Numerical estimation of volatility values from discretely observed diffusion data

Jakša Cvitanić *
Boris Rozovskii †
Ilya Zaliapin ‡

August 29, 2005

*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 grants DMS 00-99549 and DMS 04-03575.

†Department of Mathematics , USC, 3620 S Vermont Ave, MC 2532, Los Angeles, CA 90089-1113. Ph: (213) 740-6117. E-mail: rozovsky@usc.edu.

‡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.

Contents

Abstract

We consider a Black-Scholes type model, but with volatility being a Markov Chain process. Assuming that the stock price is observed at discrete, possibly random times, the goal is to estimate the current volatility value. The model parameters, that is, the possible volatility values and transition probabilities, are estimated using the Multiscale Trend Analysis method of Zaliapin, Gabrielov and Keilis-Borok [17], adapted to our framework. Once these are given, the volatility is estimated using the filtering formula of Cvitanić, Liptser and Rozovskii [3]. Our numerical implementation shows that the estimation is of very high quality under a range of conditions.

Key Words and Phrases: Volatility, estimation, filtering, high-frequency data.
MSC2000 classification: 91B28, 91B40, 93E20

1 Introduction

The goal of this paper is to develop a framework for estimating the current value of volatility from stock prices which are observed discretely, possibly at random times. In particular, we have in mind high-frequency, "tick-by-tick" stock data.

Our model is a Black-Scholes type dynamics for the stock price, but with volatility that follows an independent Markov Chain process. The stock price observation times are modeled as a Cox process — a doubly stochastic Poisson process with intensity that can depend on the current volatility value. Thus, higher volatility may go together with more frequent observations. Given the volatility model, we can estimate its current value using the filtering formula developed in Cvitanić et al. [3]. (In fact, that paper develops the formula for the case of the volatility being a general jump-diffusion process and the observation times having a general structure) 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 we do by adapting the so-called Multiscale Trend Analysis of Zaliapin et al. [17].

The bulk of the paper consists of an analysis of the performance of the estimation procedure on simulated and real data. We show that the procedure performs very well in a variety of circumstances. 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.

There have been different approaches to estimating stochastic volatility from financial data. Maybe the most famous approach is ARCH-GARCH type modeling, where the price and the volatility are modeled as discrete time-series processes; see Gourieroux [8] for a survey. Since volatility is not an observed process, it is natural to try using filtering methods to estimate it from observed stock prices. In discrete-time framework, with fixed observation intervals, this was done in Elliott et al. [4]. In continuous-time models with possibly random observation times, the filtering approach is used in Frey and Runggaldier [6]. (See Runggaldier [16] for an up-to-date survey.). They obtain numerical approximation to the optimal filter. In Cvitanić et al. [3], the continuous-time optimal filter is obtained explicitly, and it is their formula that we use in this paper. See also Gallant and Tauchen [7], who develop an approximating algorithm in continuous time when the volatility is also a diffusion process, Malliavin and Mancino [12] for a nonparametric approach, as well as Fouque et al. [5], Rogers and Zane [15], and Kallianpur and Xiang [11] for still other approaches. For a recent general approach to statistical estimation of diffusion processes with discrete observations, see Ait Sahalia and Mykland [2]. The model we use, with Markov Chain process for volatility, is usually called "a regime switching model" and has been used increasingly often in the option pricing literature. See, for example, Guo and Zhang [9]. Platania and Rogers (2004) calibrate a Markov Chain model for observation intervals between data points. Johannes

and Polson (2003) provide a nice survey of Markov Chain Monte Carlo methods for financial data.

In addition to filtering the volatility values, as mentioned above, we also present a systematic way of deciding which possible values the volatility process can attain, and what the corresponding probabilities are, using the ideas of the Multiscale Trend Analysis (MTA) of Zaliapin et al. [17]. This is in contrast to the majority of existing filtering or Hidden Markov Chain literature, where this decision is often somewhat arbitrary and ad hoc. This preliminary stage is related to the power variation estimates of volatility, as surveyed in Barndorff-Neielsen, Graversen and Shephard (2003), for example. As in the literature on power variation approach, we use a variation process to estimate possible volatility values. However, while in that 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.

While our numerical implementation shows that the framework we use is practical and efficient, there are potential drawbacks, such as the fact that we assume that the volatility process is independent of the Brownian Motion driving the stock process. Also, we consider only the estimation of the current volatility, and not the prediction of future volatility. We leave these issues for future research.

We describe the model and the filtering formula in section 2, present a method for the numerical implementation in section 3, and we analyze the filter performance in section 4. Section 5 provides the description of the MTA method for estimating a priori parameters of the model, section 6 describes the whole algorithm, section 7 provides examples with simulated and real data, and section 8 compares the filter to a simple estimation based on returns of the risky asset. In Appendix we provide proofs and a brief description of the MTA method.

2 Filtering Approach to the Estimation Problem

We consider the problem of estimating the current value v_t of the volatility $(v_s)_{s\geq 0}$ of the log-price process

$$\log S_t =: X_t = X_0 + \int_0^t \left(\mu - \frac{v_s^2}{2}\right) ds + \int_0^t v_s \, dB_s, \tag{2.1}$$

where B_t is a Brownian motion, volatility v_t is positive, and drift μ is assumed constant (for simplicity). We also assume that the process S_t is observed at discrete random times $0 < T_1 < T_2 < \ldots < T_N < T$. In Cvitanić *et al.* [3], general filtering formulas are given

for the conditional expectation $E_t[f(v_t)]$ given the sigma-algebra $\{\mathcal{G}_t\}$ generated by the observations $X_k := X_{T_k}$ and T_k up to time t.

In this paper we focus on the special case when v_t is a homogeneous Markov jump process taking values in the finite alphabet $\mathcal{A} = \{a_1, \ldots, a_M\}$ with the intensity matrix $\Lambda = (\lambda_{ij})_{i,j=1\ldots,M}$ and the initial distribution $p_i = P(v_0 = a_i)$, $i = 1, \ldots, M$. We assume that the counting process

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

is a doubly stochastic Poisson (Cox) process, with intensity $n(v_t)$, for some bounded function n. In this case the results of Cvitanić *et al.* [3] provide the formula for the conditional probabilities of the volatility state at the observation times

$$p_i(T_k) := P(v_{T_k} = a_i | \mathcal{G}_{T_k}), \quad i = 1, \ldots, M$$

in the recursive form

$$p_i(T_k) = \frac{n(a_i) \sum_j r_{ji}(T_k - T_{k-1}, X_k - X_{k-1}) p_{ji}(T_k - T_{k-1}) p_j(T_{k-1})}{\sum_{i,j} n(a_i) r_{ji}(T_k - T_{k-1}, X_k - X_{k-1}) p_{ji}(T_k - T_{k-1}) p_j(T_{k-1})}. \tag{2.2}$$

Here we denoted

$$p_{ji}(t) = P(v_t = a_i \,|\, v_0 = a_j);$$

and

$$r_{ji}(t, z) = E\left(e^{-\int_0^t n(v_u) du} \rho^j_{0,t}(z) | v^j_t = a_i\right), \tag{2.3}$$

where $\rho_{s,t}(y)$ is the density function of the normal distribution with mean

$$m(s, t) = \int_s^t \left(\mu - \frac{v_u^2}{2}\right) du$$

and variance

$$\sigma^2(s, t) = \int_s^t v_u^2 du.$$

The process v_t starting from a_j is denoted by v^j_t, and $\rho^j_{0,t}(z)$ is obtained by substituting v^j_s for v_s in $\rho_{0,t}(z)$.

Implementation of the recursive filter (2.2) requires some a priori information about the volatility process v_t and observational grid T_k. Specifically, one has to know volatility alphabet $\mathcal{A}$, transitional intensities Λ, initial probabilities p_i, and observation intensities $\mathcal{N} = n(a_i)$, $i = 1, \ldots, M$. In practice, the a priori values of these parameters are either provided by market experts or estimated from the asset price dynamics over the past. In this paper we explore the second approach developing and testing a statistical technique for estimating a priori values of the filter input parameters in (2.2). Before proceeding with this goal, we describe a numerical implementation of the filter (2.2) and explore its performance in case of known values of input parameters.

3 Numerical implementation of the filter

This section describes numerical implementation of the filter (2.2) with known a priori values of input parameters. In this case, at each step of the recursive filtering one needs to know the conditional probabilities $p_{ji}(t)$ and to compute the conditional expectations r_{ji}, which depend on the volatility. Below we introduce a discrete-time approximation d_n to the continuous-time volatility process v_t; this will provide an analytical approximation to $p_{ji}(t)$, and facilitate the Monte Carlo estimation of r_{ji}.

3.1 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

$$Q_{ij} = \begin{cases} \lambda_{ij}\,\Delta, & i \neq j \\ 1 - \sum_{i \neq k} \lambda_{ik}\,\Delta, & i = j. \end{cases} \tag{3.1}$$

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:

$$\widehat{p}_{ji}(t) = P(d_{m_t} = a_i \,|\, d_0 = a_j) = [e_j \times Q^{m_t}](i), \tag{3.2}$$

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 $\left(v_s^j \,|\, v_t^j = a_i\right)$ 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

$$\begin{aligned} &P(d_n = a_k \,|\, d_{n-1} = a_{k'}, d_{m_t} = a_i) = \\ &\quad = \frac{P(d_n = a_k \,|\, d_{n-1} = a_{k'})\, P(d_{m_t} = a_i \,|\, d_n = a_k)}{\sum_{m=1}^{M} P(d_n = a_m \,|\, d_{n-1} = a_{k'})\, P(d_{m_t} = a_i \,|\, d_n = a_m)}. \end{aligned} \tag{3.3}$$

Here

$$P(d_n = a_k \,|\, d_{n-1} = a_{k'}) = [e_{k'} \times Q](k); \tag{3.4}$$

$$P(d_{m_t} = a_i \,|\, d_n = a_k) = [e_k \times Q^{N-n}](i). \tag{3.5}$$

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 $[0,\, t)$.

3.2 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

$$\delta_k := (T_k - T_{k-1})\ ,\quad \Delta_k := (X_{T_k} - X_{T_{k-1}}), \tag{3.6}$$

and

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

we see that in estimating r_{ij} of (2.3), we can use

$$\rho^j_{0,\delta_k}(\Delta_k) = \frac{1}{\sqrt{2\pi A_k^j}} \exp\left\{-\frac{\left(\Delta_k - \mu\delta_k + \frac{1}{2}A_k^j\right)^2}{2A_k^j}\right\}. \tag{3.7}$$

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

$$A_k^j := \sum_{i=1}^{N_k} a_{(i)}^2 (u_i - u_{i-1}), \tag{3.8}$$

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 $v_{\delta_k}^j = a_i$ of (2.3) implies that $a_{(N_k)} = a_i$.

Similarly,

$$\int_0^{\delta_k} n(v_u^j) du = \sum_{i=1}^{N_k} n(a_{(i)})(u_i - u_{i-1}). \tag{3.9}$$

We estimate r_{ij} by simulating independent realizations of d_n on $[0, \delta_k)$ and using equations (3.8) and (3.9) with $\{v_t^j\}$ replaced by $\{d_n \,|\, d_0 = a_j\}$.

4 Filter performance

Here we illustrate the performance of the filter (2.2) implemented as described in Sect. 3.

4.1 Measures of estimation quality

We use three alternative "score functions" for a formal assessment of the filter estimation quality. One of them compares the posterior distribution of volatility values with a reference one; the other two evaluate the distance between the estimated and true volatility.

The first score is based on the likelihood-ratio:

$$S_{\rm LR} = \frac{1}{N}\sum_{k=1}^{N} \log \frac{p_{a_{(k)}}}{\pi_{a_{(k)}}}, \tag{4.1}$$

where N is the number of observations, $p_{a_{(k)}}$ denotes the a posteriori probability that the volatility is $a_{(k)}$ at k-th step, and $\pi_{a_{(k)}}$ is the corresponding reference (a priori) probability for this value. In our case, it is natural to set $\pi_i = p_i$, so that $S_{\rm LR}$ compares our performance with the random guessing based on the a priori distribution of volatility values. A random guessing according to initial probabilities p_i corresponds to $S_{\rm LR} = 0$; positive score values indicate that our performance is better than random; negative that it is worse than random (it is better to guess than to apply our estimation). The value of this score for the ideal prediction (probability of the true value is 1 at every step) is $-\frac{1}{N}\sum_k \log \pi_k$; in the case $\pi_k = 1/M$ this gives $\log M$. Below we use the base-M logarithm in the definition (4.1) which makes the best possible score $S_{\rm LR} = 1$.

An alternative score function is based on the empirical frequencies

$$q_{ij} := \frac{\sum_k \delta(\widehat{v}_{T_k} - a_j, v_{T_k} - a_i)}{\sum_k \delta(v_{T_k} - a_i)}, \quad i,j = 1,\ldots,M. \tag{4.2}$$

Here $\delta(\cdot,\cdot)$ is a discrete delta function, $\widehat{v}_{T_k}$ is an estimation of volatility based on posterior probabilities $p_i(T_k)$ (see for instance (7.1)). The diagonal elements of q_{ij} correspond to the instants with correct volatility estimates, while the off-diagonal elements to the instants with wrong estimates. Clearly, an ideal volatility estimation (correct volatility estimate at every step) corresponds to the diagonal form of q_{ij}. The score function is defined as

$$S_{\rm tr} = \frac{{\rm Trace}(q_{ij})}{M} = \frac{\sum_{i=1}^{M} q_{ii}}{M}. \tag{4.3}$$

An ideal estimation corresponds to $q_{ij} = \delta(i-j)$ and would give the maximum score $S_{\rm tr} = 1$. For a "random" estimation based on any time-independent posterior distribution (say, $p_i(T_k) = p_i$ using the initial distribution) the summands that correspond to $v_{T_k} = a_i$ in the numerator of (4.2) become independent binomial random variables with probability

of success being a function of $p_j(T_1)$ (in case of using (7.1) it is just $p_j(T_1)$). Hence q_{ij} converges a.s. to $p_j(T_1)$ as $\#\{k : v_{T_k} = a_i\} \to \infty$; and the corresponding score $S_{\rm tr}$ converges to $\sum p_j(T_1)/M = 1/M$.

The last score is defined as the L^2 distance between the true volatility and its estimate:

$$S_{L^2} = \frac{\sum_k \left(v_{T_k} - \widehat{v}_{T_k}\right)^2}{\sum_k \left(v_{T_k}\right)^2}. \tag{4.4}$$

We used all the above scores in our experiments and found that they give similar assessment of the filtering quality.

Remark 4.1 Strictly speaking, the above scores $S_{\rm LR}$ and $S_{\rm tr}$ are defined under the assumption that the posterior volatility takes values in the true alphabet $\mathcal{A}$. In practice, they can be still applied if the estimated alphabet $\widehat{\mathcal{A}}$ is close in some sense to the true one. Say, if the posterior volatility alphabet has the same number of elements as the true one and their numerical values are close. We will see that this condition is satisfied in our examples below.

4.2 Performance with known a priori parameters

4.2.1 Introductory examples

This section illustrates the performance of the estimates (2.2) with known values of input parameters. Here we restrict the analysis to a two-valued volatility process, which simplifies parametrization and makes our examples more intuitive.

First, we put $a_1 = 0.1$, $a_2 = 1$, $n_1 := n(a_1) = 50$, $n_2 := n(a_2) = 100$, $\lambda_{12} = \lambda_{21} = 1$; the results of the filtering estimation are illustrated in Fig. 1. The estimate performance is almost perfect: it takes a small fraction of a time unit to reset the a posteriori probability in response to the volatility change; otherwise the value of $p_1(t)$ is either close to 1 (showing that the current volatility value is a_1) or to 0 (in case of a_2).

The power of the filter (2.2) is better demonstrated when we consider the close volatility values $a_1 = 0.3$, $a_2 = 0.4$, leaving the observational intensities $n_1 = 50$, $n_2 = 100$ (Fig. 2). In this case, the asset price trajectories that correspond to alternative volatility values are visually indistinguishable (bottom panel); still, the a posteriori probabilities depict nicely the unobserved volatility changes (top panel). One notices though that now it takes longer to recognize the volatility change.

Remark 4.2 We assume that μ, v_t, n_i, and λ_{ij} have units 1/year. It is useful to notice that daily observations correspond to about $n_i = 3 \times 10^2$ while hourly to about 8×10^3.

4.2.2 Analysis of the two-valued volatility model

Here we test systematically the filtering of the two-valued volatility process. Our simulations suggest that the filtering quality is primarily affected by the magnitude of the volatility change, that is the ratio a_1/a_2, rather than specific values of a_1 and a_2; and a similar conclusion is made for the observational intensities. This allows us to consider the alphabet $\{a_1 = 0.3,\, a_2\}$, observational intensities $\{n_1 = 50,\, n_2\}$, drift $\mu = 0.05$, and jump intensities

$$\lambda = \begin{pmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{pmatrix}$$

and vary only two parameters — a_2 and n_2 — within several orders of magnitude. All estimations have been done with true a priori values of filter input parameters.

The quality score S_{LR} is shown as a function of a_2, n_2 in Fig. 3. The quality of the filter is generally very good ($S_{\mathrm{tr}} > 0.9$, $S_{\mathrm{LR}} > 0.85$). The filter quality significantly deteriorates only when $a_2 \to a_1$ and $n_2 \to n_1$ simultaneously, that is, when the two regimes become truly undistinguishable. Notice that the filter can separate two regimes either by intensity (even when $a_1 \to a_2$), or by volatility (even when $n_1 \to n_2$), or using both of them. Our simulations also show that the filter quality improves as the observational intensities n_i increase.

Remark 4.3 Our numerical analysis using the three alternative score functions suggests that the filtering quality is robust with respect to the values of the drift μ and the jump intensities Λ, when the observation intensities n_i grow larger. In particular, it seems enough to have a several dozen of observations between consecutive volatility jumps in order to make the filter estimates almost independent of μ, Λ. We plan to explore this issue in more detail in future work.

5 Estimating a priori values of the filter parameters

In this section we consider the problem of estimating a priori values of the filter parameters — volatility alphabet $\mathcal{A}$, jump intensities Λ, initial probabilities p_i $(i, j = 1, \ldots, M)$, and observation intensities $\mathcal{N} = n(a_i)$ — using the observations of $X(t)$ on the random grid T_k.

The idea of the proposed estimation procedure is to construct the 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 $L(t)$. The practical applicability of this approach depends crucially on the

properties of the process P_t as well as on our ability to construct good piece-wise linear approximations. Below we construct process P_t such that it conforms to (5.1) and allows estimation of the volatility using a single realization. The problem of piece-wise linear approximation will be effectively solved by the Multiscale Trend Analysis (MTA) [17] which we briefly describe in Sect. 9.2.

5.1 Volatility alphabet

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

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

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

In other words, for large intensity n,

$$P_t \sim tv\sqrt{\frac{n}{2}}. \tag{5.4}$$

Proof: See Appendix 9.1.

Remark 5.1 The proposition is also true for intervals of the form $[t_1, t_2]$. Thus, as $n \to \infty$ and the intervals between observations become smaller, we have for the local value of volatility v and intensity n,

$$P_{t_2} - P_{t_1} = \sum_{k:t_1 \le T_k \le t_2} |\Delta_k| \approx (t_2 - t_1)v\sqrt{\frac{n}{2}}.$$

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

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

within the respective intervals.

Remark 5.2 A similar proof shows that in the case of fixed intervals of length δ between observations

$$\frac{\sqrt{\delta}\, P_t}{t} - v\sqrt{\frac{2}{\pi}} \xrightarrow{\text{a.s.}} 0, \text{ as } \delta \to 0. \tag{5.6}$$

Thus, the distinct slopes of a piece-wise linear P_t are (asymptotically) given by

$$s_i = s_i(a_i) = a_i\sqrt{\frac{2}{\pi\,\delta}}. \tag{5.7}$$

Remark 5.3 Barndorff-Neielsen, Graversen and Shephard [1] 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

$$\sqrt{\delta}P_t \xrightarrow{P} \sqrt{\frac{2}{\pi}}\int_0^t v_s ds\ , \quad \text{as } \delta \to 0. \tag{5.8}$$

Remark 5.4 In applications it is important to know the rapidity of convergence in (5.3). Roughly speaking, it is controlled by the parameter $\kappa := b/\sqrt{n}$ with $b = r/v - v/2$. Figure 4 shows the relative error of the estimation (5.4) as a function of κ. This plot was obtained using numerical simulations within a broad range of model parameters:

$$10^{-2} \le r \le 10^2; \quad 10^{-2} \le v \le 10^2; \quad 10^{-3} \le n \le 10^4.$$

The figure shows that κ cearly controls the quality of estimation; and that the quality is very good (relative error $< 3\%$) as soon as $\kappa < 0.1$.

If we consider a piece-wise linear process $L(t)$ with slopes defined as in (5.5), (5.7), 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

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

In case of regular observational grid with step δ we similarly obtain using (5.7)

$$\widetilde{a}_i^{\,\text{Regular}} = s_i\sqrt{\frac{\pi\,\delta}{2}},\ i = 1, \ldots, N_L. \tag{5.10}$$

From (5.9),(5.10) 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 $\widehat{v}_t$. The problem of automated binning (which reduces to estimating the number of modes in an empirical distribution) is not a trivial one, especially when N_L is small; this problem is beyond the scope of the present paper.

Note that parameters $n_i, i = 1, \ldots, N_L$, in (5.9), should also be estimated from the data. Suppose that i-th segment of $L(t)$ has duration τ_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

$$\widehat{n}_{(i)} = \frac{m_i}{\tau_i}. \tag{5.11}$$

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 [17] (see Appendix 9.2). We demonstrate below that the MTA estimation based on Proposition 5.1 is very efficient within a broad range of model parameters.

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 $\widehat{v}_{T_k} = a_j$ and $\widehat{v}_{T_{k-1}} = a_i$:

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

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

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

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

$$\widehat{p}_i = \frac{\sum_k \tau_{ik}}{\sum_j \sum_k \tau_{jk}}, \quad i = 1, \ldots, \widehat{M}. \tag{5.12}$$

$$\widehat{\lambda}_{ij} = \frac{\tau_{ii} + \tau_{ij}}{m_{ij}}, \quad i, j = 1, \ldots, \widehat{M},\ i \neq j. \tag{5.13}$$

$$\widehat{n}_i = \frac{\sum_k m_{ik}}{\sum_k \tau_{ik}}, \quad i = 1, \ldots, \widehat{M}. \tag{5.14}$$

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 5.5 We introduced two different estimators for observation intensity n_i given by Eqs. (5.11) and (5.14). The estimate (5.11) 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 (5.14) produces $\widehat{M}$ estimated values using the posterior coarse alphabet $\{\widehat{a}_i\}$.

5.3 Example

Here we illustrate the alphabet estimation procedure introduced above using the process X_t defined by equation (2.1) with $\mu = 0.05$, two-valued volatility alphabet $\{\sqrt{2\,\mu},\, 2\sqrt{2\,\mu}\} \approx \{0.316,\, 0.632\}$, 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. 5a; the shaded areas depict intervals with $v_t = a_1$. Figure 5b 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 constructing the hierarchy of piece-wise linear decompositions $L_k(t)$ for P_t. The MTA spectrum—the fitting error E_k (in L^2) as a function of the number N_k of linear segments within L_k—is shown in Fig. 6. Recall from [17] that for a pure random walk

$$E_k = \frac{E_0}{N_k}, \tag{5.15}$$

which 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. 6). The approximation $L_{13}(t)$ is shown in Fig. 5b 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 (5.9); the raw estimate $\widetilde{v_t}$ is shown in Fig. 5c; 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. 5c: 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%).

We emphasize that the most delicate part of the above scheme is detecting jumps of v_t, while estimating the volatility within fixed intervals is a much simpler problem. Thus, the better the piece-wise linear approximation $L_k(t)$, the better the final estimation. A natural question is how sensitive is the described algorithm to the choice of the MTA level k. To answer this question we repeat the estimation procedure at the second corner point, $k = 25$ (see Fig. 6). The results are shown in Fig. 7; the estimation now depicts correctly even the tiny volatility interval at $t \approx 15$ while the alphabet values are estimated with the error less than 2%.

Clearly, any decomposition level between $k = 13$ and $k = 25$ will result in a good estimate of the volatility alphabet, with relative error less that 3%. We emphasize that the number N_k of segments within the piece-wise linear approximation varies here almost twice: from $N_{13} = 22$ to for $N_{25} = 42$; thus the algorithm is very stable to the particular choice of the piece-wise linear approximation.

6 The combined algorithm

In this section we merge the results of the above Sects. 2, 3, and 5 to construct the combined volatility estimation procedure; it is described by the following sequence of steps:

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

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 (5.9) and (5.11) or (5.10) 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. (5.12).

Step 3. Estimate a priori transitional intensities using Eq. (5.13).

Step 4. Estimate time-dependent volatility using the filter Eq. (2.2) with a priori parameters from Steps 1,2,3.

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

7 Examples

Here we apply our combined algorithm to three price series. First, we apply it to a synthetic asset price with known true volatility structure and show that the alphabet estimation and the filtering give very good results. Next, we analyze the daily dynamics of General Electric shares traded at NYSE during 1962–2004. Finally, we estimate the volatility of intraday trades for IBM during Nov. 1, 1990 – Jan. 11, 1991. In the latter cases neither we know the true volatility, nor we can be sure that the asset dynamics are governed by (2.1). Nevertheless, we give some indications that the proposed procedure gives reasonable and stable results.

7.1 Synthetic prices

We simulate the log-price dynamics using (2.1) with $\mu = 0.05$, $\{a_i\} = \{0.1, 0.3, 0.5\}$, $p_i = 1/3$, and

$$\Lambda = \begin{pmatrix} -20 & 20 & 0 \\ 40 & -60 & 20 \\ 60 & 0 & -60 \end{pmatrix} [1/\text{day}].$$

Thus, the changes in volatility occur about 20-60 times a calendar day (calendar day equals to 24 hours), or, the duration of each constant volatility interval varies roughly from 30 mins to 1 hour; the higher the volatility the shorter the time to the volatility jump. The duration of the synthetic series is $168 = 21 \times 8 = 7 \times 24$ hours, which corresponds to 7 calendar days (24 hours a day) or 21 business days (8 hours a day).

The intensity of observations is $\{n_i\} = \{1, 3, 5\}[1/\text{min}]$, which corresponds to a synthetic time series of 17,575 observations. A fragment of this series is shown in Fig. 8 a. The first 151 hours were used to estimate the filter input parameters, and the filter was applied during the last 17 hours. We do not illustrate here steps 1.1–1.3 of our algorithm since the results are very similar to what was shown in Figs. 5–7 above. The histogram of initial estimates $\{\widetilde{a}_i\}$ of the volatility alphabet is shown in Fig. 8 c. The three-modal structure of the volatility alphabet is clearly detected; the corresponding estimates of the filter input parameters are

$$\widehat{a}_i = \{0.1109, 0.3057, 0.5045\}, \quad \widehat{n}_i = \{1.035, 3.003, 4.994\}[1/\text{min}],$$

$$\widehat{\Lambda} = \begin{pmatrix} -31.30 & 15.15 & 16.15 \\ 37.42 & -76.75 & 39.33 \\ 43.62 & 167.16 & -210.78 \end{pmatrix} [1/\text{day}].$$

The estimates of the alphabet and observational intensities are very close to the true values, while the estimate of the transitional intensities are of inferior quality. Nevertheless, as was mentioned earlier, this latter estimation will produce a very little effect on the overall quality of the filtering. Next, we apply the filter (2.2) to obtain the a posteriori probabilities. They are illustrated in panel b, which shows the dynamics of $p_1(t)$ and $p_3(t)$ within the subinterval of panel a. The values of $p_2(t) = 1 - p_1(t) - p_3(t)$ are not shown for visual convenience. One can see that the a posteriori probabilities depict very closely the true structure of volatility. It takes 3-5 observations to detect the volatility change, and there are only two false detection within this interval (at $t \approx 160.6$, 161.7 where $p_1 > p_{2,3}$). A similar picture is observed at other times. These results suggest that in order to make a decision about current volatility on the basis of a posteriori probabilities one can use the simplest rule $v_T = a_k$ with $k := \text{argmax}_i\ [p_i(T)]$. A more stable way to define the posterior volatility is

to use the expectation with respect to the posterior distribution:

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

The a posteriori volatility $\widehat{v}_t$ is then obtained by binning the posterior expectation (7.1) into $\widehat{M}$ separate values

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

The corresponding values of the three measures of estimation quality (Sect. 4.1) are

$$S_{\mathrm{LR}} = 0.754, \quad S_{\mathrm{tr}} = 0.859, \quad S_{L^2} = 0.032.$$

Recalling that the ideal prediction would correspond to $S_{\mathrm{LR}} = S_{\mathrm{tr}} = 1$, $S_{L^2} = 0$ we conclude that the filtering is of a very good quality. To further illustrate the overall quality of the filter we show in panel d the empirical frequencies q_{ij} (4.2). The matrix is very close to diagonal, which confirms the reliability of our estimation.

7.2 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 [18]. 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. 9a. 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. 9a). MTA spectrum for the process P_t of (5.2) is shown in Fig. 9b. 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 $17 \le k \le 90$. The results of our estimation are stable with respect to particular choice of the level for analysis. Figure 9c 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:

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

The above estimates are used as inputs for the filter (2.2). The posterior probabilities $p_i(t)$, $i = 2, 3$, during 1998-1999 are shown in Fig. 10a. 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. 9 c,d (say, choosing $\{\widehat{a}_i\} = \{0.05, 0.08, 0.15\}$, *etc.*).

The posterior expectation of the volatility at instant T_k is defined by (7.1). It can be used to define the posterior volatility, which takes values from the alphabet $\{\widehat{a}_i\}$. Specifically, we smooth the posterior expectation using kernel $K(t)$:

$$v_t^\kappa := \int K(t-s)\, E(v_s)\, ds, \tag{7.3}$$

and bin it into $\widehat{M}$ distinct values

$$\widehat{v}_t := \{\widehat{a}_i,\, i = \mathrm{argmin}_k |v_t^\kappa - \widehat{a}_k|\}. \tag{7.4}$$

We use the kernel $K = I_{[-84 \text{ days}, 0]}$ to obtain the volatility estimate $\widehat{v}_t$ shown in the Fig. 11. Panel a juxtaposes the periods of high volatility ($\widehat{v}_t = 0.15$) and the largest absolute returns $|\Delta_t| > 0.015$. One sees the perfect agreement between the filter estimation and actual increases of price jitter. Panel b zoomes up a 10-year period around the market crash of 19 Oct., 1987; here the shadowing (dark, light, none) depicts intervals with three distinct volatility values (large, medium, small). Again, the correspondence between actual price variations and the filter estimation is very good.

In contrast to the synthetic models, where the true volatility structure is known and can be compared to the filter estimation, in this realistic example we do not know what the true volatility is. To assess the quality of the filter in this case, we use empirical averages E_i, $i = 1, \ldots, M$ of absolute returns $|\Delta_k|$ of the log-price within intervals of constant a posteriori volatility:

$$E_i = \frac{\sum |\Delta_{k+1}|\, \delta(v_{T_k} - a_i)}{\sum \delta(v_{T_k} - a_i)}. \tag{7.5}$$

The corresponding values for the GE data are

$$\begin{aligned} E_1 &:= 3.5 \times 10^{-3} \pm 6.6 \times 10^{-5}, \\ E_2 &:= 4.3 \times 10^{-3} \pm 5.0 \times 10^{-5}, \\ E_3 &:= 6.8 \times 10^{-3} \pm 11.1 \times 10^{-5}. \end{aligned}$$

The positive correlation between the typical price returns and the filter volatility estimate is obvious. This indirect quality analysis allows us to conclude that the suggested filter

estimation procedure might be effective for realistic data sets, which might deviate from the model (2.1) with Markov jump volatility.

Remark 7.1 At first glance, the reader could ask why do we need the filtering estimate, if we can simply use estimation based only on price variations. We do a comparison of that type in section 8, showing that, in general, the filtering procedure is more stable and efficient.

7.3 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 τ_i larger than 2 hours were replaced with random times $\tilde{\tau}_i$ 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 MTA spectrum for the corresponding process P_t is shown in Fig. 12a; it can be coarsely approximated as

$$\log E_k \propto 3/2 \log N_k.$$

To emphasize the existence of corner points, we show the normalized fitting error $E_k \times N_k^{1.4}$ as a function of the number N_k of segments in Fig. 12b. One sees clearly three corner points that correspond roughly to the scales of 16 hours, 2.5 hours, and 40 minutes (scale is defined here as the average duration of segments within a piece-wise linear approximation). We choose the rightmost corner point $k_0 = 300$ with the scale of 40 minutes for further analysis. The piece-wise linear approximation at this level has 493 segments; the histogram of the corresponding initial alphabet estimates $\widetilde{a}_i$ (Eqs. (5.9) and (5.11)) is shown in Fig. 13. While there is no striking mutimodal 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:

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

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

The filtering results are illustrated in Fig. 14, 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 by binning the posterior expectation (7.1) into $\widehat{M}$ separate values via (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.

8 Filtering vs. variation estimates

Here we compare the filter with an *ad hoc* estimation of the volatility based on absolute returns.

8.1 Filter estimation

Recall that the filter calculates the posterior volatility distribution $p_i(t) = P(v_t = \widehat{a}_i)$, $i = 1, \ldots, \widehat{M}$, where $\widehat{a}_i$ comprise the estimate of the volatility alphabet. We consider the best least-square volatility estimation

$$\tilde{v}_t^{\mathrm{filter}} = \sum_{i=1}^{\widehat{M}} p_i(t)\, \widehat{a}_i.$$

8.2 Alternative estimation

An alternative estimation, $\tilde{v}^{\mathrm{smooth}}$, uses the fact that the local slope of the process P_t (the sum of absolute returns) is asymptotically proportional to the volatility v_t (see (5.9)). Specifically,

for a given $w > 0$ we define

$$\tilde{v}_t^{\text{smooth}} = \sqrt{\frac{2}{n_t}} \frac{P_t - P_{t-w}}{w} = \sqrt{\frac{2}{n_t}} \frac{1}{w} \sum_{k\,:\,T_k, T_{k-1} \in [t-w,\,t]} |X_{T_k} - X_{T_{k-1}}|,$$

where

$$n_t = \frac{\#\{k \,:\, T_k \in [t-w,\, t]\}}{w}.$$

The choice of w depends on the structure of the volatility process, particularly, the jump intensities Λ, and the observation intensities $\mathcal{N}$. One wants to have as many observations within the window $[t-w,\, t]$ as possible, yet keeping the length w as short as possible in order not to contaminate the estimation by periods when the window covers a volatility jump. That is, we want w such that

$$\min_i n_i \gg \frac{1}{w} \gg \max_i l_i,$$

where $l_i = \sum_{i \neq k} \lambda_{ik}$ is the intensity of jumps from state i.

8.3 Binning

Notice that when we have the alphabet estimates $\widehat{\mathcal{A}} = (\widehat{a}_i)$, the values of $\tilde{v} = \tilde{v}_t^{\text{filter}}$, $\tilde{v}_t^{\text{smooth}}$ can be binned to produce the final estimates

$$v_t = \{\widehat{a}_i,\, i = \operatorname{argmin}_k |\tilde{v}_t - \widehat{a}_k|\}. \tag{8.1}$$

We work below with such binned estimates v_t^{filter}, v_t^{smooth}.

8.4 Input parameters

We apply the volatility estimates v^{filter} and v^{smooth} to the synthetic time series with parameters

$$\mu = 0.05, \quad \mathcal{A} = \{0.1,\, 0.5\}, \quad \Lambda = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} [1/\text{hr}]; \quad \mathcal{N} = \{n_i\} = \{1,\, 5\}[1/\text{min}]$$

modeled during 7 days (168 hours), which consists of 31059 observations. In order to calculate the volatility estimates we need to specify the values $\widehat{a}_i$ of the alphabet used for binning (in case of the filter, this alphabet is also used to produce the posterior volatility distribution.)

We use four different versions of the alphabet $\widehat{\mathcal{A}}$:

$$\begin{aligned}
\widehat{\mathcal{A}}_1 &= \mathcal{A} \\
\widehat{\mathcal{A}}_2 &= \mathcal{A} + .05 = \{0.15,\, 0.55\} \\
\widehat{\mathcal{A}}_3 &= \mathcal{A} \cup 0.3 = \{0.1,\, 0.3,\, 0.5\} \\
\widehat{\mathcal{A}}_3 &= \mathcal{A} \cup \{0.05,\, 0.2,\, 0.3,\, 0.4,\, 0.6\} = \{0.05,\, 0.1,\, 0.2,\, 0.3,\, 0.4,\, 0.5,\, 0.6\}
\end{aligned}$$

The associated parameters used by the filter are $\widehat{\mu} = \mu$,

$$\widehat{\Lambda}_1 = \widehat{\Lambda}_2 = \Lambda; \quad \widehat{\Lambda}_3 = \begin{pmatrix} -1 & 0.5 & 0.5 \\ 0.5 & -1 & 0.5 \\ 0.5 & 0.5 & -1 \end{pmatrix} \mathrm{hr}^{-1};$$

$$\widehat{\Lambda}_4 = \left\{\widehat{\lambda}_{ij}\right\} \,:\, \lambda_{ij} = \begin{cases} 0.5 & 0 < |i-j| \le 2 \\ 0 & |i-j| > 2 \end{cases};$$

$$\widehat{\mathcal{N}}_1 = \widehat{\mathcal{N}}_2 = \mathcal{N}; \quad \widehat{\mathcal{N}}_3 = \{1,\, 3,\, 5\}\,\mathrm{min}^{-1};$$

$$\widehat{\mathcal{N}}_4 = \{0.5,\, 1,\, 2,\, 3,\, 4,\, 5,\, 6\}\,\mathrm{min}^{-1}.$$

The parameter w of the estimation v^{smooth} was varied each time from from 30 sec. to 1 hour.

8.5 Performance

The sum of squared errors is used to measure the quality of our estimates:

$$E^{\mathrm{filter}} = \frac{\sum_k \left(v_{T_k} - v_{T_k}^{\mathrm{filter}}\right)^2}{\sum_k v_{T_k}^2}$$

and similarly for E^{smooth}. The errors of the filter estimation are

$$E_1^{\mathrm{filter}} = 0.008; \quad E_2^{\mathrm{filter}} = 0.022; \quad E_3^{\mathrm{filter}} = 0.01; \quad E_4^{\mathrm{filter}} = 0.017.$$

This demonstrates good quality and stability of the filter with respect to the input alphabet and other parameters.

The results for v^{smooth} are shown in Fig. 15. The best estimation is always achieved at $w \approx 4$ min, confirming that the choice of this parameters is only dictated by Λ and $\mathcal{N}$, and is independent of $\mathcal{A}$. The corresponding best fitting errors are

$$E_1^{\mathrm{smooth}} = 0.027; \quad E_2^{\mathrm{smooth}} = 0.048; \quad E_3^{\mathrm{smooth}} = 0.036; \quad E_4^{\mathrm{smooth}} = 0.044,$$

which is 2-3 times larger than the filtering errors. We notice also that the minima in Fig. 15 are quite sharp, which means that a slight change of parameter w from its optimal value will cause a significant decrease in the estimation quality.

Acknowledgments: We are grateful to David Vere-Jones for his advice on effective numerical modeling of Markov jump processes. We also thank Wharton Research Data Services for providing the data for General Electric company.

9 Appendix

9.1 Proof of Proposition 5.1

We start with the following

Lemma 9.1 *If volatility v and intensity n are constant within the interval $[T_{k-1}, T_k]$ then, given T_{k-1}, T_k, the conditional expectation E_k of $|\Delta_k|$ is given by*

$$E_k\left(|\Delta_k|\right) = v\sqrt{\frac{2\,\delta_k}{\pi}}\,\exp\left\{-\beta_k^2\right\} + v\,b\,\delta_k\,\mathrm{erf}(\beta_k), \tag{9.1}$$

where

$$b := \frac{\mu}{v} - \frac{v}{2} = \frac{1}{v}\left(\mu - \frac{v^2}{2}\right), \qquad \beta_k := \frac{b\sqrt{\delta_k}}{\sqrt{2}}, \qquad \delta_k := T_k - T_{k-1} \tag{9.2}$$

and

$$\mathrm{erf}(a) := \frac{2}{\sqrt{\pi}}\int_0^a \exp\left\{-z^2\right\} dz. \tag{9.3}$$

The unconditional expectation can be written as

$$E\left(|\Delta_k|\right) = \frac{v}{\sqrt{2n}}\left(1 + \frac{b^2}{2\,n}\right)^{-3/2} + v\,b\,E[\delta_k\,\mathrm{erf}(\beta_k)]. \tag{9.4}$$

Proof: The proof of (9.1) is straightforward, integrating against normal density. Similarly, the first term on the right-hand side of (9.4) is obtained integrating (9.1) against the exponential density with intensity n.

We now proceed with the proof of Proposition 5.1. Define a set B_k, for a given, but arbitrary $\varepsilon > 0$:

$$B_k = \left\{\left|\sum_{i=1}^{k}(|\Delta_i| - E\,|\Delta_i|)\right| > \sqrt{n}t\varepsilon\right\}$$

and denote by N_t the Poisson process of the number of observations by time t. Let $0 \le \delta < 1/4$ be a fixed number. We have

$$P(B_{N_t}) \le P\left(|N_t - tn| \le t\, n^{3/4+\delta}, B_{N_t}\right) + P\left(|N_t - tn| > t\, n^{3/4+\delta}\right). \tag{9.5}$$

We recall that [1]

$$E(N_t - nt)^4 = nt(1 + 3nt) \tag{9.6}$$

and use Markov inequality to find that the second term on the RHS of (9.5) satisfies

$$\begin{aligned} P\left(|N_t - tn| > t\, n^{3/4+\delta}\right) &= P\left(|N_t - tn|^4 > t^4\, n^{3+4\delta}\right) \\ &\le \frac{nt(1+3nt)}{t^4\, n^{3+4\delta}} = \frac{1}{t^3\, n^{2+4\delta}} + \frac{3}{t^2\, n^{1+4\delta}}. \end{aligned} \tag{9.7}$$

The sum of these terms over $n = 1, \ldots, \infty$ is finite. As for the first term on the RHS of (9.5), it is less than

$$P(\cup_{k\,:\,|k-tn| \le t\, n^{3/4+\delta}} B_k).$$

Since $|\Delta_i| - E\,|\Delta_i|$ are independent random variables with mean zero, using Kolmogorov's inequality we get that this is smaller than

$$\frac{1}{nt^2\varepsilon^2} \sum_{k\,:\,|k-tn| \le t\, n^{3/4+\delta}} Var\,|\Delta_k|.$$

It is easily checked that if $b = 0$

$$Var\,|\Delta_k| = \frac{v^2}{2n},$$

and if b is not zero, it is still of the order $1/n$. This shows that the first term on the RHS of (9.5) is of the order $n^{-5/4+\delta}$, and the sum of such terms over $n = 1, \ldots, \infty$ is finite. Therefore

$$\sum_{n=1}^{\infty} P(B_{N_t}) < \infty$$

which, by Borel-Cantelli lemma, implies that

$$\frac{1}{t\sqrt{n}} \sum_{i=1}^{N_t} \left(|\Delta_i| - E\,|\Delta_i|\right) = \frac{P_t}{t\sqrt{n}} - \frac{N_t}{t\sqrt{n}} E\,|\Delta_1| \xrightarrow{\text{a.s.}} 0, \text{ as } n \to \infty.$$

Using (9.6) and Markov inequality we find

$$P\left(\left|\frac{N_t}{nt} - 1\right| > \varepsilon\right) = P(|N_t - nt|^4 > (nt\varepsilon)^4) \le \frac{nt(1+3nt)}{(nt\varepsilon)^4}.$$

[1] See http://mathworld.wolfram.com/PoissonDistribution.html, for example.

The sum of these over $n = 1, \ldots, \infty$ is finite, so that, by Borel-Cantelli lemma, $N_t/(nt)$ converges to one almost surely. This implies

$$\frac{P_t}{t\sqrt{n}} - \sqrt{n}E\left|\Delta_1\right| \xrightarrow{\text{a.s.}} 0, \quad \text{as } n \to \infty.$$

Finally, since $\mathrm{erf}(\cdot)$ is a bounded nonnegative function and since $E[\delta_k] = 1/n$, we see from (9.4) that

$$\sqrt{n}E\left|\Delta_1\right| \to \frac{v}{\sqrt{2}}, \text{ as } n \to \infty.$$

This completes the proof. ■

9.2 Multiscale Trend Analysis

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. 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 $X(t)$. The larger the scale at which the trend is observed, the higher the corresponding node in the tree (referred to as *hierarchy*). The root (considered to be the top of the hierarchy) 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. Specific features of the topological and metric structure of M_X reflect self-affine (global and local) properties of $X(t)$, facilitate effective correlation analysis of time series observed at irregular non-coincident time grids, and define invariants of the fractional Brownian motions [17].

9.2.1 Representing time series as a tree

Here we outline the construction of the hierarchical tree M_X. Its zero level (root of the tree) is formed by the best least-square linear approximation $L_0(t)$ of $X(t)$ over the entire observational interval $[t_1, t_N]$; thus $L_0(t)$ consists of a single linear segment. We define the approximation error for $L_0(t)$ as

$$E_0 = \sum_{i=1}^{N} \left|X(t_i) - L_0(t_i)\right|^2. \tag{9.8}$$

At the next step we want to evaluate the most prominent deviations of $X(t)$ from the global linear fit $L_0(t)$ and depict them by the first-level approximation, $L_1(t)$; it will consist of N_1 linear segments and have the approximation error E_1 defined similar to (9.8). Here we face a two-fold problem: how to determine the optimal N_1, and how to construct the optimal

piece-wise linear approximation with N_1 segments. We only discuss here the theoretical approach to resolving these issues; the details of a corresponding computationally effective algorithm are given in [17].

We introduce the quality measure for an arbitrary piecewise linear approximation $L(t;\, n,\, e)$ consisting of $n > 1$ linear segments and having total fitting error e:

$$H(n, e) = \frac{\log(e/E_0)}{n-1}. \tag{9.9}$$

We define $L_1(t)$ as the approximation $L(t; N^*, E^*)$ that maximizes $H(n, e)$:

$$H(N^*, E^*) = \max_{n,e} H(n, e). \tag{9.10}$$

In the plane $(n,\, \log(e/E_0))$ the global linear approximation $L_0(t)$ corresponds to the point $p_0 = (1, 0)$ whereas an arbitrary piecewise approximation $L_k(t)$ corresponds to the point $p_k = (N_k,\, \log(E_k/E_0))$, $N_k > 1$, $E_k < E_0$. The slope of the linear segment $[p_0, p_k]$ shows the increase of the fitting quality per one additional segment of approximation. By the criterion (9.9)-(9.10) we choose the approximation with the maximal quality increase.

Remark 9.1 The criterion (9.9)-(9.10) can be compared to the famous Akaike Information Criterion (AIC) widely used to resolve trade-offs between quality and complexity of statistical models. In our settings, the AIC would minimize the score

$$AIC = \log(e) + \alpha\, n \to \min \tag{9.11}$$

with the common choice $\alpha = 2/N$. In contrast to AIC, we do not specify α in advance but instead look for such α that resolves (9.11) with minimal $n > 1$. In other words we choose the AIC-optimal decomposition with minimal number of segments.

To construct the next level of the hierarchy, $L_2(t)$, we apply the above procedure to one of N_1 segments of the first-level approximation $L_1(t)$, and leave all other segments unchanged. The choice of the segment to be decomposed is dictated by the maximization of the corresponding error decrease $E_1 - E_2$. The N_2 segments of this approximation comprise the second level of the hierarchy. Similarly, at each consecutive step, we apply the above procedure to one of the segments of the current approximation $L_k(t)$ to form the next approximation $L_{k+1}(t)$. By construction, each segment of the approximation $L_{k+1}(t)$ is embedded into or coincides with an appropriate segment from $L_k(t)$; this is why they form a well-defined hierarchy.

As a result, we obtain a series of piece-wise linear approximations $L_k(t)$, $k = 0, 1, \ldots, d$ of $X(t)$. The approximations are of increasing detail: the larger the index k the larger the number N_k of linear segments within $L_k(t)$, the smaller the fitting error E_k, and the smaller their average duration.

9.2.2 MTA spectrum

Important element of our hierarchical analysis is the MTA spectrum: a graph showing the fitting error E_k of approximation $L_k(t)$ as a function of the number N_k of its linear segments. By definition, this function is monotonously decreasing. For purely self-affine time series characterized by a single Hurst exponent H the error E_k and N_k are connected via [17]

$$E_k = E_0 \, N_k^{-2H}. \tag{9.12}$$

This power-law relationship reflects the absence of characteristic scale in agreement with the well-known general properties of self-affine time series [13]. Often it is advantageous to consider the dependence of E_k on N_k in bi-logarithmic scale, where the relation (9.12) takes a linear form:

$$\log E_k = \log E_0 - 2H \, \log N_k.$$

Noticeably, a typical log-price trajectory that obeys (2.1) is not a pure self-affine scale-free series. The volatility jumps create a characteristic scale in the price dynamics: The MTA spectrum is governed by the volatility structure while we consider long trends (longer than the average duration of intervals of constant volatility); and by pure Brownian motion at short trends. Accordingly, a typical MTA spectrum for the observed trajectories is not linear but 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 ($H = 1/2$).

Generally, a break of the MTA slope from a higher to a lower absolute value characterizes time series that describe a well-shaped motion accompanied by a random noise; further detail and supporting numerical simulations can be found in [17]. The existence of a corner point signals that the dynamics of a trajectory changes at this scale; hence we interpret the scales larger than k_0 (larger segment durations) as corresponding to the volatility structure, while smaller to the random noise. The corner point can be formally defined as the point of the maximal deviation from the best linear approximation to the spectrum.

References

[1] 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.

[2] 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.

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

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

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

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

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

[8] Gourieroux, C., (1997), *A*RCH Models and Financial Applications, Springer.

[9] Gui, X. and Q. Zhang (2004), Closed-form solutions for perpetual American put options with regime switching. *Stochastic Processes and their Applications*, 109(2), 167–187.

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

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

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

[13] Mandelbrot, B., 1982. The Fractal Geometry of Nature. W. H. Freeman & Co., New York, xii + 461 + xvi pp.

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

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

[16] Runggaldier, W.J., (2004), 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. .

[17] Zaliapin, I., A. Gabrielov, and V. Keilis-Borok (2004), Multiscale Trend Analysis. *Fractals*, **12**, 275-292. doi:10.1142/S0218348X04002604.

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

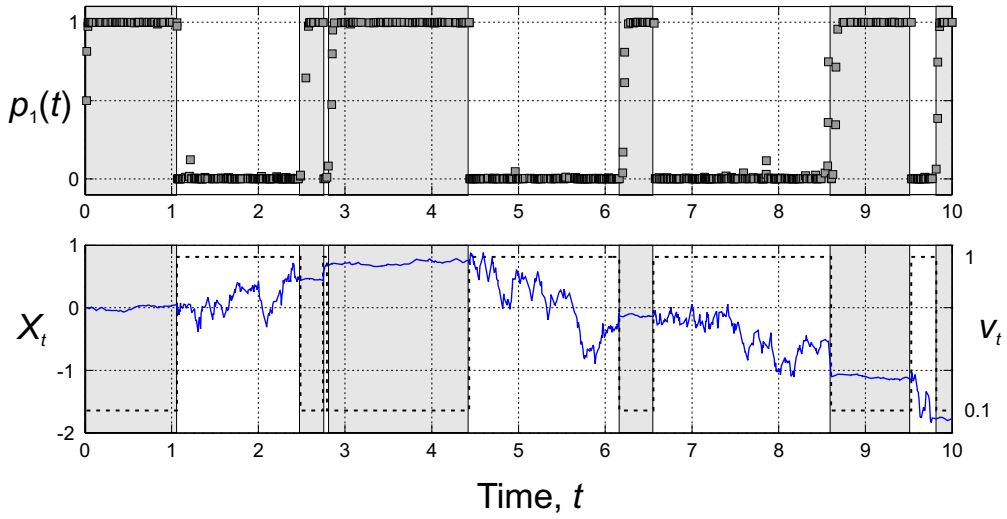

Figure 1: Filtering two-valued volatility. The posterior probability $p_1(t) = P(v_t = a_1)$ is shown in the top panel. Bottom panel shows the observed process X_t (solid line) and unobserved volatility v_t (dashed line). Volatility alphabet is $\{a_1 = 0.1, a_2 = 1.0\}$; jump intensities are $\lambda_{12} = \lambda_{21} = 1$; observational intensities are $n(a_1) = 50$, $n(a_2) = 100$. Shaded intervals correspond to $v_t = 0.1$. Drift $\mu = 0.05$; if it is interpreted as 5% per year then the conditional time units are years.

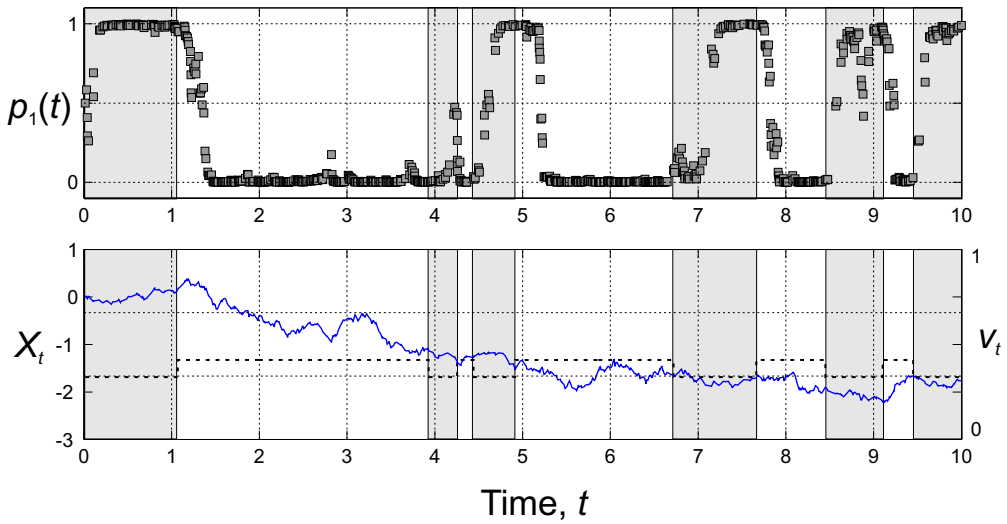

Figure 2: Filtering two-valued volatility. Notations are the same as in Fig. 1. Volatility alphabet is $\{a_1 = 0.3, a_2 = 0.4\}$; jump intensities are $\lambda_{12} = \lambda_{21} = 1$; observational intensities are $n(a_1) = 50$, $n(a_2) = 100$. Shaded intervals correspond to $v_t = 0.3$.

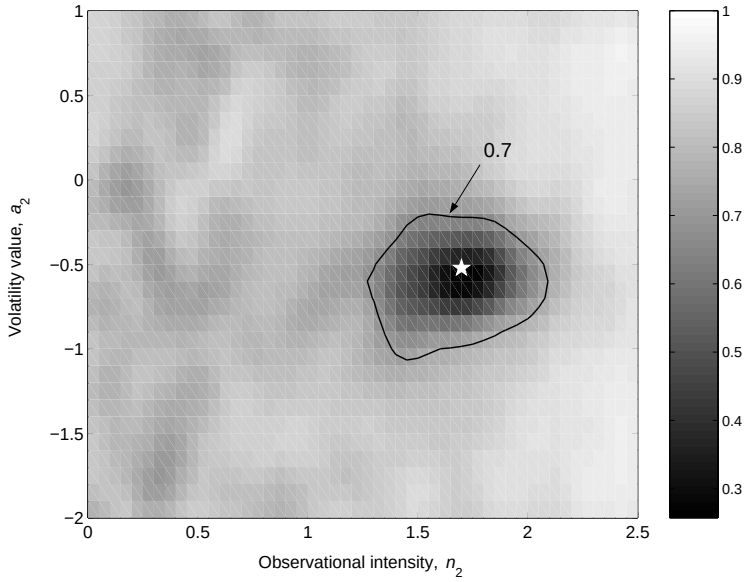

Figure 3: Filter performance. Quality of the filter estimation measured by the score S_{LR} for two-valued model described in Sect. 4. An ideal estimation would yield $S_{\mathrm{LR}} = 1$, a random guess $S_{\mathrm{LR}} = 0$. A level $S_{\mathrm{LR}} = 0.7$ is shown by the solid line.

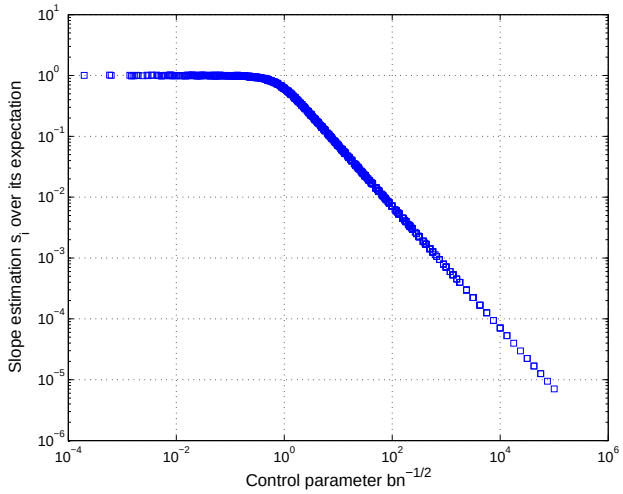

Figure 4: Slope approximation s_i (5.5) over the Monte Carlo slope expectation as a function of parameter $|b|/\sqrt{n}$. Clearly, this parameter controls the quality of approximation.

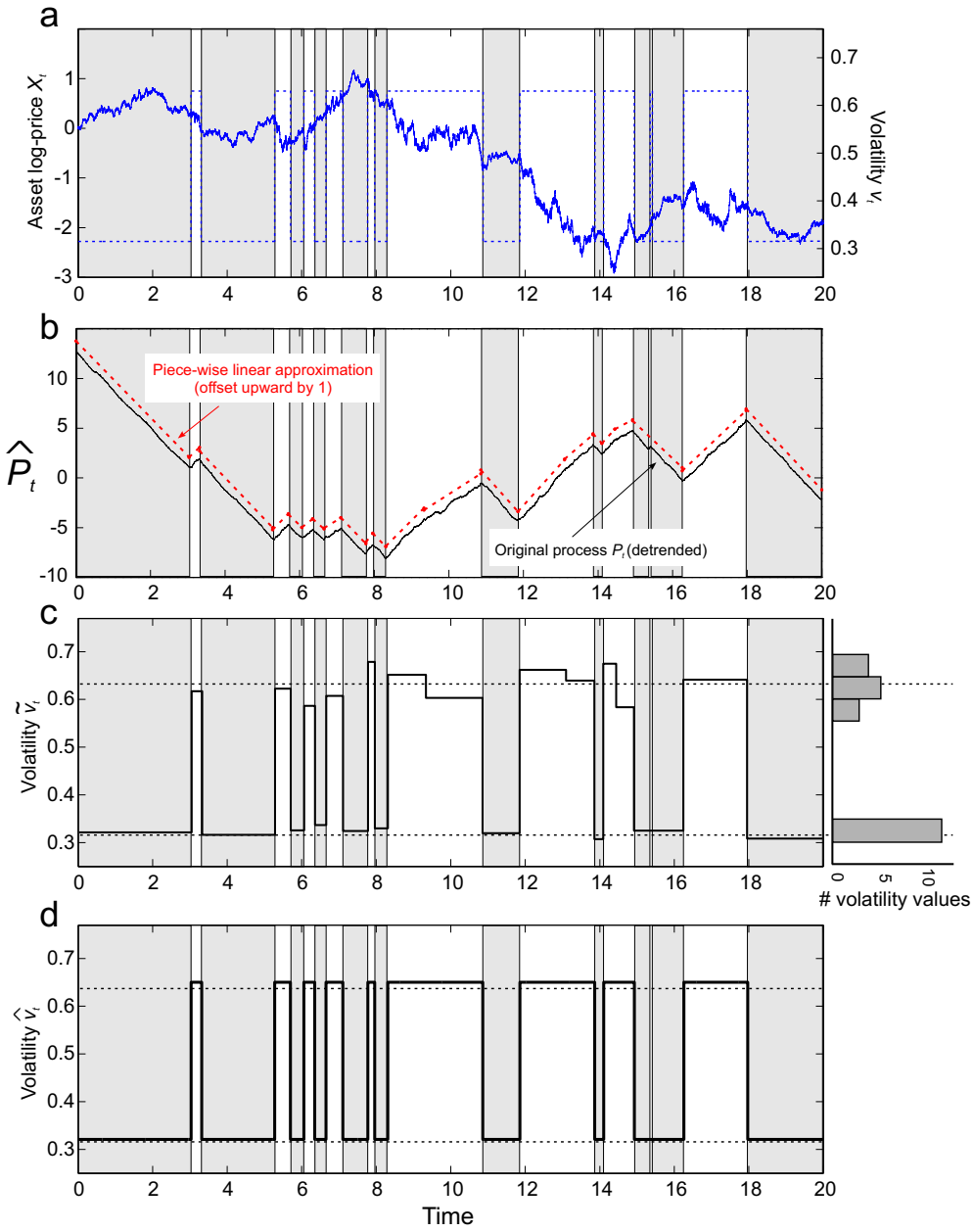

Figure 5: Example of estimating a priori values of filter input parameters. a) Asset log-price X_t (solid line, left axis) and its two-valued volatility v_t (dashed line, right axis). Parameters of the process are $\{a_i\} \approx \{0.316, 0.632\}$, $\mu = 0.05$, $\lambda_{12} = \lambda_{21} = 1$, $n(a_1) = n(a_2) = 10^3$, $p_i = 1/2$. 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. 6). 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 $v_t = a_1 \approx 0.316$.

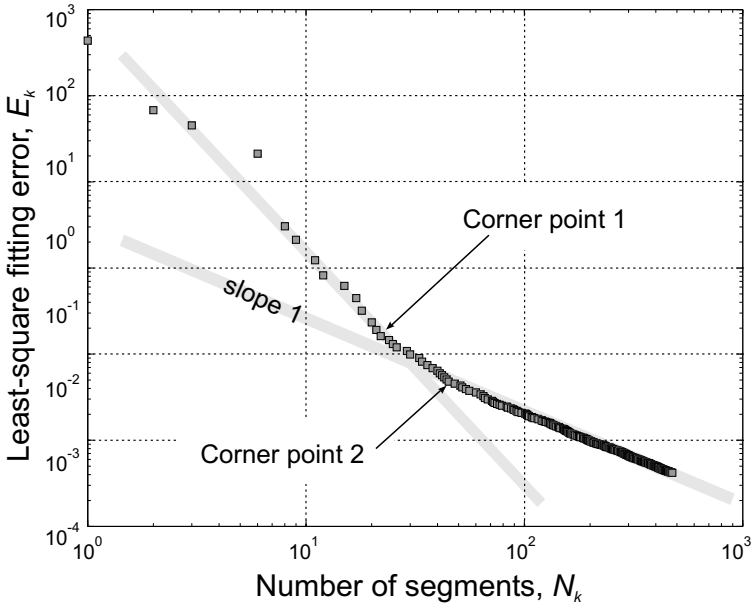

Figure 6: MTA spectrum for the process illustrated in Fig. 5a. Shaded lines depict two scaling regions with the transition zone between two *corner points* marked in the figure. The right scaling region has the slope -1, which corresponds to a self-affine random walk with no persistence. The left region 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. 5b.

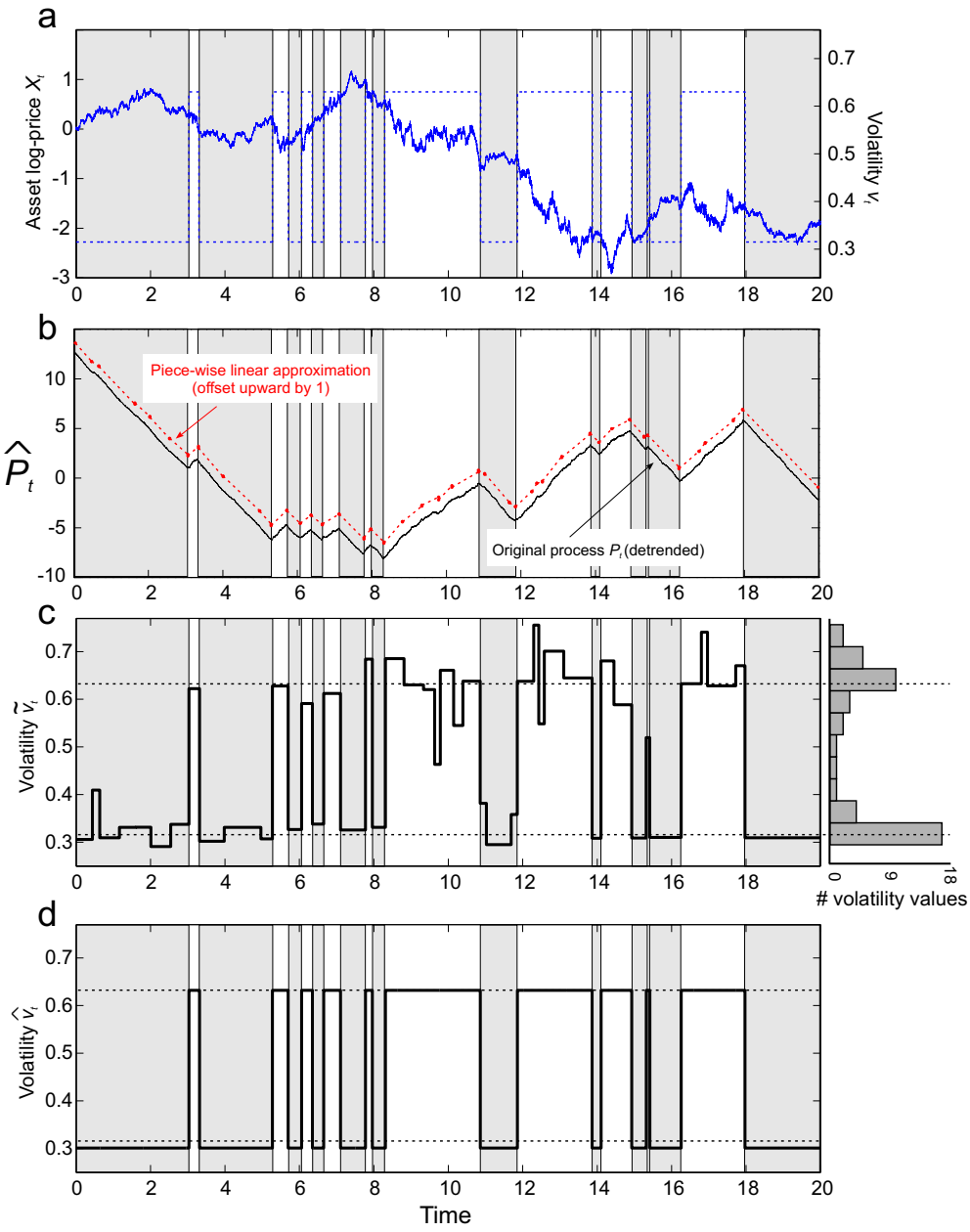

Figure 7: Example of estimating a priori values of filter input parameters. Notations are the same as in Fig. 5 except that analysis corresponds to the corner point 2 of the MTA spectrum (Fig. 6).

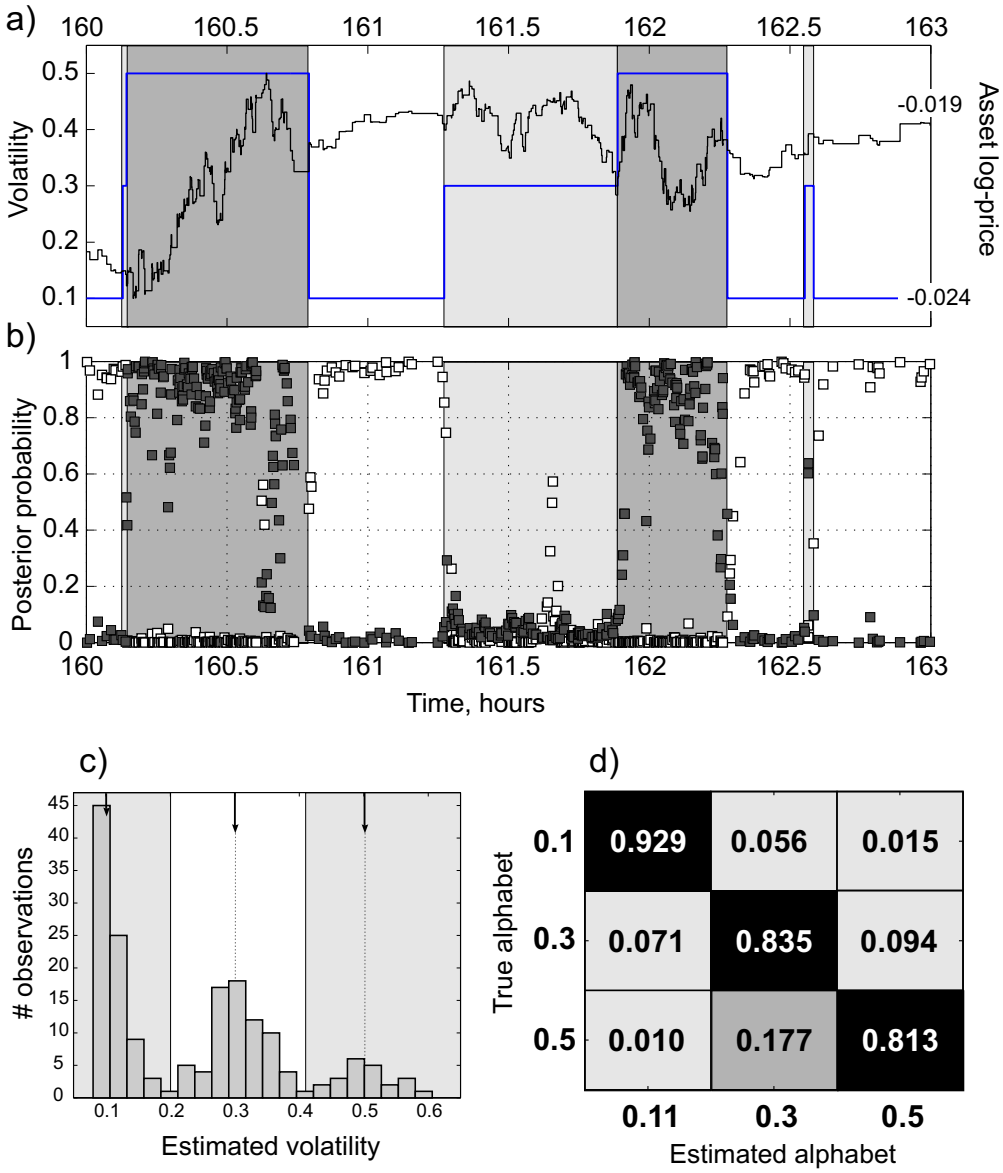

Figure 8: Filtering synthetic asset price. a) Asset log-price X_t (right axis) and true unobserved volatility v_t (left axis). Distinct volatility values are depicted by shadows: dark for $v_t = 0.5$, light for $v_t = 0.3$, none for $v_t = 0.1$. b) Aposteriori probabilities $p_3(t)$ (dark squares) and $p_1(t)$ (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 $\widehat{a}_i$. d) Empirical frequencies q_{ij} of Eq.(4.2). An ideal estimation would give a pure diagonal matrix. Our estimation is indeed very good.

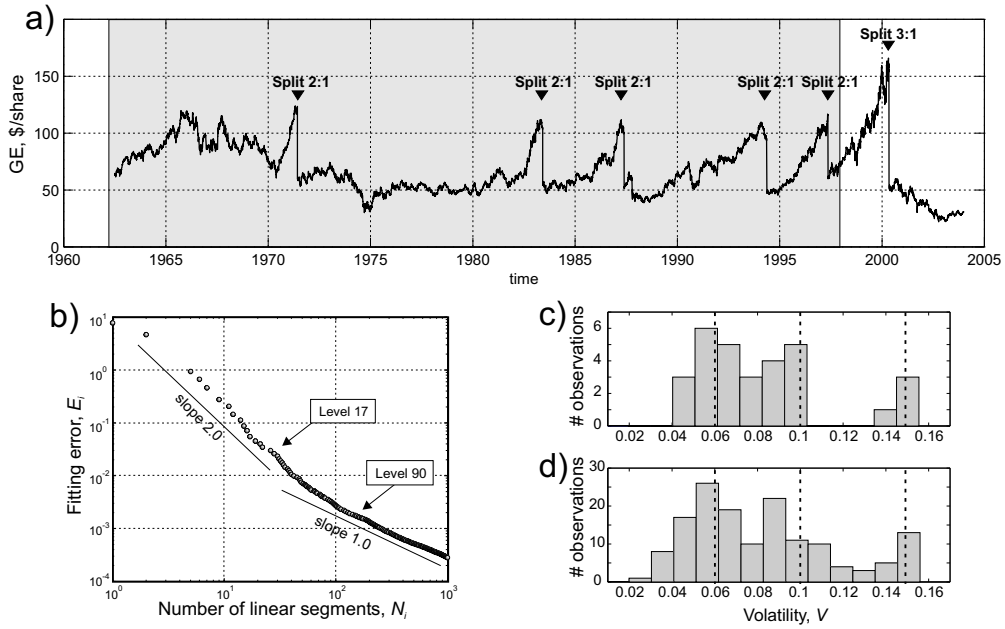

Figure 9: Estimating volatility for General Electric company during 1962-1998. a) Asset price S_t during 1962-2004; market splits are depicted by solid arrows. The shaded interval 1962-1998 is used for alphabet estimation. b) MTA spectrum for the process P_t that corresponds to GE log-price dynamics. 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.)

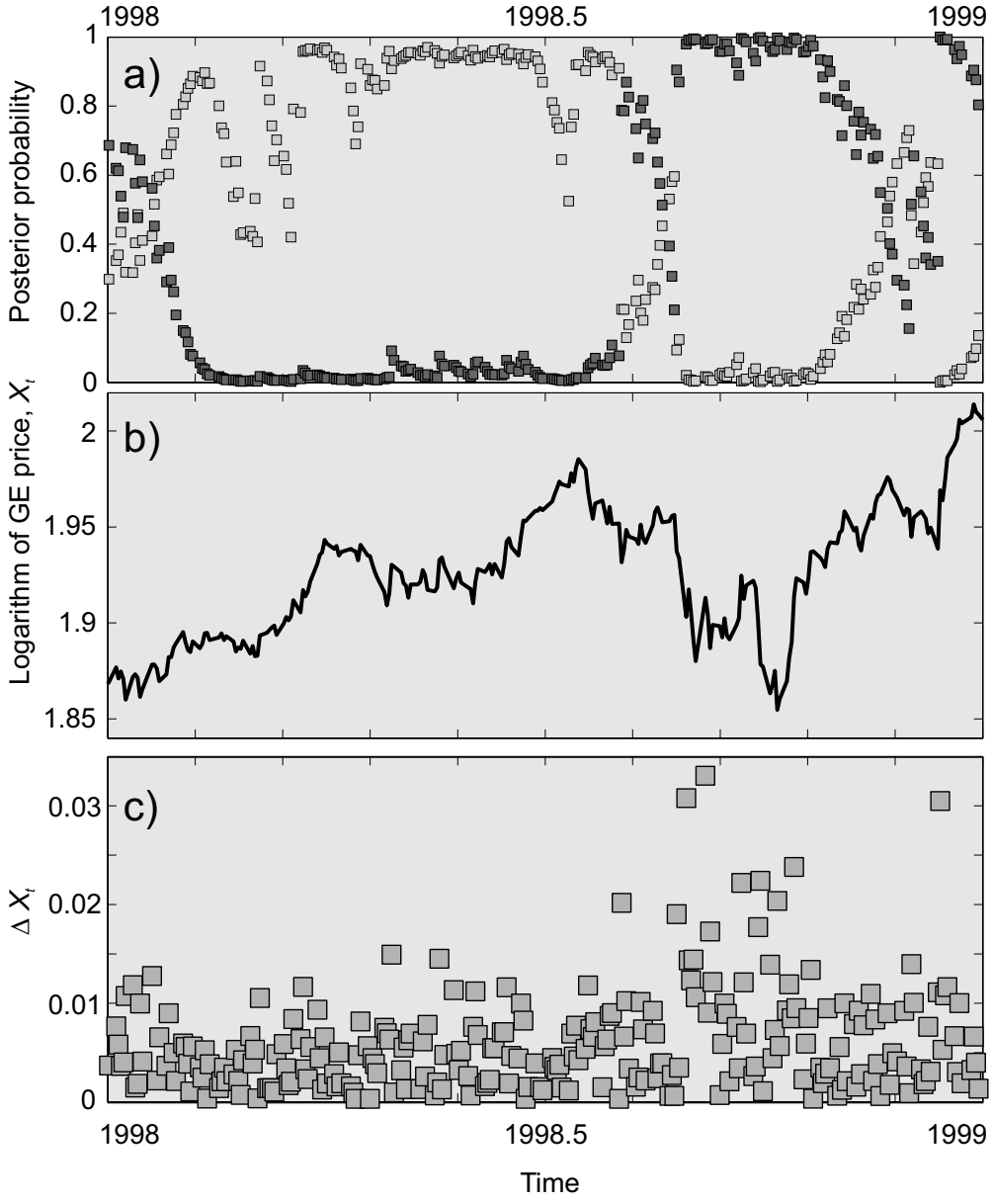

Figure 10: Estimating volatility for General Electric company during 1998-1999. a) Posterior probabilities $p_i(t)$, $i = 2$ (light squares) and $i = 3$ (dark squares) that correspond to volatility values $v_2 = 0.1$ and $v_3 = 0.15$. b) Dynamics of the log-price X_t. c) Absolute returns $|\Delta_t|$ of the log-price X_t.

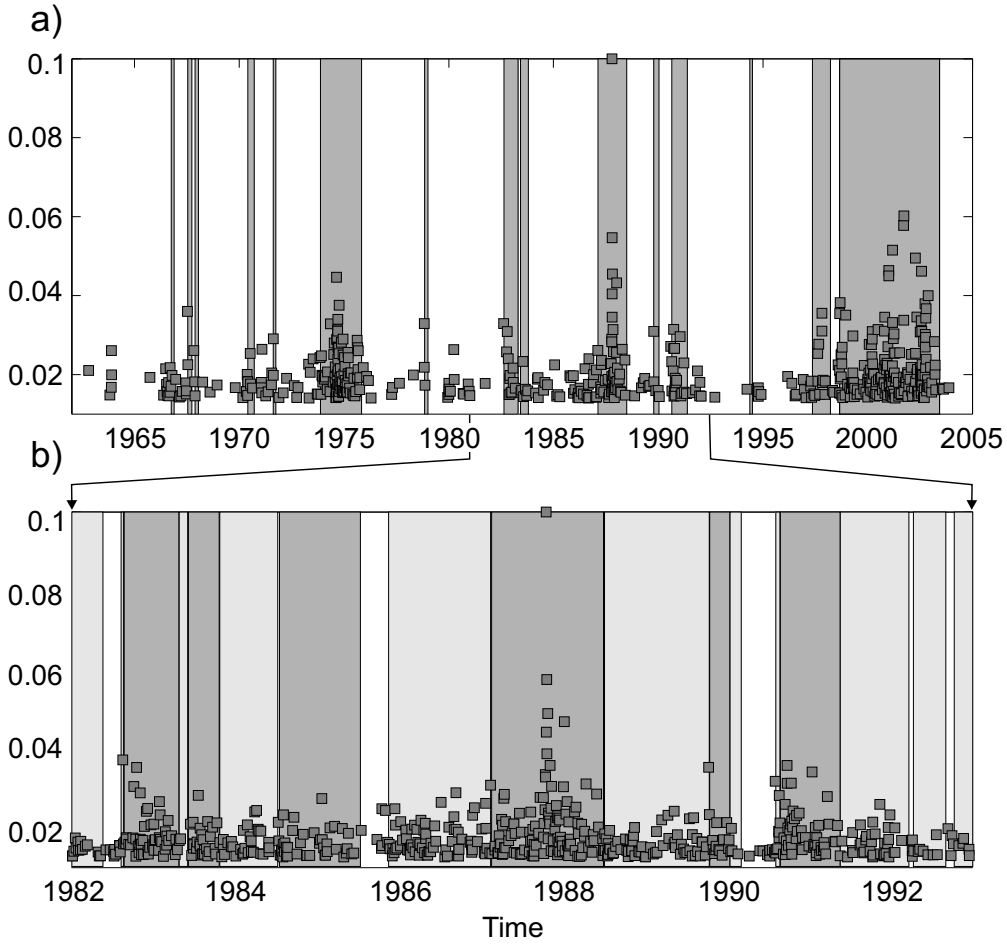

Figure 11: Filtering volatility for General Electric company during 1964-2004. a) Intervals with the largest value of the posterior volatility ($\widehat{v_t} = 0.15$) (shaded intervals) and absolute returns $|\Delta_t| > 0.015$ of the log-price (squares). One observes the largest returns only within the intervals recognized as high-volatility. b) Zoom of the interval 1982-1992, around the market crash of Oct. 19, 1987. Dark shadow depicts intervals of the largest posterior volatility ($\widehat{v_t} = 0.15$), light shadow — medium ($\widehat{v_t} = 0.1$), no shadow — lowest ($\widehat{v_t} = 0.06$).

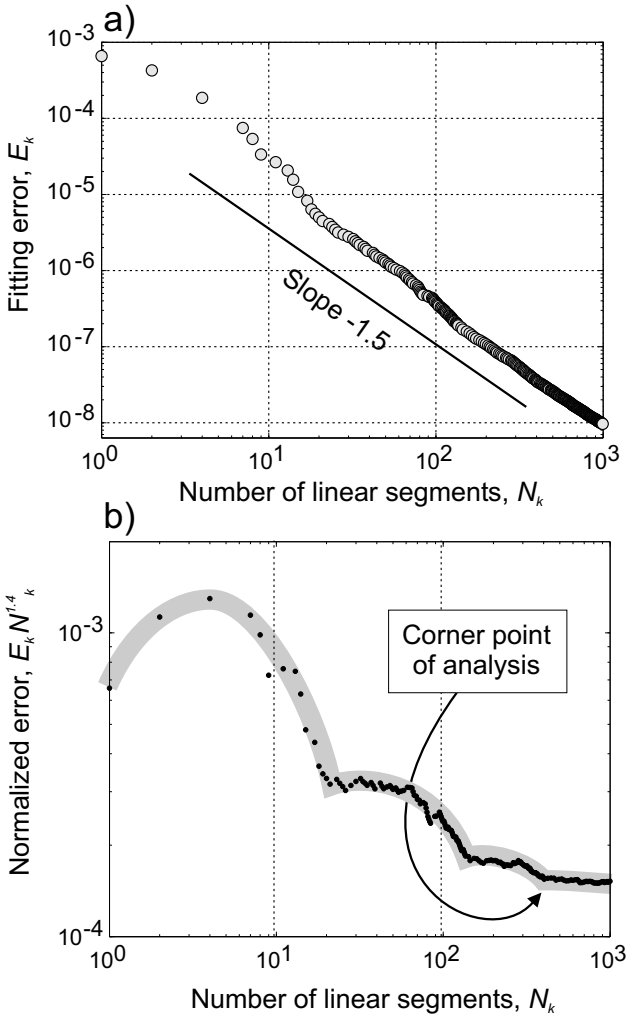

Figure 12: Estimating volatility for IBM company during November 1, 1990 – January 11, 1991. a) MTA spectrum suggests a coarsely self-similar structure with Hurst exponent $H = 1.5$; b) Renormalized spectrum – $E_k \times N_k^{1.4}$ as a function of N_k. The gray strip depicts existence of three corner points. We choose the rightmost one for our analysis. See Sect. 7.3 for details.

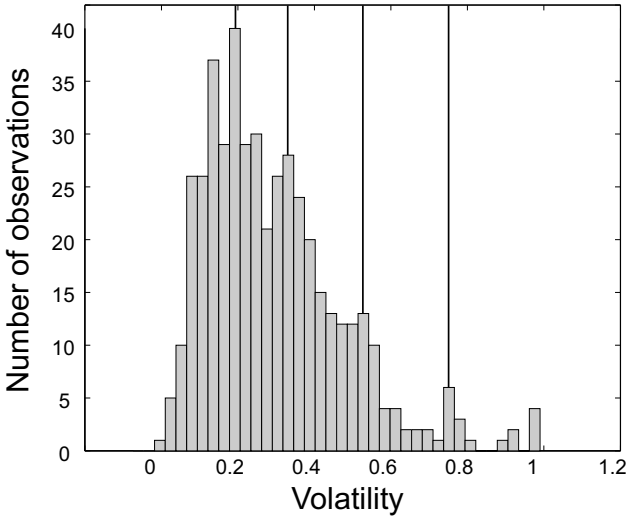

Figure 13: Estimating volatility for IBM company during November 1, 1990 – January 10, 1991. Histogram of initial estimates of volatility alphabet values $\widehat{a}_i$, $i = 1, \ldots, N_{k_0} = 493$, $k_0 = 300$ that correspond to the corner point shown in Fig. 12b.

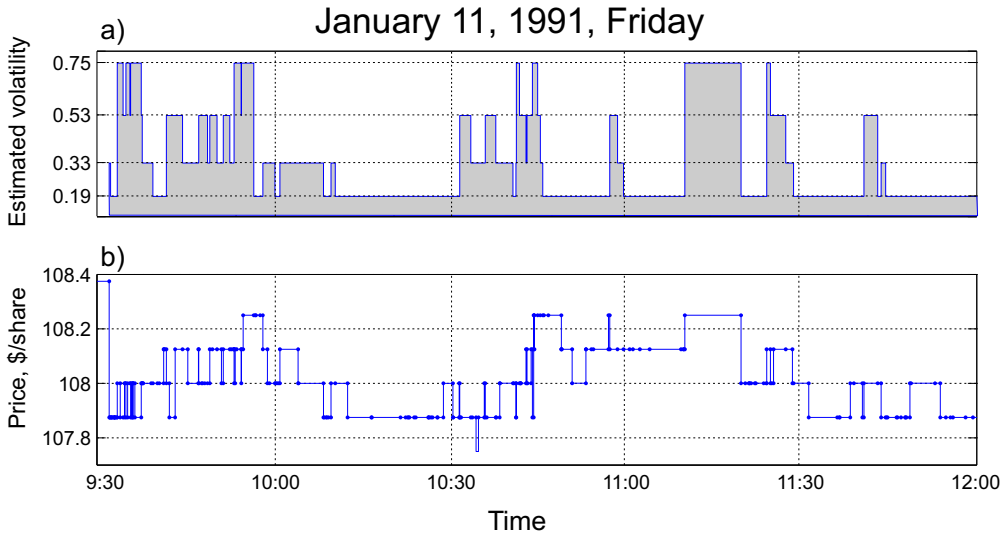

Figure 14: Filtering volatility for IBM company during January 11, 1991. a) Posterior volatility $\widehat{v}_t$ of Eq. (7.2); b) Price dynamics during the same time interval. See discussion in Sect. 7.3.

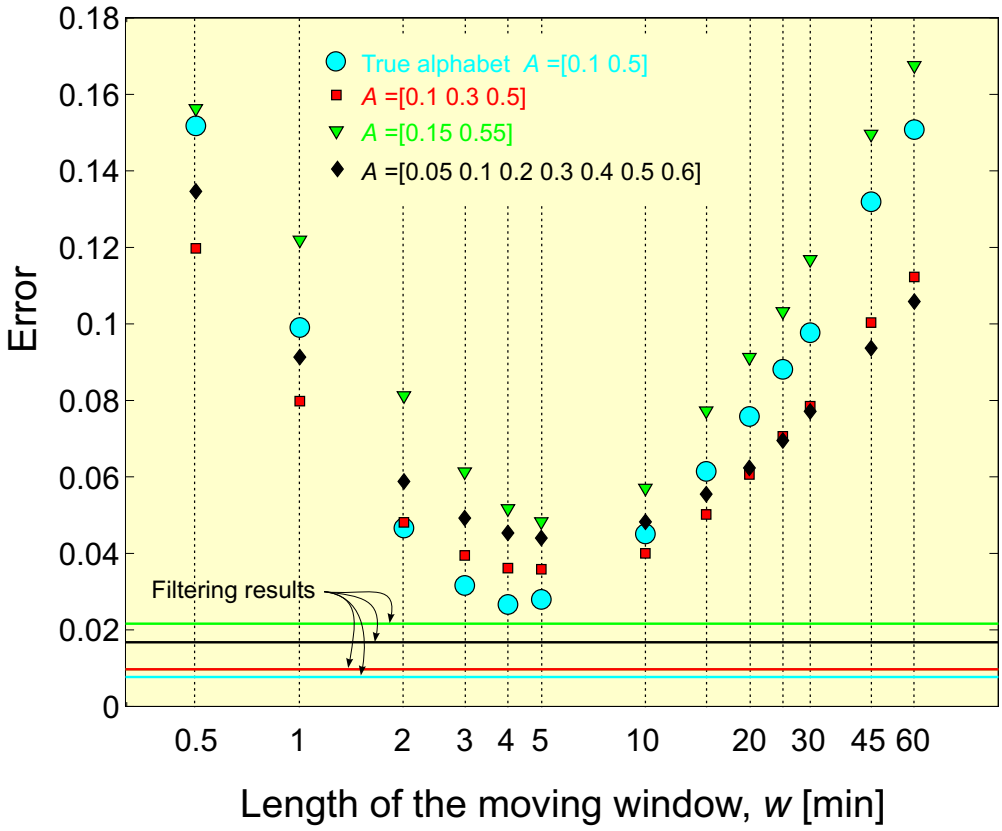

Figure 15: Comparison between alternative volatility estimations. Symbols show the estimation error of an *ad hoc* procedure based solely on absolute returns as a function of its parameter w; different symbols refer to different apriori alphabet. Horizontal lines indicate corresponding errors for the filtering estimate. See section 8 for details.