Bernoulli **12**(6), 2006, 1019–1043

Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach

LAN ZHANG

Department of Finance, University of Illinois at Chicago, 601 S Morgan Street (M/C 168), 2430 University Hall, Chicago IL 60607, USA. E-mail: lanzhang@uic.edu.

With the availability of high-frequency financial data, nonparametric estimation of the volatility of an asset return process becomes feasible. A major problem is how to estimate the volatility consistently and efficiently, when the observed asset returns contain error or noise, for example, in the form of microstructure noise. The issue of consistency has been addressed in the recent literature. However, the resulting estimator is not efficient. In work by Zhang, Myland and Aït-Sahalia, the best estimator converges to the true volatility only at the rate of $n^{-1/6}$. In this paper, we propose an estimator, the multi-scale realized volatility (MSRV), which converges to the true volatility at the rate of $n^{-1/4}$, which is the best attainable. We show a central limit theorem for the MSRV estimator, which permits intervals to be set for the true integrated volatility on the basis of the MSRV.

Keywords: consistency; dependent noise; discrete observation; efficiency; Itô process; microstructure noise; observation error; rate of convergence; realized volatility

1. Introduction

This paper is concerned with how to estimate volatility nonparametrically and efficiently.

With the availability of high-frequency financial data, nonparametric estimation of volatility of an asset return process becomes feasible. A major problem is how to estimate the volatility consistently and efficiently, when the observed asset returns are noisy. The issue of consistency has been addressed in the recent literature. However, the resulting estimator is not efficient. In Zhang *et al.* (2005) the best estimator converges to the true volatility only at the rate of $n^{-1/6}$. In this paper, we propose an estimator which converges to the true volatility at the rate of $n^{-1/4}$, which is the best attainable. The new estimator remains consistent when the observation noise is dependent. We call the estimator the *multi-scale realized volatility* (MSRV).

To demonstrate the idea, let $\{Y\}$ be the observed log prices of a financial instrument. The observations take place on a grid of time points $\mathcal{G}_n = \{t_{n,i}, i = 0, 1, 2, \ldots, n\}$ spanning the time interval $[0, T]$. For the purposes of asymptotics, we shall let $\mathcal{G}_n$ become dense in $[0, T]$ as $n \to \infty$. The $\mathcal{G}_n$s need not be nested.

Suppose that the $\{Y_{t_{n,i}}\}$ are noisy, and the corresponding true (latent) log prices are $\{X_{t_{n,i}}\}$. Their relation can be modelled as

1350–7265

$$Y_{t_{n,i}} = X_{t_{n,i}} + \epsilon_{t_{n,i}}, \tag{1}$$

where $t_{n,i} \in \mathcal{G}_n$. The noise $\epsilon_{t_{n,i}}$ will be assumed to be independent of X and independently and identically distributed (i.i.d.).

The model in (1) is quite realistic, as evidenced by the existence of microstructure noise in the price process (Brown 1990; Zhou 1996; Corsi *et al.* 2001).

We further assume that the true log prices $\{X\}$ satisfy the equation

$$\mathrm{d}X_t = \mu_t \,\mathrm{d}t + \sigma_t \,\mathrm{d}B_t, \tag{2}$$

where B_t is a standard Brownian motion. Typically, the drift coefficient μ_t and the diffusion coefficient σ_t are stochastic in the sense that

$$\mathrm{d}X_t(\omega) = \mu(t, \omega)\mathrm{d}t + \sigma(t, \omega)\mathrm{d}B_t(\omega). \tag{3}$$

Throughout this paper, we use the notation in (2) to denote (3). By the model in (3) we mean that $\{X\}$ follows an Itô process. A special case is that $\{X\}$ is Markov, where $\mu_t = \mu(t, X_t)$ and $\sigma_t = \sigma(t, X_t)$. In financial literature, σ_t is called the instantaneous volatility of X.

Our goal is to estimate $\int_0^T \sigma_t^2 \,\mathrm{d}t$, where T can be a day, a month, or some other time horizon. For simplicity, we call $\int_0^T \sigma_t^2 \,\mathrm{d}t$ the integrated volatility, and denote it by

$$\langle X, X \rangle = \int_0^T \sigma_t^2 \,\mathrm{d}t.$$

The general question is how to estimate $\int_0^T \sigma_t^2 \,\mathrm{d}t$ nonparametrically, if one can only observe the noisy data $Y_{t_{n,i}}$ at discrete times $t_{n,i} \in \mathcal{G}_n$. $\mathcal{G}_n$ is formally defined in Section 5.

To the best of our knowledge, there are two types of nonparametric estimator for $\int_0^T \sigma_t^2 \,\mathrm{d}t$ in the current literature. The first type, the simpler of the two, is to sum up all the squared returns in $[0, T]$:

$$[Y, Y]^{(n,1)} = \sum_{t_{n,i} \in \mathcal{G}_n, i \geqslant 1} (Y_{t_{n,i}} - Y_{t_{n,i-1}})^2, \tag{4}$$

this estimator is generally called *realized volatility* or *realized variance*. However, it has been reported that realized volatility using high-frequency data is not desirable (see, for example, Brown 1990; Zhou 1996; Corsi *et al.* 2001). The reason is that it is not consistent, even if the noisy observations Y are available continuously. Under discrete observations, the bias and the variance of the realized volatility are of the same order as the sample size n.

A slight modification of (4) is to use the sum of squared returns from a 'sparsely selected' sample, that is, using a subgrid of $\mathcal{G}_n$. The idea is that by using sparse data, one reduces the bias and variance of the conventional realized volatility. This approach is quite popular in the empirical finance literature. However, this 'sparse' estimator is still not consistent. In addition, the choice of data to subsample and to discard is arbitrary. The behaviour of this type of estimator, and of a sufficiency-based improvement of it, is analyzed in Zhang *et al.* (2005); see also Bandi *et al.* (2006).

A second type of estimator for $\int_0^T \sigma_t^2 \,\mathrm{d}t$ is based on *two sampling scales*. As introduced by Zhang *et al.* (2005: 1402), the two-scales realized volatility (TSRV) has the form

$$\langle \widehat{X, X} \rangle^{(\mathrm{TSRV})} = [Y, Y]^{(n,K)} - 2\frac{n-K+1}{nK}[Y, Y]^{(n,1)}, \tag{5}$$

where

$$[Y, Y]^{(n,K)} = \frac{1}{K} \sum_{t_{n,i} \in \mathcal{G}_n, i \geqslant K} (Y_{t_{n,i}} - Y_{t_{n,i-K}})^2, \tag{6}$$

for K a positive integer. Thus the estimator in (5) averages the squared returns from sampling every data point ($[Y, Y]_T^{(n,1)}$) and those from sampling every Kth data point ($[Y, Y]_T^{(n,K)}$). Its asymptotic behaviour was derived when $K \to \infty$ as $n \to \infty$. The TSRV estimator has many desirable features, including asymptotic unbiasedness, consistency, and asymptotic normality.[1] However, its rate of convergence is not satisfactory. For instance, the best estimator in Zhang *et al.* (2005) converges to $\int_0^T \sigma_t^2 \, \mathrm{d}t$ at the rate of $n^{-1/6}$.

In this paper, we propose a new class of estimators, collectively referred to as *multi-scale realized volatility* (MSRV), which converge to $\int_0^T \sigma_t^2 \, \mathrm{d}t$ at the rate of $n^{-1/4}$. This new estimator has the form

$$\langle \widehat{X, X} \rangle^{(n)} = \sum_{i=1}^{M} \alpha_i [Y, Y]^{(n,K_i)},$$

where M is a positive integer greater than 2. Comparing to $\langle \widehat{X, X} \rangle_T^{(\mathrm{TSRV})}$, which uses two time scales (1 and K), $\langle \widehat{X, X} \rangle^{(n)}$ combines M different time scales. The weights α_i are selected so that $\langle \widehat{X, X} \rangle^{(n)}$ is asymptotically unbiased and has optimal convergence rate. The rationale is that by combining more than two time scales, we can improve the efficiency of the estimator. Interestingly, the $n^{-1/4}$ rate of convergence in our new estimator is the same as that in parametric estimation for volatility, when the true process is Markov (see Gloter and Jacod 2000a, 2000b). Thus this rate is the best attainable. Earlier related results in the same direction can be found in Stein (1987, 1990, 1993) and Ying (1991, 1993). See also Aït-Sahalia *et al.* (2005a). Related work can also be found in Curci and Corsi (2005) and Barndorff-Nielsen *et al.* (2006). For the estimating functions-based approach, there is a nice review by Bibby *et al.* (2002).

We emphasize that our MSRV estimator is nonparametric, and the true process follows a general Itô process, where the volatility could depend on the entire history of the X process plus additional randomness.

The paper is organized as follows. In Section 2 we motivate the idea of averaging over M different time scales. As we shall see, our estimator is unbiased, and its asymptotic variance comes from the noise (the $\epsilon_{t_{n,i}}$) as well as from the discreteness of the sampling times $t_{n,i}$. In Sections 3 and 4 we derive the weights a_i which are optimal for minimizing the variance that comes from noise, and we give a central limit theorem for the contribution of the noise term. A specific family of weights is introduced in Section 4. We then elaborate on the discretization error in Section 5, and show a central limit theorem for this error. Section 6 gives the central limit theorem for the MSRV estimator.

[1]A related estimator can be found in Zhou (1996). However, this estimator is not consistent. See also Hansen and Lunde (2006).

For the statement of results, we shall use the following assumptions:

Assumption 1 *Structure of the latent process. The X process is adapted to a filtration $(\mathcal{X}_t)$ and satisfies* (2), *where B_t is an $(\mathcal{X}_t)$-Brownian motion, and the μ_t and σ_t are $(\mathcal{X}_t)$-adapted processes which are continuous almost surely. Also both processes are bounded above by a constant, and σ_t is bounded away from zero. We write $\mathcal{X} = \mathcal{X}_T$.*

As a technical matter, we suppose that there is a σ-field $\mathcal{N}$ and a continuous finite-dimensional local martingale (M_t), so that $\mathcal{X}_t = \sigma(M_s, 0 \leqslant s \leqslant t) \vee \mathcal{N}$.

Assumption 2 *Structure of the noise. The $\epsilon_{t_{n,i}}$ are i.i.d. with* $\mathrm{E}[\epsilon] = 0$ *and* $\mathrm{E}[\epsilon^4] < \infty$*. The $\epsilon_{t_{n,i}}$ are also independent of $\mathcal{X}$.*

These assumptions are not minimal for all results. In terms of the structure of the process, see Section 5 in Jacod and Protter (1998) and Proposition 1 in Mykland and Zhang (2006) for examples of statements where the μ and σ processes are not assumed to be continuous. For the methodology to incorporate dependence into the noise structure, see Aït-Sahalia *et al.* (2005b). Our current assumptions, however, provide a set-up with substantial generality without overly complicating the proofs.

The final item in Assumption 1 is standard for the type of limit result that we discuss; see similar conditions in Jacod and Protter (1998), Zhang (2001), Mykland and Zhang (2006) and Zhang *et al.* (2005).

2. Motivation: Averaging the observations of $\langle X, X\rangle$

In Zhang *et al.* (2005) it was observed that by combining the square increments of the returns from two time scales, the resulting two-scale estimator $\widehat{\langle X, X_T\rangle}^{(\mathrm{TSRV})}$ in (5) improves upon the realized volatility, which uses only one time scale, as in (4). The improvement is about reducing both the bias and the variance.

If the two-scale estimator is better than the one-scale estimator, a natural question arises as to estimators combining more than two time scales. This question motivates the present paper. In this section we briefly go through the main argument.

To proceed, recall definition (6) of $[Y, Y]^{(n,K)}$, and set, similarly,

$$[X, \epsilon]^{(n,K)} = \frac{1}{K} \sum_{t_{n,i} \in \mathcal{G}_n, i \geqslant K} (X_{t_{n,i}} - X_{t_{n,i-K}})(\epsilon_{t_{n,i}} - \epsilon_{t_{n,i-K}}) \tag{7}$$

and

$$[\epsilon, \epsilon]^{(n,K)} = \frac{1}{K} \sum_{t_{n,i} \in \mathcal{G}_n, i \geqslant K} (\epsilon_{t_{n,i}} - \epsilon_{t_{n,i-K}})^2.$$

Under (1), one can decompose $[Y, Y]^{(n,K)}$ into

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

We consider estimators of the form

$$\widehat{\langle X, X\rangle}^{(n)} = \sum_{i=1}^{M} \alpha_i [Y, Y]^{(n,K_i)}, \tag{8}$$

where the α_i are the weights to be determined. A first intuitive requirement is obtained by noting that

$$\mathrm{E}(\widehat{\langle X, X\rangle}^{(n)} | X \text{ process }) = \sum_{i=1}^{M} \alpha_i [X, X]^{(n,K_i)} + 2\mathrm{E}\epsilon^2 \sum_{i=1}^{M} \alpha_i \frac{n+1-K_i}{K_i}. \tag{9}$$

Since the $[X, X]^{(n,K_i)}$ are asymptotically unbiased for $\langle X, X\rangle$ (Zhang *et al.* 2005), it is natural to require that

$$\sum_{i=1}^{M} \alpha_i = 1 \quad \text{and} \quad \sum_{i=1}^{M} \alpha_i \frac{n+1-K_i}{K_i} = 0. \tag{10}$$

A slight redefinition will now make the problem more transparent. Let

$$\begin{aligned} a_1 &= \alpha_1 - \left[(n+1)\left(\frac{1}{K_1} - \frac{1}{K_2}\right)\right]^{-1}, \\ a_2 &= \alpha_2 - (a_1 - \alpha_1), \\ a_i &= \alpha_i, \qquad \text{for } i \geqslant 3. \end{aligned} \tag{11}$$

Our conditions on the αs are now equivalent to the following.

Condition 1. $\sum a_i = 1$.

Condition 2. $\sum_{i=1}^{M} (a_i / K_i) = 0$.

To understand the estimator $\widehat{\langle X, X\rangle}^{(n)}$ in terms of the a_i, consider the following asymptotic statement. Here, and everywhere below, we allow a_i, K_i and M to depend on n (i.e. they have the form $a_{n,i}$, $K_{n,i}$ and M_n), though sometimes the dependence on n is suppressed in the notation. We obtain the following proposition, proved in Section 8.

Proposition 1. *Suppose that* $K_{n,1}$ *and* $K_{n,2}$ *are* $O(1)$ *as* $n \to \infty$. *Under Assumptions 1 and 2,*

$$\widehat{\langle X, X\rangle}^{(n)} = \sum_{i=1}^{M} a_i [Y, Y]^{(n,K_i)} - 2\mathrm{E}\epsilon^2 + O_p(n^{-1/2}). \tag{12}$$

To further analyze the terms in (12), write

$$[Y, Y]^{(n,K)} = [X, X]^{(n,K)} + \frac{2}{K} \sum_{i=0}^{n} \epsilon_{t_{n,i}}^2 + U_{n,K} + V_{n,K}, \tag{13}$$

where $U_{n,K}$ is the main error term,

$$U_{n,K} = -\frac{2}{K}\sum_{i=K}^{n} \epsilon_{t_{n,i}} \epsilon_{t_{n,i-K}}, \tag{14}$$

and $V_{n,K}$ will be part of a remainder term and is given by

$$V_{n,K} = 2[X, \epsilon]^{(n,K)} - \frac{1}{K}\sum_{i=0}^{K-1} \epsilon^2_{t_{n,i}} - \frac{1}{K}\sum_{i=n-K+1}^{n} \epsilon^2_{t_{n,i}}.$$

We now can see the impact of Condition 2: to wit, from equation (12),

$$\begin{aligned}
\widehat{\langle X, X\rangle}^{(n)} &= \sum_{i=1}^{M} a_i [X, X]^{(n,K_i)} + \underbrace{2\sum_{i=1}^{M} \frac{a_i}{K_i} \sum_{j=0}^{n} \epsilon^2_{t_{n,j}}}_{=0} + \sum_{i=1}^{M} a_i U_{n,K_i} \\
&\quad + \sum_{i=1}^{M} a_i V_{n,K_i} - 2\mathrm{E}\epsilon^2 + O_p(n^{-1/2}) \\
&= \sum_{i=1}^{M} a_i [X, X]^{(n,K_i)} + \sum_{i=1}^{M} a_i U_{n,K_i} + R_n + O_p(n^{-1/2}),
\end{aligned} \tag{15}$$

where R_n is the overall remainder term, $R_n = \sum_{i=1}^{M} a_i V_{n,K_i} - 2\mathrm{E}\epsilon^2$. Thus, apart from the contribution of this remainder term, Condition 2 removes the bias term due to $\sum \epsilon^2_{n,i}$, not only in expectation but also almost surely. We emphasize this to stress that though we have assumed that the $\epsilon_{t_{n,i}}$ are i.i.d., our estimator is quite robust to the nature of the noise. As before, Condition 1 ensures that the first term in (15) will be asymptotically unbiased for $\langle X, X\rangle$.

Furthermore, for $i \neq l$, the U_{n,K_i} and U_{n,K_l} are uncorrelated. Since U_{n,K_i} and U_{n,K_l} are also the end-points of zero-mean martingales, they are asymptotically independent as $n \to \infty$. Finally, the last term R_n is treated separately in the proof of Theorem 4. For now, we focus on the terms other than the V_{n,K_i}.

If one presupposes Condition 2, and that R_n is comparatively small, it is as if we observe

$$[X, X]^{(K_i)} + U_{n,K_i}, \qquad i = 1, \ldots, M.$$

In the ideal world of continuous observations (i.e. if we take $[X, X]^{(K_i)}$ to stand in for $\langle X, X\rangle$), Condition 2 makes it possible to obtain M (almost) independent measurements of $\langle X, X\rangle$. This motivates the form of the MSRV estimator.

Our aim is to use Conditions 1 and 2 to construct optimal weights a_i. We proceed to investigate what happens if we just take $[X, X]^{(K_i)} \approx \langle X, X\rangle$ in Sections 3 and 4. From Section 5 on, we consider the more exact calculation that follows from $[X, X]^{(K_i)} = \langle X, X\rangle + O_p((n/K_i)^{-1/2})$.

3. Asymptotics for the noise term

As above, to obtain meaningful asymptotics, we let all quantities depend on n, thus $a_i = a_{n,i}$, $M = M_n$, $K_i = K_{n,i}$, $[Y, Y]^{(K)} = [Y, Y]^{(n,K)}$, etc. Sometimes the dependence on n is suppressed in the notation. All results are proved in Section 8.

Consider, first, the noise term

$$\zeta_n = \sum_{i=1}^{M_n} a_{n,i} U_{n,K_{n,i}}. \tag{16}$$

The variance of ζ_n is as follows.

Proposition 2 (Variance of the noise term). *Set*

$$\gamma_n^2 = 4 \sum_{i=1}^{M_n} \left(\frac{a_{n,i}}{K_{n,i}} \right)^2.$$

Suppose that the $\epsilon_{t_{n,i}}$ are i.i.d. with mean zero and $\mathrm{E}\epsilon^2 < \infty$, *and that $M_n = o(n)$ as $n \to \infty$. Then*

$$\mathrm{var}(\zeta_n) = \gamma_n^2 n (\mathrm{E}\epsilon^2)^2 (1 + o(1)). \tag{17}$$

Also, γ_n^2 is minimized, subject to Conditions 1 and 2, by choosing

$$a_{n,i} = \frac{K_{n,i}(K_{n,i} - \overline{K}_n)}{M_n \, \mathrm{var}(K_n)} \tag{18}$$

where

$$\overline{K}_n = \frac{1}{M_n} \sum_{i=1}^{M_n} K_{n,i} \quad \text{and} \quad \mathrm{var}(K_n) = \frac{1}{M_n} \sum_{i=1}^{M_n} K_{n,i}^2 - \left(\frac{1}{M_n} \sum_{i=1}^{M_n} K_{n,i} \right)^2.$$

The resulting minimal value of γ_n is

$$\gamma_n^{*2} = \frac{4}{M_n \, \mathrm{var}(K_n)}. \tag{19}$$

Since the $U_{n,K}$ are end-points of martingales, by the martingale central limit theorem (Hall and Heyde 1980: Chapter 3), we obtain more precisely the following:

Theorem 1. *Suppose that the $\epsilon_{t_{n,i}}$ are i.i.d., with* $\mathrm{E}\epsilon^2 < \infty$, *and that $M = M_n = o(n)$ as $n \to \infty$. Suppose that $\max_{1 \leq i \leq M_n} |a_{n,i}/(i\gamma_n)| \to 0$ as $n \to \infty$. Then $\zeta_n/(n^{1/2}\gamma_n) \to N(0, \mathrm{E}(\epsilon^2)^2)$ in law, both unconditionally and conditionally on $\mathcal{X}$.*

4. A class of estimators, and further asymptotics for the noise term

In this section we develop a class of weights $a_{n,i}$ which we shall use in the rest of the paper. The precise form of the weights is given in Theorem 2. The rest of this section is motivational in nature.

In the following and for the rest of the paper, assume that all scales $i = 1, \ldots, M$ are used, which is to say that $K_{n,i} = i$. In this case, $\overline{K}_n = (M_n + 1)/2$ and $\operatorname{var}(K_n) = (M_n^2 - 1)/12$, and the optimal weights from Proposition 2 are then given by

$$a_{n,i} = 12\frac{i}{M_n^2}\frac{i/M_n - 1/2 - 1/(2M_n)}{1 - 1/M_n^2}. \qquad (20)$$

The minimum variance is given through $y_n^{*2} = 48/[M_n(M_n^2 - 1)]$, so that

$$\operatorname{var}(\zeta_n) = \frac{48n(\mathrm{E}\epsilon^2)^2}{M_n(M_n^2 - 1)}.$$

The form (20) motivates us to consider weights of the form

$$a_{n,i} = \frac{1}{M_n} w_{M_n}\left(\frac{i}{M_n}\right), \qquad i = 1, \ldots, M_n, \qquad (21)$$

as this gives rise to a tractable class of estimators. We specifically take

$$w_M(x) = xh(x) + M^{-1}xh_1(x) + M^{-2}xh_2(x) + M^{-3}xh_3(x) + o(M^{-3}), \qquad (22)$$

where h and h_1 are functions independent of M. The reason for considering this particular functional form, where $w_M(x)$ must suitably vanish at zero, is that Condition 2 translates roughly into a requirement that $\int_0^1 w_M(x)/x\,\mathrm{d}x$ be approximately zero.

In terms of conditions on the function h, Conditions 1 and 2 imply that we have to make the following requirements on h:

Condition 3. $\int_0^1 xh(x)\mathrm{d}x = 1$.

Condition 4. $\int_0^1 h(x)\mathrm{d}x = 0$.

With slightly stronger requirements on h, we can show that (15) holds more generally.

Theorem 2. *Let $h_0 = h$, and suppose that for $i = 0, \ldots, 2$, h_i is $3 - i$ times continuously differentiable on $[0, 1]$, and that h_3 is continuous on $[0, 1]$. Suppose that h satisfies Conditions 3 and 4. Also assume that*

$$\int_0^1 h_1(x)\mathrm{d}x + \frac{1}{2}(h(1) - h(0)) = 0,$$

$$\int_0^1 h_2(x)\mathrm{d}x + \frac{1}{2}(h_1(1) - h_1(0)) + \frac{1}{12}(h'(1) - h'(0)) = 0, \quad (23)$$

$$\int_0^1 h_3(x)\mathrm{d}x + \frac{1}{12}(h_1'(1) - h_1'(0)) = 0.$$

Let the $a_{n,i}$ *be given by* (21) *and* (22)*, where* $o(M^{-3})$ *is uniform in* $x \in [0, 1]$. *Finally, suppose that the* $\epsilon_{t_{n,i}}$ *are i.i.d., with* $\mathrm{E}\epsilon^2 < \infty$. *Then approximation* (15) *remains valid, up to* $o_p(n/M_n^3)$.

The final class of estimators. Our estimation procedure will in the following be using weights $a_{n,i}$ which satisfy the description in Theorem 2.

Remark 1 Comments on Theorem 2. By adding terms in (22), one can make the approximation in (15) as good as one wants, up to $O_p(n^{-1/2})$. We will later use $M_n = O(n^{1/2})$, which is why we have chosen the given number of terms in (22). Also, it should be noted that the approximation to Condition 2 has to be much finer than to Condition 1, since we are seeking to make

$$\sum_{i=1}^{M} \frac{a_i}{K_i} \sum_{i=0}^{n} \epsilon_{t_{n,i}}^2 = n\left(\sum_{i=1}^{M} \frac{a_i}{K_i}\right) \mathrm{E}\epsilon^2 (1 + o_p(1))$$

negligible for asymptotic purposes.

As we shall see, the specific choices for h_1, h_2, and h_3 do not play any role in any of the later expressions for asymptotic variance. A simple choice of h_1 which satisfies (23) is given by $h_1(x) = -h'(x)/2$, with $h_2(x) = h_2$ and $h_3(x) = h_3$, both constants. In this case, $h_2 = -(h'(1) - h'(0))/6$ and $h_3 = (h''(1) - h''(0))/24$. With this choice, one obtains

$$a_{n,i} = \frac{i}{M_n^2} h\left(\frac{i}{M_n}\right) - \frac{1}{2}\frac{i}{M_n^3} h'\left(\frac{i}{M_n}\right) + \frac{i}{M_n^3} h_2 + \frac{i}{M_n^4} h_3. \quad (24)$$

For the noise-optimal weights in (20) at the end of Section 3, h takes the form

$$h_\zeta^*(x) = 12\left(x - \frac{1}{2}\right). \quad (25)$$

Under this choice, the $a_{n,i}$ given by (24) is identical to the one in (20), up to a negligible multiplicative factor of $(1 - M_n^{-2})^{-1}$.

The following corollary to Theorem 1 is now immediate, since

$$\gamma_n^2 = 4M_n^{-3} \int_0^1 h(x)^2 \,\mathrm{d}x (1 + o(1)), \qquad \text{as } n \to \infty.$$

Corollary 1. *Suppose that the* $\epsilon_{t_{n,i}}$ *are i.i.d., with* $\mathrm{E}\epsilon^2 < \infty$*, and that* $M = M_n = o(n)$ *as* $n \to \infty$*. Also assume that the* $a_{n,i}$ *are given by* (21)*, and that the conditions of Theorem 2 are satisfied. Then* $(M_n^3/n)^{1/2}\zeta_n \to N(0, 4\mathrm{E}(\epsilon^2)^2 \int_0^1 h(x)^2\,\mathrm{d}x$ *in law, both unconditionally and conditionally on* $\mathcal{X}$.

5. Asymptotics of the discretization error

We have obtained the optimal weights as far as reducing the noise is concerned. However, as in (15), there remain two types of error: the *discretization error*, due to the fact that the observations only take place at discrete time points; along with the *residual* R_n, which also will turn out to not quite vanish. We study these in turn, and then state a result for the total asymptotics for the MSRV estimator.

For the discretization error, we need some additional concepts.

Definition 1. *Let* $0 = t_{n,0} < t_{n,1} < \ldots < t_{n,n} = T$ *be the observation times when there are n observations. We refer to* $\mathcal{G}_n = \{t_{n,0}, t_{n,1}, \ldots, t_{n,n}\}$ *as a 'grid' or 'partition' of* $[0, T]$*. Following Section 2.6 of Mykland and Zhang (2006), the asymptotic quadratic variation of time (AQVT)* $H(t)$ *is defined by*

$$H(t) = \lim_{n\to\infty} \frac{n}{T} \sum_{t_{n,i+1} \leqslant t} (t_{n,i} - t_{n,i-1})^2, \tag{26}$$

provided the limit exists.

We assume that

$$\max_{1\leqslant i\leqslant n} |t_{n,i+1} - t_{n,i}| = O\left(\frac{1}{n}\right), \tag{27}$$

whence every subsequence has a subsequence so that the AQVT exists. From an applied point of view, there is little loss in assuming the existence of the AQVT; see the argument in Zhang *et al.* (2005: 1411).

Note that from (27), $H(t)$ is Lipschitz continuous provided it exists. We give the following change-of-variable rule for the AQVT:

Lemma 1 *(Change of variables in the AQVT). Assume* (27) *and that the AQVT* $H(t)$ *exists. Let* $G : [0, T] \to [0, T']$ *be increasing and Lipschitz continuous. Set* $u_{n,i} = G(t_{n,i})$*. Then*

$$K(u) = \lim_{n\to\infty} \frac{n}{T'} \sum_{u_{n,i} \leqslant u} (u_{n,i} - u_{n,i-1})^2$$

exists, and

$$TH'(t)G'(t) = T'K'(G(t)) \tag{28}$$

almost everywhere on $[0, T]$.

The following result is also useful and illustrative.

Lemma 2. *Assume the conditions of Lemma 1. Then* $K(T') = T'$ *if and only if*

$$\sum_{i=0}^{n} \left(u_{n,i} - u_{n,i-1} - \frac{T'}{n} \right)^2 = o(n^{-1}). \tag{29}$$

Remark 2. The importance of these two lemmas is that one can compare irregular and 'almost equidistant' sampling. If $H'(t)$ exists, is continuous, and is bounded below by a constant $c > 0$, one can define

$$G(t) = \frac{T'}{T} \int_0^t H'(s)^{-1} \, \mathrm{d}s.$$

Suppose that $G(T) = T'$, and consider the process $\tilde{X}_u = X^{(-1)}_{G(u)}$. This process satisfies the same regularity conditions as those that we impose on X, and, furthermore, the sampling times $u_{n,i} = G(t_{n,i})$ are close to equidistant in the sense of (29). A further implication of this is discussed later in Remark 3.

Define η as the non-negative square root of

$$\eta^2 = T \int_0^T H'(t) \sigma_t^4 \, \mathrm{d}t. \tag{30}$$

Note that η is invariant under the transformation in Lemma 1. Finally, we define 'stable convergence'.

Definition 2. *If* Z_n *is a sequence of* $\mathcal{X}$*-measurable random variables, then* Z_n *converges stably in law to* Z *as* $n \to \infty$ *if there is an extension of* $\mathcal{X}$ *such that for all* $A \in \mathcal{X}$ *and for all bounded continuous* g*,* $\mathrm{E} I_A g(Z_n) \to \mathrm{E} I_A g(Z)$ *as* $n \to \infty$.

For further discussion of stable convergence, see Rényi (1963), Aldous and Eagleson (1978), Hall and Heyde (1980: 56), Rootzén (1980) and Jacod and Protter (1998: 169–170). It is a useful device in operationalizing asymptotic conditionality. There is some choice in what one takes as the σ-field $\mathcal{X}$ in this definition.

We can now state the main theorem for the asymptotic behaviour of finitely many of the $[X, X]^{(K)} = [X, X]^{(n,K)}$.

Theorem 3 *(Central limit theorem for the discretization error in* $[X, X]^{(K)}$*). Suppose that the structure of* X *follows Assumption 1. Also suppose that the observation times* $t_{n,i}$ *are non-random and satisfy* (27)*, and that the AQVT* $H(t)$ *exists and is continuously differentiable. Assume that* $\min_{0 \leq t \leq T} H'(t) > 0$*. Let* $M_n \to \infty$ *as* $n \to \infty$*, with* $M_n = o(n)$*. Let* $(K_{n,1}, \ldots, K_{n,L})/M_n \to (\kappa_1, \ldots, \kappa_L)$ *as* $n \to \infty$*. Let* Γ *be an* $L \times L$ *matrix with* (I, J)*th entry given by*

$$\Gamma_{I,J} = \frac{2}{3}\min(\kappa_I, \kappa_J)\left(3 - \frac{\min(\kappa_I, \kappa_J)}{\max(\kappa_I, \kappa_J)}\right), \tag{31}$$

and let Z be a normal random vector with covariance matrix Γ. Let Z be independent of $\mathcal{X}$. Then, as $n \to \infty$ the vector

$$\left(\frac{n}{M_n}\right)^{1/2}([X, X]^{(n,K_{n,1})} - \langle X, X\rangle, \ldots, [X, X]^{(n,K_{n,L})} - \langle X, X\rangle)$$

converges stably in law to ηZ.

Remark 3. Even in the scalar ($L = 1$) case, the result in Theorem 3 is a gain over our earlier Theorem 3 in Zhang *et al.* (2005: 1401). To characterize the asymptotic distribution we use an AQVT which is independent of choice of scale and coincides with the original object introduced in Mykland and Zhang (2006: Section 2.6). This is unlike the time variation measure used in Zhang *et al.* (2005: Section 3.4), and Theorem 3 provides a substantial simplification of the asymptotic expressions. To do this, we have used the approach described above in Remark 2.

It is conjectured that the regularity conditions for Theorem 3 can be reduced to those of Proposition 1 of Mykland and Zhang (2006), but investigating this is beyond the scope of this paper.

As a corollary to Theorem 3, we now finally obtain the asymptotics for the discretization part of the MSRV as follows.

Corollary 2 *(Central limit theorem for the discretization error in the MSRV). Let $a_{n,i}$ satisfy* (21) *and* (22)*, and let the conditions of Theorem 2 be satisfied. Further, make Assumption 1. Also suppose that the observation times $t_{n,i}$ are non-random and satisfy* (27)*, and that the AQVT $H(t)$ exists and is continuously differentiable. Assume that* $\min_{0\leq t\leq T} H'(t) > 0$. *Let $M_n \to \infty$ as $n \to \infty$, with $M_n/n = o(1)$ and $M_n^3/n \to \infty$. Set*

$$\eta_h^2 = \frac{4}{3}T\eta^2\int_0^1 \mathrm{d}x\int_0^x h(y)h(x)y^2(3x - y)\mathrm{d}y. \tag{32}$$

Then

$$\left(\frac{n}{M_n}\right)^{1/2}\left(\sum_{i=1}^{M_n} a_{n,i}[X, X]^{(n,i)} - \langle X, X\rangle\right) \to \eta_h Z \tag{33}$$

stably in law, where Z is standard normal and independent of $\mathcal{X}$.

Remark 4. Note that the condition $M_n^3/n \to \infty$ is present because we have not imposed too many conditions on h; if it were necessary, the assumption could be removed by considering a slightly smaller class of hs.

6. Overall asymptotics for the MSRV estimator

There are two main sources of error in the MSRV. On the one hand, we saw in Corollary 1 that if M_n time scales are used, the part of $\widehat{\langle X, X\rangle}^{(n)} - \langle X, X\rangle$ which is due purely to the noise ϵ can be reduced to have order $O_p(n^{1/2}M_n^{-3/2})$. At the same time, Corollary 2 shows that the pure discretization error is of order $O_p(n^{-1/2}M_n^{1/2})$. To balance these two terms, the optimal M_n is therefore of order

$$M_n = O(n^{1/2}), \tag{34}$$

assuming that the remainder term in (15) does not cause problems, which is indeed the case. This leads to a variance–variance trade-off, and the rate of convergence for the MSRV estimator is then $\widehat{\langle X, X\rangle}^{(n)} - \langle X, X\rangle = O_p(n^{-1/4})$. This result is an improvement on the two-scales estimator, for which the corresponding rate is $O_p(n^{-1/6})$. We embody this in the following result.

Theorem 4. *Let $a_{n,i}$ satisfy* (21) *and* (22)*, and let the conditions of Theorem 2 be satisfied. Further, make Assumptions 1 and 2. Also suppose that the observation times $t_{n,i}$ are non-random and satisfy* (27)*, and that the AQVT $H(t)$ exists and is continuously differentiable. Assume that* $\min_{0\leqslant t\leqslant T} H'(t) > 0$. *Suppose that $M_n/n^{1/2} \to c$ as $n\to\infty$. Let Z be a standard normal random variable independent of $\mathcal{X}$. Set*

$$\nu_h^2 = 4c^{-3}(\mathrm{E}\epsilon^2)^2\int_0^1 h(x)^2\,\mathrm{d}x + c\frac{4}{3}T\eta^2\int_0^1 \mathrm{d}x\int_0^x h(y)h(x)y^2(3x-y)\mathrm{d}y$$
$$+ 4c^{-1}\operatorname{var}(\epsilon^2)\int_0^1\int_0^y xh(x)h(y)\mathrm{d}x\,\mathrm{d}y + 8c^{-1}\mathrm{E}\epsilon^2\int_0^1\int_0^1 h(x)h(y)\min(x, y)\mathrm{d}x\,\mathrm{d}y\langle X, X\rangle. \tag{35}$$

Then

$$n^{1/4}\left(\widehat{\langle X, X\rangle}^{(n)} - \langle X, X\rangle\right) \to \nu_h Z, \tag{36}$$

stably in law, as $n\to\infty$.

For the noise optimal h-function from equation (25) (cf. equation (20)), we can now calculate the value of the asymptotic variance of the MSRV. Note that if $h(x) = 12(x - 1/2)$, we obtain

$$\int_0^1 \mathrm{d}x\int_0^x h(y)h(x)y^2(3x-y)\mathrm{d}y = \frac{39}{35},$$

$$\int_0^1\int_0^y xh(x)h(y)\mathrm{d}x\,\mathrm{d}y = \frac{3}{5},$$

$$\int_0^1\int_0^1 h(x)h(y)\min(x, y)\mathrm{d}x\,\mathrm{d}y = \frac{6}{5}.$$

Hence, in this case, the asymptotic variance becomes

$$\nu_h^2 = 48c^{-3}(\mathrm{E}\epsilon^2)^2 + \frac{52}{35}cT\eta^2 + \frac{12}{5}c^{-1}\mathrm{var}(\epsilon^2) + \frac{48}{5}c^{-1}\mathrm{E}\epsilon^2\langle X, X\rangle. \tag{37}$$

7. Conclusion

In this paper we have introduced the multi-scale realized volatility and shown a central limit theorem (Theorem 4) for this estimator. This permits the setting of intervals for the true integrated volatility on the basis of the MSRV. As a consequence of our result, it is clear that the MSRV is rate efficient, with a rate of convergence of $O_p(n^{-1/4})$.

In terms of the general study of realized volatilities, Section 5 also shows further properties of the asymptotic quadratic variation of time, as earlier introduced by Zhang (2001) and Mykland and Zhang (2006). In particular, Theorem 3 shows that one can use the regular one-step AQVT also for multi-step realized volatilities, thus improving on Theorems 2 and 3 in Zhang *et al.* (2005: 1401).

Finally, note that most of the arguments we have used also hold when the noise process $\epsilon_{t_{n,i}}$ is no longer i.i.d. One can, for example, model this process as being stationary (but with mean zero). If the process is sufficiently mixing, this will change the asymptotic variance of the MSRV, but not the consistency, nor the convergence rate of $O_p(n^{-1/4})$; see, for example, Chapter 5 of Hall and Heyde (1980) for the basic limit theory for dependent sums. However, we have not sought to develop the specific conditions for the central limit theorem to hold in the case where the process is mixing.

8. Proofs of results

Note that, for ease of notation, we sometimes suppress the dependence on n. For example, $a_i = a_{n,i}$, $M = M_n$, $K_i = K_{n,i}$, $[Y, Y]^{(K)} = [Y, Y]^{(n,K)}$, etc. Also, in this section we write t_i for $t_{n,i}$, to avoid notational clutter.

Proof of Proposition 1. Write

$$\begin{aligned}\widehat{\langle X, X\rangle}^{(n)} &= \sum_{i=1}^{M} a_i[Y, Y]^{(n,K_i)} + (\alpha_1 - a_1)([Y, Y]^{(n,K_1)} - [Y, Y]^{(n,K_2)}) \\ &= \sum_{i=1}^{M} a_i[Y, Y]^{(n,K_i)} - 2\mathrm{E}\epsilon^2 + O_p(n^{-1/2}),\end{aligned} \tag{38}$$

where the final approximation follows from Lemma 1 in Zhang *et al.* (2005: 1398).

Proof of Proposition 2. Since $U_{n,K_{n,i}}$ and $U_{n,K_{n,l}}$ are uncorrelated ($i \neq l$) zero-mean martingales,

$$
\begin{aligned}
\operatorname{var}(\zeta_n) &= \sum_{i=1}^{M_n} a_{n,i}^2 \operatorname{var}(U_{n,K_{n,i}}) \\
&= 4\sum_{i=1}^{M_n}\left(\frac{a_{n,i}}{K_{n,i}}\right)^2 (n - K_{n,i} + 1)(\mathrm{E}\epsilon^2)^2 \\
&= \gamma^2 n(\mathrm{E}\epsilon^2)^2(1 + o(1)), \qquad (39)
\end{aligned}
$$

showing equation (17). The last transition in (39) follows because $M_n = o(n)$.

We minimize γ_n^2, subject to the constraints in Conditions 1 and 2. This is established by setting

$$
\frac{\partial}{\partial a_{n,i}}\left[\gamma_n^2 + \lambda_1\left(\sum a_{n,i} - 1\right) + \lambda_2\left(\sum \frac{a_{n,i}}{K_{n.i}}\right)\right] = 8\frac{a_{n,i}}{K_{n,i}^2} + \lambda_1 + \frac{\lambda_2}{K_{n,i}}
$$

to zero, resulting in $a_{n,i} = -\frac{1}{8}(\lambda_1 K_{n,i}^2 + \lambda_2 K_{n,i})$. One can determine the λs by solving

$$
\begin{aligned}
1 &= \sum_{i=1}^{M_n} a_{n,i} = -\frac{1}{8}\left(\lambda_1 \sum_{i=1}^{M_n} K_{n,i}^2 + \lambda_2 \sum_{i=1}^{M_n} K_{n,i}\right), \\
0 &= \sum_{i=1}^{M_n} \frac{a_i}{K_{n,i}} = -\frac{1}{8}\left(\lambda_1 \sum_{i=1}^{M_n} K_{n,i} + M_n\lambda_2\right).
\end{aligned}
$$

This leads to

$$
\lambda_1 = -\frac{8}{M_n \operatorname{var}(K_n)} \quad \text{and} \quad \lambda_2 = \frac{8\overline{K}_n}{M_n \operatorname{var}(K_n)},
$$

where $\overline{K}_n$ and $\operatorname{var}(K_n)$ are as given in rhe statement of the proposition. This shows the rest of the proposition. □

Proof of Theorem 1. Assume without loss of generality that $K_i = i$ for $i = 1, \ldots, M$. To avoid notational clutter, we write a_i for $a_{n,i}$. Note that ζ_n is the end-point of a martingale. We show that $\zeta_n/(n^{1/2}\gamma_n)$ satisfies the conditions of the version of the martingale central limit theorem which is stated in Corollary 3.1 of Hall and Heyde (1980: 58–59). The result then follows. Note that we shall take, in the notation of Hall and Heyde (1980), $\mathcal{F}_{n,j}$ to be the smallest σ-field making ϵ_{t_i}, $i = 1, \ldots, j$, and the whole X_t process, measurable.

We start with the Lindeberg condition. For given δ, define $f_\delta(x) = \mathrm{E}(\epsilon^2 x^2 I_{\{|\epsilon x| > \delta\}})$. Also set

$$
r_n(x) = \mathrm{E} f_{\delta n^{1/2}}\left(-\frac{1}{\gamma_n}\sum_{i=1}^{M_n \wedge j} \frac{2a_i}{i}\epsilon_{t_i}\right), \qquad \text{for } \frac{j-1}{n} \leqslant x < \frac{j}{n}.
$$

We then obtain

$$\sum_{j=1}^{n} \mathrm{E}\left(\epsilon_{t_j}^2 \left(-\frac{1}{n^{1/2}\gamma_n}\sum_{i=1}^{M_n\wedge j}\frac{2a_i}{i}\epsilon_{t_{j-i}}\right)^2 I_{\{|\epsilon_{t_j}(-(1/n^{1/2}\gamma_n)\sum_{i=1}^{M_n\wedge j}(2a_i/i)\epsilon_{t_{j-i}})|>\delta\}}\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \mathrm{E} f_{\delta n^{1/2}}\left(-\frac{1}{\gamma_n}\sum_{i=1}^{M_n\wedge j}\frac{2a_i}{i}\epsilon_{t_{j-i}}\right)$$

$$= \int_0^1 r_n(x)\mathrm{d}x \qquad \text{(since the } \epsilon_{t_i} \text{ are i.i.d.)}$$

$$\to 0 \text{ as } n \to \infty, \tag{40}$$

where the last transition is explained in the next paragraph. By Chebychev's inequality, the conditional Lindeberg condition in Corollary 3.1 of Hall and Heyde (1980) is thus satisfied.

The last transition in (40) is because of the following. First, fix $x \in [0, 1)$, and let j_n be the corresponding j in the definition of $r_n(x)$. Let

$$Z_n = -\frac{1}{\gamma_n}\sum_{i=1}^{M_n\wedge j_n}\frac{2a_i}{i}\epsilon_{t_i},$$

so that $r_n(x) = \mathrm{E} f_{\delta n^{1/2}}(Z_n)$. Note that Z_n is a sum of independent random variables which satisfies the Lindeberg condition:

$$\sum_{i=1}^{M_n\wedge j_n} \mathrm{E}\left(\frac{-2a_i}{i\gamma_n}\epsilon_{t_i}\right)^2 I_{\{|-(2a_i/i\gamma_n)\epsilon_{t_i}|>\delta\}} = \sum_{i=1}^{M_n\wedge j_n} f_\delta\left(\frac{-2a_i}{i\gamma_n}\right) \to 0$$

as $n \to \infty$, since $\max_i |a_i/i\gamma_n| \to 0$. The ensuing asymptotic normality of Z_n (if necessary by going to subsequences of subsequences) shows that $r_n(x) \to 0$ as $n \to \infty$. Since $0 \leqslant r_n(x) \leqslant 1$, the final transition in (40) follows by dominated convergence.

We now turn to the sum of conditional variances in the corollary in Hall and Heyde (1980):

$$\sum_{j=1}^{n} \mathrm{E}\left(\epsilon_{t_j}^2 \left(-\frac{1}{n^{1/2}\gamma_n}\sum_{i=1}^{M_n\wedge j}\frac{2a_i}{i}\epsilon_{t_{j-i}}\right)^2 |\mathcal{F}_{n,j-1}\right)$$

$$= \mathrm{E}(\epsilon^2)\frac{1}{n\gamma_n^2}\sum_{j=1}^{n}\left(\sum_{i=1}^{M_n\wedge j}\frac{2a_i}{i}\epsilon_{t_{j-i}}\right)^2$$

$$= 1 + o_p(1). \tag{41}$$

The last transition is obvious by appealing to M-dependence. A rigorous but tedious proof is obtained by splitting the sum into main terms of the type $\epsilon_{t_i}^2$ and cross-terms of the form $\epsilon_{t_i}\epsilon_{t_j}$ $(i \neq j)$.

In view of (40) and (41), Theorem 1 is proved by using Corollary 3.1 and the remarks following it in Hall and Heyde (1980: 58–59). □

Proof of Theorem 2. We need to show that

$$\sum_{i=1}^{M} \frac{a_{n,i}}{K_{n,i}} \sum_{i=0}^{n} \epsilon_{t_i}^2 = o_p(n/M_n^3),$$

in other words, we need

$$\sum_{i=1}^{M-n} \frac{a_{n,i}}{K_{n,i}} = o(M_n^{-3}).$$

By Taylor expansion,

$$\begin{aligned}
\frac{1}{M}\sum_{i=1}^{M} h\left(\frac{i}{M}\right) &= \int_0^1 h(x)\mathrm{d}x + \frac{1}{2M^2}\sum_{i=1}^{M} h'\left(\frac{i}{M}\right) - \frac{1}{3!M^3}\sum_{i=1}^{M} h''\left(\frac{i}{M}\right) \\
&\quad + \frac{1}{4!M^4}\sum_{i=1}^{M} h'''\left(\frac{i}{M}\right) + o(M^{-3}) \\
&= \int_0^1 h(x)\mathrm{d}x + \frac{1}{2M}(h(1)-h(0)) + \frac{1}{12M^3}\sum_{i=1}^{M} h''\left(\frac{i}{M}\right) \\
&\quad - \frac{1}{24M^4}\sum_{i=1}^{M} h'''\left(\frac{i}{M}\right) + o(M^{-3}) \\
&= \int_0^1 h(x)\mathrm{d}x + \frac{1}{2M}(h(1)-h(0)) + \frac{1}{12M^2}(h'(1)-h'(0)) + o(M^{-3}),
\end{aligned} \tag{42}$$

where the last line follows by iterating the first line. By similar argument on h_1, h_2 and h_3,

$$
\begin{aligned}
\frac{1}{M}\sum_{i=1}^{M}\left(\frac{i}{M}\right)^{-1} w_M\left(\frac{i}{M}\right) &= \frac{1}{M}\sum_{i=1}^{M} h\left(\frac{i}{M}\right) + \frac{1}{M^2}\sum_{i=1}^{M} h_1\left(\frac{i}{M}\right) \\
&\quad + \frac{1}{M^3}\sum_{i=1}^{M} h_2\left(\frac{i}{M}\right) + \frac{1}{M^4}\sum_{i=1}^{M} h_3\left(\frac{i}{M}\right) + o(M^{-3}) \\
&= \int_0^1 h(x)\mathrm{d}x + \frac{1}{M}\left(\int_0^1 h_1(x)\mathrm{d}x + \frac{1}{2}(h(1)-h(0))\right) \\
&\quad + \frac{1}{M^2}\left(\int_0^1 h_2(x)\mathrm{d}x + \frac{1}{2}(h_1(1)-h_1(0)) + \frac{1}{12}(h'(1)-h'(0))\right) \\
&\quad + \frac{1}{M^3}\left(\int_0^1 h_3(x)\mathrm{d}x + \frac{1}{12}(h_1'(1)-h_1'(0))\right) + o\left(\frac{1}{M^3}\right) \\
&= o\left(\frac{1}{M^3}\right),
\end{aligned}
$$

by (23). This shows the result. □

Proof of Lemma 1. To get the rigorous statement, we proceed as follows. Every subsequence has a further subsequence for which $K(u)$ exists, and this K is obviously Lipschitz continuous. We will show that (28) holds. Since this equation is independent of subsequence, the result will have been proved.

Let B_t be a standard Brownian motion, and let $\tilde{B}_t = B_{G(t)}$. By comparing the asymptotic distributions of $n^{1/2}[\sum_{t_i \leqslant t}(\tilde{B}_{t_i} - \tilde{B}_{t_{i-1}})^2 - \langle \tilde{B}, \tilde{B}\rangle_t]$ and $n^{1/2}[\sum_{u_i \leqslant u}(B_{u_i} - B_{u_{i-1}})^2 - \langle B, B\rangle_u]$, we obtain from Proposition 1 of Mykland and Zhang (2006) that

$$
T\int_0^t 2H'(s)(\langle \tilde{B}, \tilde{B}\rangle_s')^2\,\mathrm{d}s = T'\int_0^{G(t)} 2K'(v)(\langle B, B\rangle_v')^2\,\mathrm{d}v, \quad \text{for all } t \in [0.T].
$$

Since $\langle B, B\rangle_v' = 1$ and $\langle \tilde{B}, \tilde{B}\rangle_s' = G'(s)$ almost everywhere, equation (28), and hence the lemma, follows. □

Proof of Lemma 2. Set $\delta_{n,i} = u_{n,i} - u_{n,i-1} - T'/n$. Then

$$
\begin{aligned}
\frac{n}{T'}\sum_i (u_{n,i} - u_{n,i-1})^2 &= \frac{n}{T'}\sum_i\left(\frac{T'}{n} + \delta_{n,i}\right)^2 \\
&= T' + 2\sum_i \delta_{n,i} + \frac{T'}{n}\sum_i \delta_{n,i}^2.
\end{aligned}
$$

Since $\sum_i \delta_{n,i} = 0$, the lemma follows by letting $n \to \infty$. □

Proof of Theorem 3. Following Lemmas 1 and 2 and Remark 2, we can assume without loss of generality that the $t_{n,i}$ satisfy (in place of the $u_{n,i}$) the equation (29).

Consider the scalar case ($L = 1$) first, with $K_n = K_{n,1} = M_n$. In what follows, all prelimiting quantities are subscripted by n, and we suppress the n for ease of notation (except when it seems necessary). We now refer to Theorems 2 and 3 in Zhang *et al.* (2005: 1401). Use the notation Δt_i, h_i and η_n as in that paper, and let $\overline{\Delta t} = T/n$. (Note that the usage of η in this paper is different from that of Zhang *et al.* (2005). Also define

$$\tilde{h}_i = \frac{4}{K\overline{\Delta t}} \sum_{j=1}^{(K-1)\wedge i} \left(1 - \frac{j}{K}\right)^2 \overline{\Delta t} \quad \text{and} \quad \tilde{\eta}_n^2 = \sum_i \tilde{h}_i \sigma_{t_i}^4 \overline{\Delta t}.$$

Note that if we show that $\tilde{\eta}_n - \eta_n \to 0$ in probability as $n \to \infty$, we have shown the scalar version of the theorem. This is because we will then have shown that the conditions of the two theorems in Zhang *et al.* (2005) are satisfied, and that we can calculate the asymptotic variances as if $t_{n,i} = iT/n$.

To this end, note first that

$$\left|\sum_i h_i \sigma_{t_i}^4 (\Delta t_i - \overline{\Delta t})\right| \leqslant (\sigma^+)^4 \left(\sum_i h_i^2\right)^{1/2} \left(\sum_i (\Delta t_i - \overline{\Delta t})^2\right)^{1/2}$$
$$= O(n^{1/2}) \times o(n^{-1/2}) = o(1), \tag{43}$$

where the orders follow, respectively, from equation (45) in Zhang *et al.* (2005), and equation (29) in this paper. Then note that

$$\left|\sum_i (h_i - \tilde{h}_i)\sigma_{t_i}^4 \overline{\Delta t}\right| = \left|\frac{4}{K}(\sigma^+)^4 \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right)^2 \left(\sum_{l=(K-1)^+}^{n-j} (\Delta t_l - \overline{\Delta t})\right)\right|$$
$$\leqslant \frac{4}{K}(\sigma^+)^4 \sum_{j=1}^{K-1} \left(1 - \frac{j}{K}\right)^2 \times \left(\sum_i (\Delta t_i - \overline{\Delta t})^2\right)^{1/2}$$
$$= O(1) \times o(n^{-1/2}) = o(1) \tag{44}$$

where, again, the orders follow, respectively, from equation (45) in Zhang *et al.* (2005) and equation (29) in this paper.

Equations (43) and (44) combine to show that $\tilde{\eta}_n - \eta_n \to 0$ in probability as $n \to \infty$.

For the general ($L > 1$) case, first note that since μ_t, σ_t and σ_t^{-1} are bounded (Assumption 1), by Girsanov's theorem – see, for example, Section 3.5 of Karatzas and Shreve (1991: 190–201) or Section II-3b of Jacod and Shiryaev (2003: 168–170) – we can without loss of generality further suppose that $\mu_t = 0$ identically. This is because of the stability of the convergence; see the methodology in Rootzén (1980).

Now set

$$(X, X)^{(K)} = \frac{2}{K}\sum_{j=0}^{n-1}(X_{t_{j+1}} - X_{t_j}) \sum_{r=1}^{j\wedge(K-1)} (K - r)(X_{t_{j-r+1}} - X_{t_{j-r}})$$

and note that

$$[X, X]^{(n,K)} = (X, X)^{(K)} + [X, X]^{(n,1)} + O_p\left(\frac{K}{n}\right)$$

$$= (X, X)_T^{(K)} + \langle X, X\rangle + O_p(n^{-1/2}) + O_p(K/n),$$

from Proposition 1 in Mykland and Zhang (2006).

Let $M_t^{n,I}$ be the continuous martingale for which $M_T^{n,I} = (X, X)^{(I)}(n/M_n)^{1/2}$. The proof of Theorem 2 in Zhang *et al.* (2005) actually establishes that the sequence of processes $(M_t^{n,K_{n,I}})$ is C-tight in the sense of Definition VI.3.25 of Jacod and Shriyaev (2003: 351). This is because of Theorem VI.4.13 and Corollary VI.6.30, also in Jacod and Shriyaev (2003: 358, 385). The same corollary then establishes that the asymptotic distribution is as described in Theorem 3, provided we can show that, in probability,

$$\langle M^{n,K_{n,I}}, M^{n,K_{n,J}}\rangle_T \to \eta^2\Gamma, \qquad \text{as } n \to \infty. \tag{45}$$

This is because of Lévy's theorem (see Theorem 3.16 in Karatzas and Shreve 1991: 157). The stable convergence follows as in the proof of Theorem 3 of Zhang *et al.* (2005), the conditions for which have already been satisfied.

Finally, we need to show (45). As in the scalar case, we assume (29), and the same kind argument used in the scalar case carries over to show that we can take $t_{i,n} = iT/n$ for the purposes of our calculation. The computation is then tedious but straightforward, and carried out similarly to that for the quadratic variation in the proof of Theorem 2 in Zhang *et al.* (2005). Theorem 3 is thus proved. □

Proof of Corollary 2. First of all, note that since $M_n^3/n \to \infty$, $\sum_{i=1}^{M_n} a_{n,i} = o(-(n/M_n)^{1/2})$. In lieu of equation (33), it is therefore enough to prove

$$\left(\frac{n}{M_n}\right)^{1/2}\sum_{i=1}^{M_n} a_{n,i}\big([X, X]^{(n,i)} - \langle X, X\rangle\big) \to \eta_h Z. \tag{46}$$

Also, as in the proof of Theorem 3, our assumptions imply that we can take $\mu_t = 0$ identically, without loss of generality.

Since there are asymptotically infinitely many $[X, X]^{(n,i)}$ involved in equation (33), we have to approximate with a finite number of these. To this end, let $\delta > 0$ be an arbitrary number ($\delta < 1$). Let $\alpha = 1 - \delta/\sqrt{2}$. Let L be an integer sufficiently large that $2\alpha^{L-1} \leqslant \delta^2$. For $I = 1, \ldots, L$, let $\tilde{\kappa}_I = \alpha^{L-I}$, and let $\tilde{\kappa}_0 = 0$. For $i = 1, \ldots, M_n$, define $I_{i,n}$ to be the value I, $1 \leqslant I \leqslant L$, for which $i/M_n \in (\tilde{\kappa}_{I-1}, \tilde{\kappa}_I]$. Then note that, if $\|U\| = (\mathrm{E}U^2)^{1/2}$,

$$\left(\frac{n}{M_n}\right)^{1/2}\left\|\sum_{i=1}^{M_n} a_{n,i}\big([X,X]^{(n,i)}-[X,X]^{(n,I_{i,n})}\big)\right\|$$

$$\leqslant \left(\frac{n}{M_n}\right)^{1/2}\sum_{i=1}^{M_n}|a_{n,i}|\times \max_{1\leqslant i\leqslant n}\|[X,X]^{(n,i)}-[X,X]^{(n,I_{i,n})}\|. \tag{47}$$

Now let i_n be the value i, $1\leqslant i\leqslant M_n$, which maximizes $\|[X,X]^{(n,i)}-[X,X]^{(n,I_{i,n})}\|$ for given n, and let $I_n = I_{i_n,n}$.

For the moment, let N be an unbounded set of positive integers so that $(i_n/M_n, I_n/M_n)_{n\in N}$ converges. Call the limit (κ_1,κ_2). By the proofs of Theorems 2 and 3 in Zhang *et al.* (2005), $(n/M_n)([X,X]^{(n,i_n)}-[X,X]^{(n,I_n)})^2$ is uniformly integrable. By the statement of Theorem 3, it then follows that, as $n\to\infty$ through N,

$$\begin{aligned}\left(\frac{n}{M_n}\right)\mathrm{E}([X,X]^{(n,i_n)}-[X,X]^{(n,I_n)})^2 &\to \mathrm{E}\eta^2(\Gamma_{2,2}+\Gamma_{1,1}-2\Gamma_{1,2})\\ &= \mathrm{E}\eta^2 2\kappa_2\left(1-\frac{\kappa_1}{\kappa_2}\right)^2\\ &\leqslant \mathrm{E}\eta^2\delta^2\end{aligned} \tag{48}$$

by construction. Since every subsequence has a subsequence for which $(i_n/M_n, I_n/M_n)$ converges, it follows from (47) that

$$\limsup_{n\to\infty}\left(\frac{n}{M_n}\right)^{1/2}\left\|\sum_{i=1}^{M_n} a_{n,i}\big([X,X]^{(n,i)}-[X,X]^{(n,I_{i,n})}\big)\right\| \leqslant \delta(\mathrm{E}\eta^2)^{1/2}\max_{0\leqslant x\leqslant 1}|xh(x)|. \tag{49}$$

The result of Corollary 2 thus follows by computing the limit of

$$\left(\frac{n}{M_n}\right)^{1/2}\sum_{i=1}^{M_n} a_{n,i}\big([X,X]^{(n,I_{i,n})}-\langle X,X\rangle\big), \tag{50}$$

and then letting $\delta\to 0$. □

Proof of Theorem 4. The remainder term R_n from equation (15) can be written $R_n = R_{n,1}+R_{n,2}$, where

$$\begin{aligned} R_{n,1} &= \sum_{j=1}^{M_n} a_{n,j}\frac{1}{j}\left(\sum_{i=0}^{j-1}\epsilon_{t_i}^2+\sum_{i=n-j+1}^{n}\epsilon_{t_i}^2\right)-2\mathrm{E}\epsilon^2,\\ R_{n,2} &= 2\sum_{i=1}^{M_n} a_{n,i}[X,\epsilon]^{(i)}. \end{aligned} \tag{51}$$

We shall show that $M_n^{1/2}R_n$ converges in law, conditionally on $\mathcal{X}$, to a normal distribution with variance

$$4\mathrm{var}(\epsilon^2)\int_0^1\int_0^y xh(x)h(y)\mathrm{d}x\,\mathrm{d}y + 8\langle X, X\rangle \mathrm{var}(\epsilon)\int_0^1\int_0^1 h(x)h(y)\min(x, y)\mathrm{d}x\,\mathrm{d}y, \tag{52}$$

and also that, conditionally on $\mathcal{X}$, $R_n/M_n^{1/2}$ is asymptotically independent of $(M_n^3/n)^{1/2}\zeta_n$ in Corollary 1. Thus, in view of the results on the the pure noise and discretization terms in Corollaries 1 and 2, Theorem 4 will then be shown.

To show this, we show in the following that $M_n^{1/2}R_{n,1}$ and $M_n^{1/2}R_{n,2}$ are asymptotically normal given $\mathcal{X}$, with mean zero and variances given by (54) and (57), respectively. We then discuss the *joint* distribution of $(M_n^3/n)^{1/2}\zeta_n$, $M_n^{1/2}R_{n,1}$ and $M_n^{1/2}R_{n,2}$.

Asymptotic normality of $R_{n,1}$. Once $M_n < n/2$, write

$$R_{n,1} = \sum_{i=0}^{M_n-1}\epsilon_{t_i}^2\sum_{j=i+1}^{M_n}\frac{a_{n,j}}{j} + \sum_{i=0}^{M_n-1}\epsilon_{t_{n-i}}^2\sum_{j=i+1}^{M_n}\frac{a_{n,j}}{j} - 2\mathrm{E}\epsilon^2. \tag{53}$$

Hence,

$$\begin{aligned}\mathrm{var}(M_n^{1/2}R_{n,1}) &= 2M_n\,\mathrm{var}(\epsilon^2)\sum_{i=0}^{M-1}\left(\sum_{j=i+1}^{M}\frac{a_j}{j}\right)^2\\ &= 2\,\mathrm{var}(\epsilon^2)\int_0^1\left(\int_x^1 h(y)\mathrm{d}y\right)^2\mathrm{d}x + o(1)\\ &= 4\,\mathrm{var}(\epsilon^2)\int_0^1\int_0^y xh(x)h(y)\mathrm{d}x\,\mathrm{d}y + o(1),\end{aligned} \tag{54}$$

while under Theorem 2,

$$\mathrm{E}\left[\sum_{j=1}^{M_n}a_j\frac{1}{j}\left(\sum_{i=0}^{j-1}\epsilon_{t_i}^2 + \sum_{i=n-j+1}^{n}\epsilon_{t_i}^2\right)\right] = 2\mathrm{E}\epsilon^2(1 + o(M_n^{-1/2})). \tag{55}$$

Since the Lindeberg condition is also obviously satisfied, we obtain that $M_n^{1/2}R_{n,1}$ converges in law (conditionally on $\mathcal{X}$) to a normal distribution with mean zero and variance given by (54).

Asymptotic normality of the 'cross term' $R_{n,2}$. As in the proof of Theorem 3, we proceed, without loss of generality, as if X were a martingale. As in the proof of Theorem 1, we shall show that $M_n^{1/2}R_{n,2}$ satisfies the conditions of the version of the martingale central limit theorem which is stated in Corollary 3.1 of Hall and Heyde (1980), and calculate the asymptotic variance. As in the earlier proof, we shall take, in the notation of Hall and Heyde (1980), $\mathcal{F}_{n,j}$ to be the smallest σ-field making ϵ_{t_i}, $i = 1, \ldots, j$, and the whole X_t process, measurable.

Note that, from (6),

$$[X, \epsilon]^{(n,K)} = \frac{1}{K}\sum_{i=0}^{n} b_{n,i}^{(K)}\epsilon_{t_i},$$

where

$$b_{n,i}^{(K)} = \begin{cases} -(X_{t_{n,i+K}} - X_{n,t_i}), & \text{if } i = 0, \ldots, K-1, \\ (X_{t_{n,i}} - X_{t_{n,i-K}}) - (X_{t_{n,i+K}} - X_{t_{n,i}}), & \text{if } i = K, \ldots, n-K, \\ (X_{t_{n,i}} - X_{t_{n,i-K}}) & \text{if } i = n-K+1, \ldots, n. \end{cases}$$

Thus, from (51), one obtains

$$M_n^{1/2} R_{n,2} = M_n^{1/2} \sum_{i=1}^{n} \epsilon_{t_i} \sum_{j=1}^{M_n} \frac{a_{n,j}}{j} b_{n,i}^{(j)}. \tag{56}$$

Obviously, $M_n^{1/2} R_{n,2}$ is the end-point of a zero-mean martingale relative to the filtration $(\mathcal{F}_{n,j})$. The conditional variance process (in Corollary 3.1 in Hall and Heyde 1980) is given by

$$\begin{aligned} M_n \mathrm{E}(\epsilon^2) \sum_{i=1}^{n} \left(\sum_{j=1}^{M_n} \frac{a_{n,j}}{j} b_{n,i}^{(j)} \right)^2 &= M_n \operatorname{var}(\epsilon) \sum_{i=1}^{n} \sum_{j=1}^{M_n} \sum_{k=1}^{M_n} \frac{a_{n,j}}{j} \frac{a_{n,k}}{k} b_{n,i}^{(j)} b_{n,i}^{(k)} \\ &= M_n \operatorname{var}(\epsilon) \sum_{i=1}^{n} \sum_{j=1}^{M_n} \sum_{k=1}^{M_n} \frac{a_{n,j}}{j} \frac{a_{n,k}}{k} (b_{n,i}^{(j \wedge k)})^2 + o_p(1) \\ &= 2 M_n \operatorname{var}(\epsilon) \sum_{j=1}^{M_n} \sum_{k=1}^{M_n} \frac{a_{n,j}}{j} \frac{a_{n,k}}{k} (j \wedge k) [X, X]^{(j \wedge k)} + o_p(1) \\ &= 2 \int_0^1 \int_0^1 h(x) h(y) (x \wedge y) \mathrm{d}x \, \mathrm{d}y \langle X, X \rangle \operatorname{var}(\epsilon) + o_p(1), \end{aligned} \tag{57}$$

where $j \wedge k = \min(j, k)$ and where remainder terms are taken care of as in the proof of Theorem 3.

By similar methods, the Lindeberg condition is satisfied (see the discussion in the proof of Theorem 1). By Corollary 3.1 in Hall and Heyde (1980) it follows that $M_n^{1/2} R_n$ is asymptotically normal (conditionally on $\mathcal{X}$), with mean zero and variance given by (57). This is what we needed to show.

The joint distribution of $(M_n^3/n)^{1/2}\zeta_n$, $M_n^{1/2} R_{n,1}$ *and* $M_n^{1/2} R_{n,2}$. First of all, note that for all three quantities, we have satisfied the conditions of Corollary 3.1 of Hall and Heyde (1980). This is with the exception of (their equation) (3.21), where we have instead used the remarks following their corollary (and thus the convergence is conditional on $\mathcal{X}$ as opposed to stable with respect to the σ-field generated by both $\mathcal{X}$ and the ϵ_{t_i}).

In terms of joint distribution, note first that the sum of conditional covariances for each pair of the three quantities $(M_n^3/n)^{1/2}\zeta_n$, $M_n^{1/2} R_{n,1}$ and $M_n^{1/2} R_{n,2}$ converges to zero, by the same methods as above. In view of how Hall and Heyde's corollary implies their Theorem 3.2 (Hall and Heide 1980: 58), the Cramér–Wold device now implies the joint normality of $(M_n^3/n)^{1/2}\zeta_n$, $M_n^{1/2} R_{n,1}$ and $M_n^{1/2} R_{n,2}$, and also that they are asymptotically independent. Theorem 4 is then proved. □

Acknowledgements

I gratefully acknowledge the support of the National Science Foundation under grant DMS-0204639. I would like to thank the referees and the editor for suggestions which greatly improved the paper.

References

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

Aït-Sahalia, Y., Mykland, P.A. and Zhang, L. (2005b) Ultra high frequency volatility estimation with dependent microstructure noise. Technical report, Princeton University.

Aldous, D.J. and Eagleson, G.K. (1978) On mixing and stability of limit theorems. *Ann. Probab.*, **6**, 325–331.

Bandi, F.M., Russell, J.R. and Zhu, Y. (2006) Using high-frequency data in dynamic portfolio choice. *Econometric Rev.* To appear.

Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2006) Regular and modified kernel-based estimators of integrated variance: the case with independent noise. Technical report, Department of Mathematical Sciences, University of Aarhus.

Bibby, B.M., Jacobsen, M. and Sørensen, M. (2002) Estimating functions for discretely sampled diffusion-type models. In Y. Aït-Sahalia and L.P. Hansen (eds), *Handbook of Financial Econometrics*. Amsterdam: North-Holland.

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

Corsi, F., Zumbach, G., Müller, U., and Dacorogna, M. (2001) Consistent high-precision volatility from high-frequency data. *Economic Notes*, **30**, 183–204.

Curci, G. and Corsi, F. (2005) Discrete sine transform for multi-scales realized volatility measures. Technical report, University of Lugano.

Gloter, A. and Jacod, J. (2000a) Diffusions with measurement errors: I – Local asymptotic normality. Preprint PMA-623, Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VI.

Gloter, A. and Jacod, J. (2000b) Diffusions with measurement errors: II – Optimal estimators. Preprint PMA-624, Laboratoire de Probabilités et Modèles Aléatoires, Université de Paris VI.

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

Hansen, P.R. and Lunde, A. (2006) Realized variance and market microstructure noise. *J. Business Economic Statist.*

Jacod, J. and Protter, P. (1998) Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, **26**, 267–307.

Jacod, J. and Shiryaev, A.N. (2003) *Limit Theorems for Stochastic Processes*, 2nd edn. 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. (2006) ANOVA for diffusions and Ito processes. *Ann. Statist., forthcoming*, **34**(4): 000–000.

Rényi, A. (1963) On stable sequences of events. *Sankyā Ser. A*, **25**, 293–302.

Rootzén, H. (1980) Limit distributions for the error in approximations of stochastic integrals. *Ann. Probab.*, **8**, 241–251.

Stein, M. (1987) Minimum norm quadratic estimation of spatial variograms. *J. Amer. Statist. Assoc.*, **82**, 765–772.

Stein, M.L. (1990) A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process. *Ann. Statist.*, **18**, 1139–1157.

Stein, M.L. (1993) Spline smoothing with an estimated order parameter. *Ann. Statist.*, **21**, 1522–1544.

Ying, Z. (1991) Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Multivariate Anal.*, **36**, 280–296.

Ying, Z. (1993) Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.*, **21**, 1567–1590.

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

Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005) A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100**, 1394–1411.

Zhou, B. (1996) High-frequency data and volatility in foreign-exchange rates. *J. Business Econom. Statist.*, **14**, 45–52.

Received December 2004 and revised May 2006