Estimation of the regime shifts of the volatility and the trend by a variational method.

Françoise Dibos[1], Georges Koepfler[2], Christian Lopez[1]

[1] CEREMADE (URA CNRS 749), Université de Paris 9 Dauphine,

Place de Lattre de Tassigny, 75775 Paris cedex 16, France

[2] PRISME, UFR de Maths. et Info. , Univ. de Paris 5 René Descartes,

45, rue des Saints-Pères, 75270 Paris cedex 06, France

{dibos,gk,lopez}@ceremade.dauphine.fr

1 Introduction

Let $dX_t = \sigma(t, X_t)dW_t$ be a diffusion process, where W_t is the standard Brownian motion. A lot of estimators for σ^2, more or less sophisticated, may be found in the literature ([1], [3], [5] and references inside). They usually depend on *a priori* hypotheses on σ.
For instance, if $\sigma \equiv C$ on $[0, T]$ and $t_0 = 0 < t_1 < t_2 \cdots < t_n = T$ is a regular sampling of $[0, T]$, the "best" estimator of σ^2 , for n large enough, is

$$\widehat{\sigma_c}^2 = \frac{1}{n}\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}} - X_{t_i})^2}{t_{i+1} - t_i}\ .$$

Let us remark that $\widehat{\sigma_c}^2$ is the minimum point of the functional

$$J\ :\ \{S \mid S \quad \text{is a constant function on [0,T]}\} \longrightarrow \mathbb{R}$$

$$J(S) = \int_0^T (\widehat{\sigma}^2 - S)^2 dt\ ;$$

where $\widehat{\sigma}^2$ is the piecewise constant function equal to

$$\widehat{\sigma}^2(t) = \widehat{\sigma}^2(t_i, X_{t_i}) = \frac{(X_{t_{i+1}} - X_{t_i})^2}{t_{i+1} - t_i} \quad \text{for} \quad t \in [t_i, t_{i+1}[$$

For $\sigma = \sigma(t)$ kernel estimators are often used. Roughly speaking, all of them estimate $\widehat{\sigma}^2(t)$ by an average of the values $\widehat{\sigma}^2(t_i, X_{t_i})$ for t_i near t.

In this paper, we have a different purpose. We don't propose another estimator for σ^2 but we adapt a well known algorithm from signal and image processing in order to build, given a discrete observation of X_t on $[0, T]$, the best approximation in the L^2 sense of σ^2 by a piecewise constant function $\widehat{\sigma_p}^2$.
More precisely, let us suppose that $\sigma = \sigma(t)$, we are able to find the best approximation $\widehat{\sigma_p}^2$ of σ^2 with a given number of jumps and so localize these jumps.

Thus we can estimate easily the changes of the volatility regimes (see Hamilton [6] and Krolzig [8] for another way to estimate these changes).

In section one, we will present the Mumford-Shah method for finding the best piecewise constant approximation of $\widehat{\sigma_p}$ of the historical volatility $\sigma(t)$ with a given number of jumps. This will be used for the CAC40 (01/1994-01/1999) and we will able to recover the Asiatic and the Russian crises.

Moreover, we can use this estimation of the historical volatility regimes in order to price European options. Indeed, the price of an option can be estimated as the average of the prices of the option for each volatility regime, weighted by event frequencies, as soon as the historical volatility has been observed on a period long enough.

In section two, we will extend the method in order to approximate the trend by a piecewise constant function in a region where the volatility is now supposed to be constant. This can be used for "trend following" funds management.

2 Approximation of the volatility

2.1 The segmentation principle

Let us recall briefly the model proposed by Mumford and Shah ([10]), initially for image segmentation, in the one dimensional case.

Let $g: \ [0,T] \subset \mathbb{R} \longrightarrow \mathbb{R}$ a given function and $\lambda > 0$ a scale parameter. We define the segmentation of g at scale λ, as a minimum point of the functional

$$(E) \qquad J(u,K) = \int_{[0,T]} (g-u)^2 dx + \lambda \mathrm{Card}(K)$$

where the infimum is taken on $(u,K) \in \mathcal{S}$ where $\mathcal{S}$ is the set of couples (u,K) where u is constant on each connected component (i.e. region) of $[0,T] \setminus K$, K a finite subset of $[0,T]$ which represents the jump positions of u and $\mathrm{Card}(K)$ is the number of jumps.

Notice that

(i) if λ is large, $\mathrm{Card}(K)$ must be small, so the choice of λ is equivalent to the choice of the number of connected components of $[0,T] \setminus K$;

(ii) if K is given and (O_i) the connected components of $[0,T] \setminus K$, then $u|_{O_i} = g_{O_i}$ where g_{O_i} is the mean value of g on O_i.

The minimum point is obtained by a Region Growing algorithm ([2], [7]), which works as follows.

Let A be a subset of $[0,T]$ then we calculate

$$M(A) = \int_A (g-g_A)^2 dx + \lambda \mathrm{Card}(A) \ ,$$

where g_A is the mean value of g on A.

Now, let A_1 and A_2 be two neighbouring subsets of $[0,T]$,

we decide to merge A_1 and A_2 iff $M(A_1 \cup A_2) \leq M(A_1) + M(A_2) \quad (\mathcal{M})$.

2.2 The segmentation algorithm

The input of the method is the number of jumps, which is the number of regions N minus 1. In practice, the segmentation of the signal works as follows :
In the beginning each signal sample is a region. For all adjacent regions we compute the value of λ for which a merging will be allowed, by using the above merging criterion $(\mathcal{M})$. This yields an increasing list of values, we then will take the smallest lambda value and merge the associated adjacent regions. Thus we merge the regions with the lowest cost first. We update our data structure and reiterate until the desired number of jumps is reached.

2.3 Application

The application to our problem is quite easy. Let $0 = t_0 < t_1 < \cdots t_n = T$ a subdivision of $[0, T]$ (n large enough). If $\sigma = \sigma(t)$, then in (E), g is the piecewise constant function such that

$$g(t) = \widehat{\sigma}^2(t_i, X_{t_i}) = \frac{(X_{t_{i+1}} - X_{t_i})^2}{t_{i+1} - t_i} \quad \text{for} \quad t \in [t_i, t_{i+1}[\; .$$

Let us remark that if we impose no jump for u we retrieve the constant estimator $\widehat{\sigma_c}^2$.

Experimentations

Figures 1,2,3 validation of the method

We first simulate a stochastic process $dX_t = \sigma_t dW_t$; such that

$$X_{n+1} = X_n + \sigma_n \sqrt{t} \mathcal{N}(0, 1)$$

where

$$X_0 = 0,04, \quad n \in \{0, 1, \cdots, 1999\}, \quad t = \frac{1}{256}$$

and

$$\sigma_n = 0,01 \quad \text{for} \quad n \in \{0, \ldots, 999\}$$
$$\sigma_n = 0,005 \quad \text{for} \quad n \in \{1000, \ldots, 1499\}$$
$$\sigma_n = 0,012 \quad \text{for} \quad n \in \{1500, \ldots, 1999\}$$

If we segment the quadratic variation of the rate X_t in three regions, we obtain the precise localisations for the jumps and we find for σ the following values $(0,01;\ 0,005;\ 0,0117)$.

Figures 4,5,6
The CAC40 (01/1994-01/1999), the input of the algorithm, the best approximation of σ^2 by a piecewise constant functions with four jumps.We can retrieve here the Asiatic (summer 97) and Russian (summer 98) crises.
As it is usually accepted that the evolution of the CAC40 is given by the Black and Scholes equation

$$\frac{dP_t}{P_t} = \mu dt + \sigma dW_t$$

we have for constant μ and σ

$$d(\ln P_t) = (\mu - \frac{\sigma^2}{2})dt + \sigma dW_t \ ,$$

therefore, the input of the algorithm is now (fig. 5) the quadratic variation of the log price, indeed we can neglect the term $(\mu - \frac{\sigma^2}{2})dt$ with respect to the term σdW_t.

Figure 7

Study of the influence of the choice of the number of jumps.

Let (Y_i), $i \in \{0, 1 \cdots, n-1\}$, be a given signal and we look for its best approximation in the Mumford-Shah sense by a piecewise constant signal. That is, we minimize, the quadratic error, penalized by the number of jumps

$$J_\lambda(\widehat{Y}, N) = \lambda(N-1) + \sum_{j=1}^{N} \sum_{i_j \leq i < i_j+1} (Y_i - \widehat{Y_{j,N}})^2$$

where $N-1$ is the number of jumps of the piecewise constant approximation $\widehat{Y}$, $\widehat{Y_{j,N}}$ the value of $\widehat{Y}$ on the subset $i_j \leq i < i_j + 1$ and

$$\bigcup_{j=1}^{N} \{i_j \leq i < i_j + 1\} = \{0, 1, \cdots, n-1\} \ .$$

As we have noticed before, the choice of λ (the scale parameter) allows to look at the signal Y at different scales.

If λ is large, then N must be small, and so, the choice of λ is equivalent to the choice of the number $N-1$ of jumps.

One of the challenges of this method is to choose the "optimal" number of jumps. This can be done by the comparison of the mean square error

$$E(N) = \sum_{j=1}^{N} \sum_{i_j \leq i < i_j+1} (Y_i - \widehat{Y_{j,N}})^2 \ ,$$

obtained with a segmentation in N regions with the error due to a constant approximation $E(1)$.

For instance, we can choose N such that

$$10 \log_{10} \left(\frac{E(N)}{E(1)} \right) = -3 \ ,$$

that is the number of regions which allows to divide by two the error due to a constant approximation.

In figure 7, we can follow the evolution of $\log_{10} \left(\frac{E(N)}{E(1)} \right)$ with $\log_{10} N$. We have chosen the CAC40, so Y is the quadratic variation of the log price and $\widehat{Y}$ the piecewise constant approximation of the variance of the process.

3 Constant approximation of the trend

Let $dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dW_t$ be a stochastic process and let us suppose that $\sigma(t, X_t) = \sigma$ is known. Indeed using the method of section 2 we first segment the volatility of the diffusion. Thus we are able to analyze each constant volatility interval.

In order to give a good approximation of the trend μ, we first need to denoise the increments dX_t of the process and, using the same segmentation technique as before, we obtain a piecewise constant approximation of the trend.

Denoising of the increments

Estimation of μ.

Let $Z(i)$, $n \in \{0, 1, \cdots, n-1\}$ be a discretization of the increments, roughly speaking, we can consider, for Δt small enough that

$$Z[i] = f[i] + W[i]$$

where $f[i] = \mu\Delta t$ and $(W[i])$ is a white noise with standard deviation $\widetilde{\sigma} = \sigma\sqrt{t}$, thus

$$E(W[i]) = 0 \quad \forall i \in \{0, 1, \cdots, n-1\} \ ,$$

$$E(|W[i]|^2) = \widetilde{\sigma}^2 \quad \forall i \in \{0, 1, \cdots, n-1\}$$

$$E(W[i]\overline{W[j]}) = 0, \quad i \neq j \ .$$

Let $\mathcal{B} = \{g_m\}$, $0 \leq m < N$ be an orthonormal basis of $\mathbf{C}^n$, the coefficients of Z, f, W in this basis are respectively equal to

$$Z_{\mathcal{B}}[m] =< X, g_m >= \sum_{i=0}^{n-1} X[i]\overline{g_m}[i] \ ,$$

$$f_{\mathcal{B}}[m] =< f, g_m >= \sum_{i=0}^{n-1} f[i]\overline{g_m}[i] \ ,$$

$$W_{\mathcal{B}}[m] =< W, g_m >= \sum_{i=0}^{n-1} W[i]\overline{g_m}[i] \ .$$

As

$$\begin{aligned} E(W_{\mathcal{B}}[m]\overline{W_{\mathcal{B}}[k]}) &= E(\sum_{i,j} W[i]\overline{W[j]}\ \overline{g_m[i]}g_k[j] \\ &= \widetilde{\sigma}^2 \sum_i \overline{g_m[i}g_k[i] \\ &= \widetilde{\sigma}^2 < g_k, g_m >= \widetilde{\sigma}^2\delta_{k,m} \ , \end{aligned}$$

we can say that the coefficients of the noise define a white noise with variance σ^2.

Lemma1 Let D be a diagonal estimator such that

$$Z \longrightarrow DZ = \sum_{m=0}^{N-1} a[m] Z_{\mathcal{B}}[m] g_m$$

where $a[m] \in \{0,1\}$/par then

$$r(f) = \inf_D E(\sum_{m=0}^{n-1} | < f - DZ, g_m > |^2 = \sum_{m=0}^{N-1} \min\{|f_{\mathcal{B}}[m]|^2, \widetilde{\sigma}^2\}$$

Proof
From the Parseval formula

$$E(\|f - DZ\|^2) = E(\sum_{m=0}^{n-1} | < f - DZ, g_m > |^2) ,$$

and , as $Z = f + W$ and the coefficients of the noise define a white noise with variance $\widetilde{\sigma}^2$, we have,

$$E(\sum_{m=0}^{n-1} | < f - DZ, g_m > |^2) = (\sum_{m=0}^{n-1} (1 - a[m])^2 |f_{\mathcal{B}}[m]|^2 + a[m]^2 \widetilde{\sigma}^2$$

thus

$$r(f) = \sum_{m=0}^{n-1} \min\{|f_{\mathcal{B}}[m]|^2, \widetilde{\sigma}^2\}$$

As this minimum risk depends on the unknown signal, threshold estimators are usually used. We can define the hard thresholding and the soft thresholding respectively by:

$$H_T Z = \sum_{m=0}^{n_1} d_m(Z_{\mathcal{B}}[m]) g_m ,$$

where

$$d_m(x) = x \quad \text{si} \quad |x| > T, \quad d_m(x) = 0 \quad \text{si} \quad |x| \leq T .$$

and

$$S_T Z = \sum_{m=0}^{n_1} d_m(Z_{\mathcal{B}}[m]) g_m ,$$

$$d_m(x) = \begin{cases} x - T & x \geq T \\ x + T & x \leq -T \\ 0 & |x| \leq T \end{cases}$$

If, we denote $r_{H,T}(f) = E(\|f - H_T Z\|^2$ and $r_{S,T}(f) = E(\|f - S_T Z\|^2$ we know from the theorem of Donoho ([4],[9]) that for $T = \widetilde{\sigma}\sqrt{2 \ln n}$, and $N \geq 4$,

$$r(f) \leq r_{H,T}(f) \leq (2 \ln n + 1)(\widetilde{\sigma}^2 + r(f)) ,$$

$$r(f) \leq r_{S,T}(f) \leq (2 \ln n + 1)(\widetilde{\sigma}^2 + r(f)) .$$

Roughly speaking for this choice of the threshold parameter, the risks are equivalent.
We use this result in order to denoise $f[n] = \mu dt$. We choose a wavelet orthonormal basis for the decomposition of the signals.

$$Z = \sum_{j=1}^{J} (\sum_{n} < Z, \psi_{j,n} > \psi_{j,n}) + \sum_{n} < f, \phi_{J,n} > \phi_{J,n}$$

where J is the order of the decomposition, ϕ the scale function of the multiresolution analysis, ψ the wavelet and

$$\psi_{j,n} = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j n}{2^j} \right)$$

Piecewise constant approximation for the trend

Let $S_T Z$ the result obtained after the denoising of the increments with the Donoho method, if μ is supposed to be constant the best estimation for μ is

$$\widehat{\mu} = \frac{1}{n} \sum_{i=0}^{n-1} \frac{S_T Z_i}{\Delta t} \ .$$

then we find an approximation of the trend μ, by a piecewise constant function with a given number of jumps, by using, as for the volatility, the region growing algorithm where the input function is

$$\mu(t) = \frac{S_T Z_i}{\Delta t} \quad \text{for} \quad t \in [t_i, t_{i+1}[\ .$$

Experimentations, figures 8,9,10

Let $dX_t = \mu dt + \sigma dW_t$ be a stochastic process where $\sigma = 0,1$,

$$\mu = 0,5 \quad t \in [0, 1000[, \quad \mu = 0,2 \quad t \in [1001, 1499[, \quad \mu = -0,2 \quad t \in [1500, 1999[.$$

Figure 8, Z_i: discretisation of the increments ($\Delta t = 1/256$).
Figure 9, z_i denoising of Z_i. We use a Daubechies wavelet and the decomposition is performed to the order 5.
Figure 10, M-S on denoized Z_i. The result is obtained by the region growing algorithm where we impose 2 jumps.
We find the precise positions for the jumps and for μ the values $0,4778$; $0,2097$; $-0,168$.

4 Conclusion

In this paper we propose an easy to use method compared to other methods like switching Markov models.

We applied our method only to scalar diffusions but we can adapt it to multidimensional diffusion. Thus we can also estimate piecewise constant correlation which can be very useful for risk management or asset allocation.

Acknowledgement : We want to thank Danièle Florens for all valuable conversations we have had on this subject.

References

[1] Avesani, Bertrand: Jumps and diffusions in volatility: it takes two to tango (preprint)
[2] Chambolle: Image segmentation by variational methods: Mumford and Shah functional and discrete approximations. SIAM, J. Numerical Analysis, vol 32, 1995. [3] Dacunha-Castelle, Florens: Estimation of the coefficient of diffusion from discrete observations (Stochastic 19, p 263-284, 1986)
[4] Donoho: Denoising by soft thresholding (IEEE Trans, Information theory, vol.IT-41,p 613-627, May 1995)
[5] Florens: On estimating the variance of diffusion processes (Journal of Applied Probability, 30, p 790- 804, 1993)
[6] Hamilton: Time series analysis (Princeton University Press, 94)
[7] Koepfler, Lopez, Morel: A multiscale algorithm for image segmentation by variational method. (SIAM. J. Numer. Anal, 1, p 282-299, 1994)
[8] Krolzig, Markov-switching vector autoregressions (Springer, 97)
[9] Mallat: A wavelet tour of signal processing (Academic Press, 99)
[10] Mumford, Shah: Optimal approximations by piecewise smooth functions and associated variational problems.(Communications on pure and applied mathematics, vol XLII, 4, 1989).

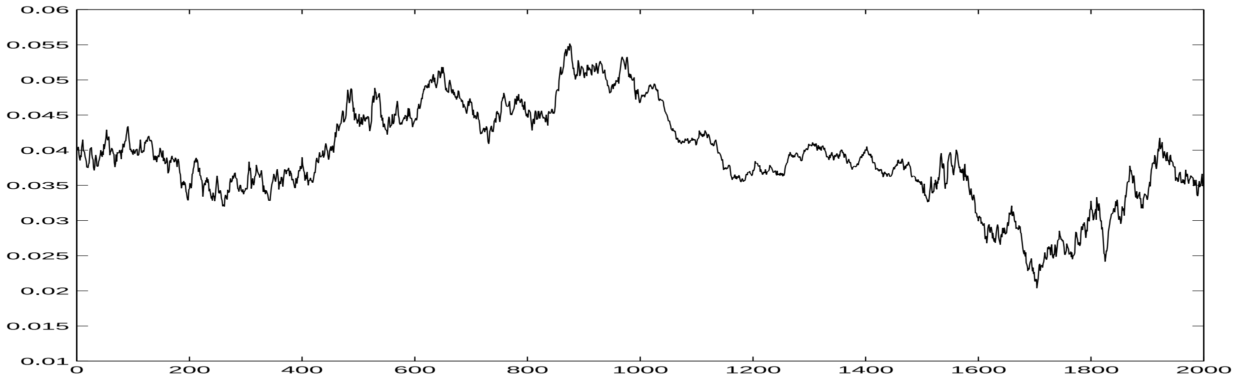

Figure 1: Simulated signal

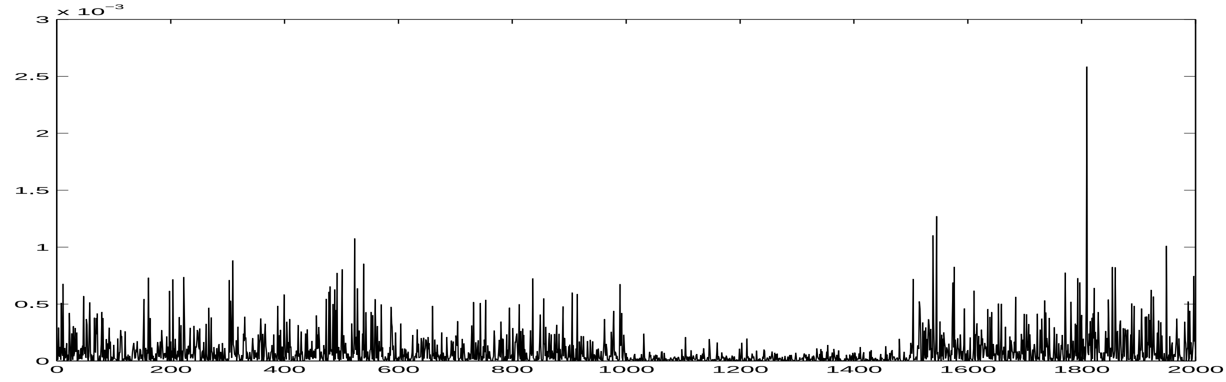

Figure 2: Quadradtic variation of signal

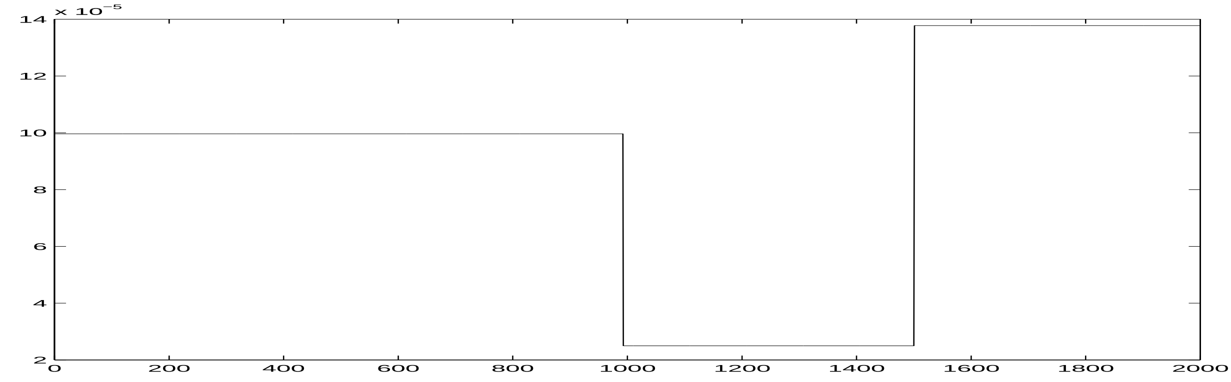

Figure 3: Segmentation in 3 regions of the quadratic variation of signal

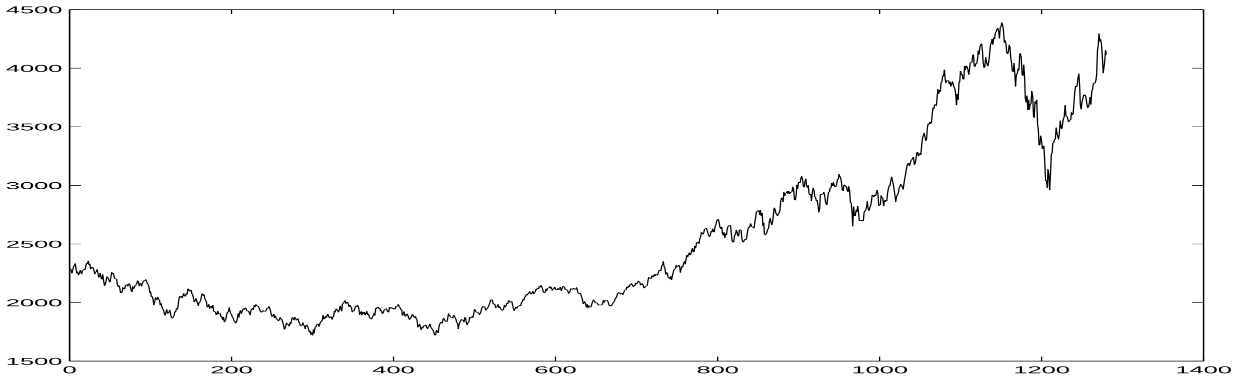

Figure 4: Real signal : Cac40 from Jan. 1994 to Jan. 1999

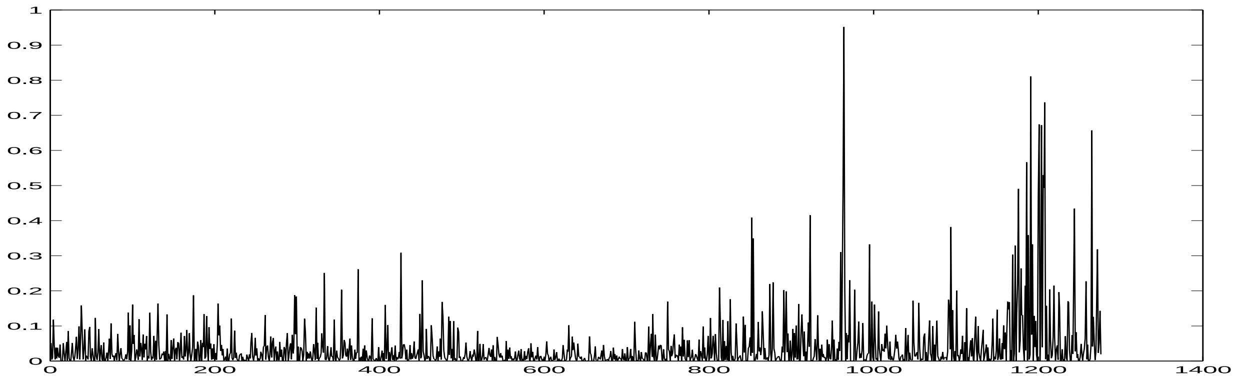

Figure 5: Quadratic variation of logprice

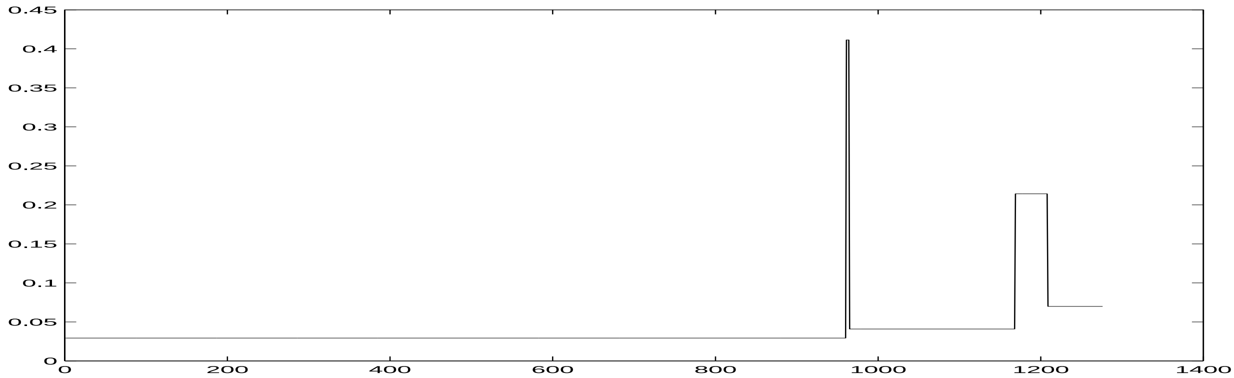

Figure 6: Segmentation in 5 regions of quadratic variation of logprice

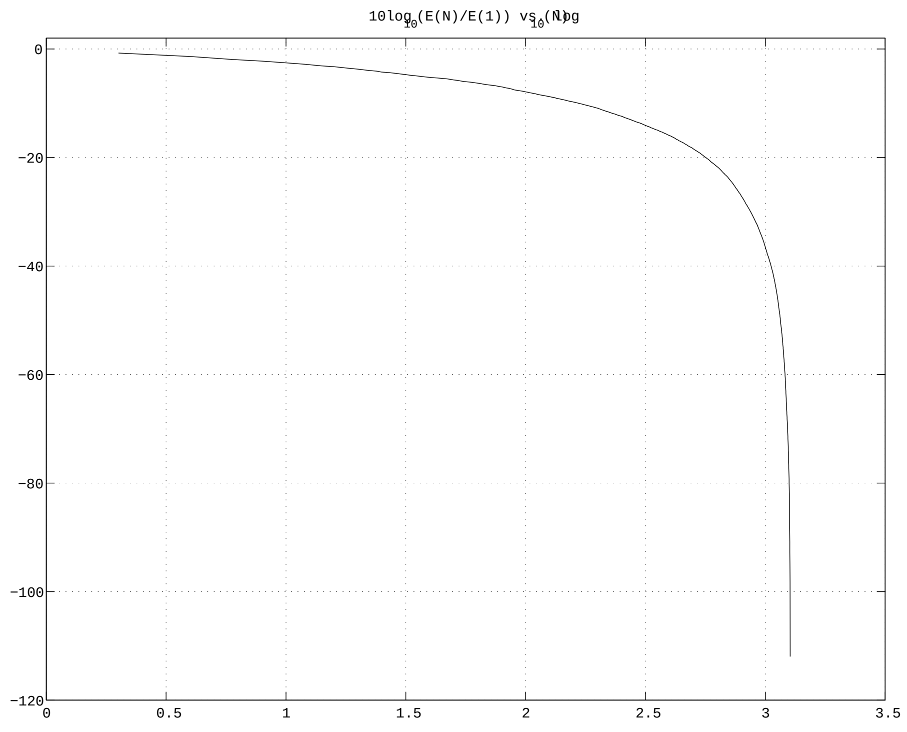

Figure 7: loglog graph for mean square error

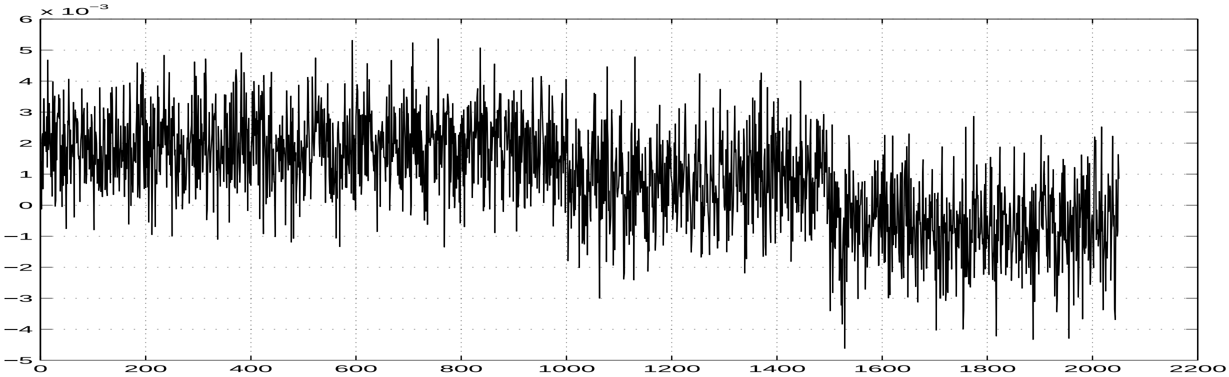

Figure 8: Discretization of increments

Figure 9: Reconstruction of smoothed wavelet transform

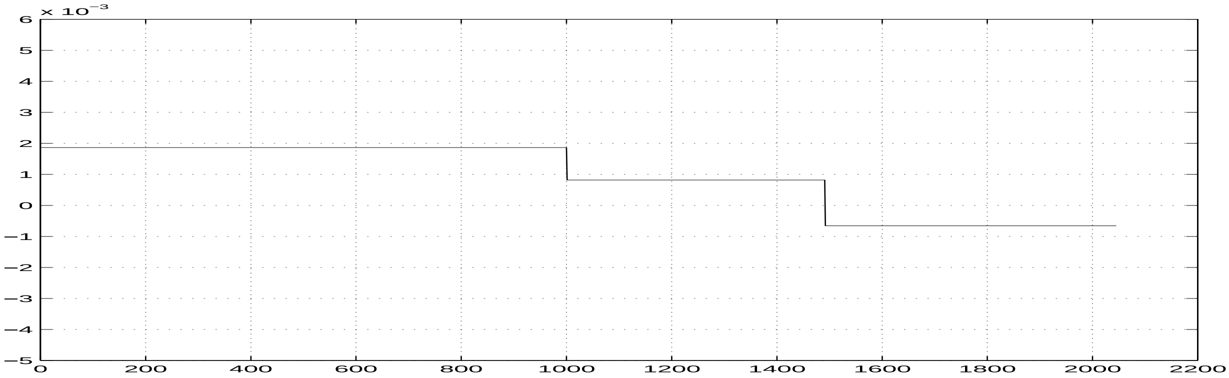

Figure 10: Segmentation of reconstruction of smoothed wavelet transform