数理解析研究所講究録 1462 巻 2006 年 1-25

Filtering volatility from data observed at random time intervals

Jakša Cvitanić *
Robert Liptser†
Boris Rozovskii ‡
Ilya Zaliapin §

October 10, 2005

AMS (2000) Subject Classifications: Primary 60G35, 91B28; secondary 62M20, 93E11.
Key Words and Phrases: nonlinear filtering, discrete observations, volatility estimation.

Abstract

We consider a continuous-time model for a stock price, which is, however, observed at discrete time intervals. The time between observations is random. We report on the formula for the optimal filter for the current value of the volatility of the stock price and we illustrate the theoretical results with a numerical example. The filter gives stable and efficient estimates of the volatility. As a preliminary step, we estimate the possible values of volatility using a variation of the Multiscale Trend Analysis (MTA) method.

*Caltech, M/C 228-77, 1200 E. California Blvd. Pasadena, CA 91125. Ph: (626) 395-1784. E-mail: cvitanic@hss.caltech.edu. Research supported in part by NSF grant DMS 04-03575.

†Department of Electrical Engineering-Systems, Tel Aviv University, 69978 Tel Aviv, Israel. E-mail:liptser@eng.tau.ac.il

‡Department of Mathematics , USC, 3620 S Vermont Ave, MC 2532, Los Angeles, CA 90089-1113. Ph: (213) 740-6117. E-mail: rozovsky@usc.edu. Research supported in part by the Army Research Office and the Office of Naval Research under the grants DAAD19-02-1-0374 and N0014-03-0027.

§Institute of Geophysics and Planetary Physics, UCLA, 3845 Slichter Hall, Los Angeles, CA 90095-1567. Ph: (310) 825-6115. E-mail: zal@ess.ucla.edu.

1 INTRODUCTION

In this paper we report on theoretical results from Cvitanić, Liptser and Rozovskii [3] and on their numerical implementation in Cvitanić, Rozovskii and Zaliapin [4]. The problem we consider is the one of estimating the current volatility value from stock price observations. The observations are discrete, possibly observed at random times. The main application we have in mind is high-frequency stock data ("tick-by-tick" data). We work in a continuous-time, Brownian motion driven model for the stock price, with stochastic volatility, independent of the driving Brownian motion process.

Related literature includes, among others, Frey and Runggaldier [7], Runggaldier [25], Elliott et al [5], Gallant and Tauchen [8], Malliavin and Mancino [20], Fouque *et al.* [6], Rogers and Zane [23], and Kallianpur and Xiong [13], Ait Sahalia and Mykland [1], Platania and Rogers [22], and Johannes and Polson [11]. There is also a rich econometrics, time-series literature on ARCH-GARCH models of stochastic volatility, presenting an alternative way to model and estimate volatility; see Gourieroux [9] for a survey.

Our work was motivated primarily by Frey and Runggaldier [7]. That paper derives a Kallianpur-Striebel type formula (see e.g. [12]) for the optimal mean-square filter of the volatility process, and investigates Markov Chain approximations for this formula. We extend this result in that we derive the exact filtering equations, which can easily be implemented.

The Frey and Runggaldier model is a natural model for stochastic volatility, but it does not quite fall in the "standard" category of diffusion or simple point processes models for which filtering results have been developed (cf. [18], [15], [24]). Thus, there was a need to develop further technical tools to deal with our problem. However, it turns out that the resulting filtering equations are simpler than in the case of continuous observations. In the latter case, the nonlinear filters are described by infinite dimensional stochastic differential equations, for example, by stochastic partial differential equations (see e.g. [24]). In contrast, in our setting, the filtering equation can be reduced to a recursive system of linked *deterministic* equations of Kolmogorov type. Moreover, at the observation times the filter is given by a simple Bayesian recursion.

In our numerical example we assume that the volatility is a Markov chain process. Before we can do the filtering, we have to decide what possible values the volatility chain can attain, and what the transition probabilities are. This preliminary stage is related to the power variation estimates of volatility, as surveyed in Barndorff-Neielsen, Graversen and Shephard [2], for example. We adapt the so-called Multiscale Trend Analysis of Zaliapin *et al.* [27], where we use a variation process to estimate possible volatility values. However, while in the power variations literature such an estimate is the final estimate of volatility, in our case it serves only as an estimate of a priori values, from which we then get a posteriori values

using filtering. Also, let us emphasize again that, unlike most of the existing work, the time intervals between observations may be random in our framework.

We show that the complete algorithm, consisting of the preliminary estimation and the filtering estimation, performs very well in a variety of circumstances, on simulated and on real data. It quickly recognizes when there is a jump in volatility value. It is also robust with respect to the given drift value, which is important, as the drift is hard to estimate in practice.

We describe the model in section 2, state the main filtering results and examples in section 3, discuss the preliminary estimation of the model parameters in section 4, the numerical implementation of the filtering formula in section 5, and the complete algorithm in section 6. We present two examples with real data in section 7.

2 THE MODEL

2.1 Observation values and observation times

We fix a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ equipped with a filtration $\mathbf{F} = (\mathcal{F}_t)_{t\geq 0}$ satisfying the "usual" conditions (see, e.g. [19]). All random processes considered in the paper are assumed to be defined on $(\Omega, \mathcal{F}, \mathsf{P})$ and adapted to $\mathbf{F}$.

We consider a stock price process $S = (S_t)_{t\geq 0}$ given by the Itô equation

$$dS_t = r(\theta_t)S_t dt + v(\theta_t)S_t dB_t \tag{2.1}$$

where $B = (B_t)_{t\geq 0}$ is a standard Brownian motion and $\theta = (\theta_t)_{t\geq 0}$ is a cádlág Markov jump-diffusion process in $\mathbb{R}$ with the generator $\mathcal{L}$. For the sake of simplicity, we assume that $r(x)$ and $v(x)$ are measurable bounded functions on $\mathbb{R}$, the initial condition S_0 is constant, and $v(x)$ and S_0 are positive.

The process $(\theta_t)_{t\geq 0}$ is referred to as the *volatility process*. It is unobservable, and the only observable quantities are the values of the log-price process $X_t = \log S_t$ taken at stopping times $(T_k)_{k\geq 0}$, so that $T_0 = 0, T_k < T_{k+1}$ if $T_k < \infty$, and $T_k \uparrow \infty$ as $k \uparrow \infty$.

According to (2.1), the log-price process is given by

$$X_t = \int_0^t \Big(r(\theta_s) - \frac{1}{2}v^2(\theta_s)\Big)ds + \int_0^t v(\theta_s)dB_s.$$

We use the abbreviated notation $X_k := X_{T_k}$. Thus, the observations are given by the sequence $(T_k, X_k)_{k\geq 0}$. The observation process $(T_k, X_k)_{k\geq 0}$ is a multivariate (marked) point process (see, e.g. [10], [16]) with the counting measure

$$r(dt, dy) = \sum_{k\geq 1} \mathrm{I}_{\{T_k<\infty\}}\delta_{\{T_k, X_k\}}(t, y)dtdy,$$

where $\delta_{\{T_k,X_k\}}$ is the Dirac delta-function on $\mathbb{R}_+ \times \mathbb{R}$.

We introduce two filtrations related to $(T_k, X_k)_{k\geq 0}$: $(\mathcal{G}(n))_{n\geq 0}$ and $(\mathcal{G}_t)_{t\geq 0}$, where

- $\mathcal{G}(n) := \sigma\{(T_k, X_k)_{k\leq n}\}$,
- $\mathcal{G}_t := \sigma(r([0,s]\times\Gamma) : s \leq t, \Gamma \in \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

It is a standard fact (see III.3.31 in [10])

$$\mathcal{G}_{T_k} = \mathcal{G}(k), \ k = 0, 1 \ldots, \tag{2.2}$$

and $\{T_k\}$ is a system of stopping times with respect to $(\mathcal{G}_t)_{t\geq 0}$.

Remark 2.1 The filtration $(\mathcal{G}_t)_{t\geq 0}$ provides more information than the filtration $\mathcal{G}_{T_k}$, namely it provides additional information about the duration between the observation times.

The paper Cvitanić, Liptser and Rozovskii [3] works out the filtering formula for general observation times, but here, for the simplicity of presentation, we will assume the following:

Assumption 2.1 The observation times $(T_k)_{k\geq 0}$ are either:

(i) the jump times of a doubly stochastic Poisson process (Cox process) with the intensity $n(\theta_t)$, or

(ii) $T_k = k\delta$, that is, the observation times are deterministic, with constant length δ of interarrival intervals.

2.2 Volatility Process

We now specify more precisely the volatility process. Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $(\mathbb{R}_+ \times \mathbb{R}, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}))$ be measurable spaces with Borel σ-algebras. The volatility process $\theta = (\theta_t)_{t\geq 0}$ is defined by the Itô equation

$$d\theta_t = b(t,\theta_t)dt + \sigma(t,\theta_t)dW_t + \int_{\mathbb{R}} u(\theta_{t-}, x)(\mu^\theta - \nu^\theta)(dt, dx), \tag{2.3}$$

where W_t is a standard Wiener process and $\mu^\theta = \mu^\theta(dt, dx)$ is a Poisson measure on $(\mathbb{R}_+ \times \mathbb{R}, \mathcal{B}(\mathbb{R}_+) \otimes \mathcal{B}(\mathbb{R}))$ with the compensator $\nu^\theta(dt, dx) = K(dx)dt$, where $K(dx)$ is a $\sigma-$finite non-negative measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We assume that $E\theta_0^2 < \infty$, the functions $b(t,z), \sigma(t,z)$, and $u(z,x)$ are Lipschitz continuous in z uniformly with respect to other variables, and

$$|b(t,z)| + |\sigma(t,z)|^2 + \int_{\mathbb{R}} |u(z,x)|^2 K(dx) \leq C(1+|z|^2).$$

It is well known that under these assumptions (2.3) possesses a unique strong solution adapted to $\mathbf{F}$, and $E\theta_t^2 < \infty$ for any $t \geq 0$.

The generator $\mathcal{L}$ of the volatility process is given by

$$\mathcal{L}f(x) := b(t,x)f'(x) + \frac{1}{2}\sigma^2(t,x)f''(x) + \int_{\mathbb{R}} \Big(f(x+u(x,y)) - f(x) - f'(x)u(x,y)\Big)K(dy).$$

Before proceeding with the assumptions and main results we shall introduce additional notation. Set

$$m(s,t) = \int_s^t \left(r(\theta_u) - \frac{1}{2}v^2(\theta_u)\right) du, \tag{2.4}$$

and

$$\sigma^2(s,t) = \int_s^t v^2(\theta_u)du \ . \tag{2.5}$$

For simplicity, it is assumed that $v^2(s,t)$ is bounded away from zero. Let us denote by $\rho_{s,t}(y)$ the density function of the normal distribution with mean $m(s,t)$ and the variance $\sigma^2(s,t)$:

$$\rho_{s,t}(y) := \frac{1}{\sqrt{2\pi}\sigma(s,t)} e^{-\frac{(y-m(s,t))^2}{2\sigma^2(s,t)}} \tag{2.6}$$

Clearly, ρ is the conditional density of the stock's log-increments $X_t - X_s$ given θ.

Let $\mathcal{F}^\theta = (\mathcal{F}_t^\theta)_{t\geq 0}$ be the right-continuous filtration generated by $(\theta_t)_{t\geq 0}$ and augmented by P-zero sets from $\mathcal{F}$. Denote by G_k^θ a regular version of the conditional distribution of $T_{k+1}$with respect to* $\mathcal{F}^\theta \vee \mathcal{G}(k)$. That is, G_k^θ is the distribution of the time of the next observation, given previous history, and given θ.

Let $N = (N_t)_{t\geq 0}$ be the counting process with interarrival times $(T_k - T_{k-1})_{k\geq 1}$, that is

$$N_t = \sum_{k\geq 1} I(T_k \leq t) \tag{2.7}$$

We also assume

Assumption 2.2 The Brownian motion B is independent of (θ, N).

3 FILTERING RESULTS

3.1 The main result

For a measurable function f on $\mathbb{R}$ such that $E|f(\theta_t)| < \infty$, define the conditional expectation estimator $\pi_t(f)$ by

$$\pi_t(f) := E\big(f(\theta_t)|\mathcal{G}_t\big) = \int_{\mathbb{R}} f(z)\pi_t(dz), \tag{3.1}$$

*Here and below $\mathcal{F}^1 \vee \mathcal{F}^2$ stands for the σ-algebra generated by the $\sigma-$algebras $\mathcal{F}^1$ and $\mathcal{F}^2$.

where $\pi_t(dz) := dP(\theta_t \leq z|\mathcal{G}_t)$ is the filtering distribution. (Note that we omit the argument θ_t of f in the estimator $\pi_t(f)$). As in the Bayesian framework, we suppose that the a priori distribution $\pi_0(dx) = P(\theta_0 \in dx)$ is given.

Let $\sigma\{\theta_{T_k}\}$ be the σ-algebra generated by θ_{T_k}. For $t > T_k$, let us define the following *structure functions*:

$$\psi_k(f;t,y,\theta_{T_k}) := E\Big(f(\theta_t)\rho_{T_k,t}(y - X_k)\phi(T_k,t)\big|\sigma\{\theta_{T_k}\} \vee \mathcal{G}_{T_k}\Big) \tag{3.2}$$

and its integral with respect to y

$$\bar{\psi}_k(f;t,\theta_{T_k}) := \int_{\mathbb{R}} \psi_k(f;t,y,\theta_{T_k})\,dy = E\Big(f(\theta_t)\phi(T_k,t)\big|\sigma\{\theta_{T_k}\} \vee \mathcal{G}_{T_k}\Big) \tag{3.3}$$

where ρ is given by (2.6) and

$$\phi(t) = n(\theta_t)\exp\left(-\int_{T_k}^{s} n(\theta_u)\,du\right) \text{ if Assumption 2.1 (i) holds} \tag{3.4}$$

$$\phi(t) = 1 \text{ if Assumption 2.1 (ii) holds.} \tag{3.5}$$

If $f \equiv 1$, the argument f in ψ and $\bar{\psi}$ is replaced by 1.

For $t \geq T_k$ and a bounded function f , define

$$\mathcal{M}_k(f;t,\pi_t) := \frac{\pi_{T_k}\left(\bar{\psi}_k(f;t)\right) - \pi_{t-}(f)\pi_{T_k}\left(\bar{\psi}_k(1;t)\right)}{\int_t^\infty \pi_{T_k}\left(\bar{\psi}_k(1;s)\right)ds}.$$

The main filtering result from Cvitanić, Liptser and Rozovskii [3] is (specialized to Assumption 2.1):

Theorem 3.1 *Under our assumptions, for every measurable bounded function f in the domain of the generator $\mathcal{L}$ such that $\int_0^t E|\mathcal{L}f(\theta_s)|ds < \infty$ for any $t \geq 0$, the following system of equations holds:*

1) For every $k = 0, 1\ldots$, at the observation times we have

$$\pi_{T_{k+1}}(f) = \frac{\pi_{T_k}(\psi_k(f;t,y))}{\pi_{T_k}(\psi_k(1;t,y))}\Bigg|_{\left\{\substack{t=T_{k+1}\\ y=X_{k+1}}\right\}} \tag{3.6}$$

Under Assumption 2.1 (i), we have between observation times:

2) For every $k = 0, 1\ldots$ and $t \in]\!]T_k, T_{k+1}[\![$,

$$d\pi_t(f) = \pi_t(\mathcal{L}f)dt - \mathcal{M}_k(f;t,\pi_t)\,dt. \tag{3.7}$$

Under Assumption 2.1 (ii), the second term is zero, that is, we have:

2) For every $k = 0, 1\ldots$ and $t \in]\!]T_k, T_{k+1}[\![$,

$$d\pi_t(f) = \pi_t(\mathcal{L}f)dt \tag{3.8}$$

Remark 3.1 Note that for high-frequency observations, it may be satisfactory to compute the volatility estimate only at price observation times. In that case we only need to use the relatively simple Bayes-type recursion formula (3.6), and not the differential equation (3.7) or (3.8).

Remark 3.2 Clearly, the "structure functions" ψ and $\bar{\psi}$ are of paramount importance for computing the posterior distribution of the volatility process. We would like to stress that these do not involve the observations and could be pre-computed "off-line" using only the a priori distribution. Then, "on-line", when the observations become available, one needs only to plug in the obtained measurements (T_k, X_k). This is important for developing efficient numerical algorithms.

Remark 3.3 Note that for almost every $\omega \in \Omega$, filtering equation (3.7) is a deterministic equation of Kolmogorov's type, rather then a stochastic partial differential equation arising in nonlinear filtering of diffusion processes. The well-posedness and regularity of such equations is well researched in the literature on second order parabolic deterministic integro-differential equations (see e.g. [17]i, [21], [14] and the references therein).

3.2 The case of the Markov chain volatility process

In this section we specialize our formulas to the case where the volatility process is modeled by a continuous time Markov chain.

We assume that the counting process is a Cox process with intensity $n(\theta_t)$, and that $\theta = (\theta_t)_{t \leq T}$ is a homogeneous Markov jump process taking values in the finite alphabet $\mathcal{A} = \{a_1, \ldots, a_M\}$ with the intensity matrix $\Lambda = (\lambda(a_i, a_j)) = (\lambda_{ij})$ and the initial distribution $p_q = P(\theta_0 = a_q)$, $q = 1, \ldots, M$. (This is one of the two models of the state process discussed in [7].) In this case,

$$\mathcal{L}f(\theta_s) = \sum_j \lambda(\theta_s, a_j) f(a_j) \quad .$$

Denote by θ_t^j the process θ_t starting from a_j, and

$$p_{ji}(t) := P(\theta_t = a_i | \theta_0 = a_j) \ , \ \pi_j(t) = P(\theta_t = a_j | \mathcal{G}_t) \ ,$$

$$r_{ji}(t, z) := E\big(e^{-\int_0^t n(\theta_u^j) du} \rho_{0,t}^j(z) | \theta_t^j = a_i\big) \ ,$$

where $\rho_{0,t}^j(z)$ is obtained by substituting θ_s^j for θ_s in $\rho_{0,t}(z)$. It follows from Theorem 3.1 (for details see Cvitanić, Liptser and Rozovskii [3]), with $f(\theta_t) := I_{\{\theta_t = a_i\}}$, that

$$\pi_i(T_k) = \frac{n(a_i) \sum_j r_{ji}(T_k - T_{k-1}, X_k - X_{k-1}) p_{ji}(T_k - T_{k-1}) \pi_j(T_{k-1})}{\sum_{i,j} n(a_i) r_{ji}(T_k - T_{k-1}, X_k - X_{k-1}) p_{ji}(T_k - T_{k-1}) \pi_j(T_{k-1})}. \tag{3.9}$$

This recursion can be easily computed.

4 Numerical implementation

In this section we consider numerical implementation of the Markov chain example from the previous section. We will estimate the a priori parameters of the chain, and then we use the filtering formulas. For simplicity, we set

$$v_t = v(\theta_t) = \theta_t.$$

4.1 Discrete approximation of v_t

We now construct a natural discrete-time Markov process approximation d_n of the volatility process v_t, with values from the alphabet $\{a_i\}_{i=1,\dots,M}$. We fix a small discrete step Δ and define the transition probability matrix $Q = (Q_{ij})_{i,j=1,\dots,M}$ for the process d_n as

$$Q_{ij} = \begin{cases} \lambda_{ij}\,\Delta, & i \neq j \\ 1 - \sum_{i \neq k} \lambda_{ik}\,\Delta, & i = j. \end{cases} \tag{4.1}$$

Here the step Δ is chosen such that $\Delta \sum_{ij} \lambda_{ij} < 1$. The finite-dimensional distributions of the process d_n converge to that of v_t as $\Delta \to 0$.

The probabilities $p_{ji}(t) = P(v_t^j = a_i)$ are estimated using the corresponding probabilities for the discrete process d_n:

$$\widehat{p}_{ji}(t) = P(d_{m_t} = a_i \,|\, d_0 = a_j) = [e_j \times Q^{m_t}](i), \tag{4.2}$$

where $m_t = \left\lfloor \frac{t}{\Delta} \right\rfloor$, e_j denotes a row-vector of length M with all zeros except for the value one at the j-th position, $[v](i)$ is the i-th element of vector v, and $\lfloor x \rfloor$ is an integer closest to x from below.

The process $(v_s^j \,|\, v_t^j = a_i)$ on $[0,\, t)$ is approximated by its discrete counterpart $(d_n \,|\, d_0 = a_j,\, d_{m_t} = a_i)$ on $[0,\, m_t)$. The one-step conditional transitional probabilities for the latter process are given by

$$\begin{aligned} &P(d_n = a_k \,|\, d_{n-1} = a_{k'}, d_{m_t} = a_i) = \\ &\quad = \frac{P(d_n = a_k \,|\, d_{n-1} = a_{k'})\, P(d_{m_t} = a_i \,|\, d_n = a_k)}{\sum_{m=1}^{M} P(d_n = a_m \,|\, d_{n-1} = a_{k'})\, P(d_{m_t} = a_i \,|\, d_n = a_m)}. \end{aligned} \tag{4.3}$$

Here

$$P(d_n = a_k \,|\, d_{n-1} = a_{k'}) = [e_{k'} \times Q](k); \tag{4.4}$$

$$P(d_{m_t} = a_i \,|\, d_n = a_k) = [e_k \times Q^{N-n}](i). \tag{4.5}$$

The only arbitrary choice in our construction is the discrete time step Δ. To approximate v_t on $[0,\, t)$ we set

$$\Delta = \min\left\{\frac{1}{100\ \max(\lambda_{ij})},\ \frac{t}{100}\right\}$$

which ensures that we have on average no less than 100 steps of d_n within each interval of constant volatility v_t, yet no less than 100 steps within $[0,\, t)$.

4.2 Monte Carlo estimation of r_{ji}

A Monte Carlo procedure used to estimate the conditional expectation r_{ji} is based on the simulations of the discrete-time process d_n defined in the previous section. Introducing the notation

$$\delta_k := (T_k - T_{k-1})\ ,\quad \Delta_k := (X_{T_k} - X_{T_{k-1}}), \tag{4.6}$$

and

$$A_k^j := \int_0^{\delta_k} \left(v_u^j\right)^2 du$$

we see that in estimating r_{ji}, we can use

$$\rho^j_{0,\delta_k}(\Delta_k) = \frac{1}{\sqrt{2\pi A_k^j}} \exp\left\{-\frac{\left(\Delta_k - r\delta_k + \frac{1}{2}A_k^j\right)^2}{2A_k^j}\right\}. \tag{4.7}$$

The only random element here is A_k^j, which can be found given a realization of v_t on $[0, \delta_k)$:

$$A_k^j := \sum_{i=1}^{N_k} a_{(i)}^2 (u_i - u_{i-1}), \tag{4.8}$$

where u_i are the times of the volatility jumps, N_k is the number of volatility jumps in the interval $[0, \delta_k)$, $v_t^j = a_{(i)}$ are the volatility values for $t \in [u_{i-1}, u_i)$ (from the alphabet $\{a_1, \ldots, a_M\}$), $u_0 = 0, u_{N_k} = \delta_k,\ a_{(1)} = a_j$. The condition $\theta_t^j = v_{\delta_k}^j = a_i$ in the definition of r_{ji} implies that $a_{(N_k)} = a_i$.

Similarly,

$$\int_0^{\delta_k} n(v_u^j) du = \sum_{i=1}^{N_k} n(a_{(i)})(u_i - u_{i-1}). \tag{4.9}$$

We estimate r_{ij} by simulating independent realizations of d_n on $[0, \delta_k)$ and using equations (4.8) and (4.9) with $\{v_t^j\}$ replaced by $\{d_n \,|\, d_0 = a_j\}$.

5 Estimating a priori values of the filter parameters

We now consider the problem of estimating a priori values of the filter parameters — volatility alphabet $\mathcal{A}$, jump intensities Λ, initial probabilities p_i, and observation intensities $\mathcal{N} = n(a_i)$, $(i, j = 1, \ldots, M)$ from observations X_{T_k}.

The idea is to find a process P_t such that

$$\Delta P_t \approx a\, v_t\, \Delta t, \tag{5.1}$$

for small Δt. The estimation of piece-wise constant volatility v_t is then equivalent to finding the optimal piece-wise linear approximation $L(t)$ to the process P_t. Distinct slopes of $L(t)$ will correspond to distinct volatility values; and the rest of the parameters can also be estimated using $L(t)$. Such a problem can be effectively solved by the Multiscale Trend Analysis (MTA) of [27].

5.1 Volatility alphabet

Consider the process P_t defined as the sum of the absolute returns between the times T_k:

$$P_t := \sum_{k:T_k<t} |\Delta_k|, \tag{5.2}$$

where $\Delta_k := X_{T_k} - X_{T_{k-1}}$. The alphabet estimation procedure is based on the following result (see Cvitanić, Rozovskii and Zaliapin [4]):

Proposition 5.1 *Suppose that the volatility v and the intensity n of observations are constant within the interval $[0, t]$.*

(i) If Assumption 2.1 (i) holds, then

$$\frac{P_t}{t\sqrt{n}} - \frac{v}{\sqrt{2}} \xrightarrow{\text{a.s.}} 0, \ \text{as } n \to \infty. \tag{5.3}$$

(i) If Assumption 2.1 (ii) holds, then

$$\frac{\sqrt{\delta}\, P_t}{t} - v\sqrt{\frac{2}{\pi}} \xrightarrow{\text{a.s.}} 0, \ \text{as } \delta \to 0. \tag{5.4}$$

Remark 5.1 The proposition is also true for intervals of the form $[t_1, t_2]$. Thus, if volatility v_t is piece-wise constant with values from the alphabet $\mathcal{A}$, and the observational intensity $\mathcal{N}$ is a function of volatility, then P_t is asymptotically a piece-wise linear function with slopes, in case (i),

$$s_i = s_i(a_i) = a_i\sqrt{\frac{n_i}{2}} \tag{5.5}$$

within the respective intervals, and with slopes, in case (ii),

$$s_i = s_i(a_i) = a_i\sqrt{\frac{2}{\pi\,\delta}}. \tag{5.6}$$

Remark 5.2 Barndorff-Neielsen, Graversen and Shephard [2] showed that if X_t is a Brownian seminartingale with v_t being a càdlàg process, and observations are made on a regular grid with a fixed step δ, then under some mild conditions on v_t

$$\sqrt{\delta}P_t \xrightarrow{P} \sqrt{\frac{2}{\pi}}\int_0^t v_s ds\ , \quad \text{as } \delta \to 0. \tag{5.7}$$

If we consider a piece-wise linear process $L(t)$ with slopes defined as in (5.5), (5.6), then the distinct volatility values a_i are uniquely determined by M distinct slopes of $L(t)$. Below we will use observations to approximate the asymptotic piece-wise linear structure of P_t. If this approximation has N_L distinct linear segments and the observations form a Poisson process, then according to (5.5) the distinct volatility values can be estimated as

$$\widetilde{a}_i^{\text{Poisson}} = s_i\sqrt{\frac{2}{n_i}},\ i = 1, \ldots, N_L. \tag{5.8}$$

In case of regular observational grid with step δ we similarly obtain using (5.6)

$$\widetilde{a}_i^{\text{Regular}} = s_i\sqrt{\frac{\pi\,\delta}{2}},\ i = 1, \ldots, N_L. \tag{5.9}$$

From (5.8),(5.9) one obtains a piece-wise constant volatility estimate $\widetilde{v}_t$ with N_L distinct values $\tilde{a}_i$. If the piece-wise linear approximation $L(t)$ is close to the piece-wise linear limit of P_t, the estimators $\widetilde{a}_i$ should have a multi-modal distribution with each mode corresponding to a single value of the true alphabet $\mathcal{A}$. To estimate the size M of the alphabet as well as its elements, the estimated volatility values $\widetilde{a}_i$, $i = 1, \ldots, N_L$, should be appropriately binned into $\widehat{M} \leq N_L$ groups $\{\widehat{a}_i\}_{i=1,\ldots,\widehat{M}}$. We denote this grouped volatility estimate by $\widehat{v}_t$. (Our binning procedure is somewhat ad-hoc.)

Note that parameters $n_i, i = 1, \ldots, N_L$, in (5.8), should also be estimated from the data. Suppose that i-th segment of $L(t)$ has duration T_i and includes m_i observations. A natural estimate of $n_{(i)}$, $i = 1, \ldots, N_L$, within the i-th segment of $L(t)$ is

$$\widehat{n}_{(i)} = \frac{m_i}{T_i}. \tag{5.10}$$

Below we use this expression to obtain initial estimates $\widetilde{a}_i$, $i = 1, \ldots, N_L$, of the alphabet values.

The main problem in constructing $L(t)$ is that we do not know a priori the intensity of volatility jumps, which would give an estimate of the number of linear segments within $L(t)$

(while the problem of constructing an optimal piece-wise linear approximation with given number of segments is well-studied). Thus, we have to resolve the tradeoff between the detail and the quality of the piece-wise linear approximation $L(t)$. In general, we want the alphabet $\{a_i\}$ (the number of distinct slopes) to be as small as possible while the approximation $L(t)$ be as close to P_t as possible; and these two goals contradict each other. This tradeoff can be effectively resolved and the approximation $L(t)$ constructed by the Multiscale Trend Analysis of [27].

5.2 Initial probabilities, observation intensities, and jump intensities

Let m_{ij} $(i,j = 1,\ldots,\widehat{M})$ denote the number of observation epochs T_k such that $\hat{v}_{T_k} = a_j$ and $\hat{v}_{T_{k-1}} = a_i$:

$$m_{ij} = \sum_{k=2}^{N} \delta(\hat{v}_{T_k} - a_j,\, \hat{v}_{T_{k-1}} - a_i),$$

where $\delta(\cdot,\cdot)$ is a discrete delta-function. Similarly we define

$$T_{ij} = \sum_{k=2}^{N} (T_k - T_{k-1})\, \delta(\hat{v}_{T_k} - a_j,\, \hat{v}_{T_{k-1}} - a_i).$$

The initial probabilities $p_i = P(v_0 = \widehat{a}_i)$, off-diagonal jump intensities $\{\lambda_{ij}\}$, $i \neq j$, and observation intensities $n_i = n(a_i)$ are estimated as

$$\widehat{p}_i = \frac{\sum_k T_{ik}}{\sum_j \sum_k T_{jk}}, \quad i = 1,\ldots,\widehat{M}, \tag{5.11}$$

$$\widehat{\lambda}_{ij} = \frac{T_{ii} + T_{ij}}{m_{ij}}, \quad i,j = 1,\ldots,\widehat{M},\ i \neq j, \tag{5.12}$$

$$\widehat{n}_i = \frac{\sum_k m_{ik}}{\sum_k T_{ik}}, \quad i = 1,\ldots,\widehat{M}. \tag{5.13}$$

After that, the diagonal jump intensities are estimated as

$$\widehat{\lambda}_{ii} = -\sum_{k \neq i} \widehat{\lambda}_{ik}, \quad i = 1,\ldots,\widehat{M}.$$

Remark 5.3 We introduced two different estimators for observation intensity n_i given by Eqs. (5.10) and (5.13). The estimate (5.10) is preliminary, it gives N_L estimated values of intensity, each corresponding to one segment of the piece-wise linear approximation $L(t)$. This is necessary to obtain a preliminary alphabet estimate $\{\widetilde{a}_i\}, i = 1,\ldots,N_L$. On the other hand, the final expression (5.13) produces $\widehat{M}$ estimated values using the posterior coarse alphabet $\{\widehat{a}_i\}$.

5.3 MTA method

Multiscale Trend Analysis (MTA) is a set of applied statistical techniques for time series analysis that operate with trends — local linear approximations — of the series $X(t)$ at different scales [27]. Formally, the time series $X(t)$ observed at finite (regular or irregular) time grid $\{t_i\}_{i=1}^N$ is represented by a tree M_X, whose nodes correspond to linear trends within $X(t)$. On average, the longer the trend, the higher the corresponding node in the tree. The root corresponds to the global linear approximation $L_0(t)$, the leaves to the elementary linear segments within $[t_i, t_{i+1}]$, and each internal node to some appropriately chosen trend on an intermediate scale.

One can use MTA tree to construct a set of piece-wise linear approximations $L_k(t)$, $k = 1, \ldots, d$, of $X(t)$ with increasing detail. It was shown in [27] that for a self-affine random walk with Hurst exponent H the fitting error E_k (in L^2) of such approximations is related to the number N_k of their linear segments as $E_k = E_0 N_k^{-2H}$. In general, the MTA *spectrum* — a graph showing E_k as a function of N_k — is a very useful tool for studying scaling properties of $X(t)$. In particular, it can be used to detect the change of self-affine scaling (for example, change of H with time or with analysis resolution). Here, we will apply MTA to the process P_t of (5.2). Noticeably, a typical P_t trajectory that corresponds to a Markov volatility model is not a pure self-affine series. The volatility jumps create a characteristic scale. Accordingly, the MTA spectrum is governed by the volatility structure while we consider approximations with long trends (longer than the average duration of intervals of constant volatility); and by pure Brownian motion at short trends. As a result, a typical MTA spectrum for the observed trajectories is characterized by a corner point k_0, at which the spectrum slope breaks from some $|s| > 1$ to $|s| = 1$; the latter corresponding to a pure Brownian walk ($H = 1/2$).

Here we illustrate the alphabet estimation procedure using an example with $r = 0.05$, two-valued volatility alphabet $\{\sqrt{2r}, 2\sqrt{2r}\} \approx \{0.316, 0.632\}$, transition intensities $\lambda_{12} = \lambda_{21} = 1$, observational intensities $n_i = 10^3$, and initial probabilities $p_i = 1/2$. A realization of the process X_t is shown in Fig. 1a; the shaded areas depict intervals with $v_t = a_1$. Figure 1b shows the process P_t, which indeed captures the time-dependent volatility structure. For visual convenience, we show here the detrended process $\widehat{P}_t$, since the monotonicity of P_t makes it difficult to distinguish between its global upward trend and piece-wise linear segments we are interested in. The piece-wise linear structure of P_t prominently overcomes the stochastic noise unavoidably present in P_t.

Next we apply the MTA to construct the set of piece-wise linear approximations $L_k(t)$ for P_t. The corresponding MTA spectrum is shown in Fig. 2. The relation $E_k = E_0/N_k$ is clearly observed for $N_k > 40$. For $N_k \leq 20$ the spectrum deviates from this line depicting presence of a non-random structure within P_t. The transition between the two scaling regimes occurs

within the interval between $N_k = 22$ and $N_k = 42$, which we denote in the figure as the corner point 1 and 2 respectively. The first corner point corresponds to the MTA level $k = 13$, the second to $k = 25$. To depict the piece-wise linear structure of P_t we first consider its piece-wise linear approximation $L_{13}(t)$ at the level $k = 13$ of the MTA decomposition; that is at the corner point 1 of the MTA spectrum (see Fig. 2). The approximation $L_{13}(t)$ is shown in Fig. 1b together with the original process P_t; recall that we extracted the global trend of P_t from both the functions. One can see that MTA correctly depicted all the major linear segments that correspond to the intervals of constant volatility.

Next we estimate the volatility alphabet using the formula (5.8); the raw estimate $\widetilde{v}_t$ is shown in Fig. 1c; the true volatility values are depicted by dashed horizontal lines. The distribution of distinct values of $\widetilde{v}_t$ is shown on the right in Fig. 1c: the bimodal structure of the distribution is obvious. The estimates $\widehat{a}_i$ of the alphabet values are obtained as the averages of $\widetilde{a}_i$ within the distinct modes. The resulting alphabet is $\{0.323, 0.647\}$, which is within 3% relative error of the true values. Next, we distribute the raw estimates $\widetilde{a}_i$ into the two bins to obtain the resulting estimate $\widehat{v}_t$ shown in panel d; indeed it is almost perfect, missing only one very short volatility interval at $t \approx 15$.

The initial probabilities are estimated as $\widehat{p}_1 = 0.56$ and $\widehat{p}_2 = 0.44$. The jump intensities as $\widehat{\lambda}_{12} = 0.97$ (2% relative error), $\widehat{\lambda}_{21} = 1.03$ (3%). The observation intensities as $\widehat{n}_1 = 985.2$ (1%), $\widehat{n}_2 = 1012.1$ (1%). These estimations are very stable with respect to choosing a particular corner point; for example, they remain within 3% relative error if we choose any point between $k = 13$ and $k = 25$.

6 The combined algorithm

Here is a description of the complete algorithm:

Input: Asset's log-prices $X(T_k)$, $T_k \leq T$.

Step 1. Estimate volatility alphabet.

1.1 Construct the process P_t of Eq. (5.2).

1.2 Construct the MTA decomposition M_P of the process P_t and find MTA spectrum (N_k, E_k), $k = 1, \ldots, d$.

1.3 Select a corner point k_0 of MTA spectrum (a point where the slope of the spectrum changes from a higher to a lower value); and consider the corresponding piece-wise linear approximation $L_{k_0}(t)$ of P_t with N_{k_0} segments.

1.4 Calculate preliminary alphabet values $\{\widetilde{a}_i\}$ applying either (5.8) and (5.10) or (5.9) to the slopes s_i, $i = 1, \ldots, N_{k_0}$ of the linear segments from $L_{k_0}(t)$.

1.5 Obtain the alphabet estimate $\widehat{M}$, $\{\widehat{a}_i\}_{i=1,\ldots,\widehat{M}}$ by binning the values $\{\widetilde{a}_i\}$ according to their multi-modal distribution.

Step 2. Estimate a priori initial probabilities using Eq. (5.11).

Step 3. Estimate a priori transitional intensities using Eq. (5.12).

Step 4. Estimate time-dependent volatility using the filter Eq. (3.9) with a priori parameters from Steps 1,2,3.

Output: Time dependent distribution $p_i(T_k)$ of volatility, $i = 1, \ldots, \widehat{M}$, $T_k \leq T$.

7 Examples

Here we apply our combined algorithm to two price series. First, we analyze the daily dynamics of General Electric shares traded at NYSE during 1962–2004. Then, we estimate the volatility of intraday trades for IBM during Nov. 1, 1990 – Jan. 11, 1991.

7.1 Daily data: General Electric

Here we estimate the volatility for General Electric company. Specifically, we consider daily closing prices provided by Wharton Research Data Services [26]. We thus assume that the observational grid is uniform with step of $\delta = 1$ day (ignoring the fact that longer intervals do exist between Fridays and Mondays as well as during holidays). The dynamics of the original prices S_t (\$/share) is shown in Fig. 4a. Below we work with the log-prices $X_t := \log_{10} S_t$. To estimate the volatility alphabet we use only the data during 1962-1998 (see Fig. 4a). MTA spectrum for the process P_t of (5.2) is shown in Fig. 4b. One sees clearly the transition from a higher absolute slope ($|s| \approx 2$) to a lower one ($|s| \approx 1$) as the number N_k of segments in our peace-wise linear decompositions increases. Transition occures within a broad interval $25 < N < 150$, which corresponds to decomposition levels $17 \leq k \leq 90$. The results of our estimation are stable with respect to particular choice of the level for analysis. Figure 4c shows the histograms of initial volatility estimates $\tilde{a}_i$ obtained at level $k = 17$. The three-modal structure with modes at about $\{0.06, 0.1, 0.15\}$ is prominent; a similar three-modal stucture is observed at level $k = 90$ (panel d). The same results are obtained at all intermediate levels $16 < k < 90$ (not shown). Thus, our analysis suggests $\widehat{M} = 3$, $\{\widehat{a}_i\} = \{0.06, 0.1, 0.15\}$, which we use to estimate initial probabilities and jump intensities:

$$\widehat{p}_i = \{0.66, 0.26, 0.07\}, \quad \widehat{\Lambda} = \begin{pmatrix} -0.66 & 0.16 & 0.50 \\ 1.21 & -1.81 & 0.60 \\ 3.58 & 2.14 & -5.72 \end{pmatrix} [1/\text{year}].$$

The above estimates are used as inputs for the filtering formula. The posterior probabilities $p_i(t)$, $i = 2, 3$, during 1998-1999 are shown in Fig. 5a. We also show for comparison the log-price X_k (panel b) and absolute returns $|\Delta_k| = |X_k - X_{k-1}|$ (panel c). During the second half of 1998 the market witnessed a significant price drop of the GE shares (panel b) associated with increased volatility nicely reflected in the dynamics of $|\Delta_t|$ (panel c). This volatility increase is captured by the posterior probabilities shown in panel a. We found (not shown) that our results are very stable with respect to the particular choice of the three-valued alphabet corresponding to the distribution of Fig. 4 c,d (say, choosing $\{\widehat{a}_i\} = \{0.05, 0.08, 0.15\}$, *etc.*).

Remark 7.1 The reader could ask why we need the filtering estimate, if we can simply use estimation based only on price variations. We do a comparison of that type in Cvitanić, Rozovskii and Zaliapin [4], showing that, in general, the filtering procedure is more stable and efficient.

7.2 Intraday data: IBM

In this section we estimate intraday volatility using the data for the IBM company during Nov. 1, 1990 – Jan. 11, 1991. We use the data prior to January 11 to estimate the filter input parameters, and then apply the filter during January 11 to estimate the volatility. The data set includes 60,328 transactions; almost all of them occur between 9:30 AM and 16:30 PM. The transaction time is reported up to a second; the average time between two consecutive transations (we call this interevent time) is 29 sec. In order to construct the process P_t we preprocessed the data in the following way. First, all interevent times T_i larger than 2 hours were replaced with random times $\tilde{T}_i$ from the empirical distribution of interevent times shorter than 2 hours. This way we removed the long gaps associated with nights, holidays, and long intraday breaks, and concentrated on the price dynamics during the business hours. Second, if several transactions with different price were reported within one second (so they have the same time tag), we separate them by 0.5 seconds; there were 6,548 such cases (10% of the data set).

The histogram of the initial alphabet estimates $\widetilde{a}_i$ (Eqs. (5.8) and (5.10)) is shown in Fig. 6. While there is no striking multimodal structure, the choice of

$$\widehat{a}_i = \{0.19, 0.33, 0.53, 0.75\}$$

seems reasonable if one wants to represent the volatility as a Markov jump process. The corresponding estimates of the filter parameters are:

$$\widehat{p}_i = \{0.51, 0.31, 0.12, 0.06\}, \quad \widehat{n}_i = \{1.62, 2.57, 3.78, 6.18\}[1/\text{min}],$$

$$\widehat{\Lambda} = \begin{pmatrix} -2.75 & 0.61 & 0.88 & 1.26 \\ 1.32 & -6.08 & 2.37 & 2.39 \\ 2.91 & 2.87 & -8.55 & 2.76 \\ 5.65 & 8.86 & 5.61 & -20.12 \end{pmatrix} \text{[1/hour]}.$$

The filtering results are illustrated in Fig. 7, where we show the estimated volatility and price of IBM shares during the morning hours on January 11, 1991. The a posteriori volatility $\widehat{v}_t$ is obtained in the following way: first, we find the expected values

$$E(v_{T_k}) := \sum_{i=1}^{\widehat{M}} p_i(T_k)\widehat{a}_i. \tag{7.1}$$

Then, we group the posterior expectations (7.1) into $\widehat{M}$ separate values

$$\widehat{v}_t := \left\{\widehat{a}_i,\, i = \text{argmin}_{k=1,\ldots,\widehat{M}} |E(v_t) - \widehat{a}_k|\right\}. \tag{7.2}$$

The filter detected four volatility bursts. Two of them (9:35AM and 11:40AM) correspond to a high trading intensity; one (9:50AM) to a rapid price increase; and one (10:40AM) to intensive price oscillations (without the net change). We see that when price changes are mild (in our example the price only changes by fixed increments of 0.125), the filter effectively uses the information on the trading intensity to make a decision about the current volatility.

References

[1] Ait Sahalia, Y. and P. Mykland (2004), Estimating Diffusions with Discretely and Possibly Randomly Spaced Data: A General Theory. *Annals of Statistics*, 32, 2186-2222.

[2] Barndorff-Nielsen, O.E., Graversen S.E. and N. Shephard (2003), Power variation & stochastic volatility: a review and some new results. *Journal of Applied Probability* 41A, 133-143.

[3] Cvitanić, J., R. Liptser, and B. Rozovskii (2004), A Filtering Approach to Tracking Volatility from Prices Observed at Random Times. Submitted.

[4] Cvitanić, J., Rozovskii, B., Zaliapin, I. (2005), Numerical estimation of volatility values from discretely observed diffusion data. Submitted.

[5] Elliott, R.J., Hunter, W.C. and Jamieson, B.M. Drift and volatility estimation in discrete time. *Jour. of Economic Dynamics & Control*, 22 (1998), 209-218.

[6] Fouque, J.-P., Papanicolaou, G. and Sircar, R., *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, (2000).

[7] Frey, R. and Runggaldier, W. A Nonlinear Filtering Approach to Volatility Estimation with a View Towards High Frequency Data. *International Journal of Theoretical and Applied Finance* 4 (2001), 199-210.

[8] Gallant, A. R., and Tauchen, G. Reprojecting Partially Observed Systems with Application to Interest Rate Diffusions. *Journal of the American Statistical Association* 93 (1998), 10-24.

[9] Gourieroux, C., *ARCH Models and Financial Applications*, Springer (1997).

[10] Jacod, J. and Shiryaev, A. N. *Limit Theorems for Stochastic Processes.* Springer-Verlag, New York, Heidelberg, Berlin, (1987).

[11] Johannes, M. and N. Polson (2003), MCMC methods for Financial Econometrics. Preprint.

[12] Kallianpur, G. and Striebel, C. Stochastic differential equations occurring in the estimation of continuous parameter stochastic processes. *Teor. Veroyatn. Primen.*, 14 (1969), 597-622.

[13] Kallianpur, G. and Xiong, J. Asset pricing with stochastic volatility. *Appl. Math. Optim.* 43 (2001), pp. 47-62.

[14] Krein, S.G. *Linear Equations in Banach Spaces.* Birkhäuser, Boston, (1982).

[15] Krylov, N.V. and Zatezalo, A. Filtering of finite-state time-non homogeneous Markov processes, a direct approach. *Applied Mathematics & Optimization* 42 (2000), 229-258.

[16] Last, G., Brandt, A. *Marked Point Processes on the Real Line: A Dynamic Approach*, Springer-Verlag, New York, 1995.

[17] Lions, J.-L. and Magenes, E. *Problèmes aux Limites Non Homogènes et Applications,* Dunod, Paris, (1968).

[18] Liptser, R.S. and Shiryaev, A.N.. *Statistics of Random Processes II. Applications*, Springer-Verlag, New York, (2000).

[19] Liptser, R.S. and Shiryayev, A.N. *Theory of Martingales.* Kluwer Acad. Publ, (1989).

[20] Malliavin, P. and Mancino, M.E. Fourier Series method for measurement of multivariate volatilities. *Finance & Stochastics* 6 (2002), 49-62.

[21] Mikulevicius, R. and Pragarauskas, H. On the Cauchy problem for certain integro-differential operators in sobolev and Hölder spaces. *Lithuanian Mathematical Journal*, 32 (1992), 238-263.

[22] Platania, A. and L.C.G. Rogers. Particle Filtering in High Frequency Data. Preprint, (2004).

[23] Rogers, L.C.G. and Zane, O. Designing and estimating models of high-frequency data. Preprint, (1998).

[24] Rozovskii, B.L. *Stochastic Evolution Systems. Linear Theory and Applications to Nonlinear Filtering*, Kluwer Acad. Publ., Dordrecht-Boston, (1990).

[25] Runggaldier, W.J. Estimation via stochastic filtering in financial market models. In : Mathematics of Finance (G.Yin and Q.Zhang eds.). Contemporary Mathematics, Vol. 351, pp.309-318. American Mathematical Society, Providence R.I., (2004) .

[26] Wharton Research Data Services available at http://wrds.wharton.upenn.edu

[27] Zaliapin, I., A. Gabrielov, and V. Keilis-Borok (2004), Multiscale Trend Analysis. *Fractals*, **12**, 275-292.

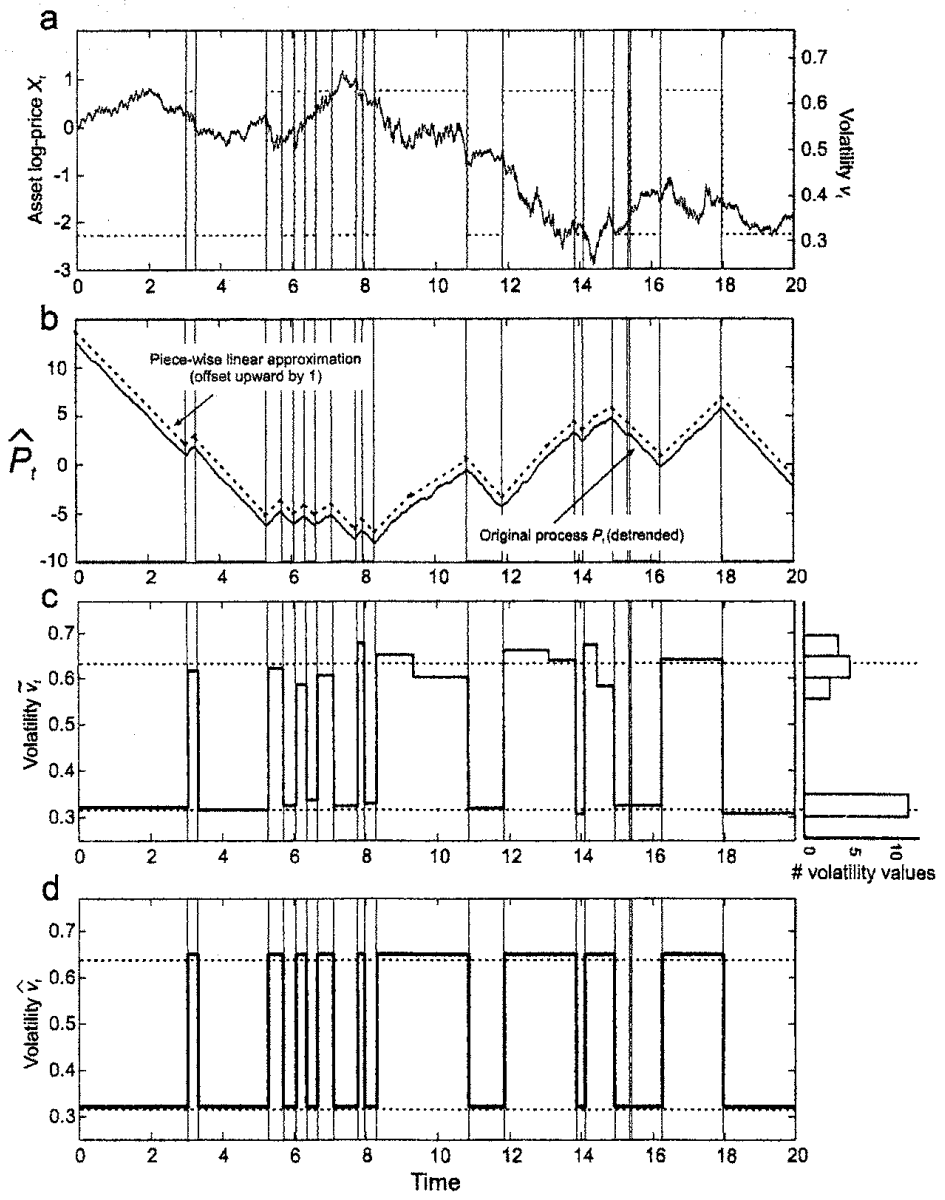

Figure 1: Example of estimating a priori values of filter input parameters. a) Asset log-price X_t (solid line, left axis) and its two-valued volatility v_t (dashed line, right axis). Parameters of the process are $\{a_i\} \approx \{0.316, 0.632\}$, $r = 0.05$, $\lambda_{12} = \lambda_{21} = 1$, $n(a_1) = n(a_2) = 10^3$, $p_i = 1/2$. b) Process P_t (solid) and its piece-wise linear approximation $L_{13}(t)$ (dashed) corresponding to the corner point 1 of MTA decomposition (see Fig. 2). The approximation is offset by 1 upward for comparison. The global linear trend of P_t is extracted from both the processes for visual convenience. c) Raw volatility estimate $\widetilde{v}_t$ (left part) and distribution of its distinct values (right part). True alphabet values are depicted by horizontal dashed lines. d) Final volatility estimate $\widehat{v}_t$. True alphabet values are depicted by horizontal dashed lines. Shaded areas in all panels correspond to intervals with $v_t = a_1 \approx 0.316$.

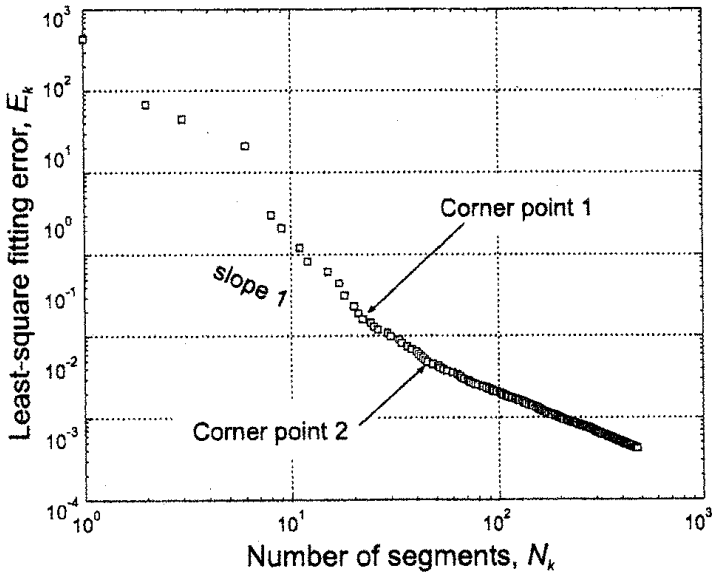

Figure 2: MTA spectrum for the process illustrated in Fig. 1a. Shaded lines depict two scaling regions with the transition zone between two *corner points* marked in the figure. The right scaling region has the slope -1, which corresponds to a self-affine random walk with no persistence. The left region deviates from this scaling depicting a non-random structure within the process P_t; this structure is due to the characteristic scales of constant volatility intervals, it can be easily seen in Fig. 1b.

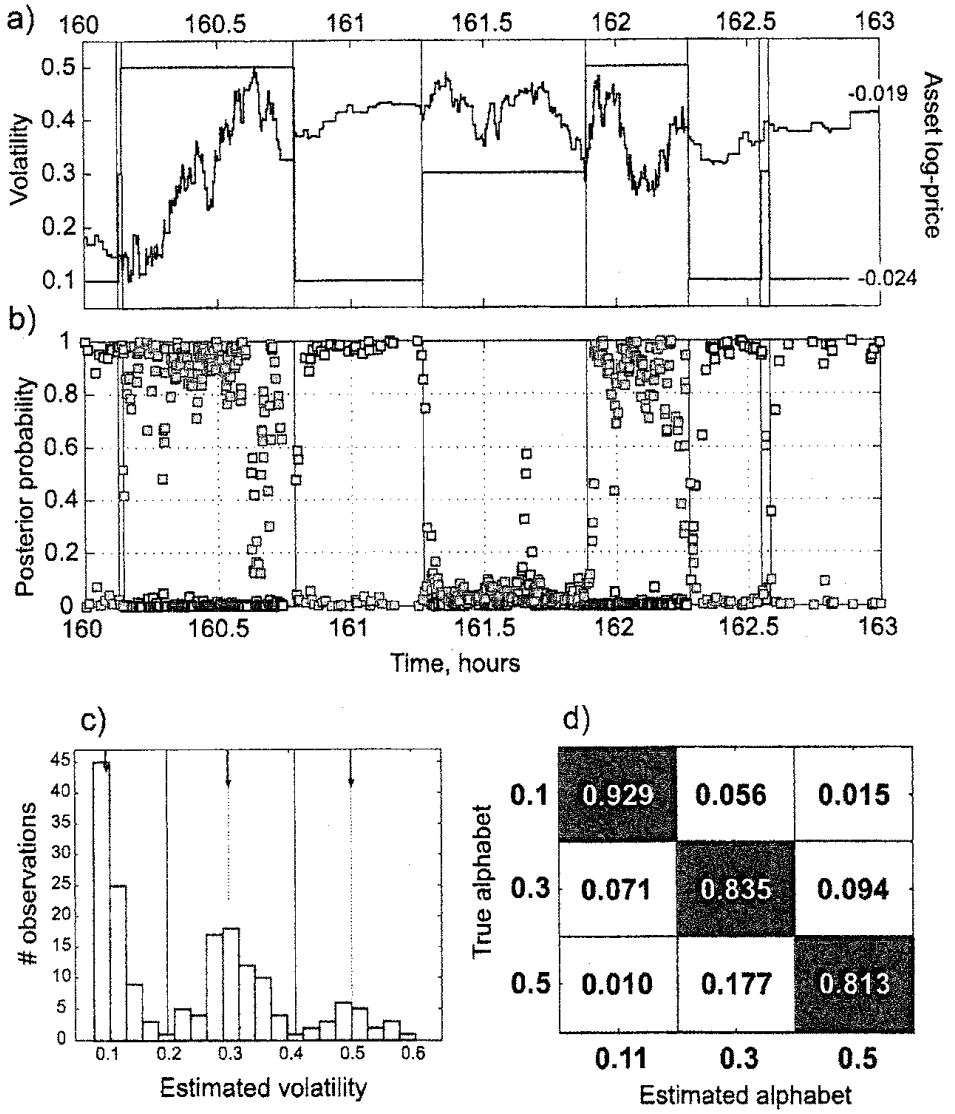

Figure 3: Filtering synthetic asset price. a) Asset log-price X_t (right axis) and true unobserved volatility v_t (left axis). Distinct volatility values are depicted by shadows: dark for $v_t = 0.5$, light for $v_t = 0.3$, none for $v_t = 0.1$. b) Aposteriori probabilities $p_3(t)$ (dark squares) and $p_1(t)$ (white squares) within the interval shown in a). c) Alphabet estimation. Histogram of initial volatility estimates $\tilde{a}_i$ clearly has a three-modal structure. Dashed lines depict true volatility values. Shadows depict three groups used to define $\widehat{a}_i$.

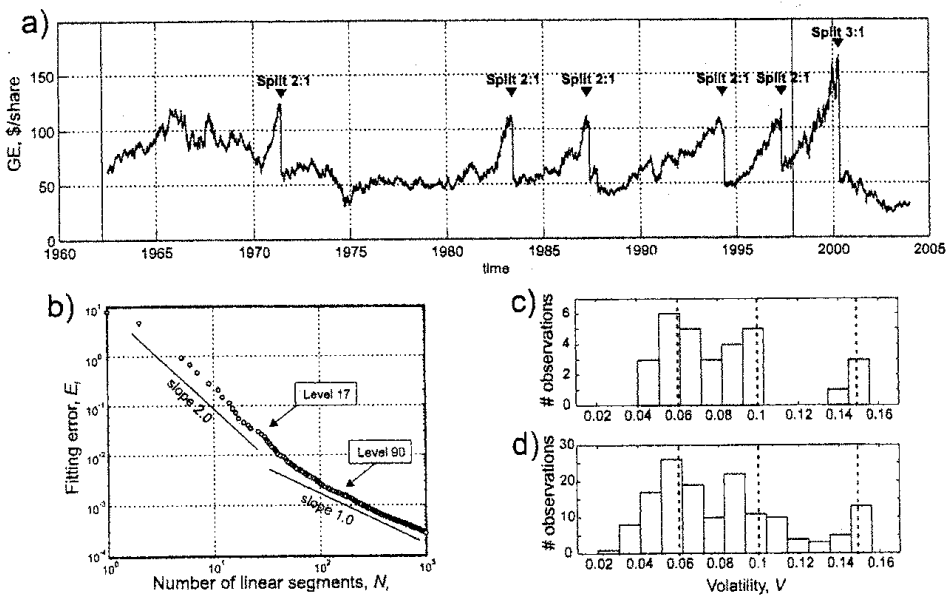

Figure 4: Estimating volatility for General Electric company during 1962-1998. a) Asset price S_t during 1962-2004; market splits are depicted by solid arrows. The shaded interval 1962-1998 is used for alphabet estimation. b) MTA spectrum for the process P_t that corresponds to GE log-price dynamics. The transition from a higher slope ($s \approx -2$) to a lower one ($s \approx -1$) as N increases is obvious; it occurres between levels $k = 17$ and $k = 90$. c),d) Histogram of initial volatility estimates $\tilde{a}_i$ at level $k = 17$ (panel c) and $k = 90$ (panel d). Three-modal structure is prominent within this broad range of levels. Similar results are obtained at all intermediate levels (not shown.)

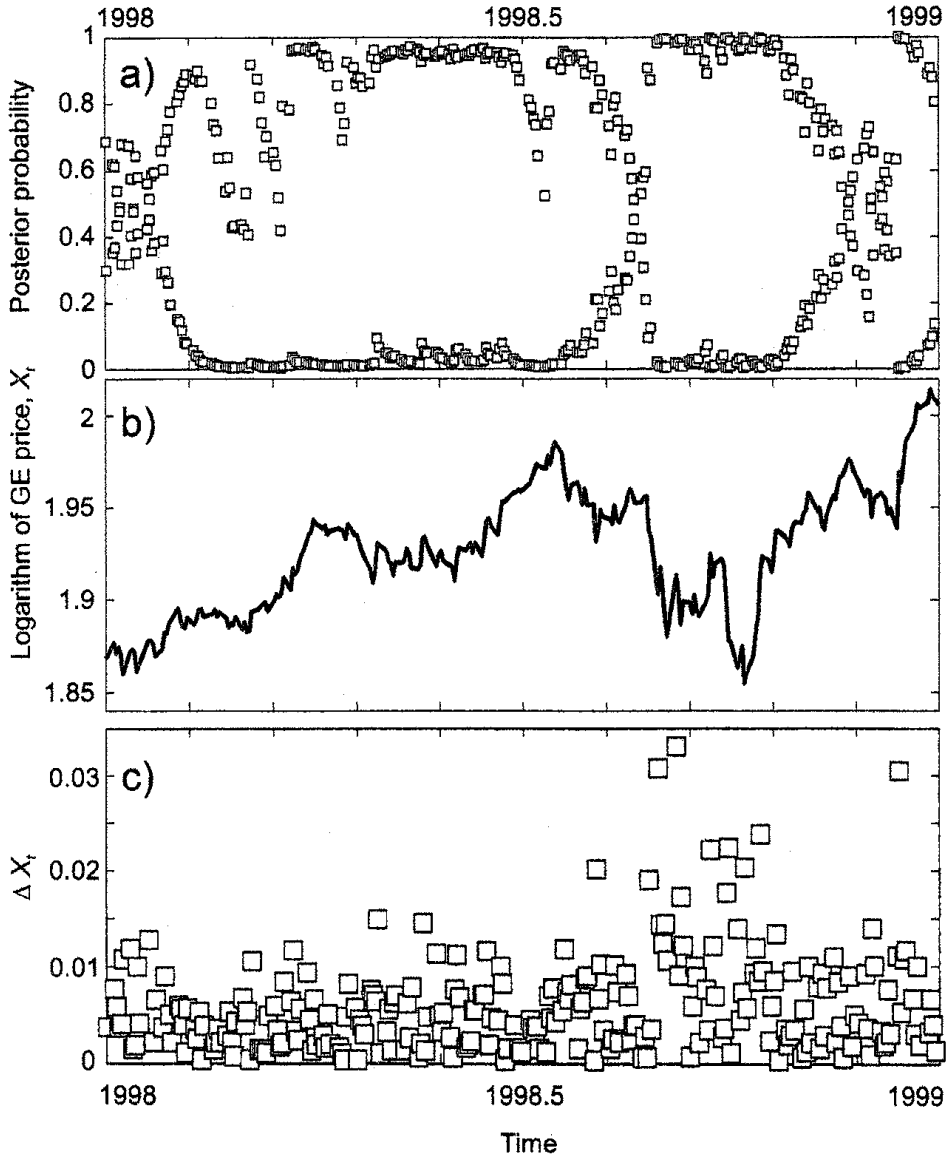

Figure 5: Estimating volatility for General Electric company during 1998-1999. a) Posterior probabilities $p_i(t)$, $i = 2$ (light squares) and $i = 3$ (dark squares) that correspond to volatility values $v_2 = 0.1$ and $v_3 = 0.15$. b) Dynamics of the log-price X_t. c) Absolute returns $|\Delta_t|$ of the log-price X_t.

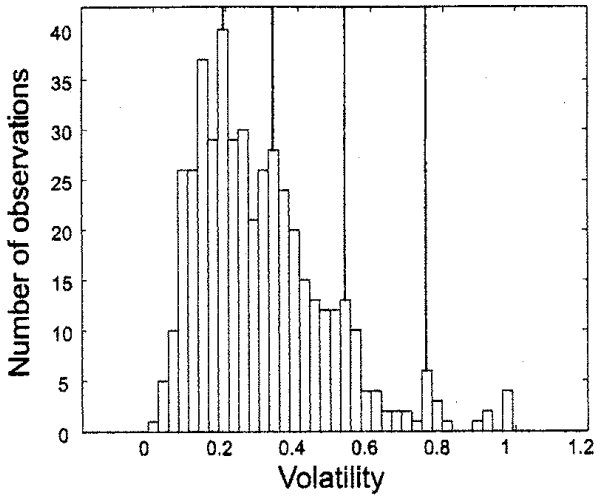

Figure 6: Estimating volatility for IBM company during November 1, 1990 – January 10, 1991. Histogram of initial estimates of volatility alphabet values $\widehat{a}_i$, $i = 1, \ldots, N_{k_0} = 493$, $k_0 = 300$.

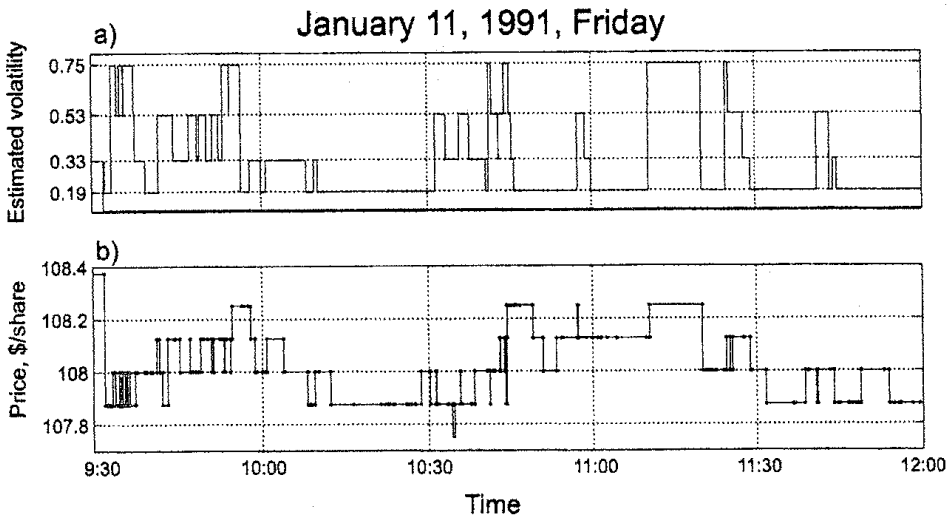

Figure 7: Filtering volatility for IBM company during January 11, 1991. a) Posterior volatility $\widehat{v}_t$ of Eq. (7.2); b) Price dynamics during the same time interval. See discussion in Sect. 7.2.