \Big| \int_0^T \sigma_t^2\,dt, \int_0^T \sigma_t^4\,dt\right)\right]. \quad (73)$$

The rows marked "relative" report the corresponding results in percentage terms, that is, for

$$\frac{\text{estimator} - \langle X,X\rangle_T}{\langle X,X\rangle_T}.$$

Note that although the convergence of the discretization error is stable as opposed to conditional, (73) seems like a reasonable stand-in for the theoretical quantity under consideration.

In each case, "small sample" represents the average value over the M simulated paths, and "asymptotic" refers to the value predicted by our theory. The noise parts of the asymptotic variances are taken to include the relevant second-order terms, as given by (20), (37), and (60), yielding spreads like (in the latter case) (61). The small-sample variance is computed by binning the results of the simulations according to the values of the pair $(\int_0^T \sigma_t^2\,dt, \int_0^T \sigma_t^4\,dt)$, calculating the variance within each bin, and then averaging across the bins, weighted by the number of elements in each bin.

Comparing the rows showing small-sample results to their asymptotic predictions, we see that the asymptotic theory provides a good approximation to the first two moments of the small-sample distribution of the five estimators. Comparing across columns, we see that the differences between the five estimators are quite large. As expected, the naive fifth-best strategy produces disastrous results. The first-best strategy results in a large decrease of root MSE (RMSE) relative to the others, including in particular the fourth-best strategy, which is the approach currently used in most of the empirical literature. But, at least for the parameter values used here, determining the optimal sampling frequency (the sparse optimal, or third-best, strategy) results in a modest improvement over the arbitrary selection of a sampling frequency (the sparse, or fourth-best, strategy). Using all of the data as in the subsampled and averaged estimator (second-best), but especially the bias correction by combining two time scales as in our first-best approach, is where the major practical improvements reside.

It may seem surprising that the variance of the estimator is reduced so much from the second to first-best estimator. After all, if one sets $c = Kn^{-2/3}$, then the asymptotic normal term in (7) becomes

$$\frac{1}{n^{1/6}}\Bigg[\underbrace{\frac{4}{c^2}(E\epsilon^4)}_{\text{due to noise}} + \underbrace{c\frac{4T}{3}\int_0^T \sigma_t^4\,dt}_{\text{due to discretization}}\Bigg]^{1/2} Z_{\text{total}}.$$

total variance

This is quite similar to the error term in (9). However, the c for the second-best estimator is chosen by (8), whereas for the first-best case it is given by (10). The two optima are very different because for the second-best estimator, it is chosen by a bias–variance trade-off, whereas for the first-best estimator, it is chosen to minimize the asymptotic variance. From this standpoint, it is quite reasonable for the first-best estimator to also have substantially better asymptotic variance. In fact, from (8) and (10), it is easy to see that for the optimal choices,

$$c^*_{\text{second best}} = n^{1/3}2^{-1/3}\left(\frac{E\epsilon^4}{E(\epsilon^2)^2}\right)^{1/3} c^*_{\text{first best}},$$

whence the total variance is in fact quite different for the two estimators.

Figure 1 shows the small-sample and asymptotic RMSE of the $[Y,Y]_T^{(sparse)}$ estimator as a function of the subsampling interval, illustrating the minimization problem studied in Section 2.3. Finally, Figure 2 shows the standardized distribution of the first-best estimator obtained from the simulations (histogram) and the corresponding asymptotic distribution predicted by our theory (solid line). Due to the effect of bias, the distributions of the fifth- to second-best estimators are irrelevant, because they cannot be used to set intervals for the volatility.

8. DISCUSSION

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

Table 2. Monte Carlo Simulations for the Five Estimation Strategies

	Fifth-best $[Y,Y]_T^{(all)}$	*Fourth-best* $[Y,Y]_T^{(sparse)}$	*Third-best* $[Y,Y]_T^{(sparse,\,opt)}$	*Second-best* $[Y,Y]_T^{(avg)}$	*First-best* $\widehat{\langle X,X\rangle}_T^{(adj)}$
Small-sample bias	1.1699×10^{-2}	3.89×10^{-5}	2.18×10^{-5}	1.926×10^{-5}	2×10^{-8}
Asymptotic bias	1.1700×10^{-2}	3.90×10^{-5}	2.20×10^{-5}	1.927×10^{-5}	0
Small-sample variance	1.791×10^{-8}	1.4414×10^{-9}	1.59×10^{-9}	9.41×10^{-10}	9×10^{-11}
Asymptotic variance	1.788×10^{-8}	1.4409×10^{-9}	1.58×10^{-9}	9.37×10^{-10}	8×10^{-11}
Small-sample RMSE	1.1699×10^{-2}	5.437×10^{-5}	4.543×10^{-5}	3.622×10^{-5}	9.4×10^{-6}
Asymptotic RMSE	1.1700×10^{-2}	5.442×10^{-5}	4.546×10^{-5}	3.618×10^{-5}	8.9×10^{-6}
Small-sample relative bias	182	.61	.18	.15	−.00045
Small-sample relative variance	82,502	1.15	.11	.053	.0043
Small-sample relative RMSE	340	1.24	.37	.28	.065

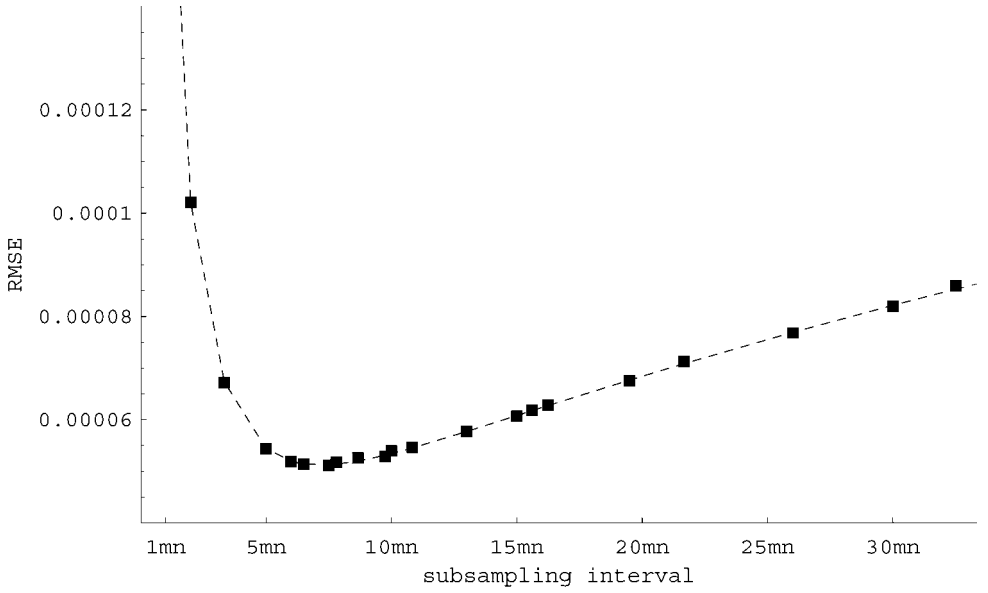

Figure 1. RMSE of the Sparse Estimator as a Function of the Subsampling Frequency (■ RMSE small sample; – – – RMSE asymptotic).

away most of the available data. We have argued that this is caused by not incorporating the noise in the model. Although 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 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 that is approximately unbiased; we have also shown how to assess the (random) variance of this estimator. We also show that for our most recommended procedure, the optimal number of multiple grids is of order $O(n^{2/3})$ (see Sec. 4). 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 multigrid extension.

One important message of the article is that any time one has an impulse to sample sparsely, one can always do better with a multigrid method, regardless of the model or the quantity being estimated.

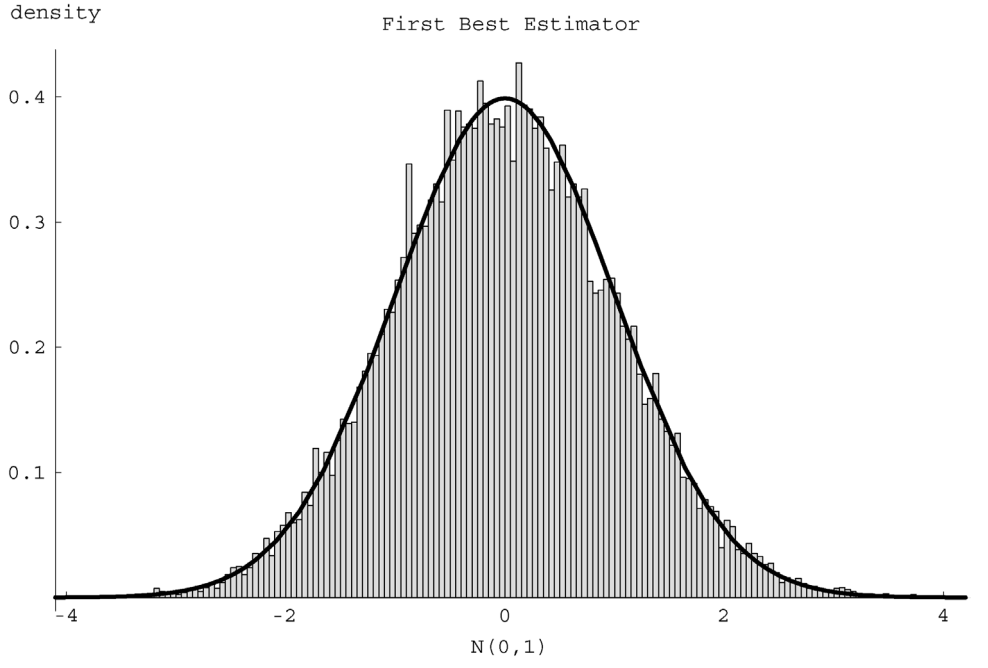

Figure 2. Asymptotic and Small-Sample Standardized Distributions of the First-Best Estimator.

Our results in this article cover the case where the latent X follows the Itô process model (1). Theorem 1 remains true under the assumption that X is a general semimartingale, that is, a process X that can also have jumps. This follows because the proof of Lemma A.2 remains true for general semimartingale X (Jacod and Shiryaev 2003, thm. I.4.47, p. 52). Also, apart from using Lemma A.2, Theorem A.1 is unrelated to the nature of the X process. As far as the effect of jumps on the discretization effect is concerned, it should be noted that consistency holds, as follows. When X is a general semimartingale, then for any sequence of grids $\mathcal{H}_n$ so that (41) holds, $[X,X]^{\mathcal{H}_n}$ converges to the continuous-time quadratic variation of X, that is,

$$\int_0^T \sigma_t^2\,dt + \sum_{0\le t\le T} (\text{jump of } X \text{ at } t^2) \tag{74}$$

(cf. Jacod and Shiryaev 2003, thm. I.4.47, p. 52). In particular, this also applies to subgrids under the regularity conditions that we have discussed. Furthermore, by extension of the same proof, $[X,X]^{(avg)}$ also converges to (74). From the same argument, so does $\widehat{\langle X,X\rangle}_T$. In this sense, the two-scales estimator is robust to jumps.

APPENDIX: PROOFS OF RESULTS

When the total grid $\mathcal{G}$ is considered, we use $\sum_{i=1}^{n-1}$, $\sum_{t_{i+1}\le T}$, and $\sum_{t_i\in\mathcal{G}}$ interchangeably in the following proofs. Also, 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

Here we calculate explicitly the variance in (19), from which the stated approximation follows. The explicit remainder term is also used for (36). Let a partition of $[0,T]$ be $0=t_0\le t_1\le\cdots\le t_n=T$. Under assumption (11),

$$\begin{aligned}\operatorname{var}\big([Y,Y]_T^{(all)}|X\big) &= \operatorname{var}\Big[\sum_{t_{i+1}\le T}\big(\Delta Y_{t_i}\big)^2\Big|X\Big]\\ &= \underbrace{\sum_{t_{i+1}\le T}\operatorname{var}\big[\big(\Delta Y_{t_i}\big)^2|X\big]}_{I_T}\\ &\quad+\underbrace{2\sum_{t_{i+1}\le T}\operatorname{cov}\big[\big(\Delta Y_{t_{i-1}}\big)^2,\big(\Delta Y_{t_i}\big)^2|X\big]}_{II_T}\end{aligned}$$

because $\Delta Y_{t_i}=\Delta X_{t_i}+\Delta\epsilon_{t_i}$ is 1 dependent given X process,

$$\begin{aligned}\operatorname{var}\big[\big(\Delta Y_{t_i}\big)^2|X\big] &= \kappa_4\big(\Delta Y_{t_i}|X\big)+2\big[\operatorname{var}\big(\Delta Y_{t_i}|X\big)\big]^2\\ &\quad+4\big[E\big(\Delta Y_{t_i}|X\big)\big]^2\operatorname{var}\big(\Delta Y_{t_i}|X\big)\\ &\quad+4E\big(\Delta Y_{t_i}|X\big)\kappa_3\big(\Delta Y_{t_i}|X\big)\\ &= \kappa_4\big(\Delta\epsilon_{t_i}\big)+2\big[\operatorname{var}\big(\Delta\epsilon_{t_i}\big)\big]^2+4\big(\Delta X_{t_i}\big)^2\operatorname{var}\big(\Delta\epsilon_{t_i}\big)\\ &\quad+4\big(\Delta X_{t_i}\big)\kappa_3\big(\Delta\epsilon_{t_i}\big)\quad\text{(under Assumption 11)}\\ &= 2\kappa_4(\epsilon)+8(E\epsilon^2)^2+8\big(\Delta X_{t_i}\big)^2E\epsilon^2,\end{aligned}$$

and because $\kappa_3(\Delta\epsilon_{t_i})=0$. The κ's are the cumulants of the relevant order,

$$\kappa_1(\epsilon)=0,\qquad \kappa_2(\epsilon)=\operatorname{var}(\epsilon)=E(\epsilon^2),$$
$$\kappa_3(\epsilon)=E\epsilon^3,\qquad\text{and}\qquad \kappa_4(\epsilon)=E(\epsilon^4)-3(E\epsilon^2)^2. \tag{A.1}$$

So $I_T=n(2\kappa_4(\epsilon)+8(E\epsilon^2)^2)+8[X,X]_T^{(all)}E\epsilon^2$. Similarly, for the covariance,

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

because of (A.1).

Thus, assuming the coefficients in (A.2),

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

Amalgamating the two expressions, one obtains

$$\begin{aligned}&\operatorname{var}\big([Y,Y]_T^{(all)}|X\big)\\ &\quad=n\big(2\kappa_4(\epsilon)+8(E\epsilon^2)^2\big)+8[X,X]_T^{(all)}E\epsilon^2\\ &\qquad+2(n-1)\big(\kappa_4(\epsilon)+2(E\epsilon^2)^2\big)\\ &\qquad-8E\epsilon^2\sum\big(\Delta X_{t_{i-1}}\big)\big(\Delta X_{t_i}\big)-4\kappa_3(\epsilon)\big(\Delta X_{t_{n-1}}-\Delta X_{t_0}\big)\\ &\quad=4nE\epsilon^4+R_n,\end{aligned} \tag{A.3}$$

where the remainder term R_n satisfies

$$\begin{aligned}|R_n| &\le 8E\epsilon^2[X,X]_T+2\big(\kappa_4(\epsilon)+2(E\epsilon^2)^2\big)\\ &\quad+8E\epsilon^2\Big|\sum\big(\Delta X_{t_{i-1}}\big)\Big|\big|\big(\Delta X_{t_i}\big)\big|+4|\kappa_3(\epsilon)|\big(\big|\Delta X_{t_{n-1}}\big|+\big|\Delta X_{t_0}\big|\big)\\ &\le 16E\epsilon^2[X,X]_T^{(all)}+2\big(\kappa_4(\epsilon)+2(E\epsilon^2)^2\big)\\ &\quad+2|\kappa_3(\epsilon)|(2+[X,X]_T)\end{aligned} \tag{A.4}$$

by the Cauchy–Schwarz inequality and because $|x|\le(1+x^2)/2$. Because $[X,X]_T^{(all)}=O_p(1)$, (19) follows.

Under slightly stronger conditions (e.g., $|\mu_t|$ and σ_t are bounded above by a constant), $\sum(\Delta X_{t_{i-1}})(\Delta X_{t_i})$ is a near-martingale and of order $O_p(n^{-1/2})$, and, similarly, $\Delta X_{t_{n-1}}-\Delta X_{t_0}=O_p(n^{-1/2})$, from which (20) follows.

Similarly, to obtain a higher-order approximation to the variance of the estimator $\widehat{\langle X,X\rangle}_T$, note that

$$\begin{aligned}\widehat{\langle X,X\rangle}_T-\langle X,X\rangle_T &= \Big([Y,Y]_T^{(avg)}-\frac{1}{K}[Y,Y]_T^{(all)}-[X,X]_T^{(avg)}\Big)\\ &\quad+\big([X,X]_T^{(avg)}-\langle X,X\rangle_T\big).\end{aligned} \tag{A.5}$$

Now recalling (20) and (37) and, in addition,

$$\begin{aligned}&\operatorname{cov}\big([Y,Y]^{(all)},[Y,Y]^{(avg)}|X\big)\\ &\quad=2\big(E\epsilon^4-(E\epsilon^2)^2\big)\Big(2\bar n-\frac{1}{K}\Big)+8E\epsilon^2[X,X]_T^{(all)}\frac{1}{K}+o_p\Big(\frac{1}{K}\Big),\end{aligned}$$

we see that

$$\begin{aligned}&\operatorname{var}\big(\widehat{\langle X,X\rangle}_T|X\big)\\ &\quad=\operatorname{var}\big([Y,Y]_T^{(avg)}|X\big)+\frac{1}{K^2}\operatorname{var}\big([Y,Y]_T^{(all)}|X\big)\\ &\qquad-\frac{2}{K}\operatorname{cov}\big([Y,Y]_T^{(avg)},[Y,Y]_T^{(all)}|X\big)\end{aligned}$$

$$
\begin{aligned}
&= 4\frac{\bar{n}}{K}E\epsilon^4 + \frac{1}{K}\big[8[X,X]_T^{(avg)}E\epsilon^2 - 2\big(E\epsilon^4 - (E\epsilon^2)^2\big)\big] + o_p\Big(\frac{1}{K}\Big)\\
&\quad + \frac{1}{K^2}\big[4\bar{n}KE\epsilon^4 + \big(8[X,X]_T^{(all)}E\epsilon^2 - 2\big(E\epsilon^4 - (E\epsilon^2)^2\big)\big) + o_p(1)\big]\\
&\quad - \frac{2}{K}\Big[2\big(E\epsilon^4 - (E\epsilon^2)^2\big)\Big(2\bar{n} - \frac{1}{K}\Big)\\
&\qquad + 8E\epsilon^2[X,X]_T^{(all)}\frac{1}{K} + o_p\Big(\frac{1}{K}\Big)\Big]\\
&= \frac{\bar{n}}{K}8(E\epsilon^2)^2 + \frac{1}{K}\big[8[X,X]_T^{(avg)}E\epsilon^2 - 2\big(E\epsilon^4 - (E\epsilon^2)^2\big)\big]\\
&\quad + o_p\Big(\frac{1}{K}\Big),
\end{aligned}
$$

because at the optimal choice, $K = cn^{2/3}$ and $\bar{n} = c^{-1}n^{1/3}$. Therefore, (60) holds.

A.2 The Relevant Central Limit Theorem

Lemma A.2. Suppose that X is an Itô process. Suppose that Y is related to X through model (4). Then under assumption (11) and definitions (16) and (32),

$$
[Y,Y]_T^{(all)} = [\epsilon,\epsilon]_T^{(all)} + O_p(1) \qquad \text{and}
$$

$$
[Y,Y]_T^{(avg)} = [\epsilon,\epsilon]_T^{(avg)} + [X,X]_T^{(avg)} + O_p\Big(\frac{1}{\sqrt{K}}\Big).
$$

Proof. (a) The one-grid case:

$$
[Y,Y]_T^{(all)} = [X,X]_T^{(all)} + [\epsilon,\epsilon]_T^{(all)} + 2[X,\epsilon]_T^{(all)}. \tag{A.6}
$$

We show that

$$
E\big(\big([X,\epsilon]_T^{(all)}\big)^2|X\big) = O_p(1) \tag{A.7}
$$

and, in particular,

$$
[X,\epsilon]_T^{(all)} = O_p(1). \tag{A.8}
$$

To see (A.7),

$$
\begin{aligned}
[X,\epsilon]_T^{(all)} &= \sum_{i=0}^{n-1}\big(\Delta X_{t_i}\big)\big(\Delta\epsilon_{t_i}\big)\\
&= \sum_{i=0}^{n-1}\big(\Delta X_{t_i}\big)\epsilon_{t_{i+1}} - \sum_{i=0}^{n-1}\big(\Delta X_{t_i}\big)\epsilon_{t_i}\\
&= \sum_{i=1}^{n-1}\big(\Delta X_{t_{i-1}} - \Delta X_{t_i}\big)\epsilon_{t_i} + \Delta X_{t_{n-1}}\epsilon_{t_n} - \Delta X_{t_0}\epsilon_{t_0}.
\end{aligned} \tag{A.9}
$$

Because $E([X,\epsilon]_T^{(all)}|X) = 0$ and ϵ_{t_i} iid for different t_i, we get

$$
\begin{aligned}
E\big(\big([X,\epsilon]_T^{(all)}\big)^2|X\big) &= \operatorname{var}\big([X,\epsilon]_T^{(all)}|X\big)\\
&= E\epsilon^2\Bigg[\sum_{i=1}^{n-1}\big(\Delta X_{t_{i-1}} - \Delta X_{t_i}\big)^2 + \Delta X_{t_{n-1}}^2 + \Delta X_{t_0}^2\Bigg]\\
&= 2[X,X]_T E\epsilon^2 - 2E\epsilon^2\sum_{i=1}^{n-1}\big(\Delta X_{t_{i-1}}\big)\big(\Delta X_{t_i}\big)\\
&\le 4[X,X]_T E\epsilon^2,
\end{aligned} \tag{A.10}
$$

by the Cauchy–Schwarz inequality, from which and from $[X,X]^{(all)}$ being of order $O_p(1)$, (A.7) follows. Hence (A.8) follows by the Markov inequality.

(b) The multiple-grid case: Notice that

$$
[Y,Y]^{(avg)} = [X,X]_T^{(avg)} + [\epsilon,\epsilon]_T^{(avg)} + 2[X,\epsilon]_T^{(avg)} \tag{A.11}
$$

(A.11) strictly follows from model (4) and the definitions of grids and $[\cdot,\cdot]_t^{(avg)}$; see Section 3.2.

We need to show that

$$
E\big(\big([X,\epsilon]_T^{(avg)}\big)^2|X\big) = O_p\Big(\frac{1}{K}\Big), \tag{A.12}
$$

in particular,

$$
[X,\epsilon]_T^{(avg)} = O_p\Big(\frac{1}{K^{1/2}}\Big) \tag{A.13}
$$

and $\operatorname{var}([X,\epsilon]_T^{(avg)}|X) = E[([X,\epsilon]_T^{(avg)})^2|X]$.

To show (A.12), note that $E([X,\epsilon]_T^{(avg)}|X) = 0$ and

$$
\begin{aligned}
E\big[\big([X,\epsilon]_T^{(avg)}\big)^2|X\big] &= \operatorname{var}\big([X,\epsilon]_T^{(avg)}|X\big)\\
&= \frac{1}{K^2}\sum_{k=1}^{K}\operatorname{var}\big([X,\epsilon]_T^{(k)}|X\big)\\
&\le \frac{4E\epsilon^2}{K}[X,X]_T^{(avg)} = O_p\Big(\frac{1}{K}\Big),
\end{aligned}
$$

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.10). Then the order follows because $[X,X]_T^{(avg)} = O_p(1)$; see the method of Mykland and Zhang (2002) for a rigorous development for the order of $[X,X]_T^{(avg)}$.

Theorem A.1. Suppose that X is an Itô process of form (1). Suppose that Y is related to X through model (4), and that (11) 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 (33), as $n \to \infty$, $(\sqrt{n}(\widehat{E\epsilon^2} - E\epsilon^2), \sqrt{\frac{K}{\bar{n}}}([Y,Y]_T^{(avg)} - [X,X]_T^{(avg)} - 2\bar{n}E\epsilon^2))$ converges in law to a bivariate normal, with mean 0 and covariance matrix

$$
\begin{pmatrix} E\epsilon^4 & 2\operatorname{var}(\epsilon^2) \\ 2\operatorname{var}(\epsilon^2) & 4E\epsilon^4 \end{pmatrix} \tag{A.14}
$$

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

Proof. By Lemma A.2, we need the distribution of $[\epsilon,\epsilon]^{(avg)}$ and $[\epsilon,\epsilon]^{(all)}$. First, we explore the convergence of

$$
\frac{1}{\sqrt{n}}\big([\epsilon,\epsilon]_T^{(all)} - 2nE\epsilon^2, [\epsilon,\epsilon]_T^{(avg)}K - 2\bar{n}KE\epsilon^2\big). \tag{A.15}
$$

Recall that all of the sampling points $t_0, t_1, \ldots, t_n$ 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 the kth grid. As before, if $t_i \in \mathcal{G}^{(k)}$, then $t_{i,-}$ and $t_{i,+}$ are the previous and next elements 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

$$
\begin{aligned}
M_T^{(1)} &= \frac{1}{\sqrt{n}}\sum_{t_i\in\mathcal{G}}\big(\epsilon_{t_i}^2 - E\epsilon^2\big),\\
M_T^{(2)} &= \frac{1}{\sqrt{n}}\sum_{t_i\in\mathcal{G}}\epsilon_{t_i}\epsilon_{t_{i-1}},\\
M_T^{(3)} &= \frac{1}{\sqrt{n}}\sum_{k=1}^{K}\sum_{t_i\in\mathcal{G}^{(k)}}\epsilon_{t_i}\epsilon_{t_i,-}.
\end{aligned} \tag{A.16}
$$

We first find the asymptotic distribution of $(M_T^{(1)}, M_T^{(2)}, M_T^{(3)})$ using the martingale central limit theorem. Then we use the result to find the limit of (A.15).

Note that $(M_T^{(1)}, M_T^{(2)}, M_T^{(3)})$ are the end points of martingales with respect to filtration $\mathcal{F}_i = \sigma(\epsilon_{t_j}, j \le i, X_t, \text{ all } t)$. 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 the proof of Lemma 1, and are different from the continuous time quadratic variations in (12)]:

$$\begin{aligned}
\langle M^{(1)}, M^{(1)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \operatorname{var}\bigl(\epsilon_{t_i}^2 - E\epsilon^2 \bigm| \mathcal{F}_{t_{i-1}}\bigr) \\
&= \operatorname{var}(\epsilon^2), \\
\langle M^{(2)}, M^{(2)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \operatorname{var}\bigl(\epsilon_{t_i}\epsilon_{t_{i-1}} \bigm| \mathcal{F}_{t_{i-1}}\bigr) \\
&= \frac{E\epsilon^2}{n} \sum_{t_i \in \mathcal{G}} \epsilon_{t_{i-1}}^2 = (E\epsilon^2)^2 + o_p(1), \qquad \text{(A.17)} \\
\langle M^{(3)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \operatorname{var}\bigl(\epsilon_{t_i}\epsilon_{t_{i,-}} \bigm| \mathcal{F}_{t_{i-1}}\bigr) \\
&= \frac{E\epsilon^2}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i,-}}^2 = (E\epsilon^2)^2 + o_p(1),
\end{aligned}$$

by the law of large numbers.

Similarly, for the predictable quadratic covariations,

$$\begin{aligned}
\langle M^{(1)}, M^{(2)} \rangle_T &= \frac{1}{n} \sum_{t_i \in \mathcal{G}} \operatorname{cov}\bigl(\epsilon_{t_i}^2 - E\epsilon^2, \epsilon_{t_i}\epsilon_{t_{i-1}} \bigm| \mathcal{F}_{t_{i-1}}\bigr) \\
&= E\epsilon^3 \frac{1}{n} \sum_{t_i \in \mathcal{G}} \epsilon_{t_{i-1}} = o_p(1), \\
\langle M^{(1)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \operatorname{cov}\bigl(\epsilon_{t_i}^2 - E\epsilon^2, \epsilon_{t_i}\epsilon_{t_{i,-}} \bigm| \mathcal{F}_{t_{i-1}}\bigr) \qquad \text{(A.18)} \\
&= E\epsilon^3 \frac{1}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i,-}} = o_p(1), \\
\langle M^{(2)}, M^{(3)} \rangle_T &= \frac{1}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \operatorname{cov}\bigl(\epsilon_{t_i}\epsilon_{t_{i-1}}, \epsilon_{t_i}\epsilon_{t_{i,-}} \bigm| \mathcal{F}_{t_{i-1}}\bigr) \\
&= \frac{E\epsilon^2}{n} \sum_{k=1}^{K} \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_{i-1}}\epsilon_{t_{i,-}} = o_p(1),
\end{aligned}$$

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

Because the ϵ_{t_i}'s are iid and $E\epsilon_{t_i}^4 < \infty$, the conditional Lindeberg conditions are satisfied. Hence by the martingale central limit theorem (see Hall and Heyde 1980, condition 3.1, p. 58), $(M^{(1)}, M^{(2)}, M^{(3)})$ 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 $\operatorname{var}(\epsilon)$, $(E\epsilon^2)^2$, and $(E\epsilon^2)^2$.

Returning to (A.15), we have that

$$\begin{aligned}
&[\epsilon, \epsilon]^{(all)} - 2nE\epsilon^2 \\
&\quad = 2 \sum_{i \ne 0,n} (\epsilon_{t_i}^2 - E\epsilon^2) + (\epsilon_{t_0}^2 - E\epsilon^2) + (\epsilon_{t_n}^2 - E\epsilon^2) - 2 \sum_{t_i > 0} \epsilon_{t_i}\epsilon_{t_{i-1}} \\
&\quad = 2\sqrt{n}\bigl(M^{(1)} - M^{(2)}\bigr) + O_p(1). \qquad \text{(A.19)}
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
&[\epsilon, \epsilon]^{(k)} - 2n_k E\epsilon^2 \\
&\quad = \sum_{t_i \in \mathcal{G}^{(k)}, t_i \ne \max \mathcal{G}^{(k)}} \bigl(\epsilon_{t_i,+} - \epsilon_{t_i}\bigr)^2 - 2n_k E\epsilon^2 \\
&\quad = 2 \sum_{t_i \in \mathcal{G}^{(k)}} \bigl(\epsilon_{t_i}^2 - E\epsilon^2\bigr) - \bigl(\epsilon_{\min \mathcal{G}^{(k)}}^2 - E\epsilon^2\bigr) - \bigl(\epsilon_{\max \mathcal{G}^{(k)}}^2 - E\epsilon^2\bigr) \\
&\qquad - 2 \sum_{t_i \in \mathcal{G}^{(k)}} \epsilon_{t_i}\epsilon_{t_i,-}, \qquad \text{(A.20)}
\end{aligned}$$

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

$$\begin{aligned}
[\epsilon, \epsilon]_T^{(avg)} K - 2\bar{n}E\epsilon^2 K &= \sqrt{n}\bigl(2M^{(1)} - 2M^{(3)}\bigr) - R \\
&= 2\sqrt{n}\bigl(M^{(1)} - M^{(3)}\bigr) + O_p\bigl(K^{1/2}\bigr), \qquad \text{(A.21)}
\end{aligned}$$

because $R = \sum_{k=1}^{K} [(\epsilon_{\min \mathcal{G}^{(k)}}^2 - E\epsilon^2) + (\epsilon_{\max \mathcal{G}^{(k)}}^2 - E\epsilon^2)]$, satisfying $ER^2 = \operatorname{var}(R) \le 4K \operatorname{var}(\epsilon^2)$.

Because $n^{-1}K \to 0$, and because the error terms in (A.19) and (A.20) are uniformly integrable, it follows that

$$(A.15) = 2\bigl(M^{(1)} - M^{(2)}, M^{(1)} - M^{(3)}\bigr) + o_p(1). \qquad \text{(A.22)}$$

Hence (A.15) is also asymptotically normal with covariance matrix

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

By Lemma A.2, and as $n^{-1}K \to 0$,

$$\frac{1}{\sqrt{n}}\bigl([Y, Y]_T^{(all)} - 2nE\epsilon^2, K\bigl([Y, Y]_T^{(avg)} - [X, X]_T^{(avg)} - 2\bar{n}E\epsilon^2\bigr)\bigr) \Bigm| X$$

is asymptotically normal,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \begin{pmatrix} [Y, Y]_T^{(all)} - 2nE\epsilon^2 \\ [Y, Y]_T^{(avg)} K - 2\bar{n}KE\epsilon^2 - [X, X]_T^{(avg)} K \end{pmatrix} \Bigm| X \\
&\quad = 2 \begin{pmatrix} M^{(1)} - M^{(2)} \\ M^{(1)} - M^{(3)} \end{pmatrix} + o_p(1) \\
&\quad \xrightarrow{\mathcal{L}} 2\mathrm{N}\left(0, \begin{pmatrix} E\epsilon^4 & \operatorname{var}(\epsilon^2) \\ \operatorname{var}(\epsilon^2) & E\epsilon^4 \end{pmatrix}\right). \qquad \text{(A.23)}
\end{aligned}$$

Because

$$\begin{aligned}
\widehat{E\epsilon^2} &= \frac{1}{2n}[Y, Y]_T^{(all)} \qquad \text{and} \\
\frac{K}{\sqrt{n}} &= \sqrt{\frac{K}{\bar{n}}}(1 + o(1)). \qquad \text{(A.24)}
\end{aligned}$$

Theorem A.1 follows.

Proof of Lemma 1. For simplicity, and without loss of generality, we assume that $\mathcal{H} = \mathcal{G}$. The result follows the proofs of Lemma A.2 and Theorem A.1, but with a more exact representation of $[Y, Y]_T^{(all)} - [X, X]_T^{(all)}$. Specifically, from (A.9) and (A.19), we have that

$$\begin{aligned}
&[Y, Y]_T^{(all)} - [X, X]_T^{(all)} - 2nE\epsilon^2 \\
&\quad = [\epsilon, \epsilon]_T - 2nE\epsilon^2 + 2[X, \epsilon]_T^{(all)} \\
&\quad = 2 \sum_{i \ne 0,n} \bigl(\epsilon_{t_i}^2 - E\epsilon^2\bigr) + \bigl(\epsilon_{t_0}^2 - E\epsilon^2\bigr) + \bigl(\epsilon_{t_n}^2 - E\epsilon^2\bigr) - 2 \sum_{t_i > 0} \epsilon_{t_i}\epsilon_{t_{i-1}} \\
&\qquad + 2 \sum_{i=1}^{n-1} \bigl(\Delta X_{t_{i-1}} - \Delta X_{t_i}\bigr)\epsilon_{t_i} + 2\Delta X_{t_{n-1}}\epsilon_{t_n} - 2\Delta X_{t_0}\epsilon_{t_0}.
\end{aligned}$$

With the same filtration as in the proof of Theorem A.1, this is the sum of a martingale triangular array with increment $2(\epsilon_{t_i}^2 - E\epsilon^2) - 2\epsilon_{t_i}\epsilon_{t_{i-1}} + (\Delta X_{t_{i-1}} - \Delta X_{t_i})\epsilon_{t_i}$ for $i \neq 0, n$. Again, the conditions of the martingale central limit theorem (see Hall and Heyde 1980, chap. 3) are satisfied, by (17), whence the term is asymptotically normal, conditionally on the X process. The conditional variance is given by (20) under the conditions mentioned in the statement of the lemma. Because $E\epsilon^2$ can vary with n, the $O_p(n_{\text{sparse}}^{-1/2})$ term in (20) is actually of order $O_p(n_{\text{sparse}}^{-1/2} E\epsilon^2)$.

A.3 Asymptotics of D_T

For transparency of notation, we take $\overline{\Delta t} = T/n$; in other words, the average of the Δt_i.

Proof of Theorem 2. Note first that because we assume that μ_t and σ_t are continuous, they are locally bounded. Because here we seek to show convergence in probability, we can therefore assume without loss of generality that $|\mu_t|$ is bounded above by a nonrandom constant and also that $\infty > \sigma^+ \geq \sigma_t \geq \sigma^- > 0$, where σ^+ and σ^- are nonrandom. This is by the usual stopping time argument.

Also note that once μ_t and σ_t are bounded as we have now assumed, by Girsanov's theorem (see, e.g., Karatzas and Shreve 1991, chap. 3.5; Jacod and Shiryaev 2003, chap. II.3b), we can, without loss of generality, further suppose that $\mu_t = 0$ identically. (If the convergence in probability holds under the equivalent probability, then it also will hold under the original one.)

Now for the derivation. First, consider

$$\begin{aligned}\bigl(X_{t_i} - X_{t_{i,-}}\bigr)^2 &= \left(\sum_{t_j=t_{i,-}}^{t_{i-1}} \Delta X_{t_j}\right)^2 \\ &= \sum_{t_j=t_{i,-}}^{t_{i-1}} \bigl(\Delta X_{t_j}\bigr)^2 + 2 \sum_{t_{i-1}\geq t_k > t_l \geq t_{i,-}} \Delta X_{t_k} \Delta X_{t_l}.\end{aligned}$$

It follows that

$$\begin{aligned}&[X,X]_T^{(avg)} \\ &\quad = [X,X]_T^{(all)} + 2\sum_{i=1}^{n-1} \bigl(\Delta X_{t_i}\bigr) \sum_{j=1}^{K\wedge i} \left(1 - \frac{j}{K}\right) \bigl(\Delta X_{t_{i-j}}\bigr) + O_p(K/n).\end{aligned}$$

The remainder term incorporates end effects and has order $O_p(K/n)$ by (41) and (42), and due to the regular allocation of points to grids. Following Jacod and Protter (1998) and Mykland and Zhang (2002), $[X,X]_T^{(all)} - \langle X,X\rangle_T = O_p(1/\sqrt{n}) = o_p(\sqrt{K/n})$ too, under the assumption (17) [which is a special case of assumption (41)] and that $\inf_{t\in[0,T]} \sigma_t^2 > 0$ almost surely. In other words,

$$\begin{aligned}D_T &= \bigl([X,X]_T^{(avg)} - \langle X,X\rangle_T\bigr) \\ &= 2\sum_{i=1}^{n-1} \bigl(\Delta X_{t_i}\bigr) \sum_{j=1}^{K\wedge i} \left(1 - \frac{j}{K}\right) \bigl(\Delta X_{t_{i-j}}\bigr) + o_p\bigl(\sqrt{K\overline{\Delta t}}\bigr). \qquad \text{(A.25)}\end{aligned}$$

It follows that

$$\begin{aligned}\langle D,D\rangle_T &= 4\sum_{j=1}^{n-1} \bigl(\Delta\langle X,X\rangle_{t_i}\bigr) \left(\sum_{j=1}^{K\wedge i} \left(1 - \frac{j}{K}\right) \bigl(\Delta X_{t_{i-j}}\bigr)\right)^2 \\ &\quad + o_p(K\overline{\Delta t}) \\ &= (I) + (II) + o_p(K\overline{\Delta t}), \qquad \text{(A.26)}\end{aligned}$$

where

$$\begin{aligned}(I) &= 4\sum_{i=1}^{n-1} \bigl(\Delta\langle X,X\rangle_{t_i}\bigr) \left(\sum_{j=1}^{K\wedge i} \left(1 - \frac{j}{K}\right)^2 \bigl(\Delta X_{t_{i-j}}\bigr)^2\right) \\ &= 4\sum_{i=1}^{n-1} \sigma_{t_i}^4 \Delta t_i \left(\sum_{j=1}^{K\wedge i} \left(1 - \frac{j}{K}\right)^2 \Delta t_{i-j}\right) + o_p(K\overline{\Delta t}) \\ &= K\overline{\Delta t} \sum_{i=1}^{n-1} \sigma_{t_i}^4 h_i \Delta t_i + o_p(K\overline{\Delta t}),\end{aligned}$$

where h_i is given by (43). Thus Theorem 2 will have been shown if we can show that

$$(II) = 8\sum_{i=1}^{n-1} \Delta\langle X,X\rangle_{t_i} \zeta_i \qquad \text{(A.27)}$$

is of order $o_p(K\overline{\Delta t})$, where

$$\zeta_i = \sum_{l>r\geq 0}^{i-1} \bigl(\Delta X_{t_l}\bigr)\bigl(\Delta X_{t_r}\bigr) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+. \qquad \text{(A.28)}$$

We do this as follows. Denote $\delta^+ = \max_i |\Delta t_i| = O(\frac{1}{n})$, and set $(II)' = 8\sum_{i=1}^{n-1} \Delta t_i \sigma_{t_{i-K}}^2 \zeta_i$. Then, by Hölder's inequality, $E|(II) - (II)'| \leq \delta^+ n \| \sup_{|t-s|\leq(K+1)\delta^+} |\sigma_t^2 - \sigma_s^2| \|_2 \sup_i \|\zeta_i\|_2 = o(\delta^+ K)$ because of the boundedness and continuity of σ_t and because, applying the Burkholder–Davis–Gundy inequality twice, we have that

$$\begin{aligned}E(\zeta_i^2) &\leq E\sum_{l=1}^{i-1} \Delta\langle X,X\rangle_{t_l} \\ &\qquad \times \left(\sum_{r\geq 0}^{(l-1)\wedge(i-1)} \bigl(\Delta X_{t_r}\bigr) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+\right)^2 \\ &\leq \delta^+ (\sigma^+)^2 \\ &\qquad \times \sum_{l=1}^{i-1} E\left(\sum_{r\geq 0}^{(l-1)\wedge(i-1)} \bigl(\Delta X_{t_r}\bigr) \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+\right)^2 \\ &\leq (\delta^+)^2 (\sigma^+)^4 \\ &\qquad \times \sum_{l=1}^{i-1} \sum_{r\geq 0}^{(l-1)\wedge(i-1)} \left(\left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+\right)^2 \\ &= O((\delta^+ K)^2) \quad \text{uniformly in } i, \qquad \text{(A.29)}\end{aligned}$$

where the second-to-last transition does for the inner sum what the two first inequalitities do for the outer sum. Now rewrite

$$\begin{aligned}(II)' &= 8\sum_{l>r\geq 0}^{n-1} \bigl(\Delta X_{t_l}\bigr)\bigl(\Delta X_{t_r}\bigr) \\ &\qquad \times \sum_{i=l+1}^{n-1} \Delta t_i \sigma_{t_{i-K}}^2 \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+. \qquad \text{(A.30)}\end{aligned}$$

In the same way as in (A.29), we then obtain

$$\begin{aligned}E((II)')^2 &\leq 64(\delta^+)^2 (\sigma^+)^4 \\ &\qquad \times \sum_{l>r\geq 0}^{n-1} E\left(\sum_{i=l+1}^{n-1} \Delta t_i \sigma_{t_{i-K}}^2 \right. \\ &\qquad\qquad \left. \times \left(1 - \frac{i-l}{K}\right)^+ \left(1 - \frac{i-r}{K}\right)^+\right)^2\end{aligned}$$

$$\le 64(\delta^+)^2(\sigma^+)^4 \times (\delta^+)^2(\sigma^+)^4 \times \sum_{l>r\ge 0}^{n-1}\left(\sum_{i=l+1}^{n-1}\left(1-\frac{i-l}{K}\right)^+\left(1-\frac{i-r}{K}\right)^+\right)^2$$

$$\le 64(\delta^+)^4(\sigma^+)^8 nK^3 = O((\delta^+K)^3) = o((\delta^+K)^2), \quad \text{(A.31)}$$

by assumptions (41) and (42). The transition from the second to the last line in (A.31) is due to the fact that $\sum_{i=l+1}^{n-1}(1-\frac{i-l}{K})^+(1-\frac{i-r}{K})^+ \le K$, and $=0$ when $l-r\ge K$. Thus we have shown that $(II) = o_p(K\overline{\Delta t})$, and hence Theorem 2 has been established.

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

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

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

Proof of Theorem 3. We can show 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

$$\sup_t \left|\frac{1}{\overline{\Delta t}K}\langle D, L\rangle_t\right| \xrightarrow{p} 0. \quad \text{(A.32)}$$

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

Finally, in the case where η_n^2 does not converge, one can still use the mixed normal with variance η_n^2. This is because every subsequence of η_n^2 has a further subsequence that does converge in probability to some η^2 and hence for which the assumption (48) in Theorem 3 would be satisfied. The reason for this is that one can define the distribution function of a finite measure by

$$G_n(t) = \sum_{t_{i+1}\le t} h_i\Delta t_i. \quad \text{(A.33)}$$

Because $G_n(t)\le T\sup_i h_i$, it follows from (45) 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 $G_n\to G$, we then get that

$$\eta_n^2 = \int_0^T \sigma_t^4\, dG_n(t) \to \int_0^T \sigma_t^4\, dG(t) \quad \text{(A.34)}$$

almost surely, because we have assumed that $\langle X, X\rangle_t'$ is a continuous function of t. One then defines η^2 to be the (subsequence-dependent) right side of (A.34). To proceed further with the asymptotics, continue the foregoing subsequence and note that

$$\frac{1}{\overline{\Delta t}K}\langle D, D\rangle_t \approx \int_0^t \sigma_s^4\, dG_n(s) \to \int_0^t \sigma_s^4\, dG(s)$$

completing the proof.

[Received October 2003. Revised December 2004.]

REFERENCES

Aït-Sahalia, Y., Mykland, P. A., and 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.

Aldous, D. J., and Eagleson, G. K. (1978), "On Mixing and Stability of Limit Theorems," *The 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.

Bandi, F. M., and Russell, J. R. (2003), "Microstructure Noise, Realized Volatility and Optimal Sampling," technical report, University of Chicago, Graduate School of Business.

Barndorff-Nielsen, O. E., and Shephard, N. (2002), "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society*, Ser. B, 64, 253–280.

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.

Brown, S. J. (1990), "Estimating Volatility," in *Financial Options: From Theory to Practice*, eds. S. Figlewski, W. Silber, and M. Subrahmanyam, Homewood, IL: Business One-Irwin, pp. 516–537.

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.

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

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," unpublished doctoral thesis, Université de Marne-la-Vallée.

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

Hansen, P. R., and Lunde, A. (2004a), "Realized Variance and IID Market Microstructure Noise," technical report, Brown University, Dept. of Economics.

——— (2004b), "An Unbiased Measure of Realized Variance," technical report, Brown University, Dept. of Economics.

Heston, S. (1993), "A Closed-Form Solution for Options With Stochastic Volatility With Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

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," *The Annals of Probability*, 26, 267–307.

Jacod, J., and Shiryaev, A. N. (2003), *Limit Theorems for Stochastic Processes* (2nd ed.), New York: Springer-Verlag.

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," technical report, University of Chicago, Dept. of Statistics.

O'Hara, M. (1995), *Market Microstructure Theorys*, Cambridge, MA: Blackwell.

Oomen, R. (2004), "Properties of Realized Variance for a Pure Jump Process: Calendar Time Sampling versus Business Time Sampling," technical report, University of Warwick, Warwick Business School.

Rényi, A. (1963), "On Stable Sequences of Events," *Sankhyā*, Ser. A, 25, 293–302.

Rootzen, H. (1980), "Limit Distributions for the Error in Approximations of Stochastic Integrals," *The Annals of Probability*, 8, 241–251.

Zhou, B. (1996), "High-Frequency Data and Volatility in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 14, 45–52.

Zumbach, G., Corsi, F., and Trapletti, A. (2002), "Efficient Estimation of Volatility Using High-Frequency Data," technical report, Olsen & Associates.