Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach *

Lan Zhang

First version: August 15, 2004.
This version: December 6, 2004

Abstract

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 contain error or noise, for example, in the form of microstructure noise. The former (consistency) has been addressed heavily in the recent literature, however, the resulting estimator is not quite efficient. In Zhang, Mykland, and Aït-Sahalia (2003), the best estimator converges to the true volatility only at the rate of $n^{-1/6}$. In this paper, we propose an efficient estimator which converges to the true at the rate of $n^{-1/4}$, which is the best attainable. The estimator remains valid when the observation noise is dependent.

Some key words and phrases: consistency, dependent noise, discrete observation, efficiency, Ito process, microstructure noise, observation error, rate of convergence, realized volatility

*Lan Zhang is Assistant Professor, Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: lzhang@stat.cmu.edu. I gratefully acknowledge the support of the National Science Foundation under grant DMS-0204639.

1 Introduction

This paper is about 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 former (consistency) has been addressed heavily in the recent literature, however, the resulting estimator is not quite efficient. In Zhang, Mykland, and Aït-Sahalia (2003), the best estimator converges to the true volatility only at the rate of $n^{-1/6}$. In this paper, we propose an efficient estimator which converges to the true quantity at the rate of $n^{-1/4}$, which is the best attainable. The new estimator remains valid when the observation noise is dependent.

To fix the idea, consider $\{Y\}$ as the observed log returns, and the observations take place at the grid of time points $\mathcal{G}_n = \{t_i, i = 0, 1, 2, \cdots n\}$ that span the time interval $[0, T]$.

Suppose that $\{Y_{t_i}\}$ are noisy, the true (latent) log returns are $\{X\}$. In other words, X is the de-noised version of Y, their relation can be modeled as,

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}. \tag{1}$$

where $t_i \in \mathcal{G}_n$. The noise ϵ_{t_i}'s are independent of X. And the noise process itself can be a white noise or a dependent process. Also, if one is more familiar with the terminology of the price process $\{P_t\}$ of an asset, the true log returns $\{X_t\}$ is just a log transformation of true price process $\{P_t\}$, i.e. $X_t = \log(P_t)$.

The model in (1) is quite realistic, as evidenced by the existence of microstructure noise in the price process (early papers include Brown (1990), Zhou (1996), Corsi, Zumbach, Muller, and Dacorogna (2001)).

Suppose that the true log returns $\{X\}$ satisfies the following equation:

$$dX_t = \mu_t dt + \sigma_t dB_t \tag{2}$$

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 of

$$dX_t(\omega) = \mu(t, \omega)dt + \sigma(t, \omega)dB_t(\omega) \tag{3}$$

Through out this paper, we use the notations in (2) to denote (3). By the model in (3), we mean that $\{X\}$ follows an Itô process. A special case would be that $\{X\}$ is Markov, where $\mu_t = \mu(t, X_t)$, and $\sigma_t = \sigma(t, X_t)$. In financial literature, σ_t is often refered as the volatility of X.

Our goal is to estimate $\int_0^T \sigma_t^2 dt$, where T can be a day, a month, or other time horizon(s). For simplicity, we call $\int_0^T \sigma_t^2 dt$ the integrated volatility, and denote it by

$$< X, X >_T \triangleq \int_0^T \sigma_t^2 dt.$$

The general question is, how to nonparametrically estimate $\int_0^T \sigma_t^2 dt$, if one can only observe the noisy data Y_{t_i} at discrete times $t_i \in \mathcal{G}_n$.

To our best knowledge, there are two types of nonparametric estimators for $\int_0^T \sigma_t^2 dt$ in the current literature.

The first type, the simpler one, is to use the sum of the squared returns

$$[Y, Y]_T^{(all)} \triangleq \sum_{t_i \in \mathcal{G}_n} (Y_{t_i} - Y_{t_{i-1}})^2, \tag{4}$$

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, Zumbach, Muller, and Dacorogna (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 both of order n (sample size).

A slight modification of (4) would be to use the sum square of the "sparsely selected" returns, namely

$$[Y, Y]_T^{(sparse)} \triangleq \sum_{s_i \in \mathcal{G}_n^{\mathcal{H}}} (Y_{s_i} - Y_{s_{i-1}})^2, \tag{5}$$

where $\mathcal{G}_n^{\mathcal{H}}$ is a strict subset of $\mathcal{G}_n$, for example, if one starts with observation # 2 and then picks every subsequent 10'th data point, $\mathcal{G}_n^{\mathcal{H}} = \{t_2, t_{12}, t_{22}, \cdots\} \subset [0, T]$, that is, in (5), $s_0 = t_2, s_1 = t_{12}, s_2 = t_{22}, \cdots$. The idea is that by using sparse data, one reduces the bias and the variance of the conventional realized volatility. This approach has been quite popular in the empirical finance literature. However, this "sparse" estimator is still not consistent in nature, and which data to subsample and which to discard is arbitrary.

A second type of estimator for $\int_0^T \sigma_t^2 dt$ is based on *two sampling scales*. For example, the estimator in Zhang, Mykland, and Aït-Sahalia (2003) has the form of

$$< \widehat{X, X} >_T^{(2)} = [Y, Y]_T^{(K)} - 2\frac{\bar{n}}{n}[Y, Y]_T^{(all)}, \tag{6}$$

where

$$[Y, Y]_T^{(K)} = \frac{1}{K} \sum_{t_i \in [0,T]} (Y_{t_i} - Y_{t_{i-K}})^2,$$

with K being a positive integer. One can see that

$$[Y,Y]_T^{(all)} = [Y,Y]_T^{(1)}.$$

Thus the estimator in (6) averages the square increment in returns from sampling every data point ($[Y,Y]_T^{(1)}$) and the one from sampling every K-th data point ($[Y,Y]_T^{(K)}$). Particularly, $<\widehat{X,X}>_T^{(2)}$ is unbiased for any sample size n, and its asymptotic property was derived when $K \to \infty$ as $n \to \infty$.

Also, the estimator in Zhou (1996) and Hansen and Lunde (2004) has the form of

$$\frac{1}{k}\sum_{i=1}^{n}\Big((Y_{t_i}-Y_{t_{i-k}})^2+(Y_{t_i}-Y_{t_{i-k}})(Y_{t_{i-k}}-Y_{t_{i-2k}})+(Y_{t_{i+k}}-Y_{t_i})(Y_{t_i}-Y_{t_{i-k}})\Big)\approx 2[Y,Y]_T{}^{(2k)}-[Y,Y]_T^{(k)}$$

which is also on basis of two time scales.

The estimators based on two different time scales is unbiased and consistent, and asymptotically normal. However, the rate of convergence is not satisfactory. For an instance, the best estimator in Zhang, Mykland, and Aït-Sahalia (2003) converges to $\int_0^T \sigma_t^2 dt$ at the rate of $n^{-1/6}$.

In this paper, we propose a new class of estimators, which converges to $\int_0^T \sigma_t^2 dt$ at the rate of $n^{-1/4}$. This new estimator has the form,

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

where M is a positive integer greater than 2. Comparing to $<\widehat{X,X}>_T^{(2)}$ which uses two time scales (1 and K), $\widehat{\langle X,X\rangle}$ combines M different time scales. The weights a_i are selected so that $\widehat{\langle X,X\rangle}$ is 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 etimator is the same as the one in parametric estimation for volatility, when the true process is Markov (see Gloter and Jacod (2001)), 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, Mykland, and Zhang (2003). For the estimating functions-based approach, there is a nice review by Bibby, Jacobsen, and Sørensen (2002).

We emphasize that our estimator is nonparametric, and the true process follows a more general Ito process, where the volatility could depend on the entire history of the X process plus additional randomness. Our proposed estimator remains valid even if the noise follows a dependent process.

The paper is organized as following. In section 2, we motivate the idea of averaing over M different time scales. As we shall see, our estimator is unbiased, and its asymptotic variance comes from the noise (ζ) as well as from the discreteness of the sampling times (Υ). In Section 3, we derive the weights a_i's which are optimal for minimizing the variance that comes from noise. We then elaborate on the discretization error in Section 4 and 5. Section 6 introduces more general weights, and Section 7 deals with the overall error under these weights, Further discussions on optimal weights and optimal variance are in Section 8-9. In Section 10, we comment on the case where the noise is dependent.

2 Motivation: The Averaging of Uncorrelated Observations of $\langle X, X\rangle$

In Zhang, Mykland, and Aït-Sahalia (2003), we have observed that by combining the square increments of the returns from two time scales, the resulting two-scale estimator $<\widehat{X,X}>_T^{(2)}$ in (6) improves upon the realized volatility, which uses only one time scale, as in (4)-(5). 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 would be how about the estimator combining more than 2 time scales. This question motivates the present paper. In this section we briefly go through the main argument.

To proceed, denote the estimator on the K-th scale to be

$$[Y,Y]_T^{(K)} = \frac{1}{K}\sum_{t_i\in[0,T]}(Y_{t_i}-Y_{t_{i-K}})^2, \tag{7}$$

with K being a positive integer. From now on, we work under model (1) with ϵ being white noise process. The case of dependent noise is discussed at the end (Section 10).

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

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

We consider estimators on the form

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

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

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

Since $[X,X]^{(K_i)}$ are asymptotically unbiased for $< X,X >$ (Zhang, Mykland, and Aït-Sahalia (2003)), it is natural to require that

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

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

$$a_1 = \alpha_1 - \left[(n+1)\left(\frac{1}{K_1} - \frac{1}{K_2}\right)\right]^{-1}, \quad a_2 = \alpha_2 - (a_1 - \alpha_1) \text{ and } a_i = \alpha_i \text{ for } i \geq 3 \tag{11}$$

Our conditions on the αs are now equivalent to

Condition 1. $\sum a_i = 1$,

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

The estimator becomes

$$\begin{aligned}\widehat{\langle X, X\rangle} &= \sum_{i=1}^{M} a_i [Y,Y]^{(K_i)} + (\alpha_1 - a_1)([Y,Y]^{(K_1)} - [Y,Y]^{(K_2)}) \\ &= \sum_{i=1}^{M} a_i [Y,Y]^{(K_i)} + 2E\epsilon^2 + O_p(n^{-1/2}) \end{aligned} \tag{12}$$

where the final approximation follows from Zhang, Mykland, and Aït-Sahalia (2003).

To see the first terms in (12), write

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

where

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

and the remainder term is given by

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

Equation (12) then becomes,

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

Thus, apart from the contribution of the remainder term, Condition 2 removes the bias term due to $\sum \epsilon_i^2$, not only in expectation, but almost surely. As before, Condition 1 assures 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 $\sum_{i=1}^M a_i V_{n,K_i} - 2E\epsilon^2$ is treated separately in Lemma 1 (see the Appendix). For now, we focus on the terms other than the V_{n,K_i}'s.

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

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

Under the ideal world of continuous observations (that is, if we take $[X, X]^{(K_i)}$ to stand in for $\langle X, X\rangle$), Condition 2 makes it possible that we get M (almost) independent measurements of $\langle X, X\rangle$.

Our aim is to use Conditions 1-2 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 Section 3. From Section 4 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 Minimizing the Size of the Noise Term

Consider the noise term

$$\zeta = \sum_{i=1}^{M} a_i U_{n,K_i} \tag{16}$$

Since U_{n,K_i} and U_{n,K_l} are uncorrelated zero-mean martingales, under Conditions 1-2,

$$\begin{aligned} Var(\zeta) &= \sum_{i=1}^{M} a_i^2 Var(U_{n,K_i}|X) \\ &= 4\sum_{i=1}^{M} (\frac{a_i}{K_i})^2 (n - K_i + 1)(E\epsilon^2)^2 \\ &\approx \gamma^2 n (E\epsilon^2)^2, \end{aligned} \tag{17}$$

for $K_i << n$, where $\gamma^2 = 4\sum_{i=1}^M (\frac{a_i}{K_i})^2$.

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

$$\frac{\partial}{\partial a_i}[\gamma^2 + \lambda_1(\sum a_i - 1) + \lambda_2(\sum \frac{a_i}{K_i})] = 8\frac{a_i}{K_i^2} + \lambda_1 + \frac{\lambda_2}{K_i}$$

to zero, resulting in $a_i = -\frac{1}{8}(\lambda_1 K_i^2 + \lambda_2 K_i)$.

One can determine λ's by solving

$$(4) \qquad \begin{cases} 1 = \sum_{i=1}^M a_i = -\frac{1}{8}(\lambda_1 \sum_{i=1}^M K_i^2 + \lambda_2 \sum_{i=1}^M K_i) \\ 0 = \sum_{i=1}^M \frac{a_i}{K_i} = -\frac{1}{8}(\lambda_1 \sum_{i=1}^M K_i + M\lambda_2) \end{cases}$$

It leads to

$$\lambda_1 = -\frac{8}{MVar(K)} \quad \text{and} \quad \lambda_2 = \frac{8\bar{K}}{MVar(K)},$$

where $\bar{K} = \frac{1}{M}\sum_{i=1}^{M} K_i$ and $Var(K) = \frac{1}{M}\sum_{i=1}^{M} K_i^2 - (\frac{1}{M}\sum_{i=1}^{M} K_i)^2$.

The optimal a_i is thus given by

$$a_i = \frac{K_i(K_i - \bar{K})}{MVar(K)} \tag{18}$$

And γ^2 is minimized at

$$\gamma^{*2} = \frac{4}{MVar(K)}.$$

In a special case where $K_i = i$, $i = 1, \cdots, M$, $\bar{K} = (M+1)/2$ and $Var(K) = (M^2-1)/12$, and the minimum variance $\gamma^{*2} = \frac{48}{M(M^2-1)}$.

Overall, therefore, in the case where $K_i = i$,

$$Var(\zeta) = \frac{48n}{M(M^2-1)}(E\epsilon^2)^2 \tag{19}$$

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* $E\epsilon^4 < \infty$*, and that* $M = M_n = o(n)$ *as* $n \to \infty$*. Let the* a_i *be given optimally as above for* $1 \le i \le M$*. Then* $Var(\zeta_n)^{-1/2}\zeta_n \to N(0,1)$ *in law.* ∎

Note that when all $i = 1, ..., M$ are used, and for $K_i = i$,

$$a_i \quad = \quad 12\frac{i}{M^2}\frac{\left(\frac{i}{M} - \frac{1}{2} - \frac{1}{2M}\right)}{\left(1 - \frac{1}{M^2}\right)} \tag{20}$$

We now have obtained the optimal weights as far as reducing the noise is concerned. However, as in (15), there is another type of error, the error due to the fact that the observations only take place at discrete time points. We study the discretization error $\sum_{i=1}^{M} a_i[X,X]^{(i)} - \langle X, X\rangle$ in the next two sections.

4 Tradeoff with The Discretization

Set

$$\widetilde{\langle X, X\rangle} = \sum_{i=1}^{M} a_i[X,X]^{(i)} \tag{21}$$

Unlike the noise components $U_{n,K}$, the $[X,X]^{(i)}$ are asymptotically highly correlated. Unless a_i goes to zero fast, $\widetilde{\langle X,X\rangle} - \langle X,X\rangle$ has the same order as the worst possibility, $[X,X]^{(M)} - \langle X,X\rangle$, which is $O_p((n/M)^{-1/2})$ by Zhang, Mykland, and Aït-Sahalia (2003). Since ζ is independent of X, under Conditions 1-2, the overall error is

$$\begin{aligned} \widehat{\langle X,X\rangle} - \langle X,X\rangle &= \zeta + \widetilde{\langle X,X\rangle} - \langle X,X\rangle + O_p(M^{-1/2}) \\ &= O_p((n/M^3)^{1/2}) + O_p((n/M)^{-1/2}) + O_p(M^{-1/2}), \end{aligned} \tag{22}$$

where the last term $O_p(M^{-1/2})$ follows from Lemma 1 in the Appendix.

The optimal M is therefore of the form

$$M = O(n^{1/2}). \tag{23}$$

By the variance-variance tradeoff, the rate of convergence for our optimal estimator is then

$$\widehat{\langle X,X\rangle} - \langle X,X\rangle = O_p(n^{-1/4}). \tag{24}$$

This is an improvement on the two scales estimator, for which the corresponding rate is $O_p(n^{-1/6})$.

We spend the following sections elaborating on this result.

5 Form of the Discretization Error

We first need assumptions on our latent process. Suppose that X is an Itô process of the form (2), with drift coefficient μ_t and diffusion coefficient σ_t, both continuous almost surely. Also suppose that $|\mu_t|$ and σ_t are bounded above by a constant, and that σ_t is bounded away from zero. Assume that the sampling points are nonrandom, and that

$$\max_i \ |t_{i+1} - t_i| \ = O\left(\frac{1}{n}\right). \tag{25}$$

Note that in view of Girsanov's Theorem (see, for example, p. 190-201 of Karatzas and Shreve (1991)), under these assumptions, we can proceed as if $\mu = 0$.

To deal with the discretization error, first note that

$$[X,X]^{(K)} = (X,X)^{(K)} + [X,X]^{(1)} + O_p(K/n)$$

where

$$(X,X)^{(K)} = \frac{2}{K}\sum_{j=1}^{n-1}(X_{t_{j+1}} - X_{t_j}) \sum_{r=1}^{j\wedge(K-1)} (K-r)(X_{t_{j-r+1}} - X_{t_{j-r}}) \tag{26}$$

Thus, from Proposition 1 in Mykland and Zhang (2002),

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

By the same methods, $(X,X)^{(J)}$ and $(X,X)^{(K)}$ are joint asymptotically mixed normal, with random covariance

$$\Upsilon_{J,K} = \frac{T}{n}\left(\min(J,K)-1\right)\frac{2}{3}\left(3-\frac{\min(J,K)+1}{\max(J,K)}\right)\eta^2 \tag{27}$$

where

$$\eta^2 = \int_0^T H'(t)\sigma_t^4 dt \tag{28}$$

and $H(t)$ is the asymptotic quadratic variation of time, which is the same as $H^{(2)}(t)$ in Mykland and Zhang (2002). Note that σ_t is allowed to be random, so is η^2. The convergence is in the sense of stable convergence; for discussions of how to present limit statements formally, please refer to Mykland and Zhang (2002) and Zhang, Mykland, and Aït-Sahalia (2003).

It is easily seen from this that for the weights a_i's discussed in (20) of Section 3, the discretization error $\widetilde{\langle X,X\rangle} - \langle X,X\rangle$ is indeed of the order $O_p((n/M)^{-1/2})$ given in Section 4. We now turn to a more general class of estimators.

6 A Class of Estimators

We here develop a tractable class of weights a_i. The final form is given at the end of this section.

As a point of departure, consider estimators of the form

$$a_i = \frac{1}{M}w_M(\frac{i}{M}) = \frac{1}{M}g(\frac{i}{M}) + \frac{1}{M^2}g_1(\frac{i}{M}) + O(\frac{1}{M^3}), i = 1,...,M, \tag{29}$$

for continuous g, g_1, with g continuously differentiable. We emphasize that while $M = M_n$, g, g_1 are assumed to be independent of n. This approximately covers the noise-optimal weights in (20) at the end of Section 3, where in that case g takes the form

$$g^*_\zeta(x) = 12x\left(x-\frac{1}{2}\right). \tag{30}$$

We use the subscript "ζ" to refer to the fact that this g is only shown to be optimal for the noise.

Conditions that parallell Conditions 1-2 can be imposed on g as follows. It seems natural to require that

Condition 3. $\int_0^1 g(x)dx = 1$,

Condition 4. $\int_0^1 \frac{g(x)}{x}dx = 0$.

Since by Taylor expansion

$$\begin{aligned}\frac{1}{M}\sum_{i=1}^{M}(\frac{i}{M})^{-1}w_M(\frac{i}{M}) &= \int_0^1 x^{-1}w_M(x)dx + \frac{1}{2}\frac{1}{M}(w_M(1) - \lim_{x\to 0} x^{-1}w_M(x)) + O(\frac{1}{M^2}) \\ &= \int_0^1 x^{-1}g(x)dx + \frac{1}{M}\int_0^1 x^{-1}g_1(x)dx + \frac{1}{2}\frac{1}{M}(g(1) - \lim_{x\to 0} x^{-1}g(x)) + O(\frac{1}{M^2}),\end{aligned} \quad (31)$$

to reconcile conditions 2 and 4, we require

$$\int_0^1 x^{-1}g_1(x)dx + \frac{1}{2}(g(1) - \lim_{x\to 0} x^{-1}g(x)) = 0 \quad (32)$$

An inspection of the order of the pure noise term shows that this requirement is necessary to achieve the cancellation in equation (15) to the order required. Higher order terms are not necessary, and conditions 1 and 3 do not have to be further reconciled. Also, g_1 does not play any role in any of the expressions for asymptotic variance.

A straightforward way of implementing the above is to assume that

$$g(x) = xh(x). \quad (33)$$

The conditions 3-4 become

Condition 5. $\int_0^1 xh(x)dx = 1$,

Condition 6. $\int_0^1 h(x)dx = 0$.

A simple choice of g_1 which satisfied (32) is given by $g_1(x) = -xh'(x)/2$, so that finally one can take

$$a_i = \frac{i}{M^2}h(\frac{i}{M}) - \frac{1}{2}\frac{1}{M^2}\frac{i}{M}h'(\frac{i}{M}) \quad (34)$$

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

$$h^*_\zeta(x) = 12\left(x - \frac{1}{2}\right). \quad (35)$$

For this choice, the a_i given by (34) is identical to the one in (20), up to a multiplicative factor of $(1 - M^{-2})^{-1}$, which is negligible.

The final class of estimators. Our estimation procedure will in the following be based on equation (34), where g and h are linked by (33), where h is continuously differentiable, and satisfies conditions 5-6.

7 Joint Noise and Discretization Asymptotics

The variance of the noise is given through (17) and

$$\gamma^2 \approx 4M^{-3} \int_0^1 \frac{g(x)^2}{x^2} dx \tag{36}$$

In view of Section 5, the similar expression for the discretization variance is

$$\begin{aligned} \sum_{J,K} a_J a_K \Upsilon_{J,K} &= \sum_{K=1}^{M} a_K^2 \Upsilon_{K,K} + 2 \sum_{K=1}^{M} \sum_{J=1}^{K-1} a_J a_K \Upsilon_{J,K} \\ &\approx \frac{4}{3} T \frac{M}{n} \eta^2 \int_0^1 dx \int_0^x g(y) g(x) y \left(3 - \frac{y}{x}\right) dy \end{aligned} \tag{37}$$

The contribution from the remainder term $V_{n,K}$ (see Lemma 1 in the Appendix) is,

$$\frac{4}{M} Var(\epsilon^2) \int_0^1 \int_0^y g(x) \frac{g(y)}{y} dx dy + \frac{8}{M} Var(\epsilon) \int_0^1 \int_0^1 g(x) g(y) \frac{\min(x,y)}{xy} dx dy < X, X > \tag{38}$$

The overall asymptotic variance of $\widehat{\langle X, X \rangle} - \langle X, X \rangle$ is, therefore,

Theorem 2.

$$\begin{aligned} V &= \frac{n}{M^3} 4 (E\epsilon^2)^2 \int_0^1 \frac{g(x)^2}{x^2} dx + \frac{4}{3} T \frac{M}{n} \eta^2 \int_0^1 dx \int_0^x g(y) g(x) y \left(3 - \frac{y}{x}\right) dy \\ &+ \frac{4}{M} Var(\epsilon^2) \int_0^1 \int_0^y g(x) \frac{g(y)}{y} dx dy + \frac{8}{M} Var(\epsilon) \int_0^1 \int_0^1 g(x) g(y) \frac{\min(x,y)}{xy} dx dy < X, X > . \end{aligned}$$

Further, $\widehat{\langle X, X \rangle} - \langle X, X \rangle$ *is asymptotic mixed normal, with mean zero and the above variance.* ∎

Note that the mixed normality follows from the same methods as in Zhang, Mykland, and Aït-Sahalia (2003).

It is clear from the above that the optimal choice of M is of the order $O(n^{1/2})$, and that $V = O(n^{-1/2})$ with this choice. Specifically, if

$$M \approx c n^{1/2}, \tag{39}$$

then

$$V = n^{-1/2} v(g), \tag{40}$$

where

$$\begin{aligned} v(g) &= c^{-3} 4 (E\epsilon^2)^2 \int_0^1 \frac{g(x)^2}{x^2} dx + c \frac{4}{3} T \eta^2 \int_0^1 dx \int_0^x g(y) g(x) y \left(3 - \frac{y}{x}\right) dy \\ &+ 4 c^{-1} Var(\epsilon^2) \int_0^1 \int_0^y g(x) \frac{g(y)}{y} dx dy + 8 c^{-1} E\epsilon^2 \int_0^1 \int_0^1 g(x) g(y) \frac{\min(x,y)}{xy} dx dy < X, X > \end{aligned} \tag{41}$$

8 Overall Variance for the Weights from Section 3.

To calculate the value of the asymptotic variance, note that if $h(x) = 6(2x-1)$, we obtain

$$\int_0^1 dx \int_0^x h(y)h(x)y^2\,(3x-y)\,dy = \frac{39}{35},$$
$$\int_0^1 \int_0^y xh(x)h(y)dxdy = \frac{3}{5},$$
$$\int_0^1 \int_0^1 h(x)h(y)\min(x,y)dxdy = \frac{6}{5}.$$

Hence the asymptotic variance becomes

$$v(h) = 48c^{-3}(E\epsilon^2)^2 + \frac{52}{35}cT\eta^2 + \frac{12}{5}c^{-1}Var(\epsilon^2) + \frac{48}{5}c^{-1}E\epsilon^2 < X, X > \tag{42}$$

9 Optimal Weights

We here give the equations that the overall optimal choice of g must satisfy. Again let $g(x) = xh(x)$. We obtain $v(g) = \nu[h]$, where

$$\begin{aligned}\nu[h] &= 4c^{-3}(E\epsilon^2)^2 \int_0^1 h(x)^2dx + c\frac{4}{3}T\eta^2 \int_0^1 dx \int_0^x h(y)h(x)y^2\,(3x-y)\,dy \\ &+ 4c^{-1}Var(\epsilon^2)\int_0^1 \int_0^y xh(x)h(y)dxdy + 8c^{-1}E\epsilon^2 \int_0^1\int_0^1 h(x)h(y)\min(x,y)dxdy < X, X > \end{aligned}\tag{43}$$

To optimize, let

$$I_p[h] = \int_0^1 h(x)x^p dx \tag{44}$$

and

$$\begin{aligned}A[h,r] &= c^{-3}(E\epsilon^2)^2 8\int_0^1 h(x)r(x)dx + c\frac{4}{3}T\eta^2 \int_0^1 dx \int_0^x h(y)r(x)[2]y^2\,(3x-y)\,dy \\ &+ 4c^{-1}Var(\epsilon^2)\int_0^1\int_0^y h(y)r(x)[2]xdxdy + 8c^{-1}E\epsilon^2 \int_0^1\int_0^1 h(y)r(x)[2](\min(x,y))dxdy < X, X > \end{aligned}$$

where $h(y)r(x)[2] = h(y)r(x) + h(x)r(y)$. Note that if $r(x) = x^p$,

$$\begin{aligned}A[h,r] &= c^{-3}(E\epsilon^2)^2 8I_p[h] + c\frac{4}{3}T\eta^2 \times \\ &\quad \left(\frac{6}{(p+4)(p+3)(p+2)(p+1)}I_{p+4}[h] + \frac{3}{p+2}I_2[h] - \frac{1}{p+1}I_3[h]\right) \\ &\quad + 4c^{-1}[Var(\epsilon^2) + 4E\epsilon^2 < X, X >](\frac{1}{p+1} - \frac{I_{p+2}[h]}{(p+1)(p+2)}).\end{aligned}\tag{45}$$

A standard optimization argument yields that if h minimizes $\nu[h]$ subject to the constraints 5-6, then $A[h,r]=0$ for all r that satisfy the same contraints.

Now let $r_i(x)$, $i=0,1,2,...$ be shifted Legendre polynomials, which obey the orthogonal relationship

$$\int_0^1 r_i(x)r_j(x)dx = \frac{1}{2j+1}\delta_{ij},$$

where δ_{ij} is the Kronecker delta (see Abramowitz and Stegun (1972)). In particular, this is to say that r_i is a polynomial of order exactly i, $\int_0^1 r_i(x)r_j(x)dx = 0$ for $i \neq j$, and the first few are $r_0(x)=1$, $r_1(x)=2x-1$, $r_2(x)=6x^2-6x+1$, $r_3(x)=20x^3-30x^2+12x-1$. Our condition for optimality becomes $A[h,r_i]=0$ for $i=1,2,....$

10 Dependent noise

The above argument is based on the assumption that the ϵ_i are independent. However, if the noise is m-dependent, and one does not use $[Y,Y]^{(K)}$ for $K=1,...,m$, the noise does not affect our results. In particular, if one redefines $a_i=0$ for $i \leq m$, and by (34) for $i>m$, the asymptotic expressions are the same. m can even become large at a slow rate as $n \to \infty$ without changing the asymptotic values.

11 Appendix: Effect of the Remainder Term $V_{n,K}$

Lemma 1.

$$M^{1/2}\sum_i^M a_i V_{n,K_i} - 2E\epsilon^2 \xrightarrow{\mathcal{L}}$$

$$[4Var(\epsilon^2)\int_0^1\int_0^y g(x)\frac{g(y)}{y}dxdy + \frac{8}{M}Var(\epsilon)\int_0^1\int_0^1 g(x)g(y)\frac{\min(x,y)}{xy}dxdy < X,X >]^{1/2} N(0,1),$$

where the convergence is stable in law. ∎

Fist consider the part which is due to $\frac{1}{K}(\sum_{i=0}^{K-1}\epsilon_{t_i}^2+\sum_{i=n-K+1}^{n}\epsilon_{t_i}^2)$. Take $K_i=i$, for $M<n/2$,

$$\begin{aligned} Var\left[\sum_{j=1}^{M} a_j \frac{1}{K_j}(\sum_{i=0}^{K_j-1}\epsilon_{t_i}^2+\sum_{i=n-K+1}^{n}\epsilon_{t_i}^2)\right] &= 2Var(\sum_{j=1}^{M}\frac{a_j}{K_j}\sum_{i=0}^{K_j-1}\epsilon_{t_i}^2) \\ &= 2Var(\sum_{i=0}^{M-1}\epsilon_{t_i}^2\sum_{j=K_i+1}^{M}\frac{a_j}{K_j}) \\ &= 2Var(\epsilon^2)\sum_{i=0}^{M-1}(\sum_{j=K_i+1}^{M}\frac{a_j}{K_j})^2 \end{aligned} \quad (46)$$

With the representation (29), (46) becomes:

$$M^{-1}2Var(\epsilon^2)\int_0^1(\int_x^1\frac{g(y)}{y}dy)^2dx+o(M^{-1})=M^{-1}4Var(\epsilon^2)\int_0^1\int_0^y g(x)\frac{g(y)}{y}dxdy+o(M^{-1}) \quad (47)$$

Since, under condition 1,

$$E\left[\sum_{j=1}^{M} a_j \frac{1}{K_j}(\sum_{i=0}^{K_j-1}\epsilon_{t_i}^2+\sum_{i=n-K+1}^{n}\epsilon_{t_i}^2)\right]=2E\epsilon^2. \quad (48)$$

one can obtain that

$$M^{1/2}\left[\sum_{j=1}^{M} a_j \frac{1}{K_j}(\sum_{i=0}^{K_j-1}\epsilon_{t_i}^2+\sum_{i=n-K+1}^{n}\epsilon_{t_i}^2)-2E\epsilon^2\right]\xrightarrow{\mathcal{L}}\left(4Var(\epsilon^2)\int_0^1\int_0^y g(x)\frac{g(y)}{y}dxdy\right)^{1/2}N(0,1), \quad (49)$$

where the convergence is stable in law, N(0,1) is independent of other asymptotic terms.

We now turn to the cross term. We make the assumptions stated at the beginning of Section 4. In particular, we proceed, without loss of generality, as if X were a martingale.

$$\begin{aligned} [X,\epsilon]^{(K)} &= \frac{1}{K}\sum_{i=K}^{n}(X_{t_i}-X_{t_{i-K}})(\epsilon_{t_i}-\epsilon_{t_{i-K}}) \\ &= \frac{1}{K}\sum_{i=0}^{n} b_i^{(K)}\epsilon_{t_i}, \end{aligned}$$

where

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

It is easy to see that

$$E\left([X,\epsilon]_T^{(K)}|X \text{ process}\right)=0, \quad (50)$$

since ϵ has mean zero. Also, because ϵ is white noise proces,

$$Cov\left([X,\epsilon]_T^{(J)},[X,\epsilon]_T^{(K)}|X \text{ process}\right) = \frac{1}{JK}\sum_i^n b_i^{(J)}b_i^{(K)}Var(\epsilon) \tag{51}$$

Note that, with $J \wedge K = \min(J,K)$,

$$b_i^{(J)}b_i^{(K)} = (b_i^{(J\wedge K)})^2 + \text{ martingale increment }. \tag{52}$$

Also

$$\sum_{i=0}^n (b_i^{(K)})^2 = \underbrace{2\sum_{i=K}^n (X_{t_i}-X_{t_{i-K}})^2}_{2K[X,X]^{(K)}} + \text{ martingale term }. \tag{53}$$

It follows from (50)-(53), and a precise but tedious analysis of the martingale remainder terms, that

$$\begin{aligned} Cov\left([X,\epsilon]_T^{(J)},[X,\epsilon]_T^{(K)}|X \text{ process}\right) &= 2\frac{J\wedge K}{JK}[X,X]^{(J\wedge K)}Var(\epsilon) + \text{ martingale term} \\ &= 2\frac{J\wedge K}{JK}(<X,X>Var(\epsilon)+o_p(1)) \end{aligned} \tag{54}$$

where we also use that $[X,X]^{(K)}$ converges in probability to $<X,X>$.

Summing up

$$\begin{aligned} Var\left(\sum_{i=1}^M a_i[X,\epsilon]^{(K_i)}|X \text{ process}\right) &= 2\sum_{J=1}^M\sum_{K=1}^M a_J a_K \frac{J\wedge K}{JK}(<X,X>Var(\epsilon)+o_p(1)) \\ &= 2\sum_{J=1}^M\sum_{K=1}^M \frac{1}{M^2}g(\frac{J}{M})g(\frac{K}{M})\frac{J\wedge K}{JK}(<X,X>Var(\epsilon)+o_p(1)) \\ &= M^{-1}2\int_0^1\int_0^1 g(x)g(y)\frac{x\wedge y}{xy}dxdy <X,X>Var(\epsilon)+o_p(M^{-1}). \end{aligned} \tag{55}$$

By similar methods in Zhang, Mykland, and Aït-Sahalia (2003), Lemma 1 follows (48)-(50) and (55).

Finally, by the same methods, it is easy to see that the two components of the remainder term are asymptotically independent (given the data), and that the remainder term as a whole is asymptotically independent (again, given the data) of the pure noise and pure discretization terms.

REFERENCES

Abramowitz, M. and Stegun, I. A. (1972), *Handbook of Mathematical Functions*, New York, N.Y.: Dover.

Aït-Sahalia, Y., Mykland, P. A., and Zhang, L. (2003), "How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise," Tech. rep., to appear in *Review of Financial Studies*.

Bibby, B. M., Jacobsen, M., and Sørensen, M. (2002), "Estimating Functions for Discretely Sampled Diffusion-Type Models," in *Handbook of Financial Econometrics*, eds. Aït-Sahalia, Y. and Hansen, L. P., Amsterdam, The Netherlands: North Holland.

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

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

Gloter, A. and Jacod, J. (2001), "Diffusions with Measurement Errors: I - Local Asymptotic Normality," *ESAIM: Probability and Statistics*, 5, 225–242.

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

Hansen, P. R. and Lunde, A. (2004), "Realized Variance and IID Market Microstructure Noise," Tech. rep., Brown University, Department of Economics.

Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.

Mykland, P. A. and Zhang, L. (2002), "ANOVA for Diffusions," Tech. rep., The University of Chicago, Department of Statistics.

Stein, M. L. (1987), "Minimum Norm Quadratic Estimation of Spatial Variograms," *Journal of the American Statistical Association*, 82, 765–772.

— (1990), "A Comparison of Generalized Cross Validation and Modified Maximum Likelihood for Estimating the Parameters of a Stochastic Process," *The Annals of Statistics*, 18, 1139–1157.

— (1993), "Spline Smoothing with an Estimated Order Parameter," *The Annals of Statistics*, 21, 1522–1544.

Ying, Z. (1991), "Asymptotic Properties of a Maximum Likelihood Estimator with Data from a Gaussian Process," *Journal of Multivariate Analysis*, 36, 280–296.

— (1993), "Maximum Likelihood Estimation of Parameters under a Spatial Sampling Scheme," *The Annals of Statistics*, 21, 1567–1590.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2003), "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," Tech. rep., Carnegie-Mellon University.

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