数理解析研究所講究録
第 1620 巻 2009 年 39-52

An attempt to the real-time estimation of the spot volatility

Shigeyoshi Ogawa

Department of Mathematical Sciences, Ritsumeikan University,
Kusatsu, Shiga 525-8577, JAPAN
ogawa-s@se.ritsumei.ac.jp

1 Introduction

We are concerned with the problem of estimating temporal values of the volatility in such situation that the observation of the price process is contaminated by a high frequency noise due to microstructural causes of the system. For the case that there is no scuh noise, we have presented in the preceding article ([3]) a scheme that is simple enough to work effectively in a real-time manner. The aim of the present note is to introduce a new scheme by doing suitable modification to that old scheme so that the new one still maintains the nice property of being a real-time estimator. This is done by the method of multi-step regularization that we are to introduce now.

2 Microstructure Noise

Given the observation data over a finite interval $[0, T]$ of an asset price process, say $p(t)$, we are concerned with the problem of estimating its volatility. Here we suppose that the price process $p(t)$ is generated by the following Itô SDE,

$$dp(t) = a(t, \omega)dt + b(t, \omega)dW_t, \quad 0 \leq t \leq T \tag{1}$$

where $W(t),\ t \geq 0$ is the real Brownian motion defined on a probability space $(\Omega, \mathcal{F}, P)$, and $a(\cdot),\ b(\cdot)$ are real coefficients, measurable in (t, ω) and square integrable in t over $[0, T]$.

We also suppose that the $b(\cdot, \omega)$ is adapted to an increasing family of σ-fields $\{\mathcal{F}_t\}$ such that for any t, $\mathcal{F}_t \supset \sigma\{W_s; s \leq t\}$, $\mathcal{F}_t$ is independent of the natural σ-fields $\sigma\{W_u - W_t; u \geq t\}$. Moreover we suppose that,

$$A' = \sup_t \sqrt{E[|a(t)|^2]} < \infty, \quad \text{and} \quad B := \sup_{t \in [0,T]} E[b^4(t)] < \infty. \tag{2}$$

We will establish a numerical scheme for the estimation of the temporal values $b^2(t_k, \omega)$, at some specified set of points $T_e = \{t_k\} \in [0, T]$ by using a finite number of observed data of the process $p(t'_k)$ at the points t'_k in a set T_o somehow richer than T_e, namely $[0, T] \supset T_o \supset T_e$. As for this problem, we would like to mention that in many situations the price process $p(t)$ is always observed with a noise $Z(t)$, namely the prices of stocks are observed in the form of a random process $X(t)$ given in (3) below, where the noise $Z(t)$ is a random process of very high frequency.

$$X(t) = p(t) + Z(t), \quad 0 \leq t \leq T. \tag{3}$$

For the noise process we suppose that it has the following properties,

Hypothesis 1 *$Z(t)$ is a real process such that,*

(Z)
- (*a*) $E[Z] = 0$, $\sup_t E[Z_t^2] = A^2 < \infty$ (*A = unknown positive constant*).
- (*b*) *For any n and any set of observation epoques $\{t_1, t_2, \cdots, t_n\}$ the random variables $\{Z(t_k),\ 1 \leq k \leq n\}$ are independent.*

Remark 2.1 *As for the characterization of the noise process $Z(t)$, when looking at the literature (e.g. see the references given in [2], [4]), it seems customary to suppose that the process $Z(t)$ is independent of the price process $p(t)$; However we do not need this hypothesis in our discussion.*

The existence of such noise in a real situation could be simply recognized when we attempt to estimate the realized volatility by techniques based on quadratic variation analysis. In fact in such a situation it is visibly clear that the estimator would explode as the number of observations increases.

It may be thought that in reality this noise is due to many causes such as the error produced in quantification process of price figures or by other micro-structural causes of the system. Whatever the causes may be there arises as a mathematical subject the problem of how to make efficient estimation of the spot volatilities for the case of noise.

For the case that there is no such noise a large amount of researches has been done (see for example J.Jacod [1], J.Jacod et al [2], S.Sanfelici-M.Mancino [4] etc), among these we like to refer to our recent result [3] where we have aimed to make an estimation of the temporal (not the integrated) values of volatility and presented a scheme that can perform in a real-time manner. The reason for this advantageous property is the fact that it is based only on the idea of quadratic variation in its simplest form and nothing else is required from theory or technique. However because of this reason we readily notice that the method of quadratic variation does not work in the new situation, unless being coupled with some auxiliary procedure to eliminate the influence of noise. The present note aims to propose such a scheme.

3 Regularization

First we have to do something to reduce the noise and we intend to apply to the observed process the regularization procedure by the kernel method. For this purpose we need to collect more data than the original set $\{X(t_k)\}$. So let us introduce the set of observation points $\{t_k^i,\ 0 \leq k \leq N,\ 0 \leq i \leq 2M\}$, finer than the originally given points set $\{t_k\}$, in such way that;

$$\begin{aligned} t_k &= \frac{T}{N}k = k\Delta, \quad (0 \leq k \leq N) \quad \Delta = \frac{T}{N} \\ t_k^i &= t_k + (\frac{i}{2M} - \frac{1}{2})\Delta, \quad (0 \leq i \leq 2M). \end{aligned} \tag{4}$$

Associated with this we employ the following symbols;

$$\Delta_k^i X = X(t_{k+1}^i) - X(t_k^i), \quad \Delta_k X = \Delta_k^{M+1} X.$$

Also $\Delta_k^i p$, $\Delta_k^i Z$ are given in the same way, and for the smoothing operation we use the following notation:

$$\overline{X}^M(t_k) = \frac{1}{2M+1} \sum_{i=0}^{2M} X(t_k^i),$$

or very simply by the notation, $\overline{X}$. The quantities $\overline{p}$, $\overline{Z}$ are defined in a similar way.

Remark 3.1 (Causal Form) *The smoothing operation, $X \longrightarrow \overline{X}^M$ given above is constructed in a symmetric form around each point t_k in question, but it can be given in a causal $\overline{X}^{(-)}$ or an advanced form $\overline{X}^{(+)}$ which are given below:*

$$\overline{X}^{(+)} = \frac{1}{2M+1}\sum_{i=0}^{2M} X(t_k + \frac{i}{2M}\Delta)$$

or

$$\overline{X}^{(-)} = \frac{1}{2M+1}\sum_{i=0}^{2M} X(t_k - \frac{i}{2M}\Delta).$$

In this article we mainly discuss the estimator of the symmetric form, just for conciseness reasons. However it should be remarked that if we use a smoothing of the causal type $\overline{X}^{(-)}$ the estimator, that we are going to establish now, can be a proper real-time estimator.

Let us apply a smoothing procedure to all quantities in equation (3) to obtain the followings,

$$\overline{X}^M(t) = \overline{p}^M(t) + \overline{Z}^M(t). \tag{5}$$

Notice at this stage that,

$$E[\overline{Z}^M] = 0, \quad E[(\overline{Z}^M)^2] \leq \frac{A^2}{2M+1}. \tag{6}$$

Now take the increment over the subinterval $[t_k, t_{k+1})$ of all quantities in equation (5),

$$\Delta_k\overline{X} = \Delta_k\overline{p} + \Delta_k\overline{Z}, \tag{7}$$

where

$$\Delta_k\overline{X} = \overline{X}(t_{k+1}) - \overline{X}(t_k) = \frac{1}{2M+1}\sum_{i=0}^{2M}\Delta_k^i X. \tag{8}$$

Here the symbol Δ_k^i stands for the difference operation over the subinterval $[t_k^i, t_{k+1}^i)$, for example,

$$\Delta_k^i X = X(t_{k+1}^i) - X(t_k^i).$$

Thus we have,

$$\Delta_k \overline{X} = \frac{1}{2M+1} \sum_{i=0}^{2M} \Delta_k^i p + \Delta_k \overline{Z}. \tag{9}$$

By definition of the price process $p(t)$, we have,

$$\Delta_k^i p = \int_{t_k^i}^{t_{k+1}^i} \{a(s)ds + b(s)dW_s\},$$

and by applying the Euler-Maruyama scheme this can be expressed in the discrete form as follows;

$$\Delta_k^i p = a(t_k^i)\Delta + b(t_k^i)\Delta_k^i W + \epsilon_k^i \tag{10}$$

with,

$$\epsilon_k^i = \int_{t_k^i}^{t_{k+1}^i} \{(a(s) - a(t_k^i))ds + (b(s) - b(t_k^i))dW_s\}.$$

Now for the evaluation of the intensity of this error ϵ_k^i we need the following assumption (H) on the regularity of the coefficients $a(\cdot), b(\cdot)$,

Hypothesis 2 *The coefficient $b(\cdot)$ is Hölder continuous of order $\alpha \in (0, 1]$ in the $L^2(\Omega)$-sense,*

(H) There exists a constant L_B such that $E|b(t) - b(s)|^2 \leq L_B^2 |t - s|^{2\alpha}$.

Then under this condition it is almost immediate to see the following

Lemma 3.2 *The error term ϵ_k^i satisfies the estimate below uniformly in "i, k",*

$$E[|\epsilon_k^i|^2] \leq C_\epsilon |\Delta|^{2(\alpha \wedge 1/2)+1} \tag{11}$$

where

$$C_\epsilon = \frac{L_B^2}{2\alpha + 1} + 4A'^2$$

and the symbol "$\alpha \wedge 1/2$" stands for the minimum of the two arguments.

4 Second Regularization

We have employed the regularization procedure to reduce the power of the noise. As there still remains a fluctuating component in the quadratic variation $(\Delta_k\overline{X})^2$ we need again another regularization procedure which we will explain in this section.

From equation (9) we have,

$$(\Delta_k\overline{X})^2 = \frac{1}{(2M+1)^2}\sum_{i,j=0}^{2M}\Delta_k^i p\Delta_k^j p + (\Delta_k\overline{Z})^2 + 2\Delta_k\overline{Z}\Delta_k\overline{p}$$

and from equation (10) we also have,

$$\begin{aligned}\Delta_k^i p\Delta_k^j p &= \{a(t_k^i)\Delta + b(t_k^i)\Delta_k^i W + \epsilon_k^i\}\{a(t_k^j)\Delta + b(t_k^j)\Delta_k^j W + \epsilon_k^j\}\\ &= b(t_k^i)b(t_k^j)\Delta_k^i W\Delta_k^j W + \delta_k^{i,j}\end{aligned}$$

where

$$\begin{aligned}\delta_k^{i,j} &= a(t_k^i)a(t_k^j)\Delta^2 + \{a(t_k^i)b(t_k^j)\Delta_k^j W + a(t_k^j)b(t_k^i)\Delta_k^i W\}\Delta\\ &+\epsilon_k^i\{a(t_k^i)\Delta + b(t_k^i)\Delta_k^i W\} + \epsilon_k^j\{a(t_k^j)\Delta + b(t_k^j)\Delta_k^j W\}\\ &+\epsilon_k^i\epsilon_k^j.\end{aligned}$$

As for the quantity $\delta_k^{i,j}$ we easily get the next estimate from Lemma 3.2,

Lemma 4.1

$$E|\delta_k^{i,j}| \le C_\delta\Delta^{(\alpha\wedge 1/2)+1}$$

where

$$C_\delta = 2\sqrt{BC_\epsilon}+1 = 2\{B(\frac{L_B^2}{2\alpha+1} + 4A'^2)\}^{1/2} + 1.$$

Proof We have

$$\begin{aligned}&E[|\epsilon_k^i(a(t_k^i)\Delta + b(t_k^i)\Delta_k^i W)|] \le \{E[|\epsilon_k^i|^2](A'^2\Delta + \sqrt{B})\Delta\}^{1/2}\\ &\le C_\epsilon\Delta^{(\alpha\wedge 1/2)+1/2}\cdot\{(\sqrt{B}+1)\Delta\}^{1/2}\\ &= \sqrt{C_\epsilon(\sqrt{B}+1)}\Delta^{(\alpha+1/2)+1}.\end{aligned}$$

On the other hand, it is immediate to see that the contribution from other terms appearing in the expression of the $\delta_k^{i,j}$ is of the order $O(\Delta^{2(\alpha\wedge 1/2)+1})$. Thus, taking

$$C_\delta = 2\sqrt{C_\epsilon(\sqrt{B}+1)}+1,$$

we confirm the validity of the estimate given above. *q.e.d.*

To analyze the main term $b(t_k^i)b(t_k^j)\Delta_k^i W \Delta_k^j W$, we first notice by a simple computation the validity of the following equality for the case $i \geq j$;

$$\Delta_k^i W \Delta_k^j W = \left(1 - \frac{i-j}{2M}\right)\Delta + \theta_k^{i,j} \tag{12}$$

where,

$$\begin{aligned}\theta_k^{i,j} &= 2\int_{t_k^i}^{t_{k+1}^j}(W_s - W_{t_k^i})dW_s \\ &+\{W(t_{k+1}^i) - W(t_{k+1}^j)\}\Delta_k^j W + \{W(t_k^i) - W(t_k^j)\}\{W(t_{k+1}^j) - W(t_k^i)\}.\end{aligned}$$

Hence,

$$\begin{aligned}(\Delta_k \overline{p})^2 &= \frac{1}{(2M+1)^2}\sum_{i,j}\Delta_k^i p \Delta_k^j p \\ &= \frac{1}{(2M+1)^2}\sum_{i,j}\delta_k^{i,j} + b(t_k^i)b(t_k^j)(1 - \frac{|i-j|}{2M})\Delta + \eta_k^{i,j}\end{aligned} \tag{13}$$

where $\eta_k^{ij} = b(t_k^i)b(t_k^j)\theta_k^{ij}$.

From Lemma 4.1 and the expression (13) we see that

$$E[(\Delta_k \overline{p})^2] \leq C_p \Delta, \qquad C_p = C_\delta \Delta^{\alpha\wedge 1/2} + 2\sqrt{B}, \tag{14}$$

thus for a sufficiently large M, the inequality (14) holds for such $C_p = 3\sqrt{B}$.

Notice also that,

$$\begin{aligned}\frac{(\Delta_k \overline{X})^2}{\Delta} &= \frac{1}{(2M+1)^2}\sum_{i,j}\{b(t_k^i)b(t_k^j)(1 - \frac{|i-j|}{2M}) + \frac{\delta_k^{ij}}{\Delta} + \frac{\eta_k^{ij}}{\Delta}\} \\ &+2\frac{\Delta_k \overline{p}}{\sqrt{\Delta}}\frac{\Delta_k \overline{Z}}{\sqrt{\Delta}} + \frac{(\Delta_k \overline{Z})^2}{\Delta}.\end{aligned} \tag{15}$$

By Hypothesis **(Z)** and the definition of the $\Delta_k\overline{Z}$ we see that,

$$E\,\frac{(\Delta_k\overline{Z})^2}{\Delta} \leq 2\frac{A^2}{2M+1}\frac{1}{\Delta}.$$

In other words, with the constant $C_z = \dfrac{2A^2}{T}$ we have

$$E[\frac{(\Delta_k\overline{Z})^2}{\Delta}] \leq C_z\frac{N}{M}, \qquad E[\frac{|\Delta_k\overline{p}\Delta_k\overline{Z}|}{\Delta}] \leq C_{pz}\sqrt{\frac{N}{M}} \tag{16}$$

where $C_{pz} = \sqrt{C_pC_z}$ and

$$E\,\frac{(\Delta_k\overline{Z})^2}{\Delta} \leq C_z\frac{N}{M}.$$

In order to keep these quantities small by letting $N, M \longrightarrow \infty$ we need to introduce some condition on these two parameters M, N;

$$(C) \qquad \lim_{N,M\to\infty}\frac{N}{M} = 0, \quad i.e. \quad \frac{1}{M} = o(\frac{1}{N}).$$

To estimate the effect caused by the quantity η_k^{ij} we remark that,

$$E[\eta_k^{ij}] = 0, \quad E[(\eta_k^{ij})^2] \leq B\Delta^2 \quad (B = \sup_t E[b^4(t)])$$

hence we see that the random variables $\dfrac{\eta_k^{ij}}{\Delta}$ are bounded in $L^2(\Omega)$, but we have no reason to expect that,

$$\limsup_{N\to\infty} E\left[\frac{|\eta_k^{ij}|}{\Delta}\right] = 0.$$

However we should also notice that they have the following property,

Lemma 4.2 *For each fixed (i,j) the family of random variables $k \to \{\eta_k^{ij}\}$ are almost uncorrelated, that is;*

$$E[\eta_k^{ij}\eta_l^{ij}] = 0 \quad for \quad |k-l| \geq 2.$$

Based on this observation we apply again the regularization procedure to the quantities $\dfrac{(\Delta_k\overline{X})^2}{\Delta}$, and we are led to propose the next scheme as our estimator;

Definition 4.3 (Estimator)

$$\hat{b}^2(t_k) = \frac{G(M)^{-1}}{2L+1} \sum_{l=0}^{2L} \frac{(\Delta_{k+l-L}\overline{X})^2}{\Delta} \tag{17}$$

where

$$G(M) = \frac{1}{(2M+1)^2} \sum_{i,j}^{2M} \left(1 - \frac{|i-j|}{2M}\right) = \frac{8M^2 + 6M + 1}{3(2M+1)^2}.$$

Remark 3 The estimator that can be computed in *real time* should be given in the following form;

$$\hat{b}^2(t_k) = \frac{G(M)^{-1}}{2L+1} \sum_{i=0}^{2L} \frac{(\Delta_{k+i-2L}\overline{X}^{(-)})^2}{\Delta}. \tag{18}$$

As for the efficiency of this estimator (17) we have the following result,

Theorem 4.4 *For some positive constants C_1, C_2, C_3, C_4, independent of the parameters L, M, N the following estimate holds at every point $t_k = k\Delta$*

$$E\left[|\hat{b}^2(t_k) - b^2(t_k)|\right] \le C_1 \left(\frac{L}{N}\right)^{\alpha} + C_2 \frac{1}{N^{\alpha \wedge 1/2}} + C_3 \sqrt{\frac{N}{M}} + C_4 \frac{1}{\sqrt{L}}.$$

This statement may be seen from the discussion done up to here, however we will give its proof for sure in the following section and give especially some concrete candidates for the constants.

5 Proof of the Theorem

From the expression (15) we have,

$$\begin{aligned}
&\hat{b}^2(t_k) - b^2(t_k) \\
&= \frac{G^{-1}}{2L+1}\sum_{l=0}^{2L}\{\frac{(\Delta_{k+l-L}\overline{X})^2}{\Delta} - G(M)b^2(t_k)\} \\
&= \frac{G^{-1}}{2L+1}\sum_{l=0}^{2L}\frac{1}{(2M+1)^2}\sum_{i,j}^{2M}\left[\{b(t^i_{k+l-L})b(t^j_{k+l-L}) - b^2(t_k)\}(1-\frac{|i-j|}{2M}) + \frac{\delta^{ij}_{k+l-L}}{\Delta}\right] \\
&+\frac{G^{-1}}{(2M+1)^2}\sum_{i,j}\frac{1}{2L+1}\sum_{l=0}^{2L}\frac{\eta^{ij}_{k+l-L}}{\Delta} \\
&+\frac{G^{-1}}{2L+1}\sum_{l=0}^{2L}\left[2(\frac{\Delta_{k+l-L}\overline{p}}{\sqrt{\Delta}})(\frac{\Delta_{k+l-L}\overline{Z}}{\sqrt{\Delta}}) + \frac{(\Delta_{k+l-L}\overline{Z})^2}{\Delta}\right].
\end{aligned}$$

Since

$$|t^i_{k+l-L} - t_k| \le L\Delta \quad {}^\forall i,k,l,$$

thus by assumption (H) we have the inequality,

$$\begin{aligned}
&E[|b(t^i_{k+l-L})b(t^j_{k+l-L}) - b^2(t_k)|] \\
&\le\; E[|b(t^i_{k+l-L})||b(t^j_{k+l-L}) - b(t_k)| + |b(t_k)||(b(t^i_{k+l-L}) - b(t_k)|] \\
&\le\; 2B^{1/4}L_B^{1/2}(L\Delta)^\alpha \\
&= C_1\left(\tfrac{L}{N}\right)^\alpha
\end{aligned}$$

where, $C_1 = 2B^{1/4}L_B^{1/2}T^\alpha$ which implies that

$$\frac{G^{-1}}{2L+1}\sum_{l=0}^{2L}\sum_{i,j}^{2M}E|\{b(t^i_{k+l-L})b(t^j_{k+l-L}) - b^2(t_k)\}(1-\frac{|i-j|}{2M})| \le C_1(\frac{L}{N})^\alpha. \quad (19)$$

On the other hand, from Lemma 4.1 we have,

$$E|\delta^{ij}_{k+l-L}| \le C_\delta\Delta^{(\alpha\wedge 1/2)+1}$$

so taking the trivial estimate, $G^{-1} \le 3 \;\; (M \ge 3)$, into account we see that,

$$\frac{G^{-1}}{2L+1}\sum_{l=0}^{2L}\frac{1}{(2M+1)^2}\sum_{i,j}^{2M}E|\frac{\delta^{ij}_{k+l-L}}{\Delta}| \le C_2\frac{1}{N^{\alpha\wedge 1/2}} \quad (20)$$

with $C_2 = 3C_\delta$.

For the remaining 2 terms in the inequality, it is almost immediate to obtaint the following estimates,

$$\frac{G^{-1}}{(2M+1)^2}\sum_{i,j}E|\frac{1}{2L+1}\sum_{l=0}^{2L}\frac{\eta_{k+l-L}^{ij}}{\Delta}| \leq C_4\frac{1}{\sqrt{L}} \quad (C_4 = 2\sqrt{3B}) \tag{21}$$

and

$$\frac{1}{2L+1}\sum_{l=0}^{2L}\frac{G^{-1}}{(2M+1)^2}\sum_{i,j}E|\frac{2\Delta_{k+l-L}\overline{Z}\cdot\Delta_{k+l-L}\overline{p}}{\Delta}+\frac{(\Delta_{k+l-L}\overline{Z})^2}{\Delta}| \tag{22}$$
$$\leq C_3\sqrt{\tfrac{N}{M}} \quad \text{with} \quad C_3 = 3(C_{pz}+C_z).$$

Now summing up all inequalities (19),(20),(21), (22) we confirm the validity of the desired inequality with appropriately chosen constants as follows;

$$\begin{aligned} C_1 &= 2B^{1/4}L_B^{1/2}T^\alpha \\ C_2 &= 3C_\delta = 6\{\sqrt{B^{1/2}(\tfrac{L_B^2}{2\alpha+1}+4A'^2)}+1\} \\ C_3 &= 3(\sqrt{C_pC_z}+C_z) \geq 3\tfrac{A^2}{T}(\sqrt{B}+1/2) \\ C_4 &= 2\sqrt{3B} \end{aligned} \tag{23}$$

q.e.d.

6 Discussion

We would like to give two comments about our estimator.

6.1 Is it optimal?

In our method, the procedure of smoothing plays a very important role. So for the mathematical interest on the optimality of the estimator, we might as well start our discussion with candidates of a more general form as follows;

Let $\{w_i; 0 \leq i \leq 2L\}$ be a window of width $2L$, namely the nonnegative numbers such that, $\sum_i w_i = 1$. Given this we may think of the estimator in the following form.

$$\hat{b}_w^2(t_k) = \sum_{l=0}^{2L} w_l\frac{(\Delta_{k+l-L}\overline{X})^2}{\Delta}. \tag{24}$$

The complete way of discussion is of course to determine the optimal window $\{w_i\}$ by searching for one that will make the estimation error a minimum, but in this note we have not followed this approach. One reason is that we have discussed this question in the previous article [3] and we know that the smoothing operation given in this note is almost optimal. Another reason is that we are interested in such practical scheme that is simple enough to compute rapidly so that it can work in a real-time manner. The scheme which is proved mathematically optimal in the above sense may not satisfy this constraint.

6.2 How to determine the parameters

Theorem 4.4 tells us how to determine the parametres L, M, N. For simplicity let us take the case where $\alpha \leq 1/2$. Then we have

$$Er = E[|\hat{b}^2(t_k) - b^2(t_k)|] \leq O((\frac{L}{N})^\alpha) + O(\sqrt{\frac{N}{M}}) + O(\frac{1}{\sqrt{L}}).$$

Notice for this case, the three parameters should obey the following constraint;

$$L << N << M$$

and that the total number of points we need for the estimator is $2M \times N$.

Thus their determinations will be carried in the following procedure;

1. Fix the precision level ϵ_0.

2. The third term on the right hand side of the inequality being independent of the other parameters M, N, the parameter L should be determined by the constraint,

$$\frac{1}{\sqrt{L}} \leq \epsilon_0 \longrightarrow L \geq \frac{1}{\epsilon_0^2}.$$

3. Three terms on the right hand side should be of the same order, and thus the parameters L, M, N should satisfy the constraints,

$$(\frac{L}{N})^\alpha = \sqrt{\frac{N}{M}} = \frac{1}{\sqrt{L}}.$$

4. From this we obtain the following expressions for M, L in terms of N,

$$(\frac{L}{N})^{\alpha} = \frac{1}{\sqrt{L}} \longrightarrow L = N^{\frac{2\alpha}{2\alpha+1}}$$

and

$$\sqrt{\frac{N}{M}} = \frac{1}{\sqrt{L}} = \frac{1}{N^{\frac{\alpha}{2\alpha+1}}} \longrightarrow M = N^{\frac{2\alpha+1}{2\alpha}} = N^{1+\frac{1}{2\alpha}}.$$

5. In this case, the relation $\frac{N}{M} = N^{-\frac{1}{\alpha}}$ fits to the condition (E).

6. In this case, we have the estimate; $N \geq (\frac{1}{\epsilon_0})^{2+\frac{1}{\alpha}}$ and

$$Er \leq 3\epsilon_0 = O(\frac{1}{N^{\frac{\alpha}{2\alpha+1}}}).$$

7. Statisticians like to measure the labor spent for the estimation procedure by the amount of data used. In our case, the number of data points necessary for the estimation at one epoque t_k is $2ML$, while the total number of data points for the estimation at all points t_k $(0 \leq k \leq N)$ is $2MN$.

In particular when $\alpha = 1/2$ we expect the following precision level;

$$L = \sqrt{N}, \quad M = N^{3/2}, \quad Er = O(\frac{1}{\sqrt[4]{N}}).$$

Though we see by numerical experiments that our estimator works well, the condition $N << M$ seems restrictive. We might relax this condition so as to make our estimator still work with less labor for computation. This problem will be discussed in future work.

References

[1] J.Jacod, "Estimation de la volatilite et problèmes connexes", *Cours Bachelier* à l'I.H.P., Novembre et Decembre 2006.

[2] J.Jacod, P.Mykland, Y.Li et al, "Microstructure noise in the continuous case - ERN-2", *in Manuscript*, July 2007.

[3] S.Ogawa and K.Wakayama "On a real-time scheme for the estimation of volatility", Monte Carlo Methods and Appl., Vol.13, No.2(2007), pp.99 – 116

[4] M.Mancino and S.Sanfelici, "Robustness of Fourier estimator of integrated volatility in the presence of microstructure noise" (*preprint*) 2007