Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Estimation of time dependent volatility via local change point analysis

Mercurio, Danilo, [1] Vladimir Spokoiny [2]

[1] Weierstrass Institute

Mohrenstr. 39, 10117 Berlin, Germany
E-Mail: mercurio@wias-berlin.de
URL: www.wias-berlin.de/~mercurio

[2] Weierstrass Institute
and Humboldt University Berlin,
Mohrenstr. 39, 10117 Berlin, Germany
E-Mail: spokoiny@wias-berlin.de
URL: www.wias-berlin.de/~spokoiny

No. 904
Berlin 2004

1991 *Mathematics Subject Classification.* 62M10; Secondary 62P20.

Key words and phrases. volatility model, adaptive estimation, local homogeneity, change point.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

This paper offers a new procedure for estimation and forecasting of the volatility of financial time series. The approach is based on the assumption of local homogeneity: for every time point there exists an *interval of time homogeneity*, in which the volatility parameter can be well approximated by a constant. The procedure recovers this interval from the data using the local change point analysis. Afterwards the estimate of the volatility can be simply obtained by local averaging. We investigate the performance of the procedure both from the theoretical point of view and through Monte Carlo simulations. Then the new procedure is applied to some data sets and a comparison with the LAVE procedure from Mercurio and Spokoiny (2004) and with a standard GARCH model is also provided. Finally we discuss applications of the new method to the Value-at-Risk problem. The numerical results demonstrate a very reasonable performance of the new method.

1 Introduction

Since the seminal papers of Engle (1982) and Bollerslev (1986), modelling the dynamic features of the variance of financial time series has become one of the most active fields of research in econometrics. New models, different applications and extensions have been proposed as it can be seen by consulting for example the monographs of Engle (1995) and of Gouriéroux (1997). The main idea behind this strain of research is that the volatility clustering effect that is displayed by stock or exchange rate returns can be modelled globally by a stationary process. This approach is somehow restrictive and it does not fit some characteristics of the data, in particular the fact that the volatility process appears to be "almost integrated" as it can be seen by usual estimation results and by the very slow decay of the autocorrelations of squared returns. Other global parametric approaches have been proposed by Engle & Bollerslev (1986) and by Baillie, Bollerslev & Mikkelsen (1996) in order to include these features in the model.

However, Mikosch & Starica (2000*b*) showed that long memory effects of financial time series can be artificially generated by structural breaks in the parameters. This motivates another modelling approach which borrows its philosophy mainly from the nonparametric statistics. The main idea consists in describing the volatility clustering effect only by a locally stationary process. Therefore, only the most recent data are considered the most important for estimation and weighting schemes, which can be themselves either global or local and data driven, are suggested in order to decrease the dependence of the estimate on the older observations. Some examples of this approach can be found in Fan & Gu (2003), in Dahlhaus & Rao (2003) and in Cheng, Fan & Spokoiny (2003). Furthermore,

Mercurio & Spokoiny (2004) (referred to as MS2004 in what follows) proposes a new local adaptive volatility estimation (LAVE) of the unknown volatility from the conditionally heteroscedastic returns. The method is based on pointwise data-driven selection of the interval of homogeneity for every time point. The numerical results demonstrate a reasonable performance of the new method. In particular, it slightly outperforms the standard GARCH(1,1) approach. Härdle, Herwartz & Spokoiny (2003) extend this method to estimating the volatility matrix of the multiple returns and Mercurio & Torricelli (2001) apply the same idea in the context of a regression problem.

The aim of the present paper is to develop another procedure which, however, applies a similar idea of pointwise adaptive choice of the interval of homogeneity. The main differences between the LAVE approach from MS2004 and the new procedure is in the way of testing the homogeneity of the interval candidate and in the definition of the selected interval. In this paper we systematically apply the approach based on the local change point analysis. This means that every interval is tested on homogeneity against a change-point alternative. If the hypothesis is not rejected, a larger interval-candidate is taken. If the change point is detected, then the location of the change point is used for defining the adaptive interval while MS2004 suggested to take the latest non-rejected interval. The modified procedure allows to improve the sensitivity of the method to changes of volatility by using the more powerful likelihood ratio test statistic with the careful choice of the critical level. In addition, the use of the additional information about the location of the change point which is delivered by the change point test, helps to reduce the estimation bias. Finally, the interpretation of the procedure as a multiple test against a change point alternative leads to a very natural method of tuning the parameters of the procedure.

The change point detection problem for financial time series was considered in Mikosch & Starica (2000*a*) but they focused on asymptotical properties of the test if only one change point is present. Kitagawa (1987) applied non-Gaussian random walk modeling with heavy tails as the prior for the piecewise constant mean for one-step-ahead prediction of nonstationary time series. However, the mentioned modeling approaches require some essential amount of prior information about the frequency of change-points and their size. The new approach proposed in this article does not assume smooth or piecewise constant structure of the underlying process and does not require any prior information. The procedure proposed below in Section 3 focuses on adaptive choice of the interval of homogeneity that allows to proceed in a unified way with smoothly varying coefficient models and change-point models.

The reminder paper is organized as follows. The next section introduces the adaptive modeling procedure, then some theoretical properties are discussed in the general situation and for two particular cases: a change-point model with piecewise constant volatility and the case of a volatility function smoothly varying in time. Section 5 illustrates the performances of the new methodology by means of some simulated examples and applications to real data sets. First we address the problem of selecting the smoothing parameters and propose one solution which will systematically applied for all the examples. Section 5.2 presents some numerical results for a change-point model. In Section 5.3 we study fore-

casting ability of the new method by mean of a comparative study with the GARCH(1,1) method. Section 5.5 discusses applications of the new method to the Value at Risk problem. Section 6 collects the proofs of the main results.

2 Volatility modeling

Let S_t be an observed asset process in discrete time, $t = 1, 2, \ldots$ and R_t are the corresponding returns: $R_t = \log(S_t/S_{t-1})$. We model this process via the *conditional heteroscedasticity* assumption

$$R_t = \sigma_t \xi_t \,, \tag{2.1}$$

where ξ_t, $t \geq 1$, is a sequence of independent standard Gaussian random variables and σ_t is the *volatility* process which is in general a predictable random process, that is, $\sigma_t \sim \mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (σ-field generated by the first $t-1$ observations).

In this paper, similarly to MS2003 we focus on the problem of filtering the parameter σ_t from the past observations $R_1, \ldots, R_{t-1}$. This problem naturally arises as an important building block for many tasks of financial engineering like Value-at-Risk or Portfolio Optimization.

2.1 Parametric modeling

A *time-homogeneous* (*time-homoscedastic*) model means that σ_t is a constant. The process S_t is then a Geometric Brownian motion observed at discrete time moments. For the homogeneous model $R_t = \sigma \varepsilon_t$ with $t \in I$, the parameter $\theta = \sigma^2$ can be estimated using the maximum likelihood method:

$$\widetilde{\theta}_I = \operatorname*{argmax}_{\theta \geq 0} L_I(\theta) = \operatorname*{argmax}_{\theta \geq 0} \sum_{t \in I} \ell(R_t, \theta)$$

where $\ell(y, \theta) = -(1/2)\log(2\pi\theta) - y^2/(2\theta)$ is the log-density of the normal distribution with the parameters $(0, \theta)$. A simple algebra yields

$$\widetilde{\theta}_I = N_I^{-1} \sum_{t \in I} R_t^2, \qquad \text{and} \qquad L_I(\widetilde{\theta}_I) = -\frac{N_I}{2} \log(2\pi\widetilde{\theta}_I) - \frac{N_I}{2} \tag{2.2}$$

where N_I denotes the number of time points in I.

The assumption of normality for the innovations ε_t is often criticized in the financial literature. Our empirical examples in Section 5.3 below also indicate that the tails of estimated innovations are heavier than the normality would imply. However, the estimate $\widetilde{\theta}_I$ remains meaningful even for the non-normal innovations, it is just a quasi-likelihood approach. One can show that this approach leads to the same asymptotic quality of estimation if the distribution of the ε_t's fulfills some exponential moment conditions.

3 Adaptive choice of the interval of homogeneity

The assumption of time homogeneity is too restrictive in practical applications and it does not allow to fit well real data. In this paper we consider an approach based on the *local time-homogeneity* which means that for every time moment n there exists a historic time interval $[n-m, n[$ in which the volatility process σ_t is nearly constant. Under such a modeling, the main intention is both to describe the interval of homogeneity and to estimate the corresponding value σ_n. Our approach is based on the adaptive choice of the interval of homogeneity for the end point n. The procedure attempts to find this interval from the data by successive testing the hypothesis of homogeneity. We start from a small interval $I = [n-m_0, n[$. Then we increase the interval I and test the hypothesis of homogeneity within I against a change-point alternative. If the hypothesis is not rejected then we take a larger interval and continue this way until we detect a change-point or the largest possible interval $\mathbb{I}$ is reached. If a change-point is detected at a point ν then use it as the left end-point of the selected interval , i.e. take the interval $\widehat{I} = [\nu, n[$, otherwise take $\widehat{I} = \mathbb{I}$. Finally, estimate the volatility process from the observations R_t for $t \in \widehat{I}$ assuming the homogeneous model within $\widehat{I}$.

The procedure reads as follows.

Initialization Select the smallest interval in $\mathcal{I}$.

Iteration Select the next interval I in $\mathcal{I}$.

Testing homogeneity Test the hypothesis of homogeneity within I against a change-point alternative.

Loop If a change point is detected at $\nu \in I$, then set $\widehat{I} = [\nu, n[$. Otherwise, continue with the iteration step by choosing a larger interval.

The main ingredient of this procedure is the way of testing the hypothesis of homogeneity.

3.1 Test of homogeneity against a change-point alternative

Let I be an interval-candidate. Here we describe a change-point test within I based on the likelihood ratio test statistics.

The null hypothesis for $I = [n-m, n[$ means that the observations R_t for $t \in I$ follow the parametric model with the parameter θ. This hypothesis leads to the log-likelihood $L_I(\theta)$. We want to test this hypothesis against a change-point alternative that the parameter θ spontaneously changes in some internal point τ of the interval I. Let $\mathcal{T}(I)$ be a family of internal points within I. Every point $\tau \in \mathcal{T}(I)$ splits the interval I onto two subintervals $J = [\tau, n[$ and $J^c = I \setminus J = [n-m, \tau[$. The change point alternative means that $\theta_t = \theta$ for $t \in J$ and $\theta_t = \theta'$ for $t \in J^c$ for some $\theta \neq \theta'$. This corresponds to the log-likelihood $L_J(\theta) + L_{J^c}(\theta')$. The likelihood ratio test statistic for the change-point alternative with

the change point location at the point τ is of the form

$$\begin{aligned} T_{I,\tau} &= \max_{\theta,\theta'} \left\{ L_J(\theta) + L_{J^c}(\theta') \right\} - \max_{\theta} L_I(\theta) \\ &= L_J(\widetilde{\theta}_J) + L_{J^c}(\widetilde{\theta}_{J^c}) - L_I(\widetilde{\theta}_I) = \widehat{L}_J + \widehat{L}_{J^c} - \widehat{L}_I. \end{aligned}$$

For the considered volatility model, this test statistic can be represented in the form

$$T_{I,\tau} = N_J K(\widetilde{\theta}_J, \widetilde{\theta}_I) + N_{J^c} K(\widetilde{\theta}_{J^c}, \widetilde{\theta}_I)$$

where $K(\theta, \theta') = -0.5\,(\log(\theta/\theta') - 1 + \theta/\theta')$ is the Kullback-Leibler information for the two normal distributions with variances θ and θ'. The change-point test for the interval I is defined as the maximum of such defined test statistics over $\tau \in \mathcal{T}_I$:

$$T_I = \max_{\tau \in \mathcal{T}_I} T_{I,\tau}\,.$$

The change-point test compares this statistic with the critical value λ_I which may depend on the interval I and the nominal first kind error probability α. The hypothesis of homogeneity is rejected if $T_I \geq \lambda_I$. The way of choosing the critical value as well as the other parameters of the procedure like the set of testing intervals $\mathcal{T}_I$ is discussed in Section 3.2.

3.2 Parameters of the procedure

To start the procedure running, one has to specify some parameters. This includes the set $\mathcal{I}$ of interval-candidates, and, for every $I \in \mathcal{I}$, the set of internal points $\mathcal{T}_I$ and the critical value λ_I. First we briefly discuss how the sets $\mathcal{I}$ and $\mathcal{T}_I$ can be selected. Then we focus on the choice of the critical values λ_I.

It is useful to take the set $\mathcal{I}$ of interval-candidates in the form of an arithmetic or geometric grid. In both cases one has to fix the starting interval length m_0, that is, the first considered interval is of the form $[n - m_0, n[$. At every iteration this length is increased by adding resp. by multiplying with some fixed step. In our theoretical study we assume the maximal possible set of all intervals I with the length not less than m_0. This is a special case of an arithmetic grid with the step one. To reduce the computational burden, one can take a larger step or even a geometric grid, that is, to define the length m_k of the interval I_k as $m_k = [m_{k-1}c]$ for some $c > 1$.

For every interval $I \in \mathcal{I}$, $I = [n - m, n[$, we define $\mathcal{T}_I$ as the set of all internal points of I separated away from the end-point. More precisely, for a fixed $\rho \leq 1/3$, set $\mathcal{T}_I = \{t : n - m + \rho m \leq t \leq n - \rho m\}$. A reasonable choice for ρ is $\rho = 1/3$. The idea behind this choice is that the behavior of the log-likelihood test statistic $T_{I,\tau}$ becomes quite irregular when τ approaches the end-points of the interval I. Note also that for the points close to n, the test on change point has been already made on the earlier steps of the algorithm while for the points close to the left end point $n - m$, a test will be made at the next iterations. Our simulations results indicate that the procedure is quite stable w.r.t. the choice of the parameters like ρ and c.

In the contrary, the choice of the critical values λ_I is rather important. Larger values λ_I improve stability of the method under homogeneity but result in a low sensitivity to parameter changes while too small critical values lead to a large "false alarm" probability. The standard approach to choosing the critical values is to provide a prescribed first kind error probability, that is, in the homogeneous case, the "false alarm" probability should not exceed the given level α. Here we describe different possibilities for this choice.

We consider a homogeneous model $R_t = \sigma\varepsilon_t$ with the constant volatility $\theta = \sigma^2$ and standard Gaussian innovations ε_t. It is worth noting that the particular value θ has no influence on the behavior of the procedure (it is cancelled in the expression for the test statistic T_I) and therefore we can assume $\theta = 1$. Thus, the probability model is completely specified and its properties can be evaluated by the Monte Carlo simulation.

Define for every I a value β_I in a way that $\sum_{I\in\mathcal{I}} \beta_I = \alpha$. A reasonable proposal is

$$\beta_I = \alpha N_I^{-1} \left(\sum_{I'\in\mathcal{I}} N_{I'}^{-1} \right)^{-1}$$

We also denote $\alpha_I = \sum_{I'\in\mathcal{I}(I)} \beta_{I'}$ where $\mathcal{I}(I) = \{I' : I' \in \mathcal{I}, I' \subseteq I\}$. Note that if N_I grows exponentially with I with a factor c, then the proposed choice ensures that the ratio α/α_I remains bounded by a constant depending on c only.

Now we run the procedure for the time homogeneous data generating process. For every realization the procedure stops if a change-point is detected, that is, the values T_I are computed only if the smaller intervals I' are not rejected. Otherwise we set $T_I = \infty$. The critical value λ_I for every interval I is defined as $(1-\alpha_I)$-quantile of the such computed test statistics T_I. If the number of replications is sufficiently large, this method provides the rejection probability about α_I for every I. However, the method is computationally intensive. Several proposals to simplify this choice are discussed below. One is based on the result of Theorem 4.3 from Section 4 that suggests to apply a critical value λ_I that grows linearly with $\log(N_I)$, that is, $\lambda_I = a + b\log(N_I)$. The constants a and b might (and should) depend on the choice of the set $\mathcal{I}$ and on the nominal error level α. They also slightly depend on the choice of sets of tested intervals I and internal points $\mathcal{T}(I)$, of, more specifically, on the parameters ρ and c. So, the following method can be recommended: for fixed values λ, ρ, c, compute critical values λ_I for a few intervals I and adjust a linear relationship $\lambda_I = a + b\log(N_I)$. We continue this discussion in Section 5 where an implementation of the procedure will be discussed in details.

4 Theoretic properties

This section discusses some useful theoretical properties of the adaptively selected interval of homogeneity $\widehat{I}$ and then of the adaptive volatility estimate $\widehat{\theta}$ that corresponds to the selected interval $\widehat{I}$, that is, $\widehat{\theta} = \widetilde{\theta}_{\widehat{I}}$.

We start by discussing the "false alarm" probability, that is, the probability that a good interval in which the hypothesis of homogeneity is nearly fulfilled is rejected by the change

point test. We show that if the critical values λ_I are properly selected the procedure ensures a prescribed false alarm probability level. The standard way for proving such a result is based on the asymptotic expansion of the log-likelihood process $L(\theta)$. Here we briefly discuss this approach and then switch to the nonasymptotic one.

4.1 Asymptotic properties of the change-point test under the null

The LR test statistic T_I introduced in the previous section has nice asymptotic properties. In particular, it weakly converges under the null hypothesis to some nondegenerated distribution. Moreover, this distribution is parameter free. The idea behind the mentioned asymptotic result is that the in the homogeneous situation the likelihood ratio process can be approximated by some fixed transformation of the standard Wiener process. More precisely, define for every $0 \le \tau_1 < \tau_2 \le 1$ the interval $J = [\tau_1 n, \tau_2 n]$. Then the likelihood ratio $L_J(\theta, \theta_0)$ based on the observations from this interval for $\theta = \theta_0(1 + un^{-1/2})$ can be (strongly) approximated by the random variable $(\tau_2 - \tau_1)^{-1} \left\{ u(W_{\tau_2} - W_{\tau_1}) - u^2/2 \right\}$ where W_τ is a standard Wiener process. This yields the approximation of the distribution of the statistic $2T_I$ by the maximum of $\tau^{-1} W_\tau^2 + (1-\tau)^{-1}(W_1 - W_\tau)^2 - W_1^2$ over $\tau \in [\rho, 1-\rho]$. We do not discuss this result in more details because its applicability is restricted to the case of a large interval I while the procedure starts with a small interval I_0 even if the sample size is large.

Therefore, we need a version of this result which applies to an arbitrary sample size. We present such nonasymptotic results for two cases: one for pure homogeneous situation with a constant volatility and another one for a nearly homogeneous case.

4.2 "False alarm" probability under the null

Suppose that an interval $I = [n - m, n[$ is fixed.

Theorem 4.1. *Let $I \in \mathcal{I}$ and $\theta_t = \theta_0$ for all $t \in I$. Then it holds for every $z \ge 0$*

$$\boldsymbol{P}\left(T_I \ge 2z\right) \le 4N_I e^{-z}.$$

In particular, for every $\beta \in (0,1)$, with $z = \log(4N_I/\beta)$, it holds

$$\boldsymbol{P}\left(T_I \ge 2\log(4N_I/\beta)\right) \le \beta.$$

The next result describes the probability of rejecting a homogeneous interval by our procedure. This means that every of the test statistics $T_{I'}$ for $I' \in \mathcal{I}(I)$ does not exceed the corresponding critical value $\lambda_{I'}$. This is a multiple testing problem requiring a correction of the critical value for using the multiple test. In our theoretical study we apply the Bonferonni method: for every interval I we assign a first kind error probability β_I such that the sum of the β_I's does not exceed the prescribed value α. It is well known that the Bonferonni method is a bit conservative. Therefore, the result we present give only an upper bound for the 'false alarm' probability.

In what follows we suppose that some values β_I, $I \in \mathcal{I}$, are fixed such that

$$\alpha_I := \sum_{I' \in \mathcal{I}(I)} \beta_{I'} \leq \alpha.$$

The next result is a straightforward corollary of Theorem 4.1.

Theorem 4.2. *If $\theta_t \equiv \theta_0$ for all $t \in I$ and if $\lambda_I \geq 2\log(4N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$, then for every $I \in \mathcal{I}$*

$$\boldsymbol{P}\left(I \text{ is rejected}\right) \leq \alpha_I\,.$$

Proof. It suffices to only mention that by Theorem 4.1

$$\boldsymbol{P}\left(I \text{ is rejected}\right) \leq \sum_{I' \in \mathcal{I}(I)} \boldsymbol{P}\left(T_{I'} \geq 2(1+2\mu)\log(8N_{I'}/\beta_{I'})\right).$$

□

4.3 "False alarm" probability in the nearly homogeneous case

Here we consider a more general situation when the volatility coefficient θ_t can be approximated within I by a value θ_0 which is measurable w.r.t. $\mathcal{F}_{n-m-1}$. The violation from the homogeneity within I can be naturally measured by the values Δ_I^* and Δ_I defined by the equations

$$\Delta_I^* = \sup_{t \in I} |\theta_t/\theta_0 - 1| \qquad \text{and} \qquad \Delta_I^2 = N_I^{-1} \sum_{t \in I} |\theta_t/\theta_0 - 1|^2. \tag{4.1}$$

Note that in general Δ_I and Δ_I^* are random variables. Near homogeneity within I means that these values are small with a high probability. For every numbers $\mu \geq 0$ and $z \geq 0$, define the random event

$$A_I(\mu, z) = \{N_I \Delta_I^2 \leq \mu^2 z \text{ and } \Delta_I^* \leq 0.8\min\{\mu, 1\}\}.$$

Theorem 4.3. *It holds for each $\mu \geq 0$ and $\beta \in (0,1)$*

$$\boldsymbol{P}\left(T_I \geq 2(1+2\mu)\log(8N_I/\beta),\ A_I^*(\mu)\right) \leq \beta.$$

As a consequence of this result, if the conditions $\Delta_I^* \leq 0.8\min\{\mu,1\}$, $N_I\Delta_I^2 \leq \mu^2 z$ are fulfilled with probability one, then

$$\boldsymbol{P}\left(T_I \geq 2(1+2\mu)\log(8N_I/\beta)\right) \leq \beta.$$

An extension of Theorem 4.2 to the nearly homogeneous case is also straightforward. Let a sequence $\{\beta_I\}$ satisfying $\sum_I \beta_I = \alpha$ be fixed. For every $I \in \mathcal{I}$ define $z_I = \log(8N_I/\beta_I)$ and

$$A_I^*(\mu) = \{\Delta_{I'}^* \leq 0.8\min\{\mu,1\},\ N_{I'}\Delta_{I'}^2 \leq \mu^2 z_{I'}\ \forall I' \in \mathcal{I}(I)\} = \bigcap_{I' \in \mathcal{I}(I)} A_I(\mu, z_{I'}).$$

Theorem 4.4. *For any $\mu \geq 0$, if $\lambda_I \geq 2(1+2\mu)\log(8N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$, then for every $I \in \mathcal{I}$*

$$\boldsymbol{P}\left(I \text{ is rejected},\ A_I^*(\mu)\right) \leq \alpha_I.$$

The result of Theorem 4.4 suggests the following definition of a "good" or "ideal" interval $\mathit{I\!I}$. We say that $\mathit{I\!I}$ is good if for some fixed $\mu \geq 0$ the event $A_{\mathit{I\!I}}^*(\mu)$ meets with a high probability. It follows from Theorem 6.2 and Lemma 6.6 in the Appendix that the corresponding "oracle" estimate $\widetilde{\theta}_{\mathit{I\!I}}$ delivers with a high probability the quality of estimation of order $N_{\mathit{I\!I}}^{-1/2}$. The result in the next section claims the same rate of accuracy for the adaptive estimate $\widehat{\theta}$.

4.4 Quality of the adaptive volatility estimate

Recall that the adaptive volatility estimate $\widehat{\theta}$ at the time point n is defined as $\widetilde{\theta}_{\widehat{I}}$ where $\widehat{I}$ is the selected interval of homogeneity with the right end-point n. Here we show that the quality of the adaptive estimate is essentially the same (in order) as the quality of the "ideal" estimate corresponding to an "ideal" choice of the interval $\widehat{I}$. Let an interval $\mathit{I\!I} = [n - m^*, n[\in \mathcal{I}$ be fixed such that the event $A_{\mathit{I\!I}}^*(\mu)$ meets with a high probability. It is straightforward to see that the corresponding estimate $\widetilde{\theta}_{\mathit{I\!I}}$ delivers the accuracy of order $N_{\mathit{I\!I}}^{-1/2}$. We now aim to show that the adaptive estimate $\widehat{\theta}$ provides at least the same (in order) accuracy of estimation. In the next result we assume for the ease of exposition that the procedure is run with the maximal set $\mathcal{I}$ of all possible intervals I of length $N_I \geq m_0$.

Theorem 4.5. *Let $\mathit{I\!I}$ be a "good" interval, that is, for some fixed μ, the event $A_{\mathit{I\!I}}^*(\mu)$ meets with a positive probability. If $\lambda_I \geq 2(1+2\mu)\log(8N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$, then the adaptive estimate $\widehat{\theta}$ fulfills*

$$\boldsymbol{P}\left(\left|\log(\widehat{\theta}/\theta_0)\right| \geq C\sqrt{\lambda_{\mathit{I\!I}}/N_{\mathit{I\!I}}},\ A_{\mathit{I\!I}}^*(\mu)\right) \leq \alpha_{\mathit{I\!I}}$$

where the constant C depends on the parameter ρ of the procedure only.

4.5 Accuracy of estimation when θ_t is smooth

Suppose that the volatility function θ_t smoothly varies with t. Then the result of Theorem 4.5 can be used to state the usual nonparametric rate of estimation of the function θ. Since the volatility model $R = \theta^{1/2}\varepsilon$ has irregularity at $\theta = 0$, it is more convenient to work with the log-transform of θ. We suppose that $\log\theta_t$ is Lipschitz at n with the constant L, that is,

$$|\log\theta_n - \log\theta_t| \leq L\,|(n-t)/n|\,. \tag{4.2}$$

This condition implies for every interval $I = [n-m, n[$ that

$$\Delta_I^* \leq \max_{t\in I} |\theta_n/\theta_{n-m} - 1| \leq 2Lm/n$$

for $m \le n/2$, that is, the conditions entering in the definition of the event $A_I^*(\mu)$ are satisfied almost sure with $\mu^2 = m(2Lm/n)^2/\log(m/\beta_I)$. Selecting $m = (n/2L)^{2/3}$ provides this condition with $\mu = 1$. Due to Theorem 4.5, the adaptive estimate $\widehat{\theta}$ ensures with a high probability the quality of estimation

$$\bigl|\log(\widehat{\theta}) - \log(\theta_0)\bigr| \le C\sqrt{\lambda_I/m} \le C\sqrt{\lambda_I}(2L/n)^{1/3}$$

which yields the usual nonparametric rate of estimation for smoothness degree one. We conclude with the following result:

Theorem 4.6. *Let θ_t fulfill (4.2) with probability one. Let also $\lambda_I \ge 6\log(8N_I/\beta_I)$ for all intervals $I \in \mathcal{I}$. Define $m = [(n/2L)^{2/3}]$ and $I = [n-m, n[$. Then it holds for the adaptive estimate $\widehat{\theta}$*

$$\boldsymbol{P}\left(\bigl|\log(\widehat{\theta}) - \log(\theta_0)\bigr| \ge C\sqrt{\lambda_I}(2L/n)^{1/3}\right) \le \alpha_I$$

where the constant C depends on the parameter ρ of the procedure only.

4.6 Change-point model

A *change-point* model is described by a sequence $\nu_1 < \nu_2 < \dots$ of Markov moments with respect to the filtration $\mathcal{F}_t$ and by values $\sigma_{(1)}, \sigma_{(2)}, \dots$ where each $\sigma_{(k)}$ is $\mathcal{F}_{\nu_k}$-measurable. By definition $\sigma_t = \sigma_{(k)}$ for $\nu_k \le t < \nu_{k+1}$ and σ_t is constant for $t < \nu_1$. This is an important special case of the model (2.1). It is worth mentioning that any volatility process σ_t can be approximated by such a change-point model. For this special case, the above procedure has a very natural interpretation: when estimating at the point n we search for a largest interval of the form $[n-m, n[$ does not containing a change-point. This is doing via testing for a change-point within the interval-candidate $I = [n-m, n[$.

The construction of the procedure automatically provides the prescribed level of the first kind error probability (probability of a "false alarm"). In this section we aim to show that the procedure delivers a near optimal quality of change point detection. The quality (sensitivity) of a change-point procedure is usually measured by the mean delay between the occurrence of the change-points and its detection.

To study this property of the proposed method, we consider the case of estimation at a point n immediately after a change-point ν. The 'ideal' choice $I\!I$ of the interval of homogeneity is clearly $I\!I = [\nu, n[$. Theorem 4.5 claims that the quality of estimation of θ_n by our adaptive procedure is essentially the same as if we knew the latest change-point ν a priori. In this section we present a more detailed analysis of the change point model. In particular, we show that the change point will be detected in an early stage of the procedure provided that the magnitude of the change is sufficiently large.

Denote $m^* = |I\!I|$, that is, $m^* = n - \nu$. Let I be a larger interval containing the change-point ν, that is, $I = [\nu - m, n[= [n - m^* - m, n[$ for some m, so that $|I| = m + m^*$, and let θ (resp. θ') denote the value of parameter θ_t before (resp. after) change-point ν. To simplify the exposition we suppose below that $m = m^*$. An extension to the case when m/m^* is

bounded away from zero and infinity is straightforward. We now aim to show that such an interval I will be rejected with a high probability. It suffices to check that one particular test corresponding to a testing interval $J = I\!I$ rejects the hypothesis, that is, $T_{I,J} \geq \lambda_I$ with a high probability. The construction of the test statistic $T_{I,J}$ and Lemma 6.4 from the Appendix suggest the following measure of change from θ to θ':

$$d^2(\theta, \theta') = K(\theta, \theta_I) + K(\theta', \theta_I)$$

where $\theta_I = (\theta + \theta')/2$.

Theorem 4.7. *Let $\theta_t = \theta$ before the change-point at ν and $\theta_t = \theta'$ after it. Let $I = [\tau - 2m^*, n[$ with $m^* = n - \nu$. There exists an absolute constant C_1 such that the condition*

$$d(\theta, \theta') \geq (1 + C_1)\sqrt{\lambda_I / m^*} \tag{4.3}$$

implies

$$\boldsymbol{P}\,(I \text{ is not rejected}) \leq \beta_I\,.$$

The result of Theorem 4.7 delivers some additional information about the sensitivity of the proposed procedure to change-points. One possible question is about the minimal delay m^* between the change-point ν and the first moment n when the procedure starts to indicate this change-point by selecting an interval of type $I\!I = [\nu, n[$. Due to Theorem 4.7, the change will be "detected" with a high probability if (4.3) meets. With fixed $\theta \neq \theta'$, condition (4.3) is fulfilled if m^* is larger than a prescribed constant, that is, we need only a finite number of observations to detect a change-point. In general, m^* should be of order $d^{-2}(\theta, \theta') \asymp |\theta - \theta'|^{-2}$, if the size of the change becomes small.

Finally we discuss the quality of estimating the location of the change point by the presented procedure. Without loss of generality we can consider the change-point model with only one change at a point ν and suppose that for the interval candidate I the point ν belongs to the set of testing points $\mathcal{T}(I)$. We know from the previous result that if the size of the change is sufficiently large, then the procedure detects (with a high probability) a change point in the sense that the test statistic $T_{I,\tau}$ with $\tau = \nu$ fulfills $\boldsymbol{P}\,(T_{I,\tau} \geq \lambda_I) \approx 1$. Now we are interested to evaluate how precise our procedure estimates the location of the change point. Recall that the estimated location $\widehat{\nu}$ is the point of maximum of $T_{I,\tau}$ over all $\tau \in \mathcal{T}(I)$. Here we want to show that estimated location of the change point differs from the true location ν in typical situation at most by a finite number m.

Theorem 4.8. *Let $\theta_t = \theta$ before the change-point at ν and $\theta_t = \theta'$ after it. Let I be such that $\nu \in \mathcal{T}(I)$. There exists an absolute constant C_2 such that if*

$$K(\theta, (\theta + \theta')/2) \geq C_2 \lambda_I / m \quad \text{and} \quad K(\theta', (\theta + \theta')/2) \geq C_2 \lambda_I / m$$

then

$$\boldsymbol{P}\,(|\widehat{\nu} - \nu| > m) \leq \alpha_I\,.$$

It is worth mentioning that the conditions $K(\theta, (\theta + \theta')/2) \geq C_2\lambda_I/m$ and $K(\theta', (\theta + \theta')/2) \geq C_2\lambda_I/m$ are asymmetric w.r.t. θ, θ'. Namely, if $\theta > \theta'$ then $K(\theta, (\theta + \theta')/2) < K(\theta', (\theta+\theta')/2)$. This implies that the change from low to high volatility is easier to detect than the change for high to low volatility. All these issues are in agreement with the theory of change-point detection, see, e.g. Pollak (1985) and Brodskij & Darkhovskij (1993), and with our numerical results from Section 5.

5 Simulated results and applications

This section illustrates the performance of the proposed procedure by means of some simulated data sets and applications to real data. We aim to show that the theoretical properties of the method claimed in the previous section are nicely confirmed by the numerical results. We especially focus on the two main features of the method: stability under homogeneity and sensitivity to changes of volatility. We also discuss in more details the problem of parameter tuning for our procedure.

5.1 Parameter tuning

Here we specify the procedure which is applied both for simulated study and for applications The family of tested intervals I_k on which the procedure is performed is defined in the following way: $I_k = [n - m_k\,, n[$ where $m_k = [m_0 c^k]$ for $k = 0, 1, 2, \ldots$. Here $[x]$ means the integer part of x. The value m_0 characterizes the length of the smallest tested interval and together with the choice of ρ it determines the length of the smallest interval which can be accepted. Note that for a fixed m_0 and ρ the procedure involves the estimation of the volatility from an interval J of length $m_0\rho$. Therefore, this value should not be too small. We apply $m_0 = 15$ and $\rho = 1/3$ for our simulated examples, leading to $m_0\rho = 5$. The choice of a larger m_0 slightly decreases the sensitivity of the procedure to changes of volatility but it improves the stability and robustness of the estimator. For real (financial) data the choice $m_0 = 30$ or even $m_0 = 60$ can be recommended.

The parameter c controls the rate of grow of tested intervals I_k. The largest admissible value is $c = 2$, that is, every interval I_k is two times larger than I_{k-1}. Selecting a large c reduces the multiple testing effect for intervals I_k and the required correction of the critical values λ_I for multiple testing is not so strong, see Section 3.2. This leads to more sensitive change-point analysis for every particular interval I_k. The problem with such a rapid choice of the intervals I_k is that at some iteration an interval I_k with more that one change point can be considered, and our change point analysis may break down. Selecting c is close to one we reduce the probability of such event but the payment for multiple testing in the critical values λ_I becomes larger that results in less sensitive procedure. However, as our simulation results and applications to real data show, the overall dependence of the estimation results on the parameter c is rather minor.

Finally we discuss the choice of the critical values λ_I. We follow the suggestion of the

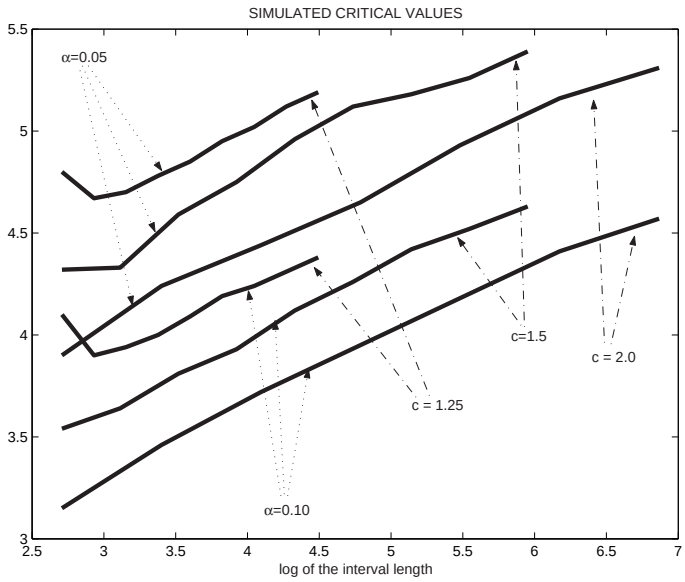

Figure 1: Critical values computed by simulation for different values of c and α.

previous section. First we define the values β_{I_k} as

$$\beta_{I_k} = \alpha m_k^{-1} \left(\sum_{\ell=1}^{\infty} m_\ell^{-1} \right)^{-1} \approx \frac{\alpha(1-c^{-1})}{c^k},$$

and the corresponding value α_{I_k} is therefore $\alpha_{I_k} \approx \alpha(1-c^{-(k+1)})$. Then we select the critical values $\lambda^*_{I_k}$ that provide the prescribed type-1-error at the level α_{I_k} for the fixed values of c and α. The results are summarized in Figure 1 where λ^*_I are plotted against the logarithm of the interval length $\log(N_I)$. It is worth mentioning that the computed values λ^*_I nicely follow the linear relationship $\lambda^*_I = a + b\log(N_{I_k})$. Moreover, the slope coefficient b is almost identical for all cases, and only the intercept slightly depend on c and α.

The results of the approximated linear regressions are shown in Table 1. The slopes of all the regression lines are about 0.35 and only the intercepts varies across c's and α's. Therefore, to ease the implementation of the procedure we suggest to define the critical values in the following form:

$$\lambda_{I_k}(c,\alpha) = \lambda_0(c,\alpha) + 0.35\log(N_{I_k}), \tag{5.1}$$

where $\lambda_0(c,\alpha)$ corresponds to the intercept reported in Table 1.

5.2 Some simulated examples

In our simulated examples we selected $c = 1.5$ and $\alpha = 0.05$ and the critical values given by equation (5.1). An extensive study which is not reported here showed that values of $c \in [1.1, 2]$ do not essentially affect the results of estimation neither on simulated, nor on

Table 1: Intercept and slope for the linear regression of the log interval length on the critical values shown in Figure 1. For $c = 1.25$ we omit the first observation in the regression.

	$\alpha = 0.05$	$\alpha = 0.10$
$c = 2.00$	3.04 0.34	2.34 0.34
$c = 1.50$	3.34 0.35	2.58 0.35
$c = 1.25$	3.61 0.35	2.93 0.32

real data, while different values of α lead to the usual trade off between type-1-error and type-2-error.

Three different jump processes are simulated, whose relative jump magnitude is 3.00, 2.00 and 1.75 respectively. Each of these processes is simulated and estimated one thousand times and the median and the quartiles of the estimates are plotted in Figure 2. We show the results for the final estimate $\widehat{\theta}$ and for the length of the selected interval $\widehat{I}$. One can see that if the size of the change is large enough, the procedure performs as if the location of the change were known. As one can expect, the sensitivity of the change point detection decreases when the magnitude of the jump becomes smaller. However, the accuracy of estimate of the volatility remains rather good even for small jumps that nicely corresponds to our theoretical results.

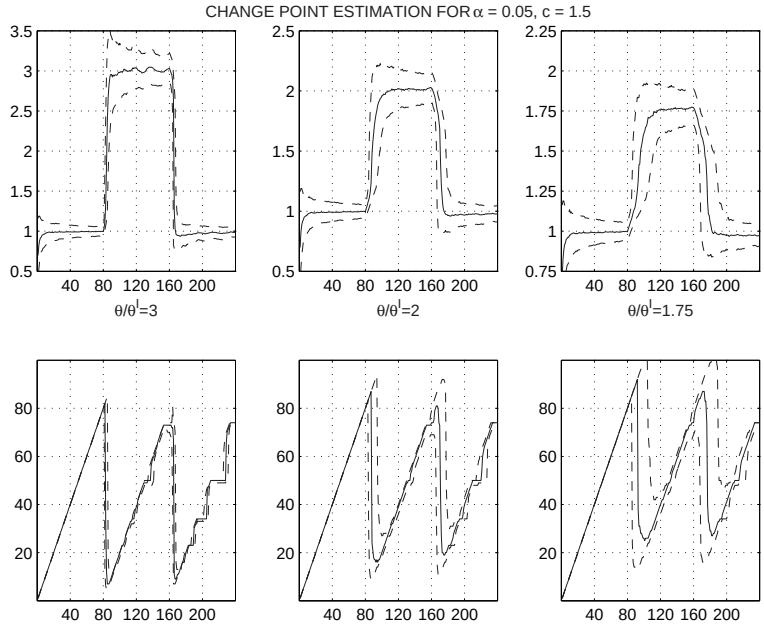

Figure 2: Pointwise median (solid line) and quartiles (dashed lines) for the estimates $\widehat{\theta}_t$ (top row) and the length of the selected interval $\widehat{I}_t$ for three jump processes with jumps of different magnitudes. The results are obtained with parameters $c = 1.5$ and $\alpha = 0.05$.

The algorithm proposed in this paper is compared with the one suggested in MS2004, with the optimized tuning parameters $\gamma = 0.5$, $M = 40$, $\lambda = 2.40$. Figure 3 shows the quartiles of estimation for the two approaches for the model with the relative jump magnitude equals 3. One can see that the new procedure outperforms the older one both with respect to the variance and to the bias of the estimator, especially for the points immediately after the changes.

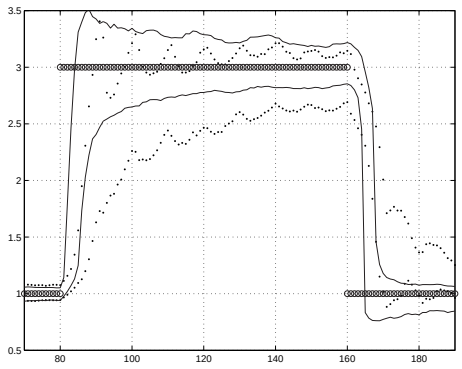

Figure 3: Comparison of the proposed estimator with the one from MS2004 for change-point model with $\theta/\theta' = 3$. Quartiles of $\widehat{\theta}$ for the new method (solid lines) and for the MS2003 (dotted lines).

Our simulation study has been done for the conditional normal model (2.1). We mentioned in Section 2.1 that this assumption is questionable as far as the real financial data is considered. To gain an impression about the robustness of the method against violation from normality we redone the simulations once again using the i.i.d. innovations from the t-distribution with five degree of freedoms. The results are shown in Figure 4. As one can expect the results are slightly worse than in the case of normal innovations, however the procedure continues to work in a quite reasonable way. The sensitivity of the procedure remains as good as for the normal innovations but a probability to reject a homogeneous interval became larger. This results in a higher variability of the estimated volatility.

5.3 Volatility estimation for some exchange rate data sets

The volatility estimation is performed on set of nine exchange rates, which are available from the web page of the US Federal Reserve. The data sets represent daily exchange rates of the US Dollar (USD) against the following currencies: Australian Dollar (AUD), British Pound (BPD) Canadian Dollar (CAD), Danish Krone (DKR), Japanese Yen (JPY), Norwegian Krone (NKR), New Zeeland Dollar (NZD), Swiss Franc (SFR) and Swedish Krone (SKR). The period under consideration goes from January the 1st 1990 to April the 7th 2000. For each time series we have 2583 observations. All the selected time series display excess kurtosis and volatility clustering.

Figure 5 and Figure 6 show the BPD/USD and JPY/USD exchange rate returns together with the volatility estimated with the parameters: $\alpha = 0.95$, $c = 1.5$ and $m_0 = 60$. The

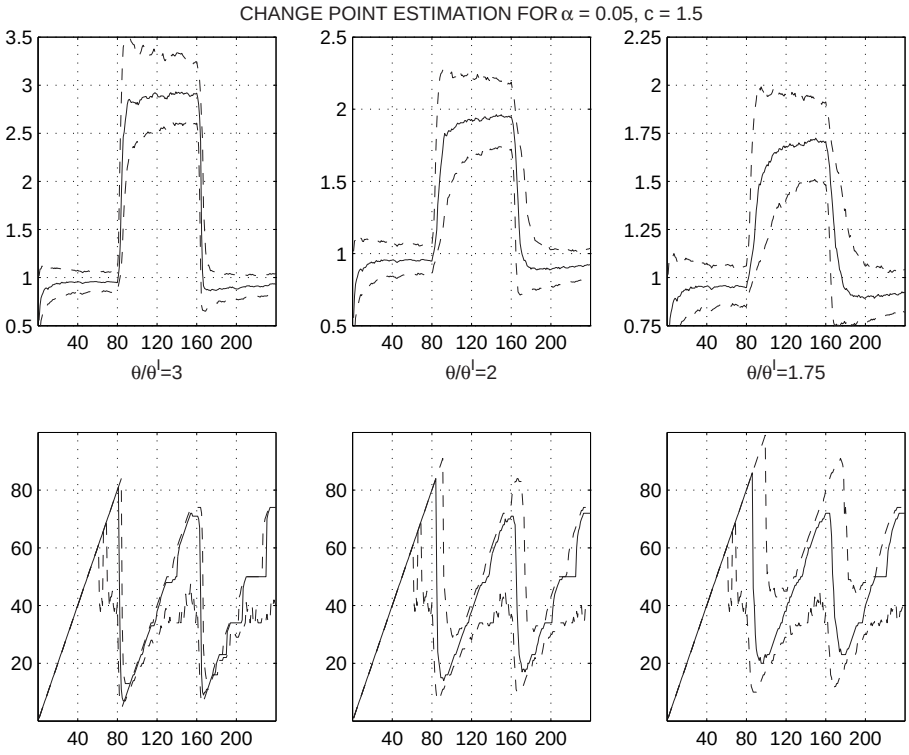

Figure 4: Estimation results with respect to jump processes with jumps of different magnitudes. The results are obtained with tuning parameters $c = 1.5$ and $\alpha = 0.05$. The conditional distribution is scaled student t_5 with 5 degrees of freedom.

choice of m_0 (which exceeds one used in the simulation) is made to improve the stability of the procedure against large shocks in the real data. The results of the estimation are in accordance with the data and the procedure seems to recognize quickly changes in the underlying volatility process.

The assumption of local homogeneity leads to the constant forecast $\widehat{\sigma}_t^2$ of the volatility σ_{t+h} for a small or moderate time horizon h. This results in the following forecast of conditional variance of the aggregated returns $R_{t+1}^2 + \ldots + R_{t+h}$:

$$V_{t,h}^{\texttt{LCPD}} := h\widehat{\sigma}_t^2.$$

Here h is the forecast horizon.

In order to assess the performance of the proposed algorithm we compare its forecasting ability with the one of the GARCH(1,1) model, which represents one of the most popular parametrization of the volatility process of financial time series. The GARCH(1,1) model is described by the following equations:

$$R_t = \sigma_t \xi_t, \quad \sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2,$$
$$\alpha > 0, \quad \beta > 0, \quad \alpha + \beta < 1, \quad \xi_t \sim N(0,1)\ \forall t.$$

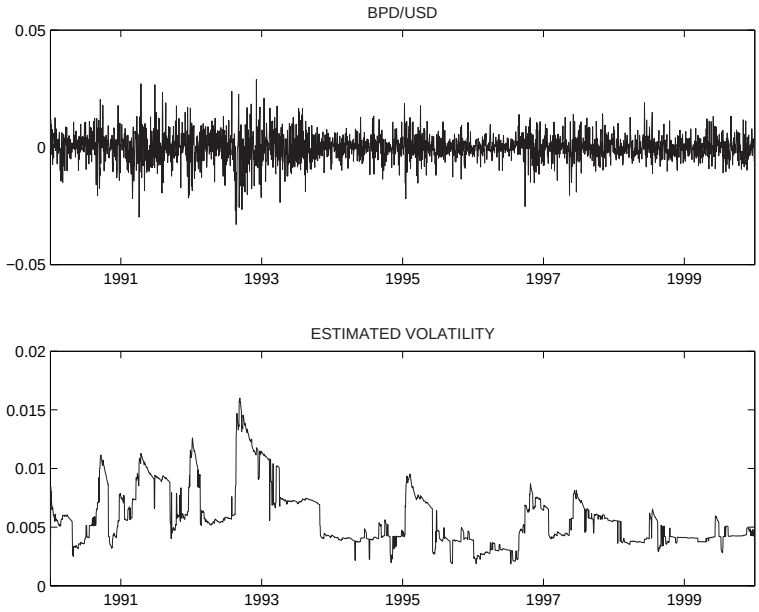

Figure 5: Returns and estimated volatility for the BPD/USD exchange rate.

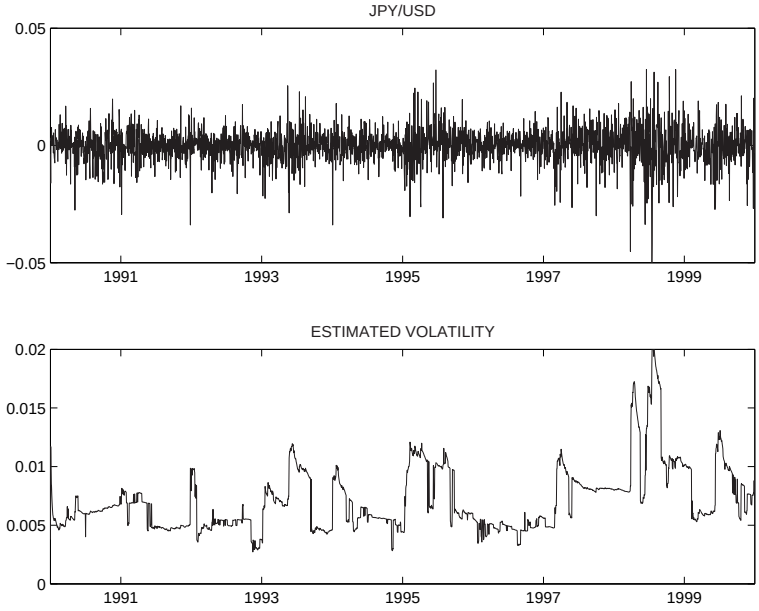

Figure 6: Returns and estimated volatility for the JPY/USD exchange rate.

The h-step ahead variance forecast of the GARCH(1,1) is given by:

$$\sigma^{2,\texttt{GARCH}}_{t+h|t} := \boldsymbol{E}_t R^2_{t+h} = \overline{\sigma}^2 + (\alpha + \beta)^h (\sigma^2_t - \overline{\sigma}),$$

where $\overline{\sigma}$ represents the unconditional volatility and $\boldsymbol{E}_t \xi$ means $\boldsymbol{E}(\xi|\mathcal{F}_t)$, see Mikosch & Starica (2000*a*). Since the returns are conditionally uncorrelated, the conditional variance

of the aggregated returns is given by the sum of the conditional variances:

$$V_{t,h}^{\texttt{GARCH}} := \boldsymbol{E}_t[(R_{t+1} + \ldots + R_{t+h})^2] = \sum_{k=1}^{h} \boldsymbol{E}_t R_{t+k}^2 = \sum_{k=1}^{h} \sigma_{t+h|t}^{2,\texttt{GARCH}}.$$

Since the assumption of constant parameter for a GARCH(1,1) model over a time interval of the considered length of about 2500 time points can be too restrictive, we considered a scrolling estimate, that is, for every date, the preceding 1000 observations are used for estimation of the GARCH parameters and then the estimated parameters are used to forecast the variance at different horizons. This method is nonadaptive in the choice of observation window but it takes advantage of a more flexible GARCH-modeling. The local change-point detection (LCPD) algorithm suggested in this paper applies a very simple local constant modeling but benefits from a data-driven choice of the interval of homogeneity.

The quality of forecasting is measured by comparing the forecasts $V_{t,h}^{\texttt{LCPD}}$ resp. $V_{t,h}^{\texttt{GARCH}}$ with the realized volatility

$$\overline{V}_{t,h} := R_{t+1}^2 + \ldots + R_{t+h}^2.$$

We apply the following mean square root error criterion (MSqE) for a time interval I:

$$\text{MSqE}_I = \sum_{t \in I} |V_{t,h}^{\texttt{LCPD}} - \overline{V}_{t,h}|^{1/2} \Big/ \sum_{t \in I} |V_{t,h}^{\texttt{GARCH}} - \overline{V}_{t,h}|^{1/2}.$$

The MSqE is considered instead of the more common MSE for robustness reasons, in this way outliers are prevented from having a strong influence on the results. The MSqE is computed for the six nonoverlapping intervals of 250 observations and the results are shown in Table 2. One can observe that both methods are comparable and the relative performance depends on the particular situation at hand. For the periods of stable volatility, the LCPD forecast is clearly better but for the periods with high volatility variation, the GARCH-method is slightly preferable.

5.4 Analysis of the standardized returns

Our model (2.1) assumes the standard normal innovations ε_t. Many empirical researches argued that this assumption is too strong and often violated, see e.g. McNeil & Frey (2000). Here we briefly discuss this issue by looking at the standardized returns $\widehat{\varepsilon}_t = R_t/\widehat{\sigma}_t$. The first observation is that even after standardization by the estimated variance, the density of standardized returns $\widehat{\varepsilon}_t$ still displays tails which are fatter than the normal. We illustrate this effect in Figure 7 where the kernel estimate of the density of the standardized returns $R_t/\widehat{\sigma}_t$ is plotted against the normal density and the scaled student t_5 density with 5 degrees of freedom. One can observe that the t-distribution delivers much better approximation to the empirical density of returns.

The volatility clustering effect, though, disappears after standardization and autocorrelations of squared returns are not significant any more, see Figure 8 the case of BPD/USD

Table 2: Relative forecasting performance MSqE on six consecutive time periods of 250 observations each.

AUD	$h=1$	**0.9616**	**0.9734**	**0.8756**	1.0978	1.1531	1.0203
	$h=5$	**0.9834**	1.0795	**0.8346**	1.1055	1.0956	**0.9968**
	$h=10$	1.0095	1.0964	**0.8586**	1.1611	1.0958	**0.9824**
CAD	$h=1$	1.1397	1.0894	**0.7473**	1.0332	1.1142	1.1177
	$h=5$	1.137	1.1169	**0.6268**	1.0405	1.0822	1.1535
	$h=10$	1.1552	1.1212	**0.5316**	1.0218	1.0651	1.1669
BPD	$h=1$	**0.7394**	**0.8475**	**0.6479**	1.0641	**0.9831**	1.0223
	$h=5$	**0.6666**	**0.8585**	**0.6006**	1.11	1.0033	1.0104
	$h=10$	**0.6123**	**0.8441**	**0.5542**	1.125	1.0185	1.0379
DKR	$h=1$	**0.8756**	**0.9980**	**0.7686**	1.022	1.0231	1.0273
	$h=5$	**0.8695**	1.0358	**0.6374**	1.0469	1.0456	1.0195
	$h=10$	**0.9001**	1.0102	**0.5933**	1.0557	**0.9994**	1.0358
JPY	$h=1$	1.1092	1.1611	**0.9721**	1.1673	1.1583	1.0327
	$h=5$	1.0628	1.1241	**0.8539**	1.1042	1.168	1.0608
	$h=10$	1.0877	1.044	**0.8016**	1.1873	1.1691	1.0607
NKR	$h=1$	**0.8776**	1.0048	**0.8810**	1.1078	1.2186	**0.9419**
	$h=5$	**0.8677**	1.017	**0.8511**	1.1061	1.2956	**0.9354**
	$h=10$	**0.9028**	1.0138	**0.7903**	1.1404	1.3232	**0.9571**
NZD	$h=1$	**0.9341**	**0.9932**	**0.9249**	1.1106	1.2433	1.1385
	$h=5$	**0.9264**	1.0052	**0.8824**	1.1463	1.1587	1.1939
	$h=10$	**0.8208**	1.0432	**0.8854**	1.179	1.165	1.2139
SFR	$h=1$	**0.9329**	1.0504	**0.8419**	**0.9564**	**0.9897**	1.0328
	$h=5$	**0.9605**	1.0874	**0.8275**	**0.9464**	**0.9721**	1.1187
	$h=10$	**0.9220**	1.0316	**0.7543**	**0.9641**	**0.9435**	1.1938
SKR	$h=1$	**0.9434**	**0.8526**	**0.7953**	1.0213	1.1042	**0.9481**
	$h=5$	**0.9438**	**0.8576**	**0.69**	1.0189	1.1097	**0.9487**
	$h=10$	**0.9532**	**0.8999**	**0.6219**	1.0704	1.1836	**0.9307**

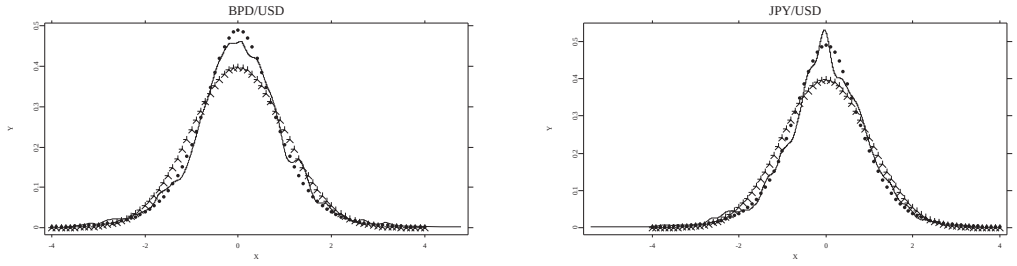

Figure 7: Kernel density estimate of exchange rate returns (solid line), normal density (x-line) and scaled student t_5 density with 5 degrees of freedom (dotted line) with fitted parameters for two exchange rate datasets.

returns. The other exchange rate examples deliver similar results. A short conclusion of this empirical study is that the standardized returns can be treated as i.i.d random variables with a distribution whose tails are fatter than normal.

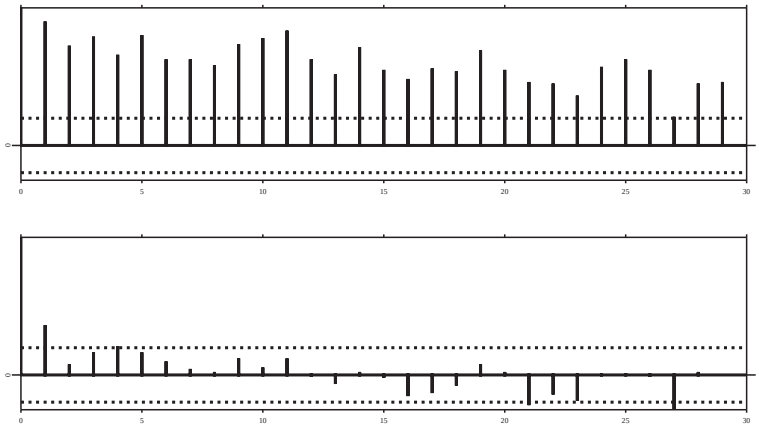

Figure 8: ACF of the absolute BPD/USD returns (upper plot) and of the standardized absolute BPD/USD returns (lower plot). Dotted straight line - the 95% significant level.

5.5 Application to Value-at-Risk

The Value-at-Risk (VaR) measures the extreme loss of a portfolio over a predetermined holding period with a prescribed confidence level $1-\alpha$. This problem can be naturally reduced to computing the quantiles of the distribution of the aggregated returns, see e.g. Fan & Gu (2003) for a nice recent overview of this topic.

Our modeling approach can be easily adapted to the VaR problem. Namely, in order to evaluate the quality of VaR estimation, for each day t, we forecast the 1% and 5% quantile of the next return R_{t+1} and of the aggregated returns $R_{t+1}+\ldots+R_{t+h} = \log(S_{t+h}/S_t)$ for the next five and 10 days assuming model (2.1) with the volatility parameter $\widehat{\sigma}_t$ estimated from the historical data R_s for $s \le t$ and using one of three different distributions for the innovations ξ_{t+h}, $h = 1, 2, \ldots, 10$: gaussian, scaled student t_5-distribution with 5 degrees of freedom and the empirical distribution $\widehat{F}_t$ of the past empirical innovations $\widehat{\varepsilon}_s$ for $s \le t$:

$$R_{t+h} = \widehat{\sigma}_t \xi_{t+h} \qquad \text{with} \qquad \xi_{t+h} \sim \mathcal{N}(0,1), \text{ or } \sqrt{5/3}\xi_{t+h} \sim t_5, \text{ or } \xi_{t+h} \sim \widehat{F}_t\,.$$

the constant volatility estimated from the historical data. Similar approaches have been applied in McNeil & Frey (2000) with the use of the GARCH(1,1) model for estimating the volatility and the extreme value theory for evaluating the distribution of the returns, while Eberlein & Prause (2002) assume the Generalized Hyperbolic Distribution for the innovations.

In order to better interpret the results, we notice that the scaled t_5 distribution has higher 5%-quantile than the ones of the gaussian at any of the considered horizons and lower 1%-quantiles. So, it is natural to expect that the gaussian distribution of innovations is slightly preferable for 5%-quantiles while t_5 does better for 1%-quantiles.

We apply the procedure to a set of nine exchange rates, which are available from the web page of the US Federal Reserve. The data sets represents daily exchange rates of the US

Dollar (USD) against the following currencies: Australian Dollar (AUD), British Pound (BPD), Canadian Dollar (CAD), Danish Krone (DKR), Japanese Yen (JPY), Norwegian Krone (NKR), New Zeeland Dollar (NZD), Swiss Franc (SFR) and Swedish Krone (SKR). The period under consideration goes from January the 1st 1990 to April the 7th 2000. The results for eight exchange rate data sets, with about 2500 observations in each one. The results (frequency of overshooting the predicted quantile for the given realizations of the returns) are given in Table 3. The first 500 observations in every time series are taken as presample for estimating the parameters. Notice that for the five and ten day horizon overlapping intervals of data are used as in Fan & Gu (2003).

We conclude that the use of the t_5 distribution for the innovations does not significantly improve the results and the VaR quality is not very good, while the application of the empirical distribution of the residuals leads to almost perfect fit of the prescribed quantiles for all considered time horizons.

Table 3: Percentage of overshooting the prescribed VaR level for nine series of exchange rate. The results are given for different nominal quantile levels, different distributions of the innovations and different values of the time horizon h.

	1% quantile									5% quantile								
	gaussian			student t_5			e.d.f.			gaussian			student t_5			e.d.f.		
h	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10	1	5	10
AUD	2.3	2.7	2.2	1.9	2.4	2.1	0.7	0.2	0.7	5.5	5.9	6.3	6.3	6.0	6.4	3.9	3.1	2.8
CAD	1.7	1.6	0.9	1.0	1.2	0.8	1.5	1.7	1.9	4.7	5.3	4.5	5.2	5.6	4.7	5.6	7.2	6.6
BPD	2.4	2.5	2.4	1.6	2.3	2.1	1.0	0.9	0.9	5.3	7.1	6.5	6.0	7.4	6.7	4.2	4.2	3.2
DKR	2.4	2.2	1.9	1.7	1.6	1.6	0.9	1.5	1.6	5.8	6.5	6.5	6.5	6.5	6.6	4.7	5.3	5.7
JPY	2.7	3.3	3.5	1.9	3.1	3.2	1.0	1.4	1.3	5.5	7.5	8.5	6.0	7.7	8.5	4.4	4.9	4.8
NKR	2.0	1.9	1.4	1.3	1.5	1.2	0.8	1.4	1.3	5.5	5.7	6.0	6.3	6.0	6.1	4.8	4.4	5.0
NZD	2.8	2.7	3.1	2.1	2.5	2.7	0.7	0.7	1.0	5.1	6.1	6.7	5.5	6.2	6.9	4.0	4.7	4.1
SFR	1.8	2.0	2.5	1.2	1.4	2.3	1.0	1.3	1.6	5.8	6.0	6.0	6.4	6.0	6.1	4.5	5.0	5.8
SKR	1.7	1.3	1.0	1.2	1.1	0.8	0.7	1.2	1.4	6.2	5.7	4.9	6.7	6.2	5.1	4.3	4.4	4.8

6 Appendix

In this section we state some results about the properties of the log-likelihood in the time-inhomogeneous volatility model and present the proofs of the results stated in previous sections.

6.1 Some properties of the log-likelihood in the homogeneous case

Let $I = [n-m, n[$ be an interval from $\mathcal{I}$. Here we present some useful results about the properties of the log-likelihood $L_I(\theta)$ and the fitted log-likelihood $\widehat{L}_I = \max_\theta L_I(\theta) = L_I(\widetilde{\theta}_I)$.

First we consider the homoscedastic situation when the volatility parameter is indeed constant within I. Denote $L_I(\theta, \theta') = L_I(\theta) - L(\theta')$ for any θ, θ'.

Theorem 6.1. *Let $\theta_t = \theta_0$ for all $t \in I$ where θ_0 is a constant or a random variable measurable w.r.t. $\mathcal{F}_{n-m-1}$. Then it holds for any θ and any $\lambda \geq 0$*

$$\boldsymbol{P}\left(L_I(\theta, \theta_0) \geq \lambda\right) \leq e^{-\lambda}$$

and

$$\boldsymbol{P}\left(L_I(\widetilde{\theta}_I, \theta_0) \geq \lambda\right) \leq 2e^{-\lambda}.$$

This result will be proved as a part of a more general result which applies if the volatility process is nearly homogeneous within I.

6.2 Properties of the log-likelihood in the nearly homogeneous case

Suppose that there exists a value θ_0 measurable w.r.t. $\mathcal{F}_{n-m-1}$ such that the values Δ_I^* and Δ_I defined by the equations (4.1) are small with a high probability. Recall the notation $A_I(\mu, z) = \{N_I \Delta_I^2 \leq \mu^2 z,\ \Delta_I^* \leq 0.8 \min\{\mu, 1\}\}$.

Theorem 6.2. *Let $\mu \geq 0$. Then it holds for any θ and any $z \geq 0$*

$$\boldsymbol{P}\left(L_I(\theta, \theta_0) \geq z + 2z\mu,\ A_I(\mu, z)\right) \leq 2N_I e^{-z}$$

and

$$\boldsymbol{P}\left(L_I(\widetilde{\theta}_I, \theta_0) \geq z + 2z\mu,\ A_I(\mu, z)\right) \leq 4N_I e^{-z}.$$

Proof. Using standard technique one can easily reduce the results of the theorem to the case when the event $A(\mu, z)$ meets almost sure, so everywhere in the proof we assume that $\Delta_I^* \leq 0.8 \min\{\mu, 1\}$ and $N_I \Delta_I^2 \leq \mu^2 z$ with probability one.

The log-likelihood ratio can be represented as

$$L_I(\theta, \theta_0) = L_I(\theta) - L_I(\theta_0) = \left(\frac{1}{2\theta_0} - \frac{1}{2\theta}\right) S_I - \frac{N_I}{2} \log(\theta/\theta_0).$$

with $S_I = \sum_{t \in I} R_t^2$.

Lemma 6.3. *For given z, there exist two values $\theta^* > \theta_0$ and $\theta_* < \theta_0$ depending on z, θ_0, N_I only such that*

$$\{L_I(\widetilde{\theta}_I, \theta_0) \geq z\} \subseteq \{L_I(\theta^*, \theta_0) \geq z\} \cup \{L_I(\theta_*, \theta_0) \geq z\}.$$

Proof. It holds

$$\begin{aligned}
\{L_I(\widetilde{\theta}_I, \theta_0) \geq z\} &= \left\{\sup_\theta \left[S_I(1/\theta_0 - 1/\theta) - N_I \log(\theta/\theta_0)\right] \geq 2z\right\} \\
&\subseteq \left\{S_I \geq \inf_{\theta > \theta_0} \frac{2z + N_I \log(\theta/\theta_0)}{1/\theta_0 - 1/\theta}\right\} \cup \left\{-S_I \geq \inf_{\theta < \theta_0} \frac{2z + N_I \log(\theta/\theta_0)}{\theta^{-1} - \theta_0^{-1}}\right\}.
\end{aligned}$$

It is straightforward to see that the function $f(u) = (2z + N_I \log(\theta/\theta_0)) \,/\, \bigl(\theta_0^{-1} - \theta^{-1}\bigr)$ attains its minimum at some point $\theta^* > \theta_0$. Therefore

$$\left\{ S \geq \inf_{\theta > \theta_0} \frac{2z + N_I \log(\theta/\theta_0)}{1/\theta_0 - 1/\theta} \right\} = \left\{ S \geq \frac{2z + N_I \log(\theta^*/\theta_0)}{1/\theta_0 - 1/\theta^*} \right\} \subseteq \{ L_I(\theta^*, \theta_0) \geq z \}.$$

Similarly

$$\left\{ -S \geq \inf_{\theta < \theta_0} \frac{2z + N_I \log(\theta/\theta_0)}{1/\theta - 1/\theta_0} \right\} \subseteq \{ L_I(\theta_*, \theta_0) \geq z \}$$

for some $\theta_* < \theta_0$. □

Lemma 6.4. *For any $z \geq 0$*

$$\boldsymbol{P}\left(\breve{L}_I(\theta) \geq z \right) \leq e^{-z}.$$

where $\breve{L}_I(\theta) = \log d\boldsymbol{P}_\theta / d\boldsymbol{P} = \sum_{t \in I} \bigl(\ell(R_t, \theta) - \ell(R_t, \theta_t) \bigr)$. Moreover, assuming the condition $\Delta_I^* \leq 0.8 \min\{\mu, 1\}$ and $N_I \Delta_I^2 \leq \mu^2 z$, it holds

$$\boldsymbol{P}\left(-\breve{L}_I(\theta) \geq 2\mu z \right) \leq e^{-z}$$

and

$$\boldsymbol{P}\left(\pm \left[\breve{L}_I(\theta) + \mathcal{K}_I(\theta) \right] \geq 2\mu z \right) \leq e^{-z}$$

where $\mathcal{K}_I(\theta) = \sum_{t \in I} K(\theta_t, \theta)$.

Proof. Since $\breve{L}_I(\theta)$ is the log-likelihood, it obviously holds

$$\boldsymbol{P}\left(\breve{L}_I(\theta) \geq z \right) \leq e^{-z} \boldsymbol{E} \exp \breve{L}_I(\theta) = e^{-z}$$

and the first assertion follows. Next,

$$\log \boldsymbol{P}\left(-\breve{L}_I(\theta) \geq 2\mu z \right) \leq -2z + \log \boldsymbol{E} \exp\left\{ -\mu^{-1} \breve{L}_I(\theta) \right\}.$$

Since R_t is progressively measurable and θ_t is predictable w.r.t. $\mathcal{F}_t$, it holds by Lemma 6.5 below

$$\begin{aligned} \boldsymbol{E} \exp\left(-\mu^{-1} \breve{L}_I(\theta) \right) &= \boldsymbol{E} \prod_{t \in I} \boldsymbol{E} \exp\bigl(\mu^{-1} \bigl(\ell(R_t, \theta_t) - \ell(R_t, \theta) \bigr) | \mathcal{F}_{t-1} \bigr) \\ &\leq \boldsymbol{E} \prod_{t \in I} \exp\bigl(|\theta_t/\theta_0 - 1|^2 / \mu^2 \bigr) \leq \boldsymbol{E} \exp\bigl(N_I \Delta_I^2 / \mu^2 \bigr) \leq e^z \end{aligned}$$

and the second assertion follows. Similarly

$$\log \boldsymbol{P}\left(\breve{L}_I(\theta) + \mathcal{K}_I(\theta) \geq 2\mu z \right) \leq -2z + \log \boldsymbol{E} \exp\left(\mu^{-1} \bigl\{ \breve{L}_I(\theta) + \mathcal{K}_I(\theta) \bigr\} \right)$$

and

$$\begin{aligned} \boldsymbol{E} e^{\left(\breve{L}_I(\theta) + \mathcal{K}_I(\theta) \right)/\mu} &= \boldsymbol{E} \prod_{t \in I} \boldsymbol{E} \exp\bigl[\mu^{-1} \bigl\{ \ell(R_t, \theta) - \ell(R_t, \theta_t) + K(\theta_t, \theta) \bigr\} | \mathcal{F}_{t-1} \bigr] \\ &\leq \boldsymbol{E} \prod_{t \in I} \exp\bigl(|\theta_t/\theta_0 - 1|^2 / \mu^2 \bigr) \leq \boldsymbol{E} \exp\bigl(N_I \Delta_I^2 / \mu^2 \bigr) \leq e^z. \end{aligned}$$

A bound for $-\breve{L}_I(\theta) - \mathcal{K}_I(\theta)$ can be proved similarly. □

Now we are prepared to complete the proof of the theorem. Indeed, $L_I(\theta,\theta_0) = \breve{L}_I(\theta) - \breve{L}_I(\theta_0)$ and Lemma 6.4 implies

$$\boldsymbol{P}\left(L_I(\theta,\theta_0) \geq \lambda + 2\mu\lambda\right) \;\leq\; \boldsymbol{P}\left(\breve{L}_I(\theta) \geq \lambda\right) + \boldsymbol{P}\left(-\breve{L}_I(\theta_0) \geq 2\mu\lambda\right) \leq 2e^{-\lambda}$$

and the result of the theorem follows by Lemma 6.3. □

6.3 Proof of Theorem 4.3

Similarly to the proof of Theorem 6.2 we reduce the general situation to the case when the conditions $N_I\Delta_I^2 \leq \mu^2\log(N_I/\beta)$ and $\Delta_I^* \leq 0.8\mu$ are fulfilled almost surely. This automatically yields $N_{I'}\Delta_{I'}^2 \leq \mu^2\log(N_I/\beta)$ and $\Delta_{I'}^* \leq 0.8\mu$ for all the subintervals J of I.

Let some point $\tau \in \mathcal{T}_I$ be fixed with the corresponding subintervals J and J^c. Then

$$T_{I,J} = L_J(\widetilde{\theta}_J,\theta_0) + L_{J^c}(\widetilde{\theta}_{J^c},\theta_0) - L_I(\widetilde{\theta}_I,\theta_0) \leq L_J(\widetilde{\theta}_J,\theta_0) + L_{J^c}(\widetilde{\theta}_{J^c},\theta_0).$$

Here $L_J(\theta,\theta_0)$ means $L_J(\theta) - L_J(\theta_0)$. We also used that $L_I(\widetilde{\theta}_I,\theta_0) \geq L_I(\theta_0,\theta_0) = 0$.

Now, it holds by Theorem 6.2 that

$$\boldsymbol{P}\left(L_J(\widetilde{\theta}_J,\theta_0) \geq (1+2\mu)\log(8N_I/\beta)\right) \leq 4\exp\bigl(-\log(8N_I/\beta)\bigr) \leq \beta/(2N_I).$$

Similarly one can bound $L_{J^c}(\widetilde{\theta}_{J^c},\theta_0)$, so that

$$\boldsymbol{P}\left(T_{I,J} \geq 2(1+2\mu)\log(8N_I/\beta)\right) \leq \beta/N_I.$$

This implies the result of the theorem because the number of testing intervals J does not exceed N_I.

6.4 Proof of Theorem 4.5

The next statement is the key step of the proof. Let an interval $I = [n-m, n[$ be accepted by the procedure. We aim to show that there exists $\tau \in \mathcal{T}_I$ such that the adaptive estimate $\widehat{\theta}$ fulfills with $J = [n-\tau, n[$

$$\left|\log(\widehat{\theta}/\widetilde{\theta}_J)\right| \leq C_0\sqrt{\lambda_I/N_I} \tag{6.1}$$

for some constant C_0 depending on the parameter ρ of the procedure. Indeed, the definition of the procedure implies for every accepted interval I and every point $\tau \in [n-m+m\rho, n-m\rho]$ that

$$T_{I,\tau} = N_J K(\widetilde{\theta}_J,\widetilde{\theta}_I) + N_{J^c}K(\widetilde{\theta}_{J^c},\widetilde{\theta}_I) \leq \lambda_I\,.$$

Since the Kullback-Leibler information K is nonnegative, this also implies $K(\widetilde{\theta}_J,\widetilde{\theta}_I) \leq \lambda_I/N_J$. Let now $\widehat{I}$ be the selected interval of the form $[n-\widehat{m}, n[$. Define $n_0 = \widehat{m}$, $n_j = [n_{j-1}/2]$, $j = 1,2,\ldots$. Because $\rho \leq 1/3$, there is some $j^* \geq 0$ such that $n_{j^*} \in [m\rho, m(1-\rho)]$. Now consider the sequence of intervals $U_j = [n-n_j, n[$ for $j = 0,\ldots,j^*$.

Since, for every $j \geq 1$, the interval U_{j-1} is accepted and U_j is one of its testing intervals, it holds $K(\widetilde{\theta}_{U_j}, \widetilde{\theta}_{U_{j-1}}) \leq \lambda_{U_{j-1}}/N_{U_j} \leq 2\lambda_{U_{j-1}}/N_{U_{j-1}}$ and, by Lemma 6.6 below, it holds $\left|\log(\widetilde{\theta}_{U_j}/\widetilde{\theta}_{U_{j-1}})\right| \leq \sqrt{12\lambda_{U_{j-1}}/N_{U_{j-1}}}$. This yields for $\widehat{\theta} = \widetilde{\theta}_{U_0}$

$$\left|\log(\widehat{\theta}/\widetilde{\theta}_{I(j^*)})\right| \leq \sum_{j=1}^{j^*} \sqrt{12\lambda_{U_{j-1}}/N_{U_{j-1}}} \leq 8\sqrt{\lambda_{U_{j^*}}/N_{U_{j^*}}}.$$

Here we have used that $N_{U_{j-1}} \geq 2N_{U_j}$ for all $j \leq j^*$ and that λ_I grows at most logarithmically with N_I. It remains to note that $N_{U_{j^*}} \geq \rho N_I$ and (6.1) follows.

By Theorem 4.4 the interval $I\!I$ will be accepted with a high probability. Moreover, in the proof of Theorem 4.4 we showed that for all testing intervals J holds with a high probability $L_J(\widetilde{\theta}_J, \theta_0) = N_J K(\widetilde{\theta}_J, \theta_0) \leq \lambda_{I\!I}/2$, which implies by Lemma 6.6 that

$$\left|\log(\widetilde{\theta}_J/\theta_0)\right| \leq \sqrt{3(1+2\mu)\lambda_{I\!I}/N_J} \leq \sqrt{3(1+2\mu)\lambda_{I\!I}/(\rho N_{I\!I})}.$$

If the interval $I\!I$ is accepted, then there is a subinterval J of $I\!I$ such that (6.1) holds, and the assertion follows from the trivial inequality

$$\left|\log(\widehat{\theta}/\theta_0)\right| \leq \left|\log(\widehat{\theta}/\widetilde{\theta}_J)\right| + \left|\log(\widetilde{\theta}_J/\theta_0)\right|.$$

6.5 Proof of Theorem 4.7

In the homogeneous situation ($\theta_t = \theta$), our choice of the critical values λ_I ensures that with the probability al least $1 - \beta_I$ it holds $L_J(\widetilde{\theta}_J, \theta) = N_J K(\widetilde{\theta}_J, \theta) \leq \lambda_I/2$ for every interval $J \in \mathcal{J}(I)$, see the proof of Theorem 4.3. In the case of the change point model we get the same bound for all intervals of homogeneity $J \in \mathcal{J}(I)$ that does contain a change point. Below in the proof we now consider the situation with $N_J K(\widetilde{\theta}_J, \theta) \leq \lambda_I/2$ for all such intervals.

Let now $I\!I = [\nu, n[$ and $J = I \backslash I\!I$, so that $\theta_t = \theta$ for $t \in J$ and $\theta_t = \theta'$ for $t \in I\!I$. We therefore assume that $K(\widetilde{\theta}_J, \theta) \leq \lambda_I/(2m^*)$ and $K(\widetilde{\theta}_{I\!I}, \theta') \leq \lambda_I/(2m^*)$.

Denote $\theta_I = (\theta + \theta')/2$. Since $N_J = N_{I\!I} = m^*$ and $N_I = 2m^*$, it also holds that

$$\widetilde{\theta}_I = (\widetilde{\theta}_{I\!I} + \widetilde{\theta}_J)/2.$$

The test statistic $T_{I,I\!I}$ can be represented as

$$T_{I,I\!I} = N_J K(\widetilde{\theta}_J, \widetilde{\theta}_I) + N_{I\!I} K(\widetilde{\theta}_{I\!I}, \widetilde{\theta}_I) = m^* d^2(\widetilde{\theta}_J, \widetilde{\theta}_{I\!I}).$$

Lemma 6.7 from below and the theorem condition imply that

$$d(\widetilde{\theta}_J, \widetilde{\theta}_{I\!I}) \geq d(\theta, \theta') - C_1\sqrt{\lambda_I/m^*} \geq \sqrt{\lambda_I/m^*}$$

and hence $T_{I,I\!I} \geq \lambda_I$. This completes the proof of the theorem.

6.6 Proof of Theorem 4.8

Let $u = \theta/\theta'$ and let $\tau \in \mathcal{T}(I)$, $\tau < \nu$. Denote $m = \nu - \tau$, $D = [\tau, \nu[$ and $J = [\tau, n[$, $J^* = [\nu, n[$, $A = I \backslash J$ and $A^* = I \backslash J^*$. We aim to bound the probability of the event $T_{I,\nu} < T_{I,\tau}$. More precisely, we intend to show that if m is sufficiently large then this probability is negligible. This particularly implies that the error of estimating the location of change point is bounded with a high probability.

Similarly to the proof of Theorem 4.7 we assume that $L_J(\widetilde{\theta}_J, \theta) = N_J K(\widetilde{\theta}_J, \theta) \le \lambda_{I'}/2$ for every $I' \in \mathcal{I}(I)$ and every $J \in \mathcal{J}(I')$ that does contain a change point. The probability of this event is not less than $1 - \alpha_I$. Obviously

$$\begin{aligned} T_{I,\nu} - T_{I,\tau} &= \widehat{L}_{A^*} + \widehat{L}_{J^*} - \left(\widehat{L}_A + \widehat{L}_J\right) \\ &= \widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J + \widehat{L}_{A^*} - \widehat{L}_A - \widehat{L}_D \,. \end{aligned}$$

Since θ_t is constant for $t \in J^*$, it holds

$$\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J \le \lambda_I \,.$$

Next we show that $\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J > \lambda_I$. This would imply that $T_{I,\tau} < T_{I,\mathrm{cp}}$ and hence $|\widehat{\nu} - \nu| > m$ is impossible. Denote $\gamma = m/(m + N^*)$ where $N^* = N_{J^*} = n - \nu$. Without loss of generality we assume that $\gamma \le 1/2$. Define $d^2_\gamma(\theta, \theta') = \gamma K(\theta, \theta_\gamma) + (1-\gamma) K(\theta', \theta_\gamma)$ with $\theta_\gamma = \gamma\theta + (1-\gamma)\theta'$. Similarly to Lemma 6.7, it holds $\widehat{L}_{J^*} + \widehat{L}_D - \widehat{L}_J = N_J d^2_\gamma(\widetilde{\theta}_{J^*}, \widetilde{\theta}_D)$ and

$$d_\gamma(\widetilde{\theta}_{J^*}, \widetilde{\theta}_D) \ge d_\gamma(\theta_{J^*}, \theta_D) - C\sqrt{\lambda_I/m} \ge K^{1/2}(\theta, (\theta+\theta')/2) - C\sqrt{\lambda_I/m}$$

for some fixed C. Now the assertion easily follows from the conditions of the theorem.

6.7 Some lemmas

In this section we collect some technical facts about the properties of the normal family with varying variance.

Lemma 6.5. *Let R be normal with parameters $(0, \theta)$ for some $\theta > 0$. Then for any θ' such that $|\theta/\theta' - 1| \le 0.8 \min\{\mu, 1\}$ it holds*

$$\log \boldsymbol{E} \exp \mu^{-1} \left\{ \ell(R, \theta) - \ell(R, \theta') \right\} \le \mu^{-2} (\theta/\theta' - 1)^2$$

and

$$\begin{aligned} \log \boldsymbol{E} \exp \mu^{-1} \left\{ \ell(R, \theta) - \ell(R, \theta') - K(\theta, \theta') \right\} &\le \mu^{-2} (\theta/\theta' - 1)^2 \\ \log \boldsymbol{E} \exp \mu^{-1} \left\{ \ell(R, \theta') - \ell(R, \theta) + K(\theta, \theta') \right\} &\le \mu^{-2} (\theta/\theta' - 1)^2 \end{aligned}$$

where $K(\theta, \theta') = -0.5 \{\log(\theta/\theta') - 1 + \theta/\theta'\}$.

Proof. Denote $\delta = \theta/\theta' - 1$. Since $\xi = \theta^{-1/2}R$ is standard normal, it holds

$$\begin{aligned}\log \boldsymbol{E} \exp \mu^{-1} \left\{\ell(R,\theta) - \ell(R,\theta')\right\} &= \frac{1}{2\mu}\log(\theta'/\theta) + \log \boldsymbol{E} \exp\left\{R^2(1/\theta' - 1/\theta)/(2\mu)\right\} \\ &= -\frac{1}{2\mu}\log(\theta/\theta') + \log \boldsymbol{E} \exp\left\{\xi^2\delta/(2\mu)\right\} \\ &= -\frac{1}{2\mu}\log(1+\delta) - \frac{1}{2}\log(1-\delta/\mu).\end{aligned}$$

Since $\left|\log(1-u) + \mu^{-1}\log(1+\mu u)\right| \le 2u^2$ for $|u| \le 0.8$ and $|\mu u| \le 0.8$, the first assertion of the lemma follows.

Similarly

$$\begin{aligned}\log \boldsymbol{E} \exp \mu^{-1} \left\{\ell(R,\theta) - \ell(R,\theta') - K(\theta,\theta')\right\} &= -\frac{\delta}{2\mu} - \frac{1}{2}\log(1-\delta/\mu) \le \delta^2/\mu^2 \\ \log \boldsymbol{E} \exp \mu^{-1} \left\{\ell(R,\theta') - \ell(R,\theta) + K(\theta,\theta')\right\} &= \frac{\delta}{2\mu} - \frac{1}{2}\log(1+\delta/\mu) \le \delta^2/\mu^2.\end{aligned}$$

□

Lemma 6.6. *The inequality $u - \log(1+u) \le 2\delta^2$ for some $\delta \ge 0$ implies for all $u \ge -1/2$ that $|\log(1+u)| \le \sqrt{6}\delta$.*

Proof. Denote $x = \log(1+u)$. Then, for $u \ge 0$, it holds $u - \log(1+u) - 1/2\log^2(1+u) = e^x - 1 - x - x^2/2 \ge 0$, that is, $\log^2(1+u) \le 2u - 2\log(1+u) \le 4\delta^2$. For $u \in [-1/2, 0]$, one similarly gets $u - \log(1+u) - (1/3)\log^2(1+u) = e^x - 1 - x - x^2/3 \ge 0$. □

The next result concerns the distance $d(u,v)$ introduced in Section 4.6: $d^2(u,v) = K(u,w) + K(v,w)$ for $w = (u+v)/2$.

Lemma 6.7. *There exists a constant C_1 such that for any positive numbers u, u_0, v, v_0, and for any $\delta \in [0,1]$, the conditions $K(u,u_0) \le \delta^2/2$, $K(v,v_0) \le \delta^2/2$ imply*

$$d(u,v) \ge d(u_0,v_0) - C_1\delta.$$

Proof. It sufficient to check that the functions $d(u,v)$, $K^{1/2}(u,u_0)$ and $K^{1/2}(v,v_0)$ have bounded partial derivatives w.r.t. the both variables u, v. We omit the details. □

References

Baillie, R., Bollerslev, T. & Mikkelsen, H. (1996), 'Fractionally integrated generalized autoregressive conditional heteroskedasticity', *J. Econometrics* **74**, 3–30.

Bollerslev, T. (1986), 'Generalised autoregressive conditional heteroskedasticity', *J. Econometrics* **31**, 307–327.

Brodskij, B. & Darkhovskij, B. (1993), *Nonparametric methods in change-point problems*, Mathematics and its Applications, Kluwer Academic Publisher, Dordrecht.

Cheng, M. Y., Fan, J. & Spokoiny, V. (2003), Dynamic nonparametric filtering with application to volatility estimation, *in* Acritas & Politis, eds, 'Recent Advances and Trends in Nonparametric Statistics', Elsevier, pp. 315–333.

Dahlhaus, R. & Rao, S. S. (2003), 'Statistical inference of time-varying arch processes', DFG Schwerpunktprogramm 1214, Nr. 31.

Eberlein, E. & Prause, K. (2002), The generalized hyperbolic model: financial derivatives and risk measures., *in* H. Geman, D. Madan, S. Pliska & T. Vorst, eds, 'Mathematical Finance-Bachelier Congress 2000', Springer Verlag, pp. 245–267.

Engle, R. (1982), 'Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation', *Econometrica* **50**, 987–1008.

Engle, R. F. & Bollerslev, T. (1986), 'Modelling the persistence of conditional variances', *Econometrc Reviews* **5**, 1–50, 81–87.

Engle, R. F., ed. (1995), *ARCH, selected readings*, Oxford University Press, Oxford.

Fan, J. & Gu, J. (2003), 'Semiparametric estimation of value at risk', *Econometrics Journal* **6**, 260–289.

Gouriéroux, C. (1997), *ARCH Models and Financial Application*, Springer-Verlag, Berlin.

Härdle, W., Herwartz, H. & Spokoiny, V. (2003), 'Time inhomogeous multiple volatility modelling', *J. Financial Econometrics* **1**, 55–95.

Kitagawa, G. (1987), 'Non-gaussian state-space modelling of nonstationary time series', *J. Amer. Statist. Assoc.* **82**, 1032–1063.

McNeil, A. J. & Frey, R. (2000), 'Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach', *J. Empirical Finance* **7**, 271–300.

Mercurio, D. & Spokoiny, V. (2004), 'Statistical inference for time-inhomogeneous volatility models', *Annals of Statistics* **32**(2).

Mercurio, D. & Torricelli, C. (2001), 'Estimation and arbitrage opportunities for exchange rate baskets'. Discussion Paper 37, Sonderforschungsbereich ğ373, Humboldt-Universität zu Berlin.

Mikosch, T. & Starica, C. (2000*a*), 'Change of structure in financial time series, long range dependence and the garch model', University of Aarhus, Arhus school of Business, Discussion Paper Nr. 58.

Mikosch, T. & Starica, C. (2000*b*), Is it really long memory we see in financial returns?, *in* P. Embrechts, ed., 'Extremes and Integrated Risk Management', Risk Books, UBS Warburg.

Pollak, M. (1985), 'Optimal detection of a change in distribution', *Annals of Statistics* **13**, 206–227.