Problems of Information Transmission, Vol. 41, No. 3, 2005, pp. 212–229. Translated from Problemy Peredachi Informatsii, No. 3, 2005, pp. 32–50.

LARGE SYSTEMS

Tracking Volatility

L. Goldentayer*, F. Klebaner, and R. Sh. Liptser*****

**Electrical Engineering Systems, Tel Aviv University, Israel*
goldlev@hotmail.com
***School of Mathematical Sciences, Monash University, Clayton, Melbourne, Australia*
fima.klebaner@sci.monash.edu.au
****Electrical Engineering Systems, Tel Aviv University, Israel*
Institute for Information Transmission Problems, RAS, Moscow
liptser@eng.tau.ac.il

Received September 3, 2004; in final form, February 24, 2005

Abstract—We propose an adaptive algorithm for tracking historical volatility. The algorithm borrows ideas from nonparametric statistics. In particular, we assume that the volatility is a several times differentiable function with a bounded highest derivative. We propose an adaptive algorithm with a Kalman filter structure, which guarantees the same asymptotics (well known from statistical inference) with respect to the sample size n, $n \to \infty$. The tuning procedure for this filter is simpler than for a GARCH filter.

1. INTRODUCTION

1.1. In the classical Black–Scholes model for financial markets, the stock price $S(t)$ is modeled as a geometric Brownian motion with the diffusion coefficient $\sqrt{v}S(t)$, where *volatility*, v, is assumed to be constant. This assumption is convenient for the "prediction" of the option price.

Contrary to this assumption, traders treat the volatility as a parameter that changes with time and whose future values have to be evaluated (predicted) for a given period of interest.

In this context, many researchers would rather interpret the volatility as a random process, $v(t)$, and study stochastic volatility models. It is natural to verify how the volatility $v(t)$ changes in time for real stock prices and attempt to select a suitable stochastic volatility model. Traditionally, it is proposed to apply generalized autoregressive conditional heteroscedasticity (GARCH) tracking algorithms (see, e.g., [1–8]) for tracking $v(t)$ from stock prices. It is known [8, p. 109] that a GARCH algorithm operates satisfactory under relatively stable market conditions but fails when highly unanticipated events leading to a significant structural change occur. Nevertheless, in many realistic settings, the simplest GARCH(p,q) algorithms, $p,q = 1,2$, are adequate for tracking volatilities even over long periods (see [9, pp. 10 and 22]). The main difficulty in implementation of GARCH comes from the multivariate minimization procedure of its parameters even for small values of $p,q = 1,2$.

In this paper, we propose a new approach for tuning the GARCH parameters. Our approach uses ideas from nonparametric statistics combined with the Kalman–Bucy filter representation of the GARCH model. This representation enables us to select a GARCH model with only one tuning parameter.

1.2. Let $S(t)$ be a stock price defined by the Black–Scholes model (see [3, 4])

$$dS(t) = \mu(t)S(t)\,dt + \sqrt{v(t)}S(t)\,dB_t, \quad 0 \le t \le T,$$

where B_t is a Brownian motion, $S(0)$ is the initial stock price, and the parameters $\mu(t)$ and $v(t)$ (*volatility*) are strictly positive deterministic functions.

Let $S(t_i)$, $i = 0, 1, \ldots, n$, be a sample, with $t_i - t_{i-1} = \dfrac{T}{n}$ $(\equiv: \Delta)$, where $t_0 = 0$. Denote by

$$X_i = \frac{1}{\Delta} \ln^2 \left(\frac{S(t_i)}{S(t_{i-1})} \right)$$

the observed data. By Itô's formula, we obtain

$$\ln \left(\frac{S(t_i)}{S(t_{i-1})} \right) = \int\limits_{t_{i-1}}^{t_i} 0.5(2\mu(s) - v(s))\, ds + \int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s,$$

and therefore

$$\begin{aligned} X_i &= \frac{1}{\Delta} \left(\int\limits_{t_{i-1}}^{t_i} 0.5(2\mu(s) - v(s)) ds + \int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s \right)^2 \\ &= \frac{1}{\Delta} \left(\int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s \right)^2 + \frac{1}{\Delta} \left(\int\limits_{t_{i-1}}^{t_i} 0.5(2\mu(s) - v(s))\, ds \right)^2 \\ &+ \frac{2}{\Delta} \int\limits_{t_{i-1}}^{t_i} 0.5(2\mu(s) - v(s))\, ds \int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s. \end{aligned} \tag{1}$$

The parameter Δ is usually small (for example, if stock prices are measured once a day for three consecutive years, then $\Delta \simeq 0.001$). For sufficiently small Δ, the dominating term in X_i is $\dfrac{1}{\Delta} \left(\int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s \right)^2$, with the mean

$$v_{i-1} = \frac{1}{\Delta} \int\limits_{t_{i-1}}^{t_i} v(s)\, ds.$$

Under some smoothness assumptions, the error $\dfrac{1}{\Delta} \int\limits_{t_{i-1}}^{t_i} (v(s) - v_{i-1})\, ds$ is sufficiently small and can be ignored.

Let us recall the structure of the GARCH filter. Following [5, 10], the filter GARCH(p, q) generates estimates $\widehat{v}_i$ of v_i by the recursion

$$\widehat{v}_i = K + \sum_{j=1}^{p} g_j \widehat{v}_{i-j} + \sum_{m=1}^{q} a_m X_{(i+1)-m}, \tag{2}$$

with appropriate initial conditions. The parameters

$$K, g_1, \ldots, g_p, a_1, \ldots, a_q, p, q$$

have to be chosen with the help of minimizing the sum of squares of (observed) innovation differences (recall that $n = \dfrac{T}{\Delta}$)

$$S_n(K, g_1, \ldots, g_p, a_1, \ldots, a_q, p, q) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \widehat{v}_{i-1})^2.$$

In contrast to (2), we propose an alternative tracking algorithm, borrowed from [11, 12], with a univariate minimizing parameter ϑ:

$$\begin{aligned}
\widehat{v}_i &= \widehat{v}_{i-1} + \frac{1}{n}\widehat{v}^{(1)}_{i-1} + \frac{q_0(\vartheta)}{n^{2(k+1)/(2k+3)}}(X_i - \widehat{v}_{i-1}),\\
\widehat{v}^{(j)}_i &= \widehat{v}^{(j)}_{i-1} + \frac{1}{n}\widehat{v}^{(j+1)}_{i-1} + \frac{q_j(\vartheta)}{n^{(2(k+1)-j)/(2k+3)}}(X_i - \widehat{v}_{i-1}), \quad j = 1,\ldots,k-1,\\
\widehat{v}^{(k)}_i &= \widehat{v}^{(k)}_{i-1} + \frac{q_k(\vartheta)}{n^{(k+2)/(2k+3)}}(X_i - \widehat{v}_{i-1}),
\end{aligned} \tag{3}$$

where $q_i(\cdot)$, $i = 0, 1, \ldots, k$, are some known functions (see (9) below) and k is defined in accordance with the "smoothness" of $v(t)$ (see (4)).

Remark 1. To avoid misunderstanding, we note that, for $k = 0$, equations (3) turn into

$$\widehat{v}_i = \widehat{v}_{i-1} + \frac{q_0(\vartheta)}{n^{2/3}}(X_i - \widehat{v}_{i-1}).$$

Initial conditions for (3) can be arbitrary (bounded). The unique parameter ϑ has to be chosen so that to minimize the innovation difference

$$S_n(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{v}_{i-1})^2.$$

It should be noted that filter (3) is also of the GARCH type. In Section 3, we give a modification of (3), which is compatible with GARCH(p, q).

For large sample size n, filter (3) is stable and admits a simple optimization procedure.

2. DESCRIPTION OF THE ESTIMATOR. QUALITY OF ESTIMATION

2.1. Assumptions and Preliminaries

We use ideas of nonparametric statistics for estimating a smooth function observed in the presence of white noise. In order to be in the framework of the above-mentioned conventions and ideas, the volatility $v(t)$ is assumed to be a smooth function. (For the setting where $v(t)$ is a piecewise constant function, a different adaptive procedures is required, see [13]).

In this paper, following [14–16], we choose the following family of functions $(f(t))_{0\le t\le T}$ (here, $f^{(0)} \equiv f$):

$$\Sigma(\beta, L) = f \left\{ \begin{array}{l} k \text{ times differentiable, } f^{(0)}, f^{(1)}, \ldots, f^{(k)};\\ |f^{(k)}(t_2) - f^{(k)}(t_1)| \le L|t_2 - t_1|^{\alpha},\ \forall\, t_1, t_2 \le T \text{ and } \alpha \in (0, 1];\\ \beta = k + \alpha,\ k = 0 \text{ is included.} \end{array} \right. \tag{4}$$

This family has appeared in various (nonstatistical) researches, while for the estimation of the regression function the family $\Sigma(\beta, L)$ appeared for the first time in [14, 15].

Assumption 1. *The volatility* $(v(t))_{0\le t\le T}$ *is a strictly positive bounded function from* $\Sigma(\beta, L)$ *with* $\beta = k + 1$.

An example of the volatility function given in Fig. 1 corresponds to the US dollar/ruble exchange rate. It is seen that fast oscillations with small amplitudes of the price provide jumps of the volatility.

Assumption 2. *The function* $\mu(t)$ *is bounded and positive.*

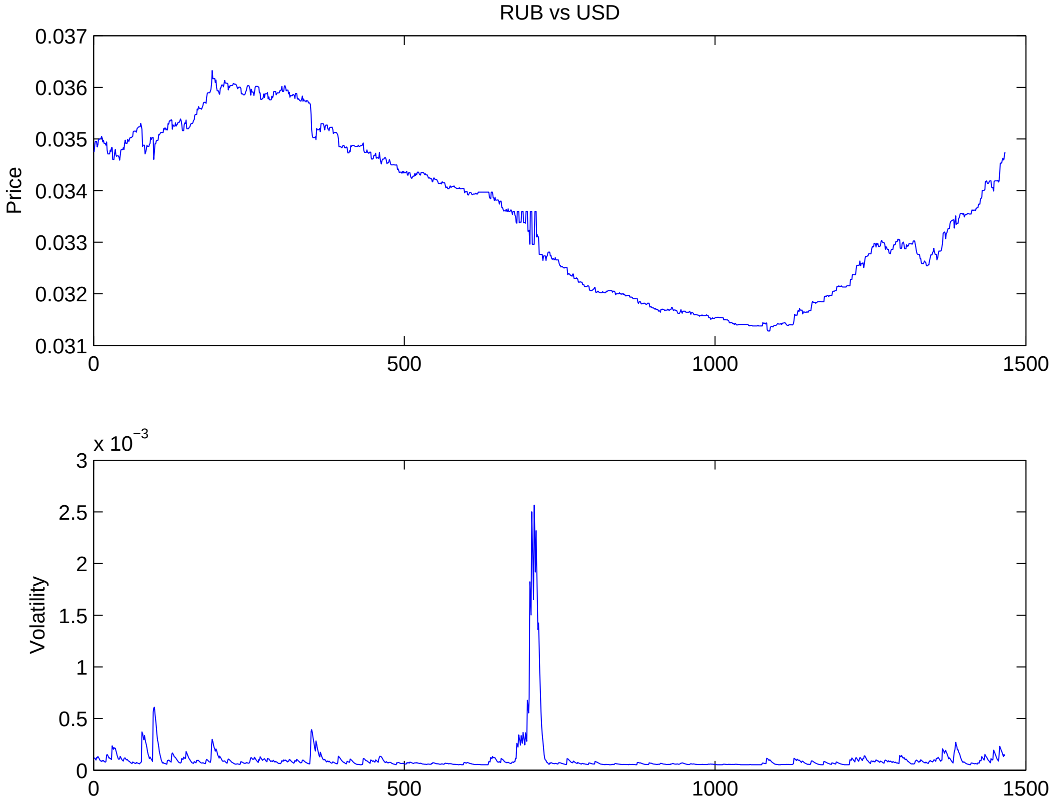

Fig. 1. Example of the volatility of the US dollar/ruble exchange rate.

We denote

$$\mu_{i-1} = \frac{1}{\Delta} \int\limits_{t_{i-1}}^{t_i} \mu(s)\, ds$$

and introduce

$$\xi_i = \frac{1}{\sqrt{v_{i-1}\Delta}} \int\limits_{t_{i-1}}^{t_i} \sqrt{v(s)}\, dB_s, \quad i \geq 1.$$

The sequence $(\xi_i)_{i\geq 1}$ is a sequence of i.i.d. $(0,1)$-Gaussian random variables. From (1), it follows that

$$X_i = 0.25\big(2\mu_{i-1} - v_{i-1}\big)^2\Delta + 2\sqrt{\Delta v_{i-1}}\Big(\mu_{i-1} - \frac{1}{2}v_{i-1}\Big)\xi_i + v_{i-1}\xi_i^2.$$

We denote

$$\eta_i = \sqrt{\Delta v_{i-1}}(2\mu_{i-1} + v_{i-1})\xi_i + v_{i-1}(\xi_i^2 - 1),$$
$$\theta_i(\Delta) = 0.25\Delta\big(2\mu_{i-1} + v_{i-1}\big)^2.$$

By Assumptions 1 and 2, we have $\theta_i(\Delta) = O(\Delta)$, while $(\eta_i)_{i\geq 1}$ is a zero-mean sequence of uncorrelated random variables with

$$\mathbf{E}\,\eta_i^2 = \Delta v_{i-1}\big(2\mu_{i-1} + v_{i-1}\big)^2 + 2v_{i-1}^2 =: \sigma_i^2,$$

where σ_i^2 are strictly positive and bounded numbers.

Thus,

$$X_i = v_{i-1} + \eta_i + \theta_i(\Delta). \tag{5}$$

Throughout the paper, we use inequalities valid up to a positive constant; this unspecified constant will always be denoted by $\mathfrak{a}$.

2.2. A Simplified Model for X_i

In this section, we replace (5) by a simpler model:

$$X_i = v_{i-1} + \eta_i. \tag{6}$$

It is known from [14, 15] (see also [16]) that there exists a kernel-type estimate $\widehat{v}_i$ of v_i, computed via $(X_i)_{1\le i\le n}$, such that, for any i,

$$\mathbf{E}(v_i - \widehat{v}_i)^2 = O\big(n^{-2(1+k)/(2k+3)}\big). \tag{7}$$

It is also known from [11] that the rate in $n\to\infty$ remains valid for the on-line estimate obtained by the recursion algorithm

$$\begin{aligned}
\widehat{v}_i &= \widehat{v}_{i-1} + \frac{1}{n}\widehat{v}^{(1)}_{i-1} + \frac{q_0}{n^{2(k+1)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}\big),\\
\widehat{v}^{(j)}_i &= \widehat{v}^{(j)}_{i-1} + \frac{1}{n}\widehat{v}^{(j+1)}_{i-1} + \frac{q_j}{n^{(2(k+1)-j)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}\big), \quad j = 1,\dots,k-1,\\
\widehat{v}^{(k)}_i &= \widehat{v}^{(k)}_{i-1} + \frac{q_k}{n^{(k+2)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}\big),
\end{aligned} \tag{8}$$

in which parameters $q_0, q_1, \dots, q_k$ satisfy the following condition:

Assumption 3. *All roots of the characteristic polynomial*

$$p^k(\lambda, \boldsymbol{q}) = \lambda^k + q_0\lambda^{k-1} + \ldots + q_{k-1}\lambda + q_k$$

are different and have negative real parts.

More exactly, the above-mentioned rate in n is preserved outside the boundary layer $i \ge \mathfrak{a}\big(n^{-1/(2k+3)}\log n\big)$, resulting from uncertainty in the initial conditions for (8) (see [11]).

2.3. Adaptive Estimator Design

Outside the above-mentioned boundary layer $i \ge \mathfrak{a}\big(n^{-1/(2k+3)}\log n\big)$, estimates $(\widehat{v}_i)_{i\ge1}$ generated by (8) obey the following property (see (7)):

$$\varlimsup_{n\to\infty} \sup_{v_i} \mathbf{E}\left(v_i - \widehat{v}_i\right)^2 n^{2(k+1)/(2k+3)} =: C(\boldsymbol{q}),$$

where the supremum is taken over all v_i satisfying Assumption 1. The parameter $C(\boldsymbol{q})$ depends on a filter gain $\boldsymbol{q}$, the vector with entries $q_0, q_1, \dots, q_k$.

Thus, preserving the rate in $n\to\infty$, one can minimize $C(\boldsymbol{q})$ with respect to $q_0, q_1, \dots, q_k$ satisfying Assumption 3. It is clear that this conditional minimization is extremely difficult. Therefore, for the case $\mathbf{E}\,\eta_i^2 \equiv \sigma^2$, the paper [12] considers the conditional minimization procedure of $C(\boldsymbol{q})$ in some restrictive class of $q_0, q_1, \dots, q_k$ satisfying Assumption 3. The choice of this class is due to the fact that (8) is nothing but a Kalman filter compatible with Assumption 3, which, in turn, provides the following parametrization for $q_1, q_2, \dots, q_k$ (see [12]):

$$\begin{aligned} q_0(\vartheta) &= U_{00}\vartheta^{1/k+1}, \\ q_1(\vartheta) &= U_{01}\vartheta^{2/k+1}, \\ &\dots\dots\dots\dots\dots\dots \\ q_k(\vartheta) &= U_{0k}\vartheta^{k/k+1}, \\ q_k(\vartheta) &= U_{0k}\vartheta, \end{aligned} \tag{9}$$

where U_{0j}, $j = 0, 1, \dots, k$, are the entries of the first column of the positive definite matrix U being the unique solution of the algebraic Riccati equation

$$aU + Ua^* + B - UA^*AU = 0$$

(* is the transposition symbol) with the matrices

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \end{pmatrix}, \qquad a = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{10}$$

of sizes $1 \times (1+k)$, $(1+k) \times (1+k)$, and $(1+k) \times 1$, respectively. It is known from [12] that

k	U_{00}	U_{01}	U_{02}	U_{03}	U_{04}
0	1	NA	NA	NA	NA
1	$\sqrt{2}$	1	NA	NA	NA
2	2	2	1	NA	NA
3	$\sqrt{4+\sqrt{8}}$	$2+\sqrt{2}$	$\sqrt{4+\sqrt{8}}$	1	NA
4	$1+\sqrt{5}$	$3+\sqrt{5}$	$3+\sqrt{5}$	$1+\sqrt{5}$	1

and that, for $k \le 4$, Assumption 3 holds true for $q_0(\vartheta), q_1(\vartheta), q_2(\vartheta), q_3(\vartheta), q_4(\vartheta)$.

Thus, in [12], we deal with the estimator

$$\begin{aligned} \widehat{v}_i(\vartheta) &= \widehat{v}_{i-1}(\vartheta) + \frac{1}{n}\widehat{v}_{i-1}^{(1)}(\vartheta) + \frac{U_{00}\vartheta^{1/k+1}}{n^{2(k+1)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}(\vartheta)\big), \\ \widehat{v}_i^{(j)}(\vartheta) &= \widehat{v}_{i-1}^{(j)}(\vartheta) + \frac{1}{n}\widehat{v}_{i-1}^{(j+1)}(\vartheta) + \frac{U_{0j}\vartheta^{(j+1)/k+1}}{n^{(2(k+1)-j)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}(\vartheta)\big), \quad j = 1, \dots, k-1, \\ \widehat{v}_i^{(k)}(\vartheta) &= \widehat{v}_{i-1}^{(k)}(\vartheta) + \frac{U_{0k}\vartheta}{n^{(k+2)/(2k+3)}}\big(X_i - \widehat{v}_{i-1}(\vartheta)\big), \end{aligned} \tag{11}$$

with one free parameter ϑ, defining the narrow class $\{\vartheta > 0\}$ where the conditional minimization of $C(\boldsymbol{q})$ is taken.

2.4. Global Adaptation

The filter in (11), corresponding to the simple model (6), is also compatible with the general one from (5):

$$X_i = v_{i-1} + \eta_i + \theta_i(\Delta).$$

In the lemma (see Appendix 1), we show that the asymptotic

$$\mathbf{E}(v_i - \widehat{v}_i)^2 = O\big(n^{-2(1+k)/(2k+3)}\big), \quad n \to \infty,$$

valid for the simple model, is also preserved for the general one.

The univariate minimization of $C(\boldsymbol{q}(\vartheta))$ with respect to ϑ guarantees the validity of Assumption 3. However, since the variance of the noise is not constant and, moreover, is unknown, a prior evaluation of $C(\boldsymbol{q}(\vartheta))$, as in [12], would be difficult. Therefore, we follow the GARCH technique adaptive method (see, e.g., [2]). In other words, we shall evaluate

$$V_n(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}(v_{i-1} - \widehat{v}_{i-1}(\vartheta))^2.$$

To this end, we use estimates $\widehat{v}_0, \widehat{v}_1, \ldots, \widehat{v}_{n-1}$ related to the sample $X_1, \ldots, X_n$ and compute

$$S_n(\vartheta) = \frac{1}{n}\sum_{i=1}^{n}\big(X_i - \widehat{v}_{i-1}(\vartheta)\big)^2.$$

We show that, for sufficiently large n, $S_n(\vartheta)$ is a consistent estimate for $V_n(\vartheta)$ in the sense that, as $n \to \infty$, $S_n(\vartheta') > S_n(\vartheta'') \Rightarrow V_n(\vartheta') > V_n(\vartheta'')$ with probability close to one.

Theorem. *For sufficiently large n, any $\vartheta' \neq \vartheta''$, and any $\varepsilon > 0$, we have*

$$P\Big(\big|[S_n(\vartheta') - S_n(\vartheta'')] - [V_n(\vartheta') - V_n(\vartheta'')]\big| > \varepsilon\Big) \le \varepsilon^{-2} O\big(n^{-(4k+5)/(2k+3)}\big).$$

Proof. Taking into account $\theta_i(\Delta) = O(\Delta) = O(n^{-1})$, we find that

$$S_n(\vartheta) = V_n(\vartheta) + \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 + \frac{2O(n^{-1})}{n}\sum_{i=1}^{n}\big(v_{i-1} - \widehat{v}_{i-1}(\vartheta)\big) + \frac{2}{n}\sum_{i=1}^{n}\big[\big(v_{i-1} - \widehat{v}_{i-1}(\vartheta)\big) + O(n^{-1})\big]\eta_i + O(n^{-2}).$$

Therefore,

$$\big[S_n(\vartheta') - S_n(\vartheta'')\big] - \big[V_n(\vartheta') - V_n(\vartheta'')\big] = \frac{2O(n^{-1})}{n}\sum_{i=1}^{n}\big[\big(v_{i-1} - \widehat{v}_{i-1}(\vartheta')\big) - \big(v_{i-1} - \widehat{v}_{i-1}(\vartheta'')\big)\big] + \frac{2}{n}\sum_{i=1}^{n}\big[\big(\widehat{v}_{i-1}(\vartheta') - \widehat{v}_{i-1}(\vartheta'')\big) + O(n^{-1})\big]\eta_i + O(n^{-2}).$$

For notational convenience, set

$$r^2 := \mathbf{E}\left(\frac{2}{n}\sum_{i=1}^{n}\big[\big(\widehat{v}_{i-1}(\vartheta') - \widehat{v}_{i-1}(\vartheta'')\big) + O(n^{-1})\big]\eta_i + \frac{2O(n^{-1})}{n}\sum_{i=1}^{n}\big(\widehat{v}_{i-1}(\vartheta') - \widetilde{v}_{i-1}(\vartheta'')\big) + O(n^{-2})\right)^2.$$

The use of $\left(\sum\limits_{\ell=1}^{3} r_1\right)^2 \le 3\sum\limits_{\ell=1}^{3} r_\ell^2$ provides

$$r^2 \le 3\left(\frac{4}{n^2}\sum_{i=1}^{n}\mathbf{E}\big[\big(\widehat{v}_{i-1}(\vartheta') - \widehat{v}_{i-1}(\vartheta'')\big) + O(n^{-1})\big]^2\,\mathbf{E}\,\eta_i^2 + \frac{\mathfrak{a}}{n^6}\,\mathbf{E}\left(\sum_{i=1}^{n}\big(\widehat{v}_{i-1}(\vartheta') - \widehat{v}_{i-1}(\vartheta'')\big)\right)^2 + O(n^{-4})\right) \equiv 3\sum_{\ell=1}^{3} r_\ell^2.$$

Applying the obvious estimates

$$\mathbf{E}\left(\widehat{v}_{i-1}(\vartheta'')-\widehat{v}_{i-1}(\vartheta')\right)^2 \le 2\,\mathbf{E}\left(v_{i-1}-\widehat{v}_{i-1}(\vartheta')\right)^2+2\,\mathbf{E}\left(v_{i-1}-\widehat{v}_{i-1}(\vartheta'')\right)^2$$
$$=\begin{cases} O(1), & i\le \mathfrak{a}(n^{-1/(2k+3)}\log n),\\ O\big(n^{-(2(k+1)/(2k+3)}\big), & i>\mathfrak{a}(n^{-1/(2k+3)}\log n),\end{cases}$$

we get the following upper bounds for r_ℓ^2, $\ell=1,2,3$:

$$r_1^2 \le O(n^{-1})\left(O(n^{-2})+\frac{1}{n}\sum_{i=1}^{n}\mathbf{E}\left(\widehat{v}_{i-1}(\vartheta')-\widehat{v}_{i-1}(\vartheta'')\right)^2\right),$$
$$r_2^2=\frac{12O(n^{-2})}{n^4}\,\mathbf{E}\left(\sum_{i=1}^{n}\left(\widehat{v}_{i-1}(\vartheta')-\widehat{v}_{i-1}(\vartheta'')\right)\right)^2,$$
$$r_3^2=O(n^{-2})\le \mathfrak{a}\big(n^{-(4k+5)/(2k+3)}\big).$$

Hence, with the help of Chebyshev's inequality, we find that, for sufficiently large n, any $\vartheta'\ne\vartheta''$, and any $\varepsilon>0$, the desired statement follows.

3. FILTERS CONTROLLED BY MULTIPLE PARAMETERS

We use the notation "Filter 0" for the filter corresponding to $k=0$ and the optimal ϑ.

We consider two GARCH filters, (1,1) and (2,2), of the type (3), which will for brevity be referred to as Filter 1 and Filter 2 respectively (for more details, see Appendix 2).

According to (22), for Filter 1 we have

$$\widehat{v}_i=\widehat{v}_{i-1}\left(1-\frac{a_1}{n}\right)+\frac{a_1K}{n}+\frac{\vartheta}{n^{2/3}}(X_i-\widehat{v}_{i-1}),$$

and for Filter 2,

$$\widehat{v}_i=\widehat{v}_{i-1}+\frac{1}{n}\widehat{v}_{i-1}^{(1)}+\frac{\sqrt{2\vartheta}}{n^{4/5}}(X_i-\widehat{v}_{i-1}),$$
$$\widehat{v}_i^{(1)}=\widehat{v}_{i-1}^{(1)}\left(1-\frac{a_1}{n}\right)-\frac{a_2}{n}\widehat{v}_{i-1}+\frac{a_2K}{n}+\frac{\vartheta}{n^{3/5}}(X_i-\widehat{v}_{i-1}),$$

where

$$0<a_1,\qquad a_2\ll n,\qquad |K|\ll n.$$

The estimates generated by Filters 1 and 2 possess the same rate in $n\to\infty$:

$$\mathbf{E}(v_i-\widehat{v}_i)^2=O\big(n^{-2(1+k)/(2k+3)}\big),\quad k=1,2.$$

Moreover, the presence of additional parameters a_1,K and a_1,a_2,K, respectively, enables us to slightly improve (by approximately 10%) the value of

$$S_n(\vartheta,K,a_1,a_2)=\frac{1}{n}\sum_{i=1}\left(X_i-\widehat{v}_{i-1}\right)^2.$$

The main adaptive parameter remains ϑ. The contribution of a_1,K or a_1,a_2,K is not essential. The latter simplifies the tuning procedure as compared to the standard tuning procedure for the classical

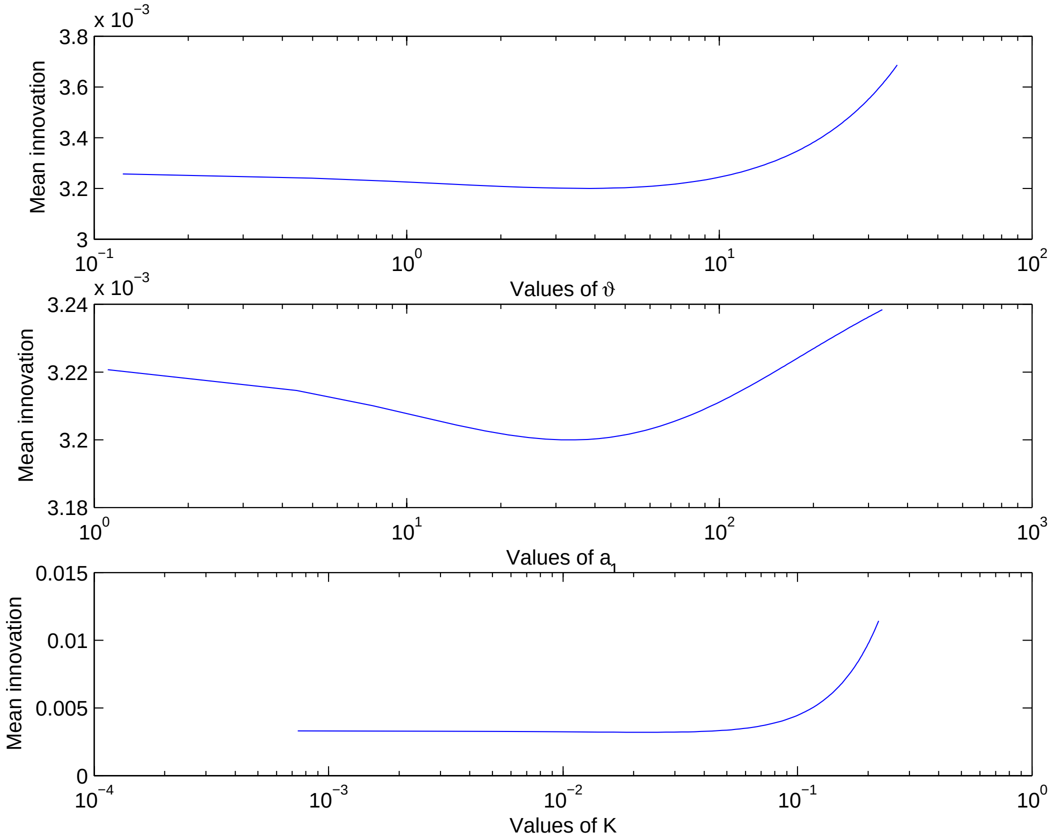

Fig. 2. Filter 1: Behavior of coefficients in the minima for the IBM stock.

GARCH(1, 1) and GARCH(2, 2) (see MATLAB GARCH Toolbox[1]) and, particularly, makes it possible to avoid local minima.

4. COMPUTER IMPLEMENTATION AND SIMULATIONS

The volatility dynamics might be essentially different for various types of assets. For example, volatility changes for stocks and risky assets are too fast, and volatility values are relatively high. On the other hand, composite indexes and exchange rates are most often characterized by slow changes and smaller volatility values. This remark points out the difficultly of finding the best filter simultaneously for all assets.

With the help of simulations, we compare the results of Filter 0, Filter 1, and Filter 2, as well as the GARCH(1, 1) and GARCH(2, 2) provided by MATLAB. Though Filters 1 and 2 are equivalent to GARCH(1, 1) and GARCH(2, 2), respectively, the comparison of numerical results show some advantage of Filters 1 and 2 due to different tuning procedures (see comment at the end of Section 3).

4.1. Tuning Procedure for Filters 1 and 2

The univariate minimization process required for Filter 0 is straightforward. For Filters 1 and 2, we used unconstrained minimization, as is given below.

[1] `http://www.mathworks.com/access/helpdesk/help/toolbox/garch/garch.shtml`.

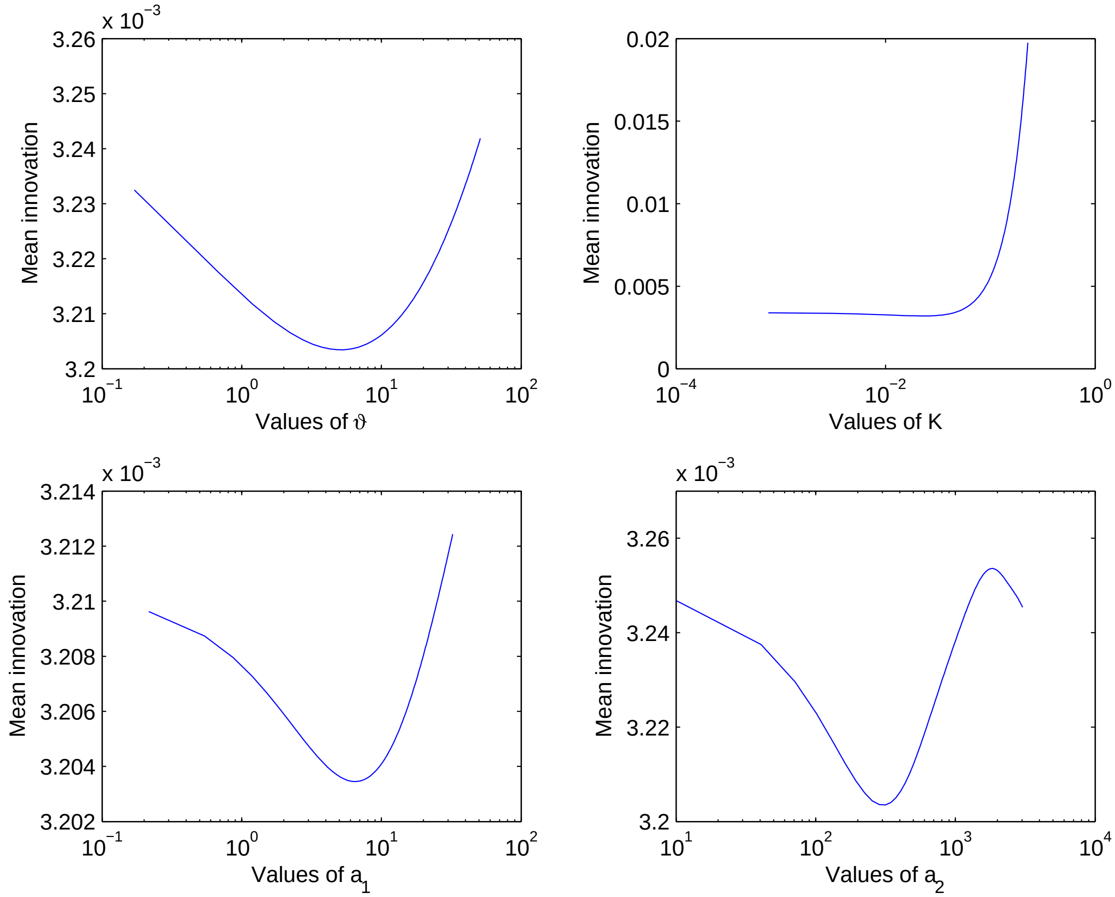

Fig. 3. Filter 2: Behavior of coefficients in the minima for the IBM stock.

<u>Filter 1</u>.

1. Set $a_1 = 0$, $K = 0$, and find $\vartheta^* = \arg\min\limits_{\vartheta} S_n(\vartheta, 0, 0)$.
2. Find $K^* = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$.
3. Find $a_1^* = \arg\min\limits_{a_1} S_n(\vartheta^*, K^*, a_1)$.
4. Local minimization of $S_n(\vartheta, K, a_1)$ in a vicinity of $(\vartheta^*, K^*, a_1^*)$.

<u>Filter 2</u>.

1. Set $a_1 = 0$, $a_2 = 0$, $K = 0$, and find $\vartheta^* = \arg\min\limits_{\vartheta} S_n(\vartheta, 0, 0, 0)$.
2. Find $K^* = \dfrac{1}{n} \sum\limits_{i=1}^{n} X_i$.
3. Find $(a_1^*, a_2^2) = \arg\min\limits_{a_1, a_2} S_n(\vartheta^*, K^*, a_1, a_2)$.
4. Local minimization of $S_n(\vartheta, K, a_1, a_2)$ in a vicinity of $(\vartheta^*, K^*, a_1^*, a_2^*)$.

The above tuning procedure consists in univariate minimization over ϑ, computation of K, and minimization over a_1 and a_2. These steps are supposed to provide some $(\vartheta^*, K^*, a_1^*, a_2^*)$ in the vicinity on the minimum point where the multidimensional minimization procedure is applied. Simulation results demonstrate that, in the vicinity of $(\vartheta^*, K^*, a_1^*, a_2^*)$, the function $S_n(\vartheta, K, a_1, a_2)$

Table 1. Average one-step prediction error for the exchange rate volatility

Currency		Filter type				
From	To	GARCH(1, 1)	GARCH(2, 2)	Filter 0	Filter 1	Filter 2
AUD	$	9.096×10^{-6}	9.090×10^{-6}	9.117×10^{-6}	9.092×10^{-6}	9.092×10^{-6}
EUR	$	5.857×10^{-6}	5.848×10^{-6}	5.869×10^{-6}	5.856×10^{-6}	5.856×10^{-6}
NIS	$	1.830×10^{-6}	1.827×10^{-6}	1.839×10^{-6}	1.827×10^{-6}	1.826×10^{-6}
RUB	$	5.026×10^{-7}	4.915×10^{-7}	5.026×10^{-7}	4.935×10^{-7}	4.880×10^{-7}
YEN	$	5.388×10^{-6}	5.372×10^{-6}	5.389×10^{-6}	5.376×10^{-6}	5.367×10^{-6}

Table 2. Average one-step prediction error for stocks

Asset name	Filter type				
	GARCH(1, 1)	GARCH(2, 2)	Filter 0	Filter 1	Filter 2
DIS	4.309×10^{-3}	4.314×10^{-3}	4.329×10^{-3}	4.284×10^{-3}	4.285×10^{-3}
HPQ	3.741×10^{-2}	3.771×10^{-2}	3.615×10^{-2}	3.608×10^{-2}	3.608×10^{-2}
IBM	3.229×10^{-3}	3.228×10^{-3}	3.217×10^{-3}	3.199×10^{-3}	3.200×10^{-3}
INTC	1.232×10^{-2}	1.230×10^{-2}	1.235×10^{-2}	1.223×10^{-2}	1.232×10^{-2}
MAT	1.899×10^{-2}	1.879×10^{-2}	1.850×10^{-2}	1.817×10^{-2}	1.847×10^{-2}
SUN	5.144×10^{-4}	5.131×10^{-4}	5.143×10^{-4}	5.135×10^{-4}	5.138×10^{-4}
TOY	6.134×10^{-3}	6.128×10^{-3}	6.158×10^{-3}	6.098×10^{-3}	6.082×10^{-3}

behaves as a concave function, and this property is preserved in a wide range around $(\vartheta^*, K^*, a_1^*, a_2^*)$. Moreover, the minimum is not sharp, so that the minimization procedure does not require high "resolution." The corresponding marginal projections of $S_n(\vartheta, K, a_1, a_2)$ are given in Figs. 2 and 3.

4.2. Exchange Rates

The USD exchange rates were used for historical volatility estimation for the period between December 1, 2001, and January 18, 2004 (i.e., $n = 1466$). Filters 1 and 2 provide estimation error similar to GARCH(1, 1) and GARCH(2, 2). The tracking accuracy in terms of $S_n(\vartheta, K, a_1, a_2)$ is given in Table 1.

4.3. Stocks

We considered some stocks of large computer manufacturers and toy and entertainment companies. The information for the period from February 24, 1999, to October 28, 2003 ($n = 1176$), was collected from Yahoo. Numerical results for $S_n(\vartheta, K, a_1, a_2)$ that correspond to the best tuning parameters show serious differences in the behavior of the filters (see Table 2). The different quality of Filters 1 and 2 and GARCH(1, 1) and (2, 2), respectively, is provided by different tuning procedures. Filter 1 provides the best quality.

4.4. Discussion of Numerical Results

The univariate minimization of Filter 0 is, practically, very fast. The tuned parameter ϑ^* gives a hint for tuning procedures of Filters 1 and 2. The multivariate designs of Filters 1 and 2 provide slightly better tracking accuracy than GARCH(1, 1) and (2, 2), respectively, especially for stock options. We attribute this effect to difficulties in tuning of the filter parameters, especially for GARCH(2, 2), which leads to local minima more often than for tuning of Filters 1 and 2.

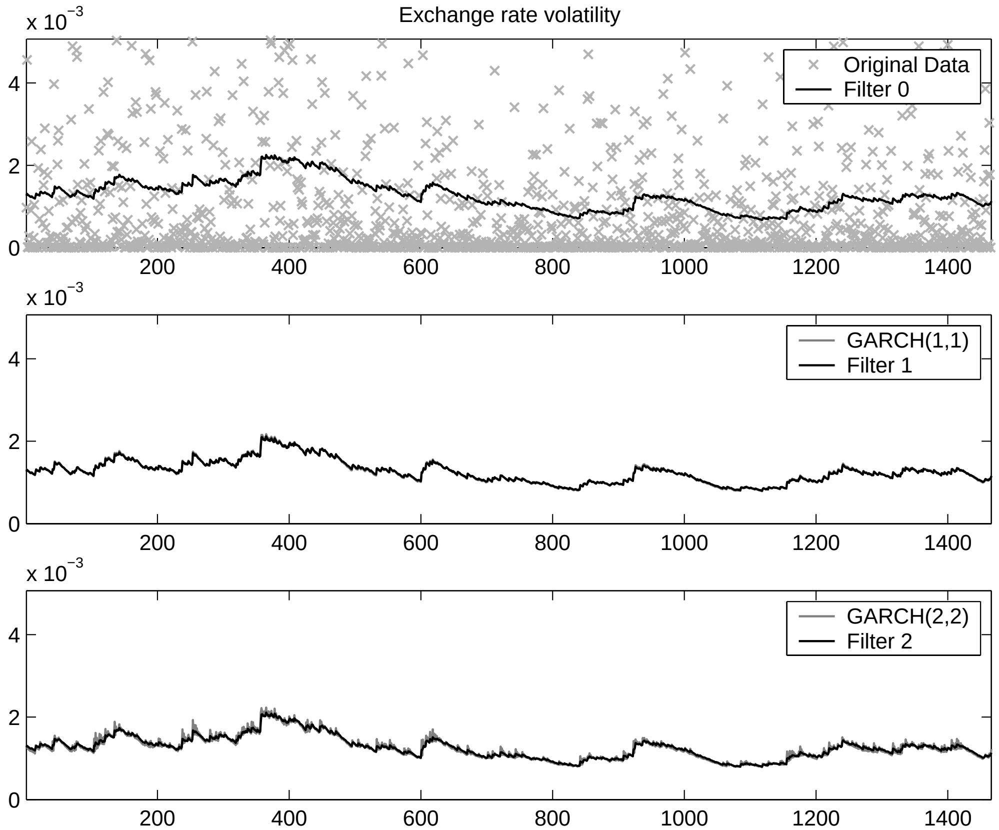

Fig. 4. Historical volatility estimation for the USD/EUR exchange rates.

Parallel to numerical results, the volatility paths obtained with the help of the above-discussed filters is given in Figs. 4 and 5.

5. CONCLUSION REMARK

The volatility estimation designs introduced in this paper are similar to the widely used GARCH algorithms. The proposed filters are stable, and their tuning is simpler. Due to that, the filters guarantee, practically, better volatility tracking. Moreover, they, being of the GARCH type, are applicable for tracking not only a smooth deterministic volatility but also a random volatility function. At least, simulation results with real data verify this property.

The authors are grateful to the anonymous referee, whose comments and advice allowed them to significantly improve the paper.

APPENDIX 1. AUXILIARY LEMMA

Lemma. *The asymptotic from* (7) *is preserved for* $\theta_i(\Delta) \not\equiv 0$.

Proof. Let $\widehat{v}_i$ and $\widetilde{u}_i$ be estimates created by (11) with and without the nuisance process $\theta_i(\Delta)$, respectively. Below we prove that

$$\mathbf{E}\,(\widehat{v}_i - \widetilde{u}_i)^2 \le O(n^{-1}). \tag{12}$$

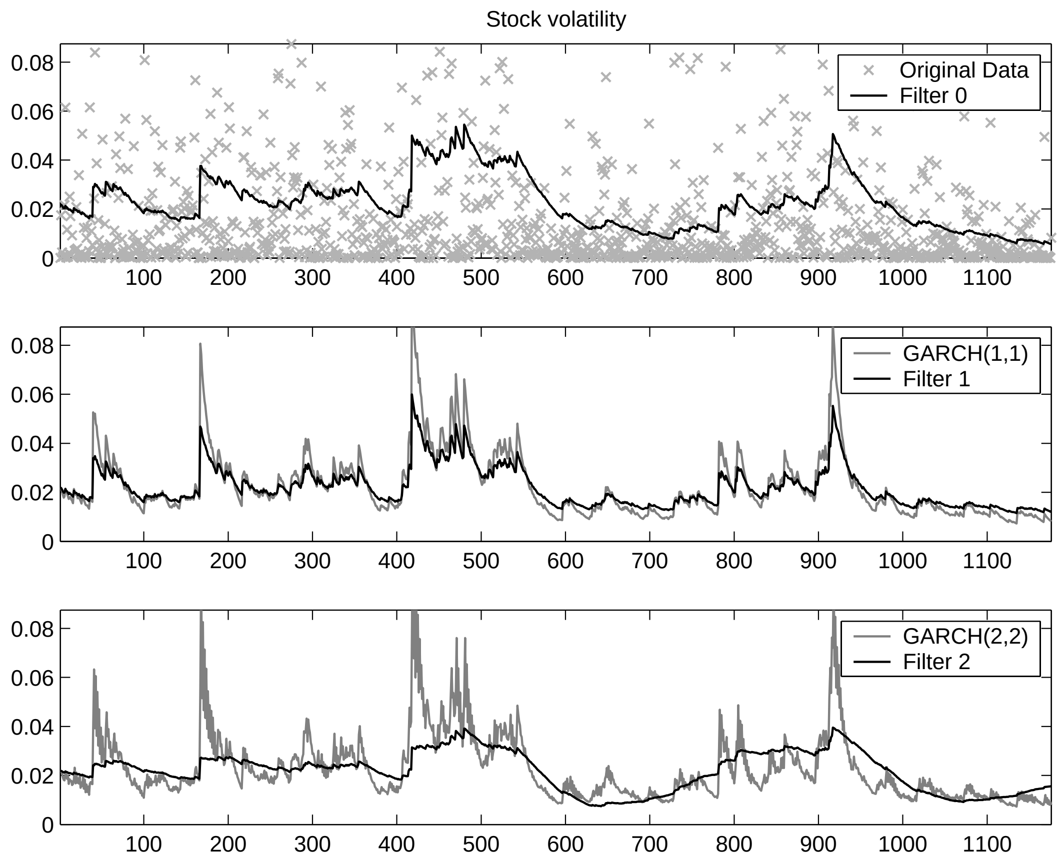

Fig. 5. Historical volatility estimation for the IBM stock.

Since

$$\frac{n^{2(k+1)/(2k+3)}}{n^2} \to 0, \quad n \to \infty,$$

the nuisance process $\theta_i(\Delta)$ does not change the rate (7) in $n \to \infty$.

For the notational convenience, we write θ_i instead of $\theta_i(\Delta)$. Set $\delta_i^{(0)} = \widehat{v}_i - \widetilde{u}_i$ and $\delta_i^{(j)} = \widehat{v}_i^{(j)} - \widetilde{u}_i^{(j)}$, $j = 1, \ldots, k$. Then we have

$$\begin{aligned} \delta_i &= \delta_{i-1} + \frac{1}{n}\delta_{i-1}^{(1)} + \frac{U_{00}\vartheta^{1/k+1}}{n^{2(k+1)/(2k+3)}}(\theta_i - \delta_{i-1}), \\ \delta_i^{(j)} &= \delta_{i-1}^{(j)} + \frac{1}{n}\delta_{i-1}^{(j+1)} + \frac{U_{0j}\vartheta^{(j+1)/k+1}}{n^{(2(k+1)-j)/(2k+3)}}(\theta_i - \delta_{i-1}), \quad j = 1, \ldots, k-1, \\ \delta_i^{(k)} &= \delta_{i-1}^{(k)} + \frac{U_{0k}\vartheta}{n^{(k+2)/(2k+2)}}(\theta_i - \delta_{i-1}), \end{aligned} \tag{13}$$

subject to the appropriate initial conditions $\delta(0) = 0$ and $\delta^{(j)}(0) = 0$, $j = 1, \ldots, k$.

Set

$$\boldsymbol{q}_n = \begin{pmatrix} U_{00}\vartheta^{1/(1+k)}n^{-2(1+k)/(2k+3)} \\ U_{01}\vartheta^{2/(1+k)}n^{-(2(1+k)-1)/((2k+3)} \\ \dots\dots\dots\dots\dots\dots \\ U_{0k}\vartheta n^{-(2(1+k)-k)/(2k+3)} \end{pmatrix}, \qquad \boldsymbol{q} = \begin{pmatrix} U_{00}\vartheta^{1/(1+k)} \\ U_{01}\vartheta^{2/(1+k)} \\ \dots\dots\dots \\ U_{0k}\vartheta \end{pmatrix}, \qquad F_i = \begin{pmatrix} \delta_i \\ \delta_i^{(1)} \\ \dots \\ \delta_i^{(k)} \end{pmatrix},$$

and recall that the matrices a and A are defined in (10). We rewrite (13) in the vector-matrix form

$$F_i = F_{i-1} + \frac{1}{n} a F_{i-1} + \boldsymbol{q}_n \theta_i - \boldsymbol{q}_n A F_{i-1},$$

where $F_0 = 0$. Set $G_i = C_n F_i$, where

$$C_n = \begin{pmatrix} n^{(1+k)/(2k+3)} & 0 & \ldots & 0 & 0 \\ 0 & n^{k/(2k+3)} & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & n^{2/(2k+3)} & 0 \\ 0 & 0 & \ldots & 0 & n^{1/(2k+3)} \end{pmatrix}$$

is a diagonal $(1+k) \times (1+k)$ matrix. Then $G_0 = 0$ and

$$G_i = G_{i-1} + \frac{1}{n} C_n a F_{i-1} + C_n \boldsymbol{q}_n \theta_i - C_n \boldsymbol{q}_n A F_{i-1}. \tag{14}$$

By directly verifying the identities $C_n a = n^{1/(2k+3)} a C_n$ and $C_n \boldsymbol{q}_n = n^{-(1+k)/(2k+3)} \boldsymbol{q}$, we have

$$\begin{gathered} \frac{1}{n} C_n a F_{i-1} = n^{-2(1+k)/(2k+3)} a G_{i-1}, \\ C_n \boldsymbol{q}_n \theta_i = n^{-(1+k)/(2k+3)} \boldsymbol{q} \theta_i. \end{gathered} \tag{15}$$

The structure of the matrix A provides the equality

$$n^{(1+k)/(2k+1)} A = A C_n.$$

From this equality and $C_n \boldsymbol{q}_n = n^{-(1+k)/(2k+3)} \boldsymbol{q}$, it follows that

$$C_n \boldsymbol{q}_n A F_{i-1} = n^{-2(1+k)/(2k+3)} \boldsymbol{q} A G_{i-1}. \tag{16}$$

Gathering now (14)–(16), we find the following recurrent equation for G_i:

$$G_i = G_{i-1} + n^{-2(1+k)/(2k+3)} (a - \boldsymbol{q} A) G_{i-1} + n^{-(1+k)/(2k+3)} \boldsymbol{q} \theta_i.$$

With the matrix $D_n = I + n^{-2(1+k)/(2k+3)} (a - \boldsymbol{q} A)$, this recurrent equation is transformed into

$$G_i = D_n G_{i-1} + n^{-(1+k)/(2k+3)} \boldsymbol{q} \theta_i.$$

Hence, due to $G_0 = 0$, we have

$$G_i = \sum_{p=1}^{i} D_n^{i+1-p} n^{-(1+k)/(2k+3)} \boldsymbol{q} \theta_p.$$

The latter and $|\theta_i| = O(n^{-1})$ provide

$$\|G_i\| \le O(n^{-1}) n^{-(1+k)/(2k+3)} \sum_{p=0}^{\infty} \|D_n^p\| = O(n^{-(3k+4)/(2k+3)} \sum_{p=0}^{\infty} \|D_n^p\|.$$

On the other hand, by Statement 2 in [11], for some positive constants c_0 and C and any positive p, we have the following estimate for $\|D_n^p\|$:

$$\|D_n^p\| \le C \exp\big(- c_0 n^{-2p(1+k)/(2k+3)} \big).$$

Consequently,

$$\|G_i\| \le O\big(n^{-(3k+4)/(2k+3)}\big)\Big(1 - e^{-c_0 n^{-2(1+k)/(2k+3)}}\Big)^{-1} = O\big(n^{-(k+2)/(2k+3)}\big).$$

Finally, by $F_i = C_n^{-1} G_i$, we find that

$$\big|\delta_i^{(j)}\big| \le n^{-(1+k-j)/(2k+3)}\|G_i\| \le O\big(n^{-1+j/(2k+3)}\big), \quad j = 0, 1, \ldots, k.$$

Hence, $|\delta_i^{(0)}| \le O(n^{-1})$, and (12) holds true.

APPENDIX 2. GARCH IN THE FORM OF FILTER (3)

The filter of the type (3) was proposed in [11] for tracking functions from $\Sigma(\beta, L)$ (see (4)). Let us now consider a set of functions from $\Sigma\big((k+1), L\big)$ such that, for some positive constants $a_1, a_2, \ldots, a_k$, roots of the polynomial

$$P(x) = x^k + a_1 x^{k-1} + \ldots + a_{k-1} x + a_k$$

have negative real parts, and

$$\big|f^{(k)}(t) + a_1 f^{(k-1)}(t) + \ldots + a_{k-1} f^{(1)} + a_k f(t)\big| \le \ell$$

for some $\ell > 0$ (for more details, see [17]).

The GARCH filter (2) is adapted to the set of functions chosen above. Moreover, this filter, being rewritten in the form (3), could be tuned in a framework of this paper.

By analogy with [18], we propose the following estimator for $f(t)$ and its derivatives $f^{(j)}(t)$, $j = 1, \ldots, k$, via observations of the process X_t with $X_0 = 0$ and

$$dX_t = f(t)\,dt + \varepsilon\,dW_t,$$

where W_t is the Wiener process:

$$\begin{aligned} d\widehat{f}(t) &= \widehat{f}^{(1)}(t)\,dt + \frac{q_0}{\varepsilon^{2/(2k+3)}}\big(dX_t - \widehat{f}(t)\,dt\big), \\ d\widehat{f}^{(j)}(t) &= \widehat{f}^{(j+1)}(t)\,dt + \frac{q_j}{\varepsilon^{2j/(2k+3)}}\big(dX_t - \widehat{f}(t)\,dt\big), \quad j = 1, \ldots, k-1, \\ d\widehat{f}^{(k)}(t) &= -\Big(a_1 \widehat{f}^{(k-1)}(t) + a_2 \widehat{f}^{(k-2)}(t) + \ldots + a_{k-1}\widehat{f}^{(1)}(t) + a_k \widehat{f}(t) + a_k K\Big)\,dt \\ &\qquad + \frac{q_k}{\varepsilon^{2k/(2k+3)}}\big(dX_t - \widehat{f}(t)\,dt\big). \end{aligned} \tag{17}$$

Now we show that, behind a boundary layer, we have

$$\begin{aligned} \mathbf{E}\,\big(f(t) - \widehat{f}(t)\big)^2 &\asymp \varepsilon^{4k/(2k+3)}, \\ \mathbf{E}\,\big(f^{(j)}(t) - \widehat{f}^{(j)}(t)\big)^2 &\asymp \varepsilon^{4(k-j)/(2k+3)}, \quad j = 1, \ldots, k. \end{aligned} \tag{18}$$

Set $\Delta(t) = f(t) - \widehat{f}(t)$ and $\Delta^{(j)}(t) = f^{(j)}(t) - \widehat{f}^{(j)}(t)$. From (17) and

$$\begin{aligned} df(t) &= f^{(1)}(t)\,dt, \\ df^{(j)}(t) &= f^{(j+1)}(t)\,dt, \quad j = 1, \ldots, k-1, \\ df^{(k)}(t) &= -\Big(a_1 f^{(k-1)}(t) + a_2 f^{(k-2)}(t) + \ldots + a_{k-1} f^{(1)}(t) + a_k f(t) + a_k K\Big)\,dt, \end{aligned}$$

we derive

$$\begin{aligned}
d\Delta(t) &= \Delta^{(1)}(t)dt - \frac{q_0}{\varepsilon^{2/(2k+3)}}(\varepsilon\, dW_t + \Delta(t)\, dt),\\
d\Delta^{(j)}(t) &= \Delta^{(j+1)}(t)\, dt - \frac{q_j}{\varepsilon^{2j/(2k+3)}}(\varepsilon\, dW_t + \Delta(t)\, dt), \quad j = 1, \dots, k-1,\\
d\Delta^{(k)}(t) &= -\left(\sum_{\ell=1}^{k-1} a_\ell \Delta^{(k-\ell)}(t) + a_k\Delta(t) + u(t)\right) dt - \frac{q_k}{\varepsilon^{2k/(2k+3)}}(\varepsilon\, dW_t + \Delta(t)\, dt).
\end{aligned}$$

Following [18], we introduce

$$\delta(t) = \frac{\Delta(t\varepsilon^{2/(2k+3)})}{\varepsilon^{2(k+1)/(2k+3)}}, \qquad \delta^{(j)}(t) = \frac{\Delta^{(j)}(t\varepsilon^{2/(2k+3)})}{\varepsilon^{2(k+1-j)/(2k+3)}}, \quad j = 1, \dots, k, \tag{19}$$

and note that

$$\begin{aligned}
d\delta(t) &= [\delta^{(1)}(t) - q_0\delta(t)]\, dt - q_0\, dW_t^\varepsilon,\\
d\delta^{(j)}(t) &= [\delta^{(j+1)}(t) - q_j\delta(t)]\, dt - q_0\, dW_t^\varepsilon, \quad j = 1, \dots, k-1,\\
d\delta^{(k)}(t) &= -\left(\sum_{\ell=1}^{k-1} \varepsilon^{2\ell/(2k+3)} a_\ell \delta^{(k-\ell)}(t) + \varepsilon^{2k/(2k+3)} a_k \delta(t)\right) dt\\
&\qquad - u(t\varepsilon^{2/(2k+3)})\, dt - q_k\delta(t)\, dt - q_k\, dW_t^\varepsilon,
\end{aligned} \tag{20}$$

where $W_t^\varepsilon = \dfrac{1}{\varepsilon^{1/(2k+3)}} W_{t\varepsilon^{2/(2k+3)}}$ is the standard Wiener process.

Set

$$D(t) = \begin{pmatrix} \delta(t) \\ \delta^{(1)}(t) \\ \dots \\ \delta^{(k)} \end{pmatrix}, \qquad U(t) = \begin{pmatrix} 0 \\ \dots \\ 0 \\ u(t\varepsilon^{2/(2k+3)}) \end{pmatrix}, \qquad Q = \begin{pmatrix} q_0 \\ q_1 \\ \dots \\ q_k \end{pmatrix},$$

and introduce the matrices a_ε and A of sizes $(k+1) \times (k+1)$ and $1 \times (k+1)$, respectively:

$$a_\varepsilon = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ -\varepsilon^{2k/(2k+3)} a_k & \dots & \dots & \dots & \dots & -\varepsilon^{2/(2k+3)} a_1 \end{pmatrix}$$

and $A = \begin{pmatrix} 1 & 0 & \dots & 0 \end{pmatrix}$. With the help of the introduced matrices, we rewrite (20) in the vector-matrix form:

$$dD(t) = (a_\varepsilon - QA)D(t)\, dt + U(t)\, dt - Q\, dW_t^\varepsilon.$$

It is known from [18] that the vector Q can be chosen such that eigenvalues of the matrix $a_0 - QA$ have negative real parts. This property is preserved for $a_\varepsilon - QA$, at least for sufficiently small ε. Henceforth, we assume this property for $a_\varepsilon - QA$ too. Consider now the Lyapunov equation (here, I is the unit matrix)

$$(a_\varepsilon - QA)P_\varepsilon + P_\varepsilon(a_\varepsilon - QA)^* + I = 0. \tag{21}$$

This equation has a unique solution, which we also denote by P_ε. It is clear that $\lim\limits_{\varepsilon\to 0} P_\varepsilon = P_0$, where P_0 is the unique solution of the Lyapunov equation

$$(a_0 - QA)P_0 + P_0(a_0 - QA)^* + I = 0.$$

Denote $\|D(t)\|^2_{P_\varepsilon} = \langle D(t), P_\varepsilon D(t)\rangle$ $(:= D^*(t)P_\varepsilon D(t))$. With the help of the Itô formula and (21), we find

$$d\|D(t)\|^2_{P_\varepsilon} = \Big(-\|D(t)\|^2 + 2\big\langle D(t), P_\varepsilon U(t)\big\rangle + \langle Q, P_\varepsilon Q\rangle\Big)\,dt + 2\big\langle D(t), P_\varepsilon Q\,dW_t\big\rangle.$$

Therefore, the function $V(t) = \mathbf{E}\,\|D(t)\|^2_{P_\varepsilon}$ is differentiable, and its derivative is

$$\dot V(t) = \mathbf{E}\left(-\|D(t)\|^2 + 2\big\langle D(t), P_\varepsilon U(t)\big\rangle + \langle Q, P_\varepsilon Q\rangle\right).$$

It is obvious that, for sufficiently small ε, positive constants c_1, c_2, and c_3 can be found such that

$$\mathbf{E}\,\|D(t)\|^2 \ge c_1 V(t), \qquad 2\,\mathbf{E}\,\langle D(t), P_\varepsilon U(t)\rangle \le c_2\sqrt{V(t)}, \qquad \langle Q, P_\varepsilon Q\rangle \le c_3.$$

Hence, $\dot V(t) \le -c_1 V(t) + c_2\sqrt{V(t)} + c_3$. The use of the inequality $\sqrt{x} \le \alpha^{-1} + \alpha x$, $\alpha \ge 0$, $x \ge 0$, with $\alpha = \dfrac{c_1}{2c_2}$, provides

$$\dot V(t) \le -0.5 c_1 V(t) + \frac{2c_2^2}{c_1} + c_3.$$

Hence, for any $t \ge 0$ we have

$$V(t) \le V(0) + \frac{2c_2^2 + c_1 c_3}{0.5 c_1^2}.$$

Obviously, this property and (19) provide (18).

For the discrete-time setting with $\sqrt{t_i - t_{i-1}} \equiv \varepsilon$, we get a filter (equivalent to that from [11])

$$\begin{aligned}
\widehat f_i &= \widehat f_{i-1} + \frac{1}{n}\widehat v^{(1)}_{i-1} + \frac{q_0}{n^{2(k+1)/(2k+3)}}(X_i - \widehat f_{i-1}),\\
\widehat f^{(j)}_i &= \widehat f^{(j)}_{i-1} + \frac{1}{n}\widehat f^{(j+1)}_{i-1} + \frac{q_1}{n^{(2(k+1)-j)/(2k+3)}}(X_i - \widehat f_{i-1}), \quad j = 1,\dots,k-1,\\
\widehat f^{(k)}_i &= \widehat f^{(k)}_{i-1}\left(1 - \frac{a_1}{n}\right) - \frac{1}{n}\left(\sum_{\ell=2}^{k-1} a_\ell \widehat f^{(k-\ell)}_{i-1} + a_k \widehat f_{i-1} + a_k K\right) + \frac{q_k}{n^{(k+2)/(2k+3)}}(X_i - \widehat f_{i-1}).
\end{aligned}$$

Finally, in the framework of this paper, we have

$$\begin{aligned}
\widehat v_i &= \widehat v_{i-1} + \frac{1}{n}\widehat v^{(1)}_{i-1} + \frac{U_{00}\vartheta^{1/k+1}}{n^{2(k+1)/(2k+3)}}(X_i - \widehat v_{i-1}),\\
\widehat v^{(j)}_i &= \widehat v^{(j)}_{i-1} + \frac{1}{n}\widehat v^{(j+1)}_{i-1} + \frac{U_{0j}\vartheta^{(j+1)/k+1}}{n^{(2(k+1)-j)/(2k+3)}}(X_i - \widehat v_{i-1}), \quad j = 1,\dots,k-1,\\
\widehat v^{(k)}_i &= \widehat v^{(k)}_{i-1}\left(1 - \frac{a_1}{n}\right) - \frac{1}{n}\left(\sum_{\ell=2}^{k-1} a_\ell \widehat v^{(k-\ell)}_{i-1} + a_k \widehat v_{i-1} + a_k K\right) + \frac{U_{0k}\vartheta}{n^{(k+2)/(2k+3)}}(X_i - \widehat v_{i-1}).
\end{aligned} \tag{22}$$

REFERENCES

1. Andersen, T., Bollerslev, T., Diebold, F.X., and Labys, P., Exchange Rate Returns Standardized by Realized Volatility Are (Nearly) Gaussian, *Multinational Finance J.*, 2000, vol. 4, pp. 159–179.
2. Baillie, R.T. and Bollerslev, T., Prediction in Dynamic Models with Time-Dependent Conditional Variances, *J. Econometrics*, 1992, vol. 52, pp. 91–113.
3. Black, F., The Pricing of Commodity Contracts, *J. Financial Economics*, 1976, vol. 9, pp. 167–179.
4. Black, F. and Scholes, M., The Pricing of Options and Corporate Liabilities, *J. Political Economics*, 1973, vol. 81, pp. 637–659.

5. Bollerslev, T., Generalized Autoregressive Conditional Heteroskedasticity, *J. Econometrics*, 1986, vol. 31, pp. 307–327.

6. Day, T.E. and Lewis, C.M., Forecasting Futures Market Volatility, *J. Derivatives*, 1993, vol. 1, pp. 33–50.

7. Duan, J.C., The GARCH Option Pricing Model, *Math. Finance*, 1995, vol. 5, no. 1, pp. 13–32.

8. Hamilton, J.D., *Time Series Analysis*, Princeton: Princeton Univ. Press, 1994.

9. Bollerslev, T., Chou, R.Y., and Kroner, K.F., ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence, *J. Econometrics*, 1992, vol. 52, pp. 5–59.

10. Engle, R., Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica*, 1982, vol. 50, pp. 987–1007.

11. Liptser, R. and Khasminskii, R., On-line Estimation of a Smooth Regression Function, *Teor. Veroyatn. Primen.*, 2002, vol. 47, no. 3, pp. 567–594.

12. Goldentayer, L. and Liptser, R., On-line Tracking of a Smooth Regression Function, *Stat. Inference Stoch. Process.*, to appear.

13. Mercurio, D. and Spokoiny, V., Statistical Inference for Time-Inhomogeneous Volatility Models, *Preprint of Weierstrass Inst. for Applied Analysis and Stochastics*, 2000, no. 583. Avaiable at `www.wias-berlin.de/publications/preprints/index-2000.html`.

14. Ibragimov, I. and Khasminskii, R., On Nonparametric Estimation of Regression, *Doklady Akad. Nauk SSSR*, 1980, vol. 252, no. 4, pp. 780–784 [*Soviet Math. Dokl.* (Engl. Transl.), 1980, vol. 21, pp. 810–814].

15. Ibragimov, I.A. and Khas'minskii, R.Z., *Asimptoticheskaya teoriya otsenivaniya*, Moscow: Nauka, 1979. Translated under the title *Statistical Estimation. Asymptotic Theory*, New York: Springer, 1981.

16. Stone, C., Optimal Global Rates of Convergence for Nonparametric Regression, *Ann. Statist.*, 1982, vol. 10, pp. 1040–1053.

17. Goldenshluger, A. and Nemirovski, A., Adaptive De-noising of Signals Satisfying Differential Inequalities, *IEEE Trans. Inform. Theory*, 1997, vol. 43, no. 3, pp. 873–889.

18. Chow, P.L., Khasminskii, R., and Liptser, R., Tracking of Signal and Its Derivatives in the Gaussian White Noise, *Stochastic Process. Appl.*, 1997, vol. 69, no. 2, pp. 259–273.