Multiscale local change point detection with with applications to Value-at-Risk

Spokoiny, Vladimir
Weierstrass-Institute,
Mohrenstr. 39,
10117 Berlin, Germany
spokoiny@wias-berlin.de

Chen, Ying
Weierstrass-Institute,
Mohrenstr. 39,
10117 Berlin, Germany
chen@wias-berlin.de

Abstract

This paper offers a new procedure for estimation and forecasting of univariate and multivariate time series with applications to volatility modeling for financial data. The approach is based on the assumption of local homogeneity: for every time point there exists a historical *interval of 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 (LCP). 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.

Keywords: volatility model, adaptive estimation, local homogeneity, change point

AMS 2000 Subject Classification: 62M10; Secondary 62P20

1 Introduction

Since the seminal papers of Engle (1982) and Bollerslev (1986), modeling 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 modeled 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 and Bollerslev (1986) and by Baillie et al. (1996) in order to include these features in the model. Furthermore, continuous time models, and in particular diffusions and jump diffusions, have also been considered; see for example Andersen et al. (2002) and Duffie et al. (2000).

However, Mikosch and Starica (2000b) showed that long memory effects of financial time series can be artificially generated by structural breaks in the parameters. This motivates another modeling 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 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 and Gu (2003), in Dahlhaus and Rao (2003) and in Cheng et al. (2003). Furthermore, Mercurio and Spokoiny (2004) (referred to as MS2004 in what follows) proposes a new local adaptive volatility estimation (LAVE) of the unknown volatility from the conditionally heteroskedastic 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 usually outperforms the standard GARCH(1,1) approach. Härdle et al. (2003) extend this method to estimating the volatility matrix of the multiple returns and Mercurio and Torricelli (2003) 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. In this paper we systematically apply the approach based on the local multiscale change point analysis. This means that a growing sequence of historical time intervals is considered and every interval is tested on homogeneity against a change point alternative. If the hypothesis is not rejected, a larger interval candidate is taken. 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. 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 and Starica (2000a) 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.

Another important feature of the proposed procedure is that it can be easily extended to multiple volatility modeling, cf. Härdle et al. (2003). Suppose that a number of financial time series is observed and the goal is to estimate the corresponding time depending volatility matrix. We again assume that the volatility matrix is nearly constant within some historical time interval which we identify from the data. The volatility matrix is estimated in a usual way from the observations which belong to the detected interval.

The theoretical study given in Sections 2 and 4 focuses on two important features of the proposed procedure: stability in the homogeneous situation and sensitivity to spontaneous changes of the model parameter(s). We particularly show that the procedure provides the optimal sensitivity to changes for the prescribed "false alarm" probability. Our main estimation result in Theorem 4.5 claims that the procedure delivers the estimation accuracy corresponding to the largest historical interval of homogeneity as if this interval were known.

The paper is organized as follows. Section 2 describes the local parametric approach

for the volatility modeling and presents some results about the accuracy of the local constant volatility estimate in the univariate and multivariate case. Section 3 introduces the adaptive modeling procedure, first in the univariate case then extending it to the multivariate situation. Theoretical properties of the procedure 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 in Section 4. 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 is systematically applied for all the examples. Section 5.1 presents some numerical results for a change point model. In Section 5.3 we study forecasting ability of the new method by mean of a comparative study with the GARCH(1,1) method. Sections 5.6 and 5.7 discuss applications of the new method to the Value at Risk problem. Finally, Section 6 collects the proofs of the main results.

2 Volatility modeling. Local parametric approach

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

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

where ε_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, σ_t is measurable with respect to $\mathcal{F}_{t-1}$ 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 MS2004 we focus on the problem of filtering the parameter $f(t) = \sigma_t^2$ 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-homoskedastic*) 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 $\sigma_t^2 \equiv \theta$ with $t \in I$, the squared returns $Y_t = R_t^2$ follow the

equation $Y_t = \theta\varepsilon_t^2$ and the parameter θ can be estimated using the maximum likelihood method:

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

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

$$L_I(\theta) = -(N_I/2)\log(2\pi\theta) - S_I/(2\theta) \tag{2.2}$$

where N_I denotes the number of time points in I and $S_I = \sum_{t\in I} Y_t$.

The volatility model is a particular case of an exponential family, so that a closed form representation for the maximum likelihood estimate $\widetilde{\theta}_I$ and for the corresponding fitted log-likelihood $L_I(\widetilde{\theta}_I)$ are available, see Polzehl and Spokoiny (2006) for more details.

Theorem 2.1. *For every interval I*

$$\widetilde{\theta}_I = S_I/N_I = N_I^{-1}\sum_{t\in I} Y_t.$$

Moreover, for every $\theta > 0$ the fitted likelihood ratio $L_I(\widetilde{\theta},\theta) = \max_{\theta'} L_I(\theta',\theta)$ with $L_I(\theta',\theta) = L_I(\theta') - L_I(\theta)$ satisfies

$$L_I(\widetilde{\theta}_I,\theta) = N_I\mathcal{K}(\widetilde{\theta}_I,\theta) \tag{2.3}$$

where

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

is the Kullback-Leibler information for the two normal distributions with variances θ' and θ.

Proof. The both results follow by simple algebra from (2.2). □

Remark 2.1. 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.

2.2 Some properties of the (quasi) MLE $\widetilde{\theta}_I$ in the homogeneous situation

This section collects some useful properties of the (quasi) MLE $\widetilde{\theta}_I$ and of the fitted log-likelihood $L_I(\widetilde{\theta}_I, \theta^*)$ in the homogeneous situation $f(t) = \theta^*$ for $t \in I$.

Theorem 2.2 (Polzehl and Spokoiny, 2005)**.** *Let $f(t) = \theta^*$ for $t \in I$. If the innovations ε_t are i.i.d. standard normal, then for any $\mathfrak{z} > 0$*

$$I\!P_{\theta^*}(L_I(\widetilde{\theta}_I, \theta^*) > \mathfrak{z}) = I\!P_{\theta^*}\Big(N_I \mathcal{K}(\widetilde{\theta}_I, \theta^*) > \mathfrak{z}\Big) \le 2e^{-\mathfrak{z}}.$$

The Kullback-Leibler divergence $\mathcal{K}$ fulfills $\mathcal{K}(\theta', \theta^*) \le I^*|\theta' - \theta^*|^2$ for any point θ' in a neighborhood of θ^*, where I^* is the maximum of the Fisher information over this neighborhood. Therefore, the result of Theorem 2.2 guarantees that $|\widetilde{\theta}_I - \theta^*| \le C N_I^{-1/2}$ with a high probability. Theorem 2.2 can be used for constructing the confidence intervals for the parameter θ^*.

Theorem 2.3. *If $\mathfrak{z}_\alpha$ satisfies $2e^{-\mathfrak{z}_\alpha} \le \alpha$, then*

$$\mathcal{E}_\alpha = \{\theta : N_I \mathcal{K}\big(\widetilde{\theta}_I, \theta\big) \le \mathfrak{z}_\alpha\}$$

is an α-confidence set for the parameter θ^*.

Theorem 2.2 claims that the estimation loss measured by $\mathcal{K}(\widetilde{\theta}_I, \theta^*)$ is with high probability bounded by $\mathfrak{z}/N_I$ provided that $\mathfrak{z}$ is sufficiently large. Similarly, one can establish a risk bound for a power loss function.

Theorem 2.4. *Let R_t be i.i.d. from $\mathcal{N}(0, \theta)$. Then for any $r > 0$*

$$I\!E_{\theta^*}\big|L_I(\widetilde{\theta}_I, \theta^*)\big|^r \equiv I\!E_{\theta^*}\big|N_I \mathcal{K}(\widetilde{\theta}_I, \theta^*)\big|^r \le \mathfrak{r}_r\,.$$

where $\mathfrak{r}_r = 2r \int_{\mathfrak{z} \ge 0} \mathfrak{z}^{r-1} e^{-\mathfrak{z}} d\mathfrak{z} = 2r\Gamma(r)$. Moreover, for every $\lambda < 1$

$$I\!E_{\theta^*} \exp\big\{\lambda L_I(\widetilde{\theta}_I, \theta^*)\big\} \equiv I\!E_{\theta^*} \exp\big\{\lambda N_I \mathcal{K}(\widetilde{\theta}_I, \theta^*)\big\} \le 2(1-\lambda)^{-1}.$$

Proof. By Theorem 2.2

$$\begin{aligned} I\!E_{\theta^*}\big|L_I(\widetilde{\theta}_I, \theta^*)\big|^r &\le -\int_{\mathfrak{z} \ge 0} \mathfrak{z}^r \, dI\!P_{\theta^*}(L(\widetilde{\theta}_I, \theta^*) > \mathfrak{z}) \\ &\le r \int_{\mathfrak{z} \ge 0} \mathfrak{z}^{r-1} I\!P_{\theta^*}(L_I(\widetilde{\theta}_I, \theta^*) > \mathfrak{z}) d\mathfrak{z} \le 2r \int_{\mathfrak{z} \ge 0} \mathfrak{z}^{r-1} e^{-\mathfrak{z}} d\mathfrak{z} \end{aligned}$$

and the first assertion is fulfilled. The last assertion is proved similarly. □

Remark 2.2. The results of Theorems 2.2 and 2.4 can be extended to the case of non-Gaussian innovations ε_t under some conditions on the exponential moments of the ε_t, see Golubev and Spokoiny (2006).

2.3 Risk of estimation in nonparametric situation. "Small modeling bias" condition

This section extends the bound of Theorem 2.4 to the nonparametric model $R_t^2 = f(t)\varepsilon_t^2$ when the function $f(\cdot)$ is not any longer constant even in a vicinity of the reference point $t^\diamond$. We, however, suppose that the function $f(\cdot)$ can be well approximated by a constant θ at all points $t \in I$.

Let $Z_\theta = d\mathbb{P}/d\mathbb{P}_\theta$ be the likelihood ratio of the underlying measure $\mathbb{P}$ w.r.t. the parametric measure $\mathbb{P}_\theta$ corresponding to the constant parameter $f(\cdot) \equiv \theta$. Then

$$\log Z_\theta = \sum_t \log \frac{p(Y_t, f(t))}{p(Y_t, \theta)}.$$

If we restrict our analysis to an interval I and denote by $\mathbb{P}_I$ resp. $\mathbb{P}_{I,\theta}$ the measure corresponding to the observations Y_t for $t \in I$, then in a similar way

$$\log Z_{I,\theta} := \log \frac{d\mathbb{P}_I}{d\mathbb{P}_{I,\theta}} = \sum_{t \in I} \log \frac{p(Y_t, f(t))}{p(Y_t, \theta)}.$$

To measure the quality of the approximation of the underlying measure $\mathbb{P}_I$ by the parametric measure $\mathbb{P}_{I,\theta}$, define

$$\Delta_I(\theta) = \sum_{t \in I} \mathcal{K}\big(f(t), \theta\big), \tag{2.4}$$

where $\mathcal{K}\big(f(t), \theta\big)$ means the Kullback-Leibler distance between two parameter values $f(t)$ and θ.

Let $\varrho(\widetilde{\theta}_I, \theta)$ be a loss function for an estimate $\widetilde{\theta}_I$ constructed from the observations Y_t for $t \in I$. Define also the corresponding risk under the parametric measure $\mathbb{P}_\theta$:

$$\mathcal{R}(\widetilde{\theta}_I, \theta) = \mathbb{E}_\theta \varrho(\widetilde{\theta}_I, \theta).$$

The next result explains how the risk bounds can be translated from the parametric to the nonparametric situations.

Theorem 2.5. *Let for some $\theta \in \Theta$ and some $\Delta \geq 0$*

$$\mathbb{E}\Delta_I(\theta) \leq \Delta. \tag{2.5}$$

Then it holds for any estimate $\widetilde{\theta}$ measurable w.r.t. $\mathcal{F}_I$

$$\mathbb{E}\log\big(1 + \varrho(\widetilde{\theta}, \theta)/\mathcal{R}(\widetilde{\theta}, \theta)\big) \leq \Delta + 1.$$

Proof. The proof is based on the following general result.

Lemma 2.6. *Let $I\!P$ and $I\!P_0$ be two measures such that the Kullback-leibler divergence $I\!E \log(dI\!P/dI\!P_0)$, satisfies*

$$I\!E \log(dI\!P/dI\!P_0) \le \Delta < \infty.$$

Then for any random variable ζ with $I\!E_0\zeta < \infty$

$$I\!E \log\bigl(1+\zeta\bigr) \le \Delta + I\!E_0\zeta.$$

Proof. By simple algebra one can check that for any fixed y the maximum of the function $f(x) = xy - x\log x + x$ is attained at $x = e^y$ leading to the inequality $xy \le x\log x - x + e^y$. Using this inequality and the representation $I\!E \log\bigl(1+\zeta\bigr) = I\!E_0\bigl\{Z\log\bigl(1+\zeta\bigr)\bigr\}$ with $Z = dI\!P/dI\!P_0$ we obtain

$$\begin{aligned} I\!E \log\bigl(1+\zeta\bigr) &= I\!E_0\bigl\{Z\log\bigl(1+\zeta\bigr)\bigr\} \\ &\le I\!E_0\bigl(Z\log Z - Z\bigr) + I\!E_0(1+\zeta) \\ &= I\!E_0\bigl(Z\log Z\bigr) + I\!E_0\zeta - I\!E_0 Z + 1. \end{aligned}$$

It remains to note that $I\!E_0 Z = 1$ and $I\!E_0\bigl(Z\log Z\bigr) = I\!E\log Z$. □

We now apply this lemma with $\zeta = \varrho(\widetilde{\theta},\theta)/\mathcal{R}(\widetilde{\theta},\theta)$ and utilize that $I\!E_0\zeta = I\!E_\theta\varrho(\widetilde{\theta},\theta)/\mathcal{R}(\widetilde{\theta},\theta) = 1$. This yields

$$\begin{aligned} I\!E_\theta\bigl(Z_{I,\theta}\log Z_{I,\theta}\bigr) &= I\!E\log Z_{I,\theta} = I\!E\sum_{t\in I}\log\frac{p(Y_t,f(t))}{p(Y_t,\theta)} \\ &= I\!E\sum_{t\in I} I\!E\Bigl\{\log\frac{p(Y_t,f(t))}{p(Y_t,\theta)}\Bigl|\mathcal{F}_{t-1}\Bigr\} = I\!E\Delta_I(\theta) \end{aligned}$$

□

This result implies that the bound for the risk of estimation $I\!E L_I^r(\widetilde{\theta}_I,\theta) \equiv N_I^r I\!E\mathcal{K}^r(\widetilde{\theta}_I,\theta)$ under the parametric hypothesis can be extended to the nonparametric situation provided that the value $\Delta_I(\theta)$ is sufficiently small.

Define for $r > 0$

$$\varrho(\widetilde{\theta}_I,\theta) = \bigl|N_I\mathcal{K}(\widetilde{\theta}_I,\theta)\bigr|^r.$$

By Theorem 2.4 $\mathcal{R}(\widetilde{\theta}_I,\theta) = I\!E_\theta\varrho(\widetilde{\theta}_I,\theta) \le \mathfrak{r}_r$.

Corollary 2.7. *Let (2.5) hold for some* θ. *For any* $r > 0$

$$\mathbb{E} \log\Bigl(1 + \bigl|N_I \mathcal{K}(\widetilde{\theta}_I, \theta)\bigr|^r / \mathfrak{r}_r\Bigr) \le \Delta + 1.$$

This result means that in the nonparametric situation under the condition (2.5) with some fixed Δ the losses $\bigl|N_I \mathcal{K}(\widetilde{\theta}_I, \theta)\bigr|^r$ are stochastically bounded. Note that this result applies even if Δ is large, however the bound is proportional to $\mathrm{e}^{\Delta+1}$ and grows exponentially with Δ.

2.4 "Small modeling bias" condition and rate of estimation

This section briefly comment on relations between the results of Section 2.3 and the usual rate results under smoothness conditions on the function $f(\cdot)$.

Let n be the parameter meaning length of the largest considered historical interval. More precisely, we assume that the function $f(\cdot)$ is smooth in the sense that for $\theta^* = f(t^\diamond)$ and any $t \ge t^\diamond - n$

$$\mathcal{K}^{1/2}\bigl(f(t), \theta^*\bigr) \le (t^\diamond - t)/n. \tag{2.6}$$

In view of the inequality $\mathcal{K}(\theta, \theta') \asymp |\theta/\theta' - 1|^2$ this condition is equivalent to the usual Lipschitz property of the rescaled function $f(t/n)$. This condition bounds the *bias* of approximating the underlying function $f(t)$ by a constant $f(t^\diamond)$ by $(t^\diamond - t)/n$. The *variance* of the estimate $\widetilde{\theta}_I$ for $I = [t, t^\diamond[$ is proportional to $1/(t^\diamond - t)$. The usual "bias-variance trade-off" means the relation "$\mathrm{bias}^2 \asymp \mathrm{variance}$", leading to $(t^\diamond - t)^3 \asymp n^2$.

Now note that (2.5) and (2.6) implies

$$\Delta_I(\theta^*) \le N_I^3 / n^2.$$

Therefore, the "small modeling bias" condition $\Delta_I(\theta) \le \Delta$ is essentially equivalent to "bias-variance trade-off". Moreover, combined with the result of Corollary 2.7 this condition lead to the following classical rate results.

Theorem 2.8. *Assume (2.6). Select* I *such that* $N_I = cn^{2/3}$ *for some* $c > 0$. *Then (2.5) holds with* $\Delta = c^3$ *and for any* $r > 0$

$$\log\Bigl(1 + \bigl|N_I \mathcal{K}(\widetilde{\theta}_I, \theta)\bigr|^r / \mathfrak{r}_r\Bigr) \le c^3 + 1.$$

This corresponds to the classical accuracy of nonparametric estimation for the Lipschitz functions, cf. Fan, Farmen and Gijbels (1998).

2.5 Extension to multiple volatility modeling

In this section we discuss how the approach local parametric can be extended to modeling of multiple time series. Let $S_t \in \mathbb{R}^d$ be a vector of observed asset processes in discrete time, $t = 1, 2, \ldots$ and R_t is the vector of the corresponding returns: $R_{t,m} = \log(S_{t,m}/S_{t-1,m})$, $m = 1, \ldots, d$. The *conditional heteroscedasticity* assumption reads in this case as

$$R_t = \Sigma_t^{1/2} \varepsilon_t \,, \tag{2.7}$$

where ε_t, $t \geq 1$, is a sequence of independent standard Gaussian random vectors and Σ_t is a symmetric $d \times d$ *volatility* matrix which is in general a predictable random process, that is, Σ_t is measurable with respect to $\mathcal{F}_{t-1}$ with $\mathcal{F}_{t-1} = \sigma(R_1, \ldots, R_{t-1})$ (σ-field generated by the first $t-1$ observations).

A *time-homogeneous* (*time-homoscedastic*) model means that the matrix Σ_t is a constant. For the homogeneous model $R_t = \Sigma^{1/2} \varepsilon_t$ with $t \in I$, the parameter Σ can be estimated from the observations R_t, $t \in I$ by maximizing the log-likelihood

$$L_I(\Sigma) = \sum_{t \in I} \ell(R_t, \Sigma)$$

where $\ell(y, \Sigma) = -(1/2)\log(\det \Sigma) - y^\top \Sigma^{-1} y/2$ for $y \in \mathbb{R}^d$ corresponds to the log-density of the normal distribution with the parameters $(0, \Sigma)$. (We skip the unimportant constant term $-(d/2)\log(2\pi)$ in the expression of the density.) The simple algebra yields

$$L_I(\Sigma) = -\frac{N_I}{2} \log(\det \Sigma) - \frac{1}{2} \sum_{t \in I} R_t^\top \Sigma^{-1} R_t = -\frac{N_I}{2} \log(\det \Sigma) - \frac{N_I}{2} \mathrm{tr}\big(\Sigma^{-1} \widetilde{\Sigma}_I\big)$$

where $\widetilde{\Sigma}_I$ be the empirical covariance matrix:

$$\widetilde{\Sigma}_I = N_I^{-1} \sum_{t \in I} R_t R_t^\top .$$

Theorem 2.9. *For every interval I, it holds*

$$\begin{aligned} \widetilde{\Sigma}_I &= \operatorname*{argmax}_{\Sigma} L_I(\Sigma) \\ L_I(\widetilde{\Sigma}_I) &= -\frac{N_I}{2} \log(\det \widetilde{\Sigma}_I) - \frac{d}{2} \end{aligned} \tag{2.8}$$

Moreover, for every $\Sigma \in \Theta$ the fitted likelihood ratio $L_I(\widetilde{\Sigma}, \Sigma) = \max_{\Sigma' \in \Theta} L_I(\Sigma', \Sigma)$ with $L_I(\Sigma', \Sigma) = L_I(\Sigma') - L_I(\Sigma)$ satisfies

$$L_I(\widetilde{\Sigma}_I, \Sigma) = N_I \mathcal{K}(\widetilde{\Sigma}_I, \Sigma) \tag{2.9}$$

where

$$\mathcal{K}(\Sigma, \Sigma_0) = -0.5\big\{\log(\det \Sigma / \det \Sigma_0) - \mathrm{tr}\big(\Sigma\Sigma_0^{-1} - I_d\big)\big\}$$

is the Kullback-Leibler information for the two normal distributions with covariance matrices Σ *and* Σ_0*, see Kullback (1959).*

Proof. The first assertion is can be proved by differentiating the expression $L_I(\Sigma)$ w.r.t. Σ. The other results follow by simple algebra. □

The important deviation result of Theorem 2.4 extends to the multiple case as well under some technical condition on the set Θ of volatility matrices Σ.

$(\boldsymbol{\Theta})$ for some constant $\mathfrak{a}$ with $0 < \mathfrak{a} \le 1$, any two matrices $\Sigma, \Sigma_0 \in \Theta$ satisfy

$$\mathfrak{a}^2 \le \Sigma^{-1/2}\Sigma_0\Sigma^{-1/2} \le \mathfrak{a}^{-2}.$$

In the univariate case this condition means that the volatility parameter has a finite range:

$$\mathfrak{a}^2 \le \theta_0/\theta \le \mathfrak{a}^{-2}$$

Theorem 2.10. *Assume* (Θ). *Let* R_t *be i.i.d. from* $\mathcal{N}(0, \Sigma^*)$ *for* $\Sigma^* \in \Theta$. *Then for any* $r > 0$

$$I\!\!E\big|L_I(\widetilde{\Sigma}_I, \Sigma^*)\big|^r \equiv I\!\!E\big|N_I\mathcal{K}(\widetilde{\Sigma}_I, \Sigma^*)\big|^r \le \mathfrak{R}_r\,,$$

where constant $\mathfrak{R}_r$ *depends on the dimension* d *and the constant* $\mathfrak{a}$ *from* (Θ) *only.*

In the nonparametric situation when the volatility matrix Σ_t is not constant any more, we again again apply the approach based on the local parametric approximation and small modeling bias condition. This condition means that the underlying volatility process Σ_t can be well approximated within the interval I by a constant matrix Σ where the quality of approximation is measured by the quantity

$$\Delta_I(\Sigma) = \sum_{t \in I} \mathcal{K}\big(\Sigma_t, \Sigma\big).$$

Similarly to the univariate case, $\Delta_I(\Sigma)$ is a random variable and the "small modeling bias" condition should bound it in a stochastic sense.

Theorem 2.11. *Assume* (Θ). *Let for some fixed matrix* Σ *and a constant* Δ

$$\mathbb{E}\Delta_I(\Sigma) \le \Delta.$$

Then for every $r > 0$

$$\mathbb{E}\log\bigl(1 + \bigl|N_I\mathcal{K}(\widetilde{\Sigma}_I, \Sigma^*)\bigr|^r/\mathfrak{R}_r\bigr) \le \Delta + 1.$$

This result, as in the univariate situation, yields the usual nonparametric quality of estimation for the local MLE $\widetilde{\Sigma}_I$.

3 Adaptive volatility estimation

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 assumption of *local time-homogeneity* which means that for every time moment $t^\diamond$ there exists a historic time interval $[t^\diamond - m, t^\diamond[$ 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 $\sigma_{t^\diamond}$.

3.1 Choice of the interval of homogeneity by local change point analysis

Our approach is based on the adaptive choice of the interval of homogeneity for the fixed end point $t^\diamond$. This choice is made by the *local (multiscale) change point detection* (LCP) algorithm described below. The procedure attempts to find this interval from the data by successive testing the hypothesis of homogeneity. Let a growing sequence of numbers $m_1 < m_2 < \ldots < m_K$ be fixed. Each m_k means the *scale* parameter describing the length of the historical time interval screened at the step k. Define a family $\{\mathcal{I}_k, k = 1, \ldots, K\}$ of nested intervals of the form $\mathcal{I}_k = [t^\diamond - m_k, t^\diamond[$ with the right edge at $t^\diamond$. The procedure starts from the smallest interval $\mathcal{I}_1$ by testing the hypothesis of homogeneity within $\mathcal{I}_1$ against a change point alternative. If the hypothesis is not rejected then we take the next larger interval $\mathcal{I}_k$ and test for a change point. We continue this way until we detect a change point or the largest considered interval $\mathcal{I}_K$ is accepted. If the hypothesis of homogeneity within some interval $\mathcal{I}_k$ is rejected and a change point is detected at a point $\widehat{\tau} \in \mathcal{I}_k$, then the estimated interval of homogeneity is defined as the latest accepted interval, that is, $\widehat{\mathcal{I}} = \mathcal{I}_{k-1} = [t^\diamond - m_{k-1}, t^\diamond[$, otherwise we take $\widehat{\mathcal{I}} = \mathcal{I}_K$. Finally, we define the estimate $\widehat{f}(t^\diamond) = \widehat{\theta}$ of the volatility parameter $f(t^\diamond) = \sigma^2_{t^\diamond}$ as $\widehat{f}(t^\diamond) = \widetilde{\theta}_{\widehat{\mathcal{I}}}$.

The main ingredient of this procedure is the homogeneity test for every interval $\mathcal{I}_k$, which is described in the next section.

3.2 Test of homogeneity against a change point alternative

Let J be a *tested* interval which has to be checked on a possible change point. For carrying out the test, we also introduce be a larger *testing* interval $I = [t', t''[$. The hypothesis of homogeneity for J means that the observations R_t follow the parametric model with the parameter θ for J itself and for the larger interval I. This hypothesis leads to the log-likelihood $L_I(\theta)$ for the observations $R_t \in I$. 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 J. Every point $\tau \in J$ splits the interval $I = [t', t''[$ onto two subintervals $I'' = [\tau, t''[$ and $I' = I \setminus I'' = [t', \tau[$. The change point alternative means that $f(t) = \theta''$ for $t \in I''$ and $f(t) = \theta'$ for $t \in I''$ for some $\theta'' \neq \theta'$. This corresponds to the log-likelihood $L_{I''}(\theta'') + L_{I'}(\theta')$. The likelihood ratio test statistic for the change point alternative with the change point location at the point τ is of the form

$$\begin{aligned} T_{I,\tau} &= \max_{\theta'',\theta'} \bigl\{ L_{I''}(\theta'') + L_{I'}(\theta') \bigr\} - \max_{\theta} L_I(\theta) \\ &= L_{I''}(\widetilde{\theta}_{I''}) + L_{I'}(\widetilde{\theta}_{I'}) - L_I(\widetilde{\theta}_I). \end{aligned}$$

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

$$T_{I,\tau} = \min_{\theta} \bigl\{ N_{I''} \mathcal{K}(\widetilde{\theta}_{I''}, \theta) + N_{I'} \mathcal{K}(\widetilde{\theta}_{I'}, \theta) \bigr\} = N_{I''} \mathcal{K}(\widetilde{\theta}_{I''}, \widetilde{\theta}_I) + N_{I'} \mathcal{K}(\widetilde{\theta}_{I'}, \widetilde{\theta}_I), \qquad (3.1)$$

see (2.3).

The change point test for the interval J is defined as the maximum of such defined test statistics over $\tau \in J$:

$$T_I = \max_{\tau \in J} T_{I,\tau}\,. \qquad (3.2)$$

The change point test compares this statistic with the critical value $\mathfrak{z}$ which may depend on the intervals J and I. The hypothesis of homogeneity is rejected if $T_I \geq \mathfrak{z}$, in this case the estimator of the change point location is defined as $\widehat{\tau} = \operatorname{argmax}_{\tau \in J_I} T_{I,\tau}$.

In the multivariate case, the corresponding test is very similar. The null hypothesis for J means that the observations R_t for $t \in I$ follow the parametric model with the volatility matrix Σ. This hypothesis leads to the log-likelihood $L_I(\Sigma)$. We want to test this hypothesis against a change point alternative that the volatility matrix Σ spontaneously changes in some internal point τ of the interval J, so that $\Sigma_t = \Sigma'$ for

$t \in I'$ and $\Sigma_t = \Sigma''$ for $t \in I''$. Similarly to the univariate case, the corresponding likelihood ratio test statistic is of the form

$$\begin{aligned} T_{I,\tau} &= \max_{\Sigma',\Sigma''} \left\{ L_{I'}(\Sigma') + L_{I''}(\Sigma'') \right\} - \max_{\Sigma} L_I(\Sigma) \\ &= L_{I'}(\widetilde{\Sigma}_{I'}) + L_{I''}(\widetilde{\Sigma}_{I''}) - L_I(\widetilde{\Sigma}_I). \end{aligned}$$

For the considered volatility model, due to (2.9), this test statistic can be represented as

$$T_{I,\tau} = N_{I'} \mathcal{K}(\widetilde{\Sigma}_{I'}, \widetilde{\Sigma}_I) + N_{I''} \mathcal{K}(\widetilde{\Sigma}_{I''}, \widetilde{\Sigma}_I)$$

where $\mathcal{K}(\Sigma, \Sigma_0) = 0.5 \left\{ \mathrm{tr}(\Sigma {\Sigma_0}^{-1} - I_d) - \log(\det \Sigma / \det \Sigma_0) \right\}$ is the Kullback-Leibler divergence for the two normal distributions with variances Σ and Σ_0.

Remark 3.1. The change point alternative suggested above is only one of possibilities to test the homogeneity assumptions. One can apply many other tests, e.g. omnibus tests against polynomials or trigonometric functions, see e.g. Hart (1998). Our choice is motivated by several reasons. First of all, it is simple to implement and does not require a special model estimation under alternative because the alternative reduces to the null hypothesis for two smaller intervals. Secondly, it has a natural interpretation and delivers an additional information about the location of the change and the length of the interval of homogeneity. Finally, it was shown in Ingster (1986), see also Horowitz and Spokoiny (2001) that a test based on the local constant alternative is powerful against smooth alternatives as well.

3.3 The procedure

This section describes the LCP procedure for the univariate case. The procedure is sequential and consists of K steps corresponding to the given growing sequence of numbers $m_0 < m_1 < \ldots < m_K$. This sequence determines the sequence of nested intervals $\mathcal{I}_0 \subset \mathcal{I}_1 \subset \ldots \subset \mathcal{I}_K$ with the right edge at the point of estimation $t^\diamond$: $\mathcal{I}_k = [t_k^\diamond, t^\diamond[= [t^\diamond - m_k, t^\diamond[$. This set of intervals leads to the set of estimates $\widetilde{\theta}_{\mathcal{I}_k}$, $k = 0, 1, \ldots, K$. For conciseness of notation, we write $\widetilde{\theta}_k$ in place of $\widetilde{\theta}_{\mathcal{I}_k}$ and N_k in place of $N_{\mathcal{I}_k} = m_k$.

The proposed adaptive method chooses an index $\widehat{k}$ of equivalently, the estimate $\widetilde{\theta}_{\widehat{k}}$ from this set. The procedure is sequential and it successively checked the intervals $\mathcal{I}_0, \mathcal{I}_1, \ldots, \mathcal{I}_k,$ on change points.

The interval $\mathcal{I}_0$ is always accepted and the procedure starts with $k = 1$. At every step k, the interval $J_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$ is tested against a change point using a larger interval

$I_k = \mathcal{I}_{k+1}$. $\mathcal{I}_k$ is accepted if the previous interval $\mathcal{I}_{k-1}$ was accepted and the test statistic $T_k = T_{\mathcal{I}_{k+1}}$ defined by (3.2) does not exceed the critical value $\mathfrak{z}_k$. In this case we set $\varkappa_k = k$. The event $\{\mathcal{I}_k \text{ is rejected}\}$ means that $T_l > \mathfrak{z}_l$ for some $l \le k$ and hence, a change point has been detected in the first k steps of the procedure. In this case $\varkappa_k = \varkappa_{k-1}$. For every k, we define a current estimate $\widehat{\theta}_k = \widetilde{\theta}_{\varkappa_k}$. The final estimate is defined as $\widehat{\theta} = \widehat{\theta}_K$ and it corresponds to the largest found interval of homogeneity. The formal definition reads as follows:

$$\varkappa = \max\{k \le K : T_l \le \mathfrak{z}_l,\ l = 1, \ldots, k\}, \qquad \widehat{\theta} = \widetilde{\theta}_{I_\varkappa}.$$

The way of choosing the critical value as well as the other parameters of the procedure like the intervals $\mathcal{I}_k$ is discussed in the next section.

3.4 Parameters of the LCP procedure

This section discusses the parameters of the LCP procedure and some implementation details.

3.4.1 Choice of the parameters $\mathfrak{z}_k$

The "critical values" $\mathfrak{z}_k$ define the level of significance for the test statistics $T_k = T_{I_k}$. A proper choice of these parameters is crucial for the performance of the procedure. We propose in this section one general approach for selecting them which is similar to the bootstrap idea in the hypothesis testing problem. Namely, we select these values to provide the prescribed performance of the procedure in the parametric situation (under the null hypothesis). An important and helpful property of the volatility parametric model $f(\cdot) \equiv \theta^*$ is that the distribution of the test statistics $T_{I,\tau}$ and T_I does not depend on the parameter value θ^*. This is a simple corollary of the fact that volatility is a scale parameter of the corresponding parametric family. However, in view of its importance for our study we state it in a separate lemma.

Lemma 3.1. *Let the return R_t follow the parametric model with the constant volatility parameter θ^*, that is, $R_t^2 = \theta^* \varepsilon_t^2$. Then the distribution of the test statistics T_k under $\mathbb{P}_{\theta^*}$ is the same for all $\theta^* > 0$.*

Proof. It suffices to notice that for every interval I the estimate $\widetilde{\theta}_I$ can be represented under $\mathbb{P}_{\theta^*}$ as

$$\widetilde{\theta}_I = N_I^{-1} \sum_{t \in I} Y_t^2 = \theta^* N_I^{-1} \sum_{t \in I} \varepsilon_t^2$$

and for each two intervals I, I', the Kullback-Leibler divergence $\mathcal{K}(\widetilde{\theta}_I, \widetilde{\theta}_{I'})$ is a function of the ratio $\widetilde{\theta}_I / \widetilde{\theta}_{I'}$. □

The result of Lemma 3.1 allows to reduce the parametric null situation to the case of a simple null consisting of one point θ^*, e.g. $\theta^* = 1$. The corresponding distribution of the observation under this measure will be denoted by $I\!P_{\theta^*}$.

For every step k, we require that in the parametric situation $f(\cdot) \equiv \theta^*$ the estimate $\widehat{\theta}_k$ is sufficiently close to the "oracle" estimate $\widetilde{\theta}_k$ in the sense that

$$I\!E_{\theta^*} \bigl| N_k \mathcal{K}\bigl(\widetilde{\theta}_k, \widehat{\theta}_k\bigr) \bigr|^r \le \alpha \mathfrak{r}_r \tag{3.3}$$

for all $k = 1, \ldots, K$ with $\mathfrak{r}_r$ from Theorem 2.4.

Note that the $\widehat{\theta}_k$ differs from $\widetilde{\theta}_k$ only if a change point is detected at the first k steps. The usual condition to any change point detector is that such "false alarms" occur with a small probability. Our condition (3.3) has the same flavor but it is a bit stronger. Namely, a false alarm at an early stage of the procedure is more crucial because it results in selecting an estimate with a high variability. Therefore, such events have to be stronger penalized than the false alarms in the final steps of the algorithm.

The values α and r in (3.3) are two global parameters. The role of α is similar to the level of the test in the hypothesis testing problem while r describes the power of the loss function. A specific choice is subjective and depends on the particular application at hand. Taking a large r and small α would result in an increase of the critical values and therefore, improves the performance of the method in the parametric situation at cost of some loss of sensitivity to parameter changes. Theorem 4.1 presents some upper bounds for the critical values $\mathfrak{z}_k$ as functions of α and r in the form $a_0 + a_1 \log \alpha^{-1} + a_2 r \log(N_K / N_k) + a_3 r \log(N_k)$ with some coefficients a_0, a_1, a_2 and a_3. We see that these bounds linearly depend on r and on $\log \alpha^{-1}$. For our applications to volatility estimation, we apply a relatively small value $r = 1/2$ which makes the procedure more stable and robust against outliers. We also apply $\alpha = 1$ although the other values in the range $[0.5, 1]$ lead to very similar results.

The set of conditions (3.3) do not directly define the critical values $\mathfrak{z}_k$. We present below one constructive method for selecting $\mathfrak{z}_k$ to provide the conditions (3.3).

3.4.2 The sequential choice

Here we present a proposal for a sequential choice of the $\mathfrak{z}_k$'s. Consider the situation after the first k steps of the algorithm. We distinguish between two cases: a change point

has been detected at some step $l \le k$ and the other case of no change point detected. In the first case we denote by $\mathcal{B}_l$ the event meaning the rejection at the step l, that is,

$$\mathcal{B}_l = \{T_1 \le \mathfrak{z}_1, \ldots, T_{l-1} \le \mathfrak{z}_{l-1}, T_l > \mathfrak{z}_l\}$$

and $\widehat{\theta}_k = \widetilde{\theta}_{\mathcal{I}_{l-1}}$ on $\mathcal{B}_l$, $l = 1, \ldots, k$. The sequential choice of the critical values $\mathfrak{z}_k$ is based on the decomposition

$$\bigl|\mathcal{K}\bigl(\widetilde{\theta}_k, \widehat{\theta}_k\bigr)\bigr|^r = \sum_{l=1}^{k} \bigl|\mathcal{K}\bigl(\widetilde{\theta}_k, \widetilde{\theta}_{l-1}\bigr)\bigr|^r \mathbf{1}\bigl(\mathcal{B}_l\bigr) \tag{3.4}$$

for every $k \le K$. Now we utilize that the event $\mathcal{B}_l$ only depends on $\mathfrak{z}_1, \ldots, \mathfrak{z}_l$. In particular, the event $\mathcal{B}_1$ means that $T_1 > \mathfrak{z}_1$ and $\widehat{\theta}_j = \widetilde{\theta}_0$ for all $j \ge 1$. We select $\mathfrak{z}_1$ as the minimal value providing that

$$\max_{j=1,\ldots,K} I\!\!E_{\theta^*} \bigl|N_j \mathcal{K}\bigl(\widetilde{\theta}_j, \widetilde{\theta}_0\bigr)\bigr|^r \mathbf{1}\bigl(T_1 > \mathfrak{z}_1\bigr) \le \alpha \mathfrak{r}_r / K. \tag{3.5}$$

Similarly, for every $k \ge 2$, select $\mathfrak{z}_k$ by considering the event $\mathcal{B}_k = \{\varkappa_k = k\}$ meaning that the false alarm occurs at the step k and $\widehat{\theta}_j = \widetilde{\theta}_{k-1}$ for all $j > k$. If $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$ have been already fixed, the event $\mathcal{B}_k$ is only controlled by $\mathfrak{z}_k$ leading to the following condition on $\mathfrak{z}_k$: this is the minimal value that provides

$$\max_{j \ge k} I\!\!E_{\theta^*} \bigl|N_j \mathcal{K}\bigl(\widetilde{\theta}_j, \widetilde{\theta}_{k-1}\bigr)\bigr|^r \mathbf{1}\bigl(\mathcal{B}_k\bigr) = \alpha \mathfrak{r}_r / K. \tag{3.6}$$

It is straightforward to check that such defined $\mathfrak{z}_k$ fulfill (3.3) in view of the decomposition (3.4).

3.5 Examples of choosing the intervals $\mathcal{I}_k$

To start the procedure running, one has to specify the set of intervals $\mathcal{I}_1, \ldots, \mathcal{I}_K$. Note, however, that this choice is not a part of the LCP procedure. The method applies whatever intervals $\mathcal{I}_k$ are selected under condition (MD), see Section 4. This section presents one example which is at the same time the default choice for our simulation study and applications.

The set $m_1, \ldots, m_K$ is defined geometrically by the rule $m_k = [m_0 a^k]$ for some fixed value m_0 and the growth rate $a > 1$. It is straightforward to see that such a proposal provides (MD). Note also that the sets J_k do not intersect for different k and every point $\tau \in [t^\diamond - m_k, t^\diamond - m_0]$ is tested as a possible location of the change point at some of the first k steps of the procedure.

Our numerical results (not reported here) indicate that the procedure is quite stable w.r.t. the choice of the parameters like m_0 and a. We apply $a = 1.25$. The other values of a in the range 1.1 to 1.3 lead to very similar results. We also apply $m_0 = 5$ or 10 which is motivated by our applications to risk management in financial engineering.

3.6 The procedure for multiple case

This section briefly comments on the extension of the method to the case of multiple volatility modeling. The procedure reads exactly as in the univariate case. One only has to adjust the definition of the test statistics $T_{I_k,\tau}$:

$$T_{I_k,\tau} = L_{I''}(\widetilde{\Sigma}_{I''}) + L_{I'}(\widetilde{\Sigma}_{I'}) - L_I(\widetilde{\Sigma}_{I_k}) = N_{I''}\mathcal{K}(\widetilde{\Sigma}_{I''}, \widetilde{\Sigma}_{I_k}) + N_{I'}\mathcal{K}(\widetilde{\Sigma}_{I'}, \widetilde{\Sigma}_{I_k}).$$

The important result of Lemma 3.1 about pivotality of the test statistics T_k extends by the same arguments from univariate to multivariate volatility modeling and the distribution of all the test statistics T_k does not depend on the underlying volatility matrix Σ^* in the parametric situation. To see this, it suffices to replace the observations R_t by the standardized returns $(\Sigma^*)^{-1/2}R_t$. Then the volatility matrix Σ^* cancels in the expression for the log-likelihood and for the risk. Unfortunately, in the multivariate case there is no closed form expression for the risk bound $\mathfrak{R}_r$. However, this bound can be computed by Monte-Carlo simulations using any particular Σ^*, e.g. $\Sigma^* = I_d$: draw Monte-Carlo samples from the simple model with i.i.d. standard normal $R_t = \varepsilon_t$, then compute estimate $\widetilde{\Sigma}_k$ for all considered intervals $\mathcal{I}_k$ and define $\mathfrak{R}_r$ as the maximum of the corresponding estimated risk $I\!E\big|N_k\mathcal{K}(\widetilde{\Sigma}_k, I_d)\big|^r$ over $k = 1, \ldots, K$, where $\widetilde{\Sigma}_k$ means $\widetilde{\Sigma}_{\mathcal{I}_k}$.

Similarly, the critical value $\mathfrak{z}_k$ can be selected by Monte-Carlo simulations from the homogeneous model with $\Sigma^* = I_d$ and $R_t = \varepsilon_t$ to provide the condition

$$I\!E_{\Sigma^*}\big|N_k\mathcal{K}\big(\widetilde{\Sigma}_k, \widehat{\Sigma}_k\big)\big|^r \le \alpha\mathfrak{R}_r \tag{3.7}$$

for all $k = 1, \ldots, K$ where $\widehat{\Sigma}_k$ is the adaptive estimate after k steps of the algorithm and $\mathfrak{R}_r$ is shown in Theorem 2.10. The sequential choice of the $\mathfrak{z}_k$'s can be applied for finding the critical values $\mathfrak{z}_k$ similarly to the univariate case.

4 Theoretical properties

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

selected interval $\widehat{\mathcal{I}}$, that is, $\widehat{\theta} = \widetilde{\theta}_{\widehat{\mathcal{I}}}$. We state the most of our results for the case of one time series. However, all the results can be extended in a straightforward way to the case of multiple volatility modeling using the results from Theorems 2.9, 2.10 and 2.11.

Our main "oracle" result claims that the final estimate $\widehat{\theta}$ delivers essentially the same quality of estimation as the estimate with the "optimal" ("ideal") choice of the interval $\mathcal{I}_{k^*}$ for estimating the volatility parameter $\theta = f(t^\diamond)$. As shown in Section 2.3 this automatically ensures the optimal estimation rate under usual smoothness conditions on the function $f(\cdot)$.

The "oracle" result is in its turn a corollary of two important properties of the procedure: "propagation" under homogeneity and "stability". The first one means that in the nearly homogeneous situation the procedure would not terminate (no "false alarm") with a high probability. In the other words, if the parametric (constant) approximation well applies in the interval $\mathcal{I}_k$ then this interval will be accepted with a high probability. The "stability" property ensures that the estimation quality will not essentially deteriorate in the steps "after propagation" when the local constant approximation is not sufficiently accurate. Typically the procedure just stops in such situations.

The results require some regularity conditions on the growth of the intervals $\mathcal{I}_k$ and the choice of testing intervals I_k. Namely, we require that the length m_k of $\mathcal{I}_k$ grows exponentially with k, and change point detection for every $J_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$ is restricted to the interior of I_k.

(*MD*) for some constants $\mathfrak{u}_0, \mathfrak{u}$ with $0 < \mathfrak{u}_0 \le \mathfrak{u} < 1$, the values $m_1, \ldots, m_K$ satisfy

$$\mathfrak{u}_0 \le m_{k-1}/m_k \le \mathfrak{u}.$$

In addition we assume that the parameter set Θ satisfies the condition (Θ) from Section 2.5. In the univariate case this condition means that the relative variation of volatility is bounded by a constant $\mathfrak{a}^2$.

We start by discussing the behavior of the procedure in the time homogeneous situation with the constant volatility parameter θ^*. In this case the properties of the resulting estimate $\widehat{\theta}$ are guaranteed by the condition (3.3). Our first results claims a possibility of selecting the critical values $\mathfrak{z}_k$ to provide (3.3) and states some upper bounds for the $\mathfrak{z}_k$'. Similar result can be stated in the local parametric situation when the homogeneity condition $f(t) = \theta^*$ is only fulfilled for some time interval I.

4.1 Behaviour under (local) homogeneity

First we consider the homogeneous situation with the constant parameter value $f(x) = \theta^*$. Our first result claims that in this situation under the condition (MD) the parameters $\mathfrak{z}_k$ can be chosen in the form $\mathfrak{z}_k = \mathfrak{z}_K + \iota(K-k)$ to fulfill the condition (3.3). The proof is given in the Appendix.

Theorem 4.1. *Assume (MD). Let $f(t) = \theta^*$ for all $t \in \mathcal{I}_K$. Then there are three constants a_0, a_1, a_2 and a_3 depending on r and $\mathfrak{u}_0$, $\mathfrak{u}$ only such that the choice*

$$\mathfrak{z}_k = a_0 + a_1 \log \alpha^{-1} + a_2 r \log(N_K/N_k) + a_3 r \log N_k$$

ensures (3.3) for all $k \le K$. Particularly, $\mathbb{E}_{\theta^*} \bigl| N_K \mathcal{K}\bigl(\widetilde{\theta}_K, \widehat{\theta}\bigr) \bigr|^r \le \alpha \mathfrak{r}_r$.

Remark 4.1. Using the general bounds on the minimum contrast, see e.g. Golubev and Spokoiny (2006), this result can be improved by removing the term $a_3 r \log N_k$. However, a careful check of the related conditions would result in a significant increase of the paper size. We, therefore, present a more simple bound whose proof is postponed until Section 6.

4.2 Behaviour under "small modelling bias" condition

Now we extend the previous result to the situation when the parametric assumption is not fulfilled any more but the deviation from the parametric structure within the considered local model is sufficiently small. At the step k the procedure operates with the interval $I_k = \mathcal{I}_{k+1}$ used for testing a change point within $\mathcal{I}_k$. Therefore, the deviation from the parametric situation can be measured for the step k by $\Delta_{\mathcal{I}_{k+1}}(\theta)$ from (2.4).

We suppose that there is a number k^* such that the modeling bias $\Delta_k(\theta) = \Delta_{\mathcal{I}_{k+1}}(\theta)$ is small for some θ and all $k \le k^*$. Consider the corresponding estimate $\widehat{\theta}_{k^*}$ obtained after the first k^* steps of the algorithm. Theorem 2.5 implies in this situation the following result.

Theorem 4.2. *Assume (MD). Let θ and k^* be such that*

$$\max_{k \le k^*} \mathbb{E} \Delta_k(\theta) \le \Delta \tag{4.1}$$

for some $\Delta \ge 0$. Then

$$\begin{aligned} \mathbb{E} \log\biggl(1 + \frac{N_{k^*}^r \mathcal{K}^r\bigl(\widetilde{\theta}_{k^*}, \widehat{\theta}_{k^*}\bigr)}{\alpha \mathfrak{r}_r}\biggr) &\le 1 + \Delta, \\ \mathbb{E} \log\biggl(1 + \frac{N_{k^*}^r \mathcal{K}^r\bigl(\widetilde{\theta}_{k^*}, \theta\bigr)}{\mathfrak{r}_r}\biggr) &\le 1 + \Delta. \end{aligned}$$

4.3 "Stability after propagation" and "oracle" results

Due to the "propagation" result, the procedure performs well as long as the "small modeling bias" condition $\Delta_k(\theta) \le \Delta$ is fulfilled. To establish the accuracy result for the final estimate $\widehat{\theta}$, we have to check that the aggregated estimate $\widehat{\theta}_k$ does not vary much at the steps "after propagation" when the divergence $\Delta_k(\theta)$ from the parametric model becomes large.

Theorem 4.3. *Suppose* (MD) *and* (Θ). *Let, for some* $k \le K$, *the interval* $\mathcal{I}_k$ *be accepted by the procedure and hence,* $\widehat{\theta}_k = \widetilde{\theta}_k$. *Then it holds*

$$N_k \mathcal{K}\bigl(\widehat{\theta}_k, \widehat{\theta}_{k+1}\bigr) \le \mathfrak{z}_k. \tag{4.2}$$

Moreover, under (MD), *it holds for every* k' *with* $k < k' \le K$

$$N_k \mathcal{K}\bigl(\widehat{\theta}_k, \widehat{\theta}_{k'}\bigr) \le \mathfrak{a}^2 c_{\mathfrak{u}}^2 \, \bar{\mathfrak{z}}_k \tag{4.3}$$

with $c_{\mathfrak{u}} = (\mathfrak{u}^{-1/2} - 1)^{-1}$ *and* $\bar{\mathfrak{z}}_k = \max_{l \ge k} \mathfrak{z}_l$.

Remark 4.2. An interesting feature of this result is that it is fulfilled with probability one, that is, the control of stability "works" not only with a high probability, it always applies. This property follows directly from the construction of the procedure.

Proof. If $\mathcal{I}_{k+1}$ is rejected then $\widehat{\theta}_{k+1} = \widehat{\theta}_k$ and the assertion (4.2) trivially follows. Now we consider the case when $\mathcal{I}_{k+1}$ is accepted yielding $\widehat{\theta}_k = \widetilde{\theta}_k$ and $\widehat{\theta}_{k+1} = \widetilde{\theta}_{k+1}$. The acceptance of $\mathcal{I}_k$ implies by definition of the procedure that $T_{I_k} \le \mathfrak{z}_k$ and, in particular, $T_{I_k,\tau} \le \mathfrak{z}_k$ with $\tau = t^\diamond - m_k$ being the left edge of $\mathcal{I}_k$. This yields, see (3.1), that

$$N_k \mathcal{K}\bigl(\widetilde{\theta}_k, \widetilde{\theta}_{k+1}\bigr) \le \mathfrak{z}_k \, .$$

and the assertion (4.2) is proved.

Now, Assumption (Θ) and Lemma 6.1 yield

$$\mathcal{K}^{1/2}\bigl(\widehat{\theta}_k, \widehat{\theta}_{k'}\bigr) \le \mathfrak{a} \sum_{j=k}^{k'-1} \mathcal{K}^{1/2}\bigl(\widehat{\theta}_j, \widehat{\theta}_{j+1}\bigr) \le \mathfrak{a} \sum_{j=k}^{k'-1} \bigl(\mathfrak{z}_j / N_j\bigr)^{1/2}.$$

The use of Assumption (MD) leads to the bound

$$\mathcal{K}^{1/2}\bigl(\widehat{\theta}_k, \widehat{\theta}_{k'}\bigr) \le \mathfrak{a}\bigl(\bar{\mathfrak{z}}_k / N_k\bigr)^{1/2} \sum_{j=k}^{k'-1} \mathfrak{u}^{(j-k)/2} \le \mathfrak{a}(1 - \sqrt{\mathfrak{u}})^{-1} \bigl(\bar{\mathfrak{z}}_k / N_k\bigr)^{1/2}$$

which proves (4.3). □

Combination of the "propagation" and "stability" statements implies the main result concerning the properties of the adaptive estimate $\widehat{\theta}$.

Theorem 4.4 ("Oracle" property)**.** *Let $\mathbb{E}\Delta_k(\theta) \le \Delta$ for some $\theta \in \Theta$ and $k \le k^*$. Then $\widehat{\theta}$ is close to the "oracle" estimate $\widetilde{\theta}_{k^*}$ in the sense that*

$$\begin{aligned} \mathbb{E}\log\Bigl(1 + \frac{\bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \widehat{\theta}_{k^*}\bigr)\bigr|^r}{\alpha \mathfrak{r}_r}\Bigr) &\le 1 + \Delta, \\ N_{k^*}\mathcal{K}\bigl(\widehat{\theta}_{k^*}, \widehat{\theta}\bigr) &\le \mathfrak{a}^2 c_{\mathrm{u}}^2 \bar{\mathfrak{z}}_{k^*}. \end{aligned}$$

The result claims the "oracle" accuracy $N_{k^*}^{-1}$ for the loss $\mathcal{K}(\widehat{\theta}, \theta)$ up to the log factor $\mathfrak{z}_{k^*}$. We state on corollary of the "oracle" result for $r = 1/2$. An extension to an arbitrary $r > 0$ is obvious.

Theorem 4.5. *Assume (MD) and (4.1) for some k^*, θ and Δ. Then*

$$\mathbb{E}\log\Bigl(1 + \frac{N_{k^*}^{1/2}\mathcal{K}^{1/2}\bigl(\widehat{\theta}, \theta\bigr)}{\mathfrak{a}\mathfrak{r}_{1/2}}\Bigr) \le \log\Bigl(1 + \frac{c_{\mathrm{u}}\sqrt{\bar{\mathfrak{z}}_{k^*}}}{\mathfrak{r}_{1/2}}\Bigr) + \Delta + \alpha + 1$$

where c_{u} is the constant from Theorem 4.3.

Proof. By Lemma 6.1 similarly to the proof of Theorem 4.3

$$\begin{aligned} \mathfrak{a}^{-1}\bigl|N_{k^*}\mathcal{K}\bigl(\widehat{\theta}, \theta\bigr)\bigr|^{1/2} &\le \bigl|N_{k^*}\mathcal{K}\bigl(\widehat{\theta}_{k^*}, \widehat{\theta}\bigr)\bigr|^{1/2} + \bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \widehat{\theta}_{k^*}\bigr)\bigr|^{1/2} + \bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \theta\bigr)\bigr|^{1/2} \\ &\le c_{\mathrm{u}}\sqrt{\bar{\mathfrak{z}}_{k^*}} + \bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \widehat{\theta}_{k^*}\bigr)\bigr|^{1/2} + \bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \theta\bigr)\bigr|^{1/2}. \end{aligned}$$

This, the elementary inequality $\log(1 + a + b) \le \log(1 + a) + \log(1 + b)$ for all $a, b \ge 0$, Lemma 2.6, Theorem 2.4, and (3.3) yield

$$\begin{aligned} &\mathbb{E}\log\Bigl(1 + (\mathfrak{a}\mathfrak{r}_{1/2})^{-1} N_{k^*}^{1/2}\mathcal{K}^{1/2}\bigl(\widehat{\theta}, \theta\bigr)\Bigr) \\ &\le \log\Bigl(1 + \frac{c_{\mathrm{u}}\sqrt{\bar{\mathfrak{z}}_{k^*}}}{\mathfrak{r}_{1/2}}\Bigr) + \mathbb{E}\log\Bigl(1 + \frac{N_{k^*}^{1/2}\mathcal{K}^{1/2}\bigl(\widetilde{\theta}_{k^*}, \widehat{\theta}_{k^*}\bigr)}{\mathfrak{r}_{1/2}} + \frac{N_{k^*}^{1/2}\mathcal{K}^{1/2}\bigl(\widetilde{\theta}_{k^*}, \theta\bigr)}{\mathfrak{r}_{1/2}}\Bigr) \\ &\le \log\Bigl(1 + \frac{c_{\mathrm{u}}\sqrt{\bar{\mathfrak{z}}_{k^*}}}{\mathfrak{r}_{1/2}}\Bigr) + \Delta + \alpha + 1 \end{aligned}$$

as required. □

Remark 4.3. Recall that by Theorem 4.2, the "oracle" choice k^* leads to the risk bound for the loss $\bigl|N_{k^*}\mathcal{K}\bigl(\widetilde{\theta}_{k^*}, \theta^*\bigr)\bigr|^{1/2}$ of the corresponding estimate $\widetilde{\theta}_{k^*}$. The adaptive choice states a similar bound but for the loss $\bigl|N_{k^*}\mathcal{K}\bigl(\widehat{\theta}, \theta^*\bigr)\bigr|^{1/2}/\bar{\mathfrak{z}}_{k^*}^{1/2}$. This means that the accuracy of the adaptive estimate $\widehat{\theta}$ is worse by factor $\sqrt{\bar{\mathfrak{z}}_{k^*}}$ which can be considered

as the payment for adaptation. Due to Theorem 4.1, $\bar{\mathfrak{z}}_{k^*}$ is bounded from above by $a_0 + a_1 \log(\alpha^{-1}) + a_2 r \log(N_K/N_{k^*}) + a_3 r \log N_{k^*}$. Therefore, the risk of the aggregated estimate corresponds to the best possible risk among the family $\{\widetilde{\theta}_k\}$ for the choice $k = k^*$ up to a logarithmic factor in the sample size. Lepski, Mammen and Spokoiny (1997) established a similar result in the regression setup for the pointwise adaptive Lepski procedure. Combining the result of Theorem 4.5 with Theorem 2.8 yields the rate of adaptive estimation $\left(n^{-1} \log n\right)^{1/(2+d)}$ under Lipschitz smoothness of the function f and the usual design regularity, see Polzehl and Spokoiny (2005) for more details. It was shown by Lepski (1990) that in the problem of pointwise adaptive estimation this rate is optimal and cannot be improved by any estimation method. This gives an indirect proof of the optimality of our procedure: the factor $\bar{\mathfrak{z}}_{k^*}$ in the accuracy of estimation cannot be removed or reduced in the rate because otherwise the similar improvement would appear in the rate of estimation.

4.4 Switching regime model

A *switching regime* model is described by a sequence $\nu_1 < \nu_2 < \ldots$ of Markov moments with respect to the filtration $\mathcal{F}_t$ and by values $\theta_1, \theta_2, \ldots$ where each θ_j is $\mathcal{F}_{\nu_j}$-measurable. By definition $\sigma_t^2 = f(t) = \theta_j$ for $\nu_j \le t < \nu_{j+1}$ 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 switching regime model. For this special case, the above procedure has a very natural interpretation: when estimating at the point $t^\diamond$ we search for a largest interval of the form $[t^\diamond - m, t^\diamond[$ does not containing a change point. More precisely, with a giving sequence of interval-candidates $\mathcal{I}_k = [t^\diamond - m_k, t^\diamond[$, we are looking for the largest among them which does not contain a change point. This is done via successive testing for a change point within the intervals $\mathcal{I}_k = [t^\diamond - m_k, t^\diamond[$.

The construction of the procedure automatically provides the prescribed risk level associated with the first kind error (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 $t^\diamond$ next after a change point ν. The "ideal" choice $\mathcal{I}_{k^*}$ among $\mathcal{I}_1, \ldots, \mathcal{I}_K$ is obviously the largest one does not containing ν. Theorem 4.5 claims that the procedure accepts with a high probability all

the intervals $\mathcal{I}_k$ for which the testing interval $I_k = \mathcal{I}_{k+1}$ does not contain the point of change ν. This particularly implies that the quality of estimation of $\theta_{t^\diamond}$ by our adaptive procedure is essentially the same as if we knew the latest change point ν a priori. Now we show that in addition the procedure rejects with a high probability the first interval $\mathcal{I}_{k^*+1}$ which contains the point of change ν provided that the magnitude of the change is sufficiently large. This fact can be treated as the sensitivity of the procedure to the changes of regime.

In our study we assume that the changes occur not too often, and there is exactly one change within $\mathcal{I}_{k^*+1}$ and moreover, within the larger interval $I_{k^*+1} = \mathcal{I}_{k^*+2}$ which is used for testing $\mathcal{I}_{k^*+1}$. Let θ' be the value of the parameter before the change and θ'' after it. The point τ splits $\mathcal{I}_{k^*+2}$ into two homogeneous intervals: $f(t) = \theta''$ for $t \in I'' = [\tau, t^\diamond[$ while $f(t) = \theta_\ell$ within the complementary interval $t \in I_{k^*+1} \setminus I''$. Define $c_1 = m_{k^*}/m_{k^*+2}$, $c_2 = m_{k^*+1}/m_{k^*+2}$. By condition (MD), $c_1 \geq \mathfrak{u}_0^2$ and $c_2/c_1 \geq \mathfrak{u}^{-1}$. The length $t^\diamond - \tau$ of the interval $[\tau, t^\diamond[$ fulfills $c_1 \leq (t^\diamond - \tau)/m_{k^*+2} \leq c_2$. Based on these considerations, define the following measure of change from θ' to θ'':

$$d^2(\theta', \theta'') = \inf_{\theta} \inf_{c \in [c_1, c_2]} \bigl\{ (1-c)\mathcal{K}(\theta', \theta) + c\mathcal{K}(\theta'', \theta) \bigr\}. \tag{4.4}$$

The following simple bound can be useful for bounding the distance $d^2(\theta', \theta'')$.

Lemma 4.6. *There is a constant $\mathfrak{b} > 0$ depending on c_1 and c_2 only such that*

$$d^2(\theta', \theta'') \geq \mathfrak{b}\bigl(\theta'/\theta'' - \theta''/\theta'\bigr)^2.$$

Proof. For every fixed $\theta', \theta'', \theta$, the expression $(1-c)\mathcal{K}(\theta', \theta) + c\mathcal{K}(\theta'', \theta)$ is a linear function of c. Therefore, its minimum w.r.t. c is attained in one of the edge points c_1, c_2, and it suffices to check the assertion only for these two values of c. Now the assertion follows directly from the definition of the Kullback-Leibler distance $\mathcal{K}(\theta', \theta)$ as a smooth function of the ratio θ'/θ with $\mathcal{K}(\theta, \theta) \equiv 0$. □

We aim to show that if the contrast $d(\theta', \theta'')$ is sufficiently large then the test statistic T_{k^*+1} will be large as well yielding that the interval $\mathcal{I}_{k^*+1}$ will be rejected with a high probability.

Theorem 4.7. *Let $f(t) = \theta'$ before the change point at ν and $f(t) = \theta''$ after it. If for some $\mathfrak{z} > 0$,*

$$d^2(\theta', \theta'') \geq \frac{2\mathfrak{a}^2}{m_{k^*+2}}(\mathfrak{z}_{k^*+1} + \mathfrak{z}) \tag{4.5}$$

then

$$I\!P\left(\mathfrak{I}_{k^*+1} \textit{ is not rejected}\right) \leq 4\mathrm{e}^{-\mathfrak{z}}\,.$$

The proof of this result is given in Section 6. The result of Theorem 4.7 delivers some additional information about the sensitivity of the proposed procedure to changes in the volatility parameter. One possible question is about the minimal delay m^* between the change point ν and the first moment $t^\diamond$ when the procedure starts to detect this change. Due to Theorem 4.7, the change will be "detected" with a high probability if (4.5) meets. With fixed $\theta' \neq \theta''$, condition (4.5) 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. All these issues are in agreement with the theory of change point detection, see, e.g. Pollak (1985) and Brodskij and Darkhovskij (1993), and with our numerical results from Section 5.

5 Simulated results and applications

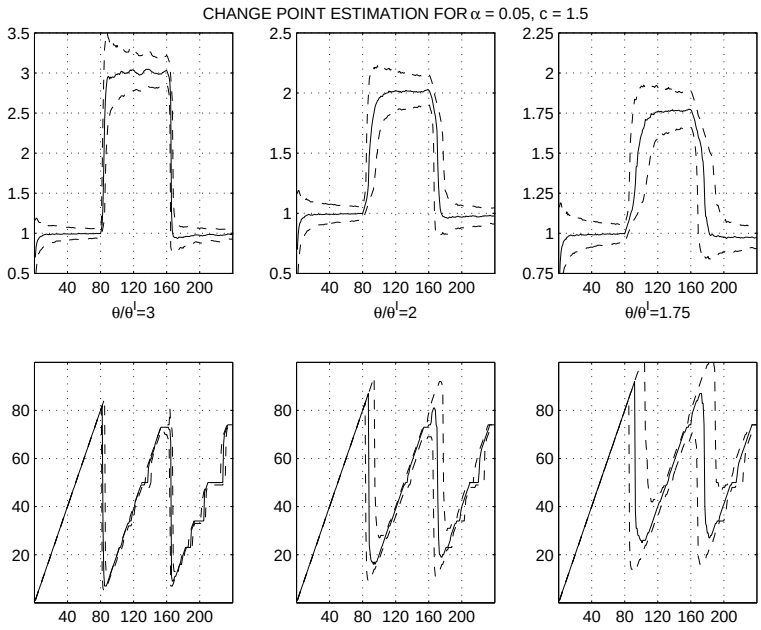
This section illustrates the performance of the proposed local change point detection (LCP) procedure by means of some simulated data sets and applications to real data. We aim to show that the theoretical properties of the method derived in the previous section are confirmed by the numerical results. We focus especially on the two main features of the method: stability under homogeneity and sensitivity to changes of volatility.

5.1 Some simulated examples. Univariate case

Three different jump processes are simulated, whose relative jump magnitude is 3.00, 2.00 and 1.75 respectively. Each process is simulated and estimated 1000 times and the median and the quartiles of the estimates are plotted in Figure 5.1. We show the results for the final estimate $\widehat{\theta}$ and for the length of the selected interval $\widehat{\mathfrak{I}}$. 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 corresponds to our theoretical results.

The algorithm proposed in this paper is compared with the LAVE procedure from MS2004 with the optimized tuning parameters $\gamma = 0.5$, $M = 40$, $\mathfrak{z} = 2.40$. Figure 5.2

Figure 5.1: Pointwise median (solid line) and quartiles (dashed lines) for the estimates $\widehat{\theta}_t$ (top row) and the length of the selected interval $\widehat{\mathfrak{I}}_t$ for three jump processes with jumps of different magnitudes. The results are obtained with parameters $c = 1.5$ and $\alpha = 0.05$.



shows the quartiles of estimation for the two approaches for the model with the relative jump magnitude 3. One can see that the new procedure outperforms LAVE both with respect to the variance and to the bias of the estimator, especially for the points immediately after the changes.

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 of the normality assumption we also simulated using i.i.d. innovations from the t_5-distribution with five degree of freedoms. The results are shown in Figure 5.3. As one can expect they 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 with normal innovations but a probability to reject a homogeneous interval became larger. This results in a higher variability of the estimated volatility.

Figure 5.2: Comparison of the proposed estimator with the one from MS2004 for change point model with $\theta/\theta' = 3$. Quartiles of $\widehat{\theta}$ for the LCP method (solid lines) and for the LAVE method from MS2004 (dotted lines); simulated trajectory of the volatility (thick line).

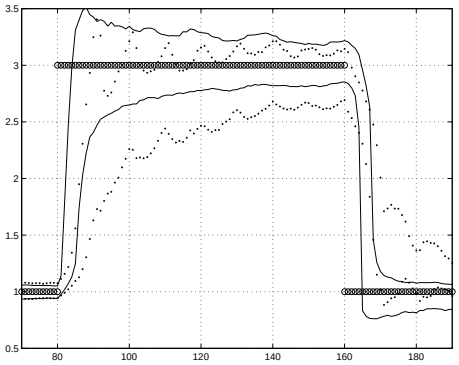
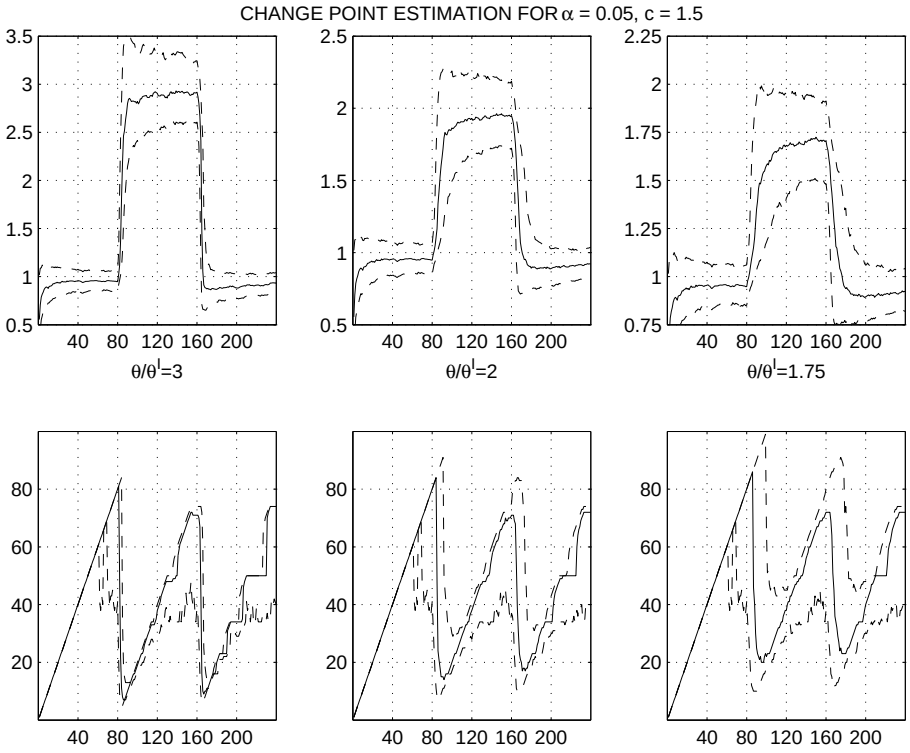


Figure 5.3: Estimation results with respect to jump processes with jumps of different magnitudes. The results are obtained with tuning parameters $c = 1.5$ and $\alpha = 0.05$. The conditional distribution is scaled student t_5 with 5 degrees of freedom.

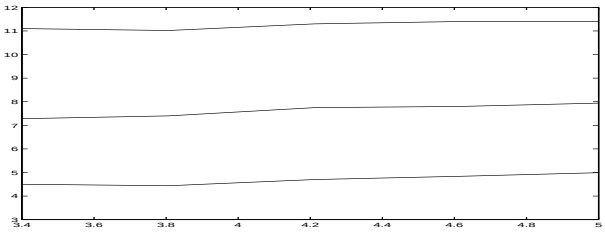


5.2 Some simulated examples. Multivariate case

The implementation of the estimator in the multivariate case is similar to the univariate case. In particular, for obtaining the critical values we again exploit the fact that the distribution of the likelihood ratio statistic under the hypothesis of homogeneity does not depend on the value of the covariance matrix, and therefore the critical values can be easily obtained by simulation.

In our implementation we again select $m_0 = 30$ and $c = 1.5$. Figure 5.4 shows the critical values as a function of the log interval length for a univariate, a bivariate and a trivariate system at a 5% level. Note that the curves indicate an approximately linear relationship and are almost parallel. This fact can be used to extrapolate critical values for larger intervals and for systems of a larger dimension.

Figure 5.4: Critical values computed by simulation for $c = 1.5$ and $\alpha = 5\%$ for systems of different dimension. From the top: trivariate, bivariate and univariate.



The log length of the interval is regressed on the simulated critical values, and as suggested by Figure 5.4 we allow for three different intercepts but only one slope coefficient. The results of the regression are displayed in Table 5.1 and they suggest the

Table 5.1: Linear approximation of the critical values

slope	intercept					
	univariate		bivariate		trivariate	
	10%	5%	10%	5%	10%	5%
0.3095	2.6103	3.3923	5.2463	6.3323	8.8363	9.9403

use of the following linear rule for selecting critical values which keep the type-1-error approximately at a 5% and depend on the dimension of the system d and on the length

of the currently tested interval $|I|$:

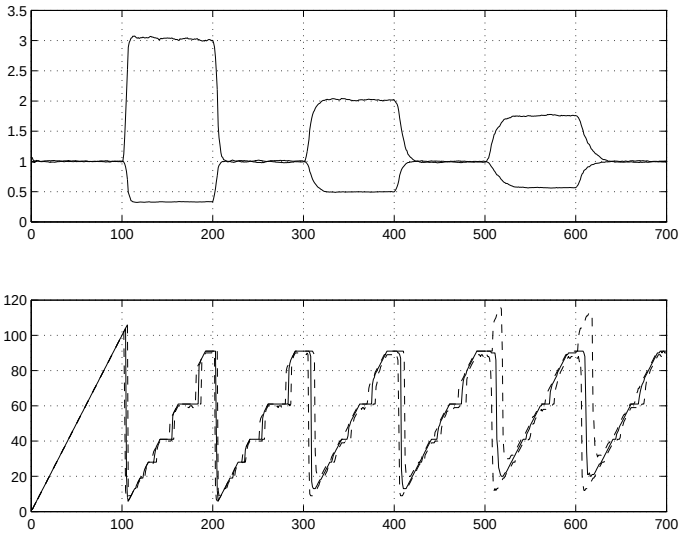
$$\mathfrak{z}(d,\,|I|) = 3.3d + 0.31\log|I|. \tag{5.1}$$

Using the critical values given by (5.1) we apply the estimation procedure on simulated data.

We consider the following bivariate example. The correlation is set to zero and the volatilities are jump processes:

$$\begin{aligned}\sigma_{1t} &= 1 + 2\boldsymbol{I}_{\{t\in[101,200]\}} + \boldsymbol{I}_{\{t\in[301,400]\}} + \frac{3}{4}\boldsymbol{I}_{\{t\in[501,600]\}} \\ \sigma_{2t} &= 1 - \frac{2}{3}\boldsymbol{I}_{\{t\in[101,200]\}} - \frac{1}{2}\boldsymbol{I}_{\{t\in[301,400]\}} - \frac{3}{7}\boldsymbol{I}_{\{t\in[501,600]\}}.\end{aligned}$$

Figure 5.5: Simulation results for a diagonal bivariate process with jumps of different magnitude. Upper plot: pointwise median over 500 simulations for the two diagonal elements of the volatility matrix. Lower plot: median estimate of the interval of time homogeneity and quartiles.



This system is simulated and estimated five hundred times and the average estimates of the volatilities, the median and the quartiles of the estimated interval of time homogeneity are plotted in Figure 5.5. As expected, the performance is very similar to the univariate case. The jump detection is quick and it is more accurate for larger jumps.

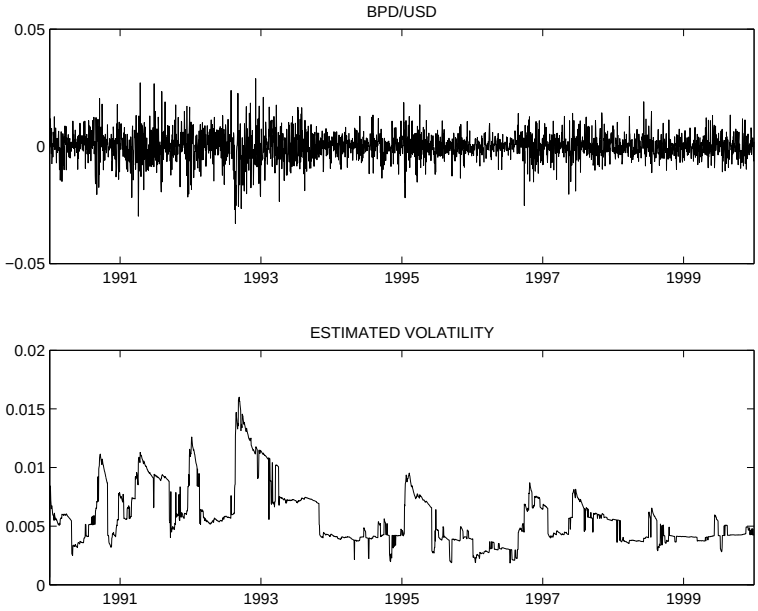
5.3 Volatility estimation for different exchange rate data sets

The volatility estimation is performed on a 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 1st, 1990 to April 7th, 2000. For each time series we have 2583 observations. All selected time series display excess kurtosis and volatility clustering.

Figure 5.6 and Figure 5.7 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 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 changes in the underlying volatility process quickly.

Figure 5.6: Returns and estimated volatility for the BPD/USD exchange rate.

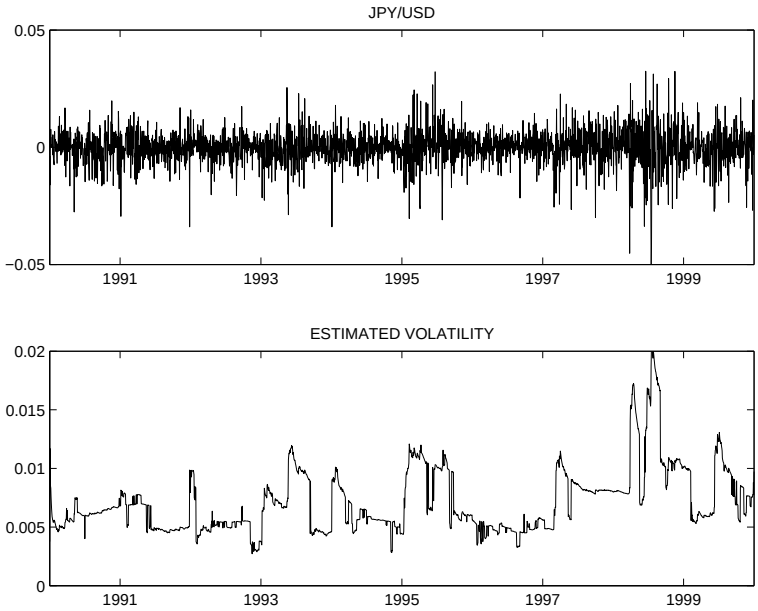


The assumption of local homogeneity leads to the constant forecast $\widehat{\sigma}_t^2$ of the volatility σ_{t+h} for a small or moderate time horizon h. This results in the following forecast of the conditional variance of the aggregated returns $R_{t+1}^2 + \ldots + R_{t+h}$:

$$V_{t,h}^{\mathtt{LCP}} := h\widehat{\sigma}_t^2,$$

with h being the forecast horizon.

Figure 5.7: Returns and estimated volatility for the JPY/USD exchange rate.



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 parameterizations of the volatility process of financial time series. The GARCH(1,1) model is described by the following equations:

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

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

$$\sigma_{t+h|t}^{2,\mathtt{GARCH}} := I\!E_t R_{t+h}^2 = \overline{\sigma}^2 + (\alpha+\beta)^h(\sigma_t^2 - \overline{\sigma}^2),$$

where $\overline{\sigma}$ represents the unconditional volatility and $I\!E_t\xi$ means $I\!E(\xi|\mathcal{F}_t)$, see Mikosch and Starica (2000a). Since the returns are conditionally uncorrelated, the conditional variance of the aggregated returns is given by the sum of the conditional variances:

$$V_{t,h}^{\mathtt{GARCH}} := I\!E_t(R_{t+1} + \ldots + R_{t+h})^2 = \sum_{k=1}^{h} I\!E_t R_{t+k}^2 = \sum_{k=1}^{h} \sigma_{t+h|t}^{2,\mathtt{GARCH}}.$$

The assumption of constant parameters for a GARCH(1,1) model over a time interval of the considered length of about 2500 time points can be too restrictive. We therefore

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 the observation window but it takes advantage of a more flexible GARCH-modeling. The LCP 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 $V_{t,h}^{\mathtt{LCP}}$ resp. $V_{t,h}^{\mathtt{GARCH}}$ 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:

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

The MSqE is considered instead of the more common MSE for robustness reasons. Actually, in this way outliers are prevented from having a strong influence on the results. The MSqE is computed for six nonoverlapping intervals of 250 observations and the results are shown in Table 5.2. One can observe that both methods are comparable and that the relative performance depends on the particular situation at hand. For periods with stable volatility the LCP forecast is clearly better, but for periods with high volatility variation the GARCH method is slightly preferable.

In order to assess the performance of the new algorithm we compare its forecasting ability with the ones for the LCP procedure from MS200 and the GARCH(1,1) procedure. The following criteria are used:

$$E_p^{CP} = \sum_{k=100}^{N} |R_k^2/\widehat{\sigma}_{k-1}^2 - 1|^p \qquad \text{and} \qquad \widetilde{E}_p^{CP} = \sum_{k=100}^{N} |R_k^2 - \widehat{\sigma}_{k-1}^2|^p$$

for $p = 0.5$, 1 and 2. Similar quantities E_p^{GARCH} were computed for the GARCH procedure. The ratios E_p^{CP}/E_p^{GARCH} and $\widetilde{E}_p^{CP}/\widetilde{E}_p^{GARCH}$ are shown in Table 5.3. One can observe that the both methods are comparable and the relative performance depends on the particular situation at hand.

5.4 Analysis of 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 and

Table 5.2: 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**

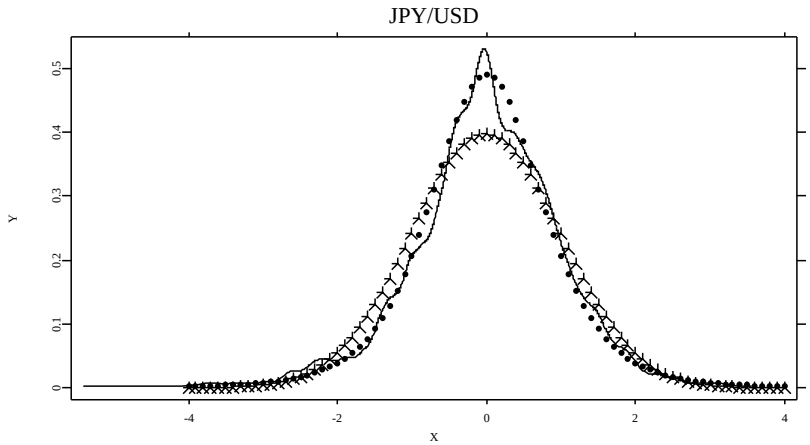
Frey (2000). Here we briefly discuss this issue by looking at the standardized returns $\widehat{\xi}_t = R_t/\widehat{\sigma}_t$. The first observation is that even after standardization by the estimated variance, the density of standardized returns $\widehat{\xi}_t$ still displays tails which are fatter than the normal. We illustrate this effect in Figure 5.8 where the kernel estimate of the density of 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 a much better approximation to the empirical density of returns.

The volatility clustering effect, though, disappears after standardization and auto-

Table 5.3: Forecasting performance relative to GARCH(1,1)

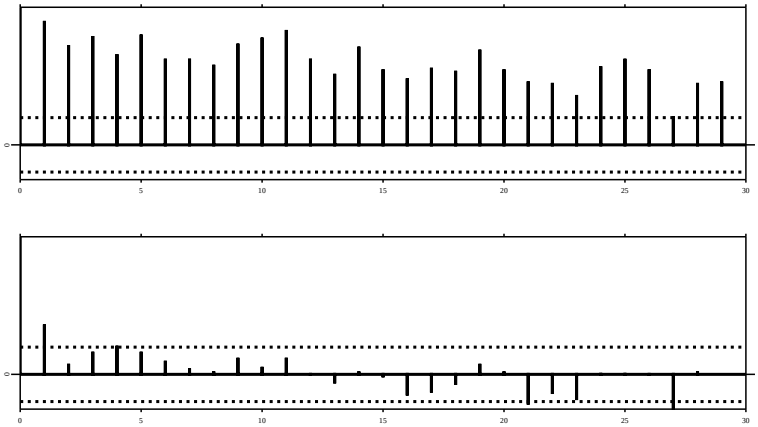
	E_p^{CP}/E_p^{GARCH}			$\widetilde{E}_p^{CP}/\widetilde{E}_p^{GARCH}$		
	$p = 0.5$	$p = 1.0$	$p = 2.0$	$p = 0.5$	$p = 1.0$	$p = 2.0$
AUD	1.0032	1.0108	**0.98093**	1.0079	1.0227	**0.94511**
CAD	1.0281	1.0515	1.0249	**0.99626**	1.0028	1.1148
BPD	1.0137	1.0585	1.0694	1.0165	1.0761	1.4696
DKR	1.0064	1.0219	1.0382	1.0059	1.0328	1.3042
JPY	1.0235	1.0582	**0.99125**	1.0021	1.0093	1.1626
NKR	**0.994**	1.0052	**0.97641**	1.0158	1.053	1.3243
NZD	**0.99301**	1.1138	1.0551	1.0293	1.1058	1.2533
SFR	1.0151	1.032	**0.98532**	**0.98715**	**0.95887**	**0.72988**
SKR	**0.97004**	**0.96783**	**0.94407**	1.02	1.0502	**0.92539**

Figure 5.8: 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.



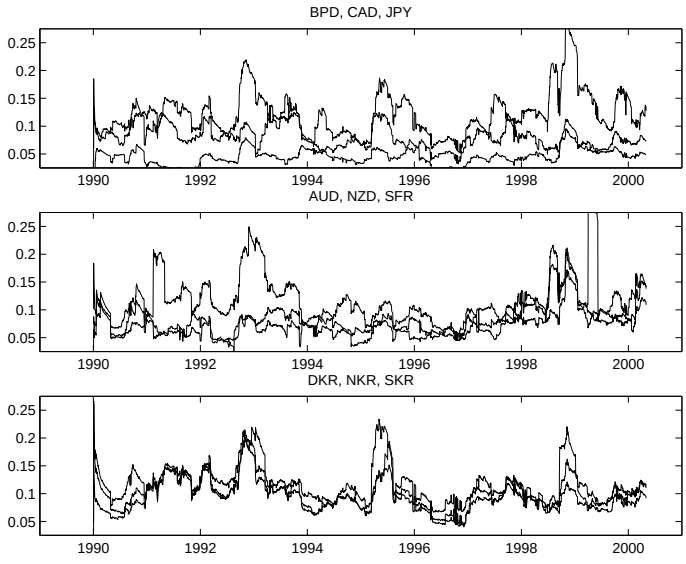
correlations of squared returns are not significant any more, see Figure 5.9 for the case of BPD/USD returns. The other exchange rate examples deliver similar results. A short conclusion of this empirical study is that the standardized returns can be treated as i.i.d. random variables with a distribution whose tails are fatter than the ones of the normal distribution.

Figure 5.9: ACF of the absolute BPD/USD returns (upper plot) and of the standardized absolute BPD/USD returns (lower plot). Dotted straight line - the 95% significance level.



5.5 Multiple volatility estimation for exchange rate datasets

Figure 5.10: Adaptive estimate of the diagonal elements of the covariance matrix of nine exchange rate data sets. The scale is annualized volatility, i.e. we plot $\sqrt{250\widehat{\Sigma}_{ii}}$.



Now we apply the multiple LCP procedure to the same set of nine exchange rates. Figure 5.10 shows the estimated diagonal elements of the covariance matrix. The upper

plot shows the values estimated by the procedure described in Section 3. Similarly to the univariate case, these estimates have been constructed only from past observations and therefore be used for forecasting and other practical applications such as Value at Risk.

Common movements, especially in the second part of the sample, can clearly be recognized, hinting that the volatilities of these processes are probably driven by some low dimensional common factor.

Figure 5.11: ACF for the absolute returns of NZD and AUD.

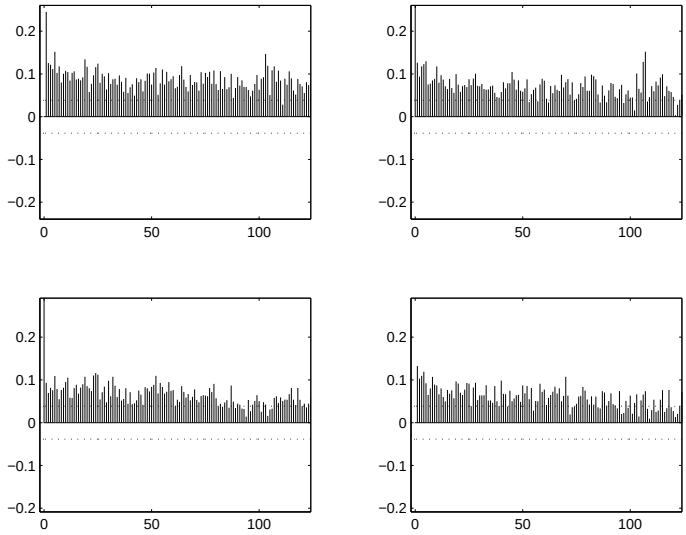
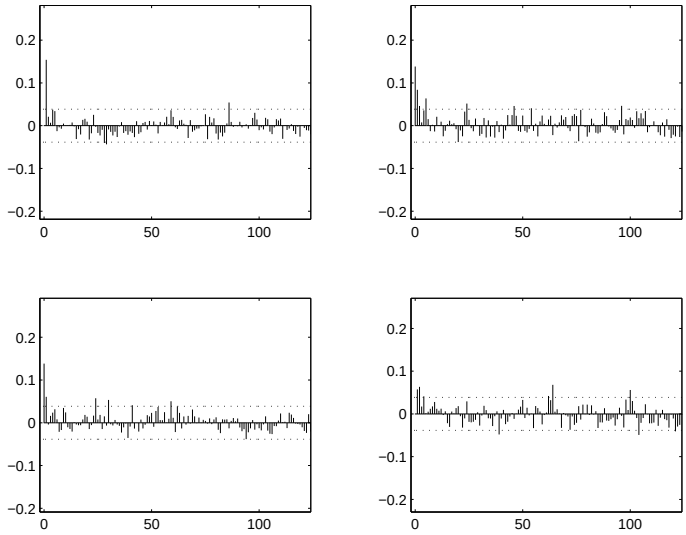


Figure 5.11 shows the multivariate autocorrelation function of the absolute returns for two exchange rate time series, New-Zealand Dollar and Australian Dollar, while, Figure 5.12 presents the multivariate autocorrelation function of the absolute standardized returns. Again, the standardized returns do not indicate significant correlations and autocorrelations. Note, that the autocorrelation of the absolute residuals has been almost completely smoothed away after standardizing by the estimated volatility matrix. The same conclusion holds for the other considered time series.

5.6 Application to Value at Risk. Univariate case

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 reduced to computing the quantiles of the distribution of aggregated returns, see e.g. Fan and Gu (2003) for a recent overview of this topic.

Figure 5.12: ACF for the absolute standardized returns of NZD and AUD.



Our modeling approach can easily be adapted to the VaR problem. Namely, one may 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 each date t in the following way. The volatility parameter $\widehat{\sigma}_t$ is estimated from the historical data R_s for $s \le t$ and one can consider different distributions for the innovations ε_t. In our study we compare the Gaussian, the scaled student t_5-distribution with 5 degrees of freedom and the empirical distribution $\widehat{F}_t$ of the past empirical innovations $\widehat{\xi}_s$ for $s \le t$, that is:

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

Similar approaches have been applied in McNeil and Frey (2000) with the use of the GARCH(1,1) model for estimating the volatility and extreme value theory for evaluating the distribution of returns, while Eberlein and 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%-quantiles than the ones of the Gaussian at any of the considered horizons and lower 1%-quantiles. Therefore the Gaussian distribution of innovations is more conservative for 5%-quantiles while the opposite is true for 1%-quantiles.

We apply the procedure to the set of nine exchange rates considered in Section 5.5 with about 2500 observations in each one. The frequency of overshooting the predicted quantile for the given realizations of the returns is given in Table 5.4. 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 and Gu (2003).

According to the requirement of the regulators (BIS, 1996), a bank has to determine its capital requirements in order to cover from market risk proportionally to the 1% quantile of the distribution of the portfolio losses over a ten day horizon. Internal models calculating this quantile are regularly monitored. The coefficient of proportionality is set to 3 for models whose performance is satisfactory (green zone) and it can be increased up to 4 by a discretionary judgment of the regulators for models which appear to estimate the quantile imprecisely (yellow zone). While, if the model performance is considered very poor, the coefficient is automatically increased to 4 (red zone).

The official criterion for the evaluation of an internal model is the statistical significance of the 1% quantile estimates of the portfolio loss distribution over a one day horizon. The prescribed procedure, called backtesting, checks, whether the observed frequency of days out of the last 250, for which the losses were larger than the value computed by the prescribed VaR procedure does not significantly deviates from the nominal level 0.01, see Deutsch (2001). Every procedure is classified as green, yellow and red. The green zone means that the empirical frequency is in agreement with the nominal probability 0.01. The yellow zone begins at the point such that the probability of exceptions for the tested VaR procedure exceeds the value 0.01 with 95% confidence interval. One can easily verify that such probability corresponds to 5 or more exceptions out of 250 days, that is, the frequency of exceptions equals 2%. Similarly, the red zone corresponds to the 99.99% level evidence that the tested procedure does not provide the required probability of exceptions. For a sample of 250 observations, this corresponds to 10 exceptions, or equivalently, 4% frequency of overshooting the VaR value.

The comparison of these requirements with our results presented in Table 5.4 shows that on average none of the procedures we tried is in the red zone, and that the procedure using e.d.f. for the residuals is always in the green zone. The use of the student t_5 distribution also allows to get the green zone results for most of the examples, while the procedure with Gaussian innovations is often in the yellow zone.

We conclude that the use of the t_5 distribution for the innovations slightly improve the results and the VaR quality is acceptable for both Gaussian and scaled student quantiles, 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 5.4: Percentage of overshooting the prescribed VaR level for nine series of exchange rates for nominal quantile levels 1% and 5%, three different distributions of innovations and time horizon $h = 1, 5, 10$.

	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

Table 5.5: 5% VaR

	5% quantile					
	Gaussian			student t_5		
	$h = 1$	$h = 5$	$h = 10$	$h = 1$	$h = 5$	$h = 10$
AUD	5.5	5.9	6.3	6.3	6.0	6.4
CAD	4.7	5.3	4.5	5.2	5.6	4.7
BPD	5.3	7.1	6.5	6.0	7.4	6.7
DKR	5.8	6.5	6.5	6.5	6.5	6.6
JPY	5.5	7.5	8.5	6.0	7.7	8.5
NKR	5.5	5.7	6.0	6.3	6.0	6.1
NZD	5.1	6.1	6.7	5.5	6.2	6.9
SFR	5.8	6.0	6.0	6.4	6.0	6.1
SKR	6.2	5.7	4.9	6.7	6.2	5.1

5.7 Value at Risk for multiple time series

Here we illustrate the same approach for the portfolio containing several assets. Suppose the portfolio consists of N assets with the vector of allocations $\psi \in I\!R^N$. The allocations are usually time dependent, that is, at time t we have $\psi_i = \psi_i(t)$ units of the assets i in the portfolio, $i = 1, \ldots, N$.

The goal is to estimate the Value at Risk for the whole portfolio. The estimated covariance matrix $\widehat{\Sigma}$ is one of the most important input for this problem. We present two possible approaches for computing VaR. One is based on the assumption that the joint distribution of the returns is multivariate normal, while the other estimates the quantiles from the empirical distribution of the returns standardized by the estimated covariance matrix and therefore can be interpreted as a version of the historical simulation approach.

Formally, for a given fixed allocation ψ, the VaR for an h day horizon and a probability level α of the portfolio $\psi^\top S_t$, is defined as the α-quantile of the distribution of the changes in the portfolio value:

$$P(\psi^\top (S_{t+h} - S_t) < \mathrm{VaR} | S_t) = \alpha.$$

For a given S_t, the estimation of the VaR of the portfolio changes can be obtained if one is able to determine the conditional distribution of the sum of the returns:

$$\sum_{i=1}^{N} \psi_i \left(\frac{S_{t+h,i} - S_{t,i}}{S_{t,i}} \right).$$

For small h the above expression can be conveniently approximated by the sum of the log-returns:

$$\sum_{i=1}^{N} \psi_i (\log(S_{t+h,i}) - \log(S_{t,i})) = \sum_{i=1}^{N} \psi_i (R_{t+1,i} + \ldots + R_{t+h,i}) = \sum_{u=1}^{h} \psi^\top R_{t+u}\,,$$

so that for the computation of the VaR we can exploit the properties of the log-returns. In particular, if we assume that the returns are normally distributed with the covariance matrix Σ_t, then

$$\psi^\top (R_{t+1} + \ldots + R_{t+h}) \sim \mathcal{N}(0, h\psi^\top \Sigma_t \psi).$$

This suggests to compute the VaR using the quantiles of the $\mathcal{N}(0, h\psi^\top \widehat{\Sigma}_t \psi)$.

Similar to the univariate case, the normal distribution assumption is not very accurate for modeling the tails of the distribution of financial returns. One possibility to

cope with this problem consists in fitting to the returns a distribution, which can better approximate the tail behavior of the data as in Eberlein and Prause (2002). We consider however another strategy, which consists in estimating the quantiles from the empirical distribution function of the standardized returns. Indeed, as shown in the previous Section, the standardized returns $\widehat{\varepsilon}_t := \widehat{\Sigma}_{t-1}^{-1/2} R_t$ are approximately independent, therefore the quantile of the distribution of $\psi^\top \sum_{u=1}^h R_{t+u}$ can be estimated by the quantile of the empirical distribution function of $\psi^\top \widehat{\Sigma}_t^{\frac{1}{2}} (\widehat{\xi}_{t+1} + \ldots + \widehat{\xi}_{t+h})$.

The estimation of VaR is performed on a data set of nine exchange rates with 2583 daily observations. The first 82 observation are used to perform the first estimation of the covariance matrix and the following 500 observations are used to perform the first estimation of the empirical distribution function. On the last 2000 observations VaR is computed every day using all the past standardized residuals for the estimation of the empirical distribution function. For the sake of comparison, VaR is also computed under Gaussian assumption on the last 2000 observations. Table 5.6 reports the frequency with which the realization of the portfolio return is lower than the estimated quantile. We consider the values of $\alpha = 0.05$ and 0.01 and the horizons of 1, 5 and 10 trading days for two different portfolios. One portfolio has equal weights for each asset for the whole period, while the other portfolio has randomly generated weights from an uniform distribution on $[0,1]^9$ for each day.

Table 5.6: Value at risk for two portfolios with 9 possible exchange rate allocations

	1% quantile						5% quantile					
	Gaussian			e.d.f.			Gaussian			e.d.f.		
horizon	1	5	10	1	5	10	1	5	10	1	5	10
equally weighted portfolio	2.3	2.0	1.9	1.0	0.8	0.5	6.0	5.8	5.5	4.0	3.5	3.3
random uniform portfolio	2.1	2.2	2.1	0.9	0.7	0.7	6.3	6.4	5.7	4.1	3.5	3.2

As expected the Gaussian model slightly overestimates the quantiles (yellow zone) and the method which relies on the estimation of the empirical distribution function performs remarkably well in particular for the 1% quantile (green zone). For the 5% quantile the method seems to be slightly conservative.

6 Appendix

In this section we present the proofs of the results stated in the previous sections.

6.1 Proof of Theorem 4.1

With the given constants $\mathfrak{z}_k$, define for $k > 1$ the random sets

$$\mathcal{A}_k = \{T_{I_k} \le \mathfrak{z}_k\}, \qquad \mathcal{A}^{(k)} = \mathcal{A}_2 \cap \ldots \cap \mathcal{A}_k\,,$$

Obviously $\widehat{\theta}_k = \widetilde{\theta}_k$ on $\mathcal{A}^{(k)}$ for all $k \le K$.

Therefore, it remains to bound the risk of $\widehat{\theta}_k$ on the complement $\overline{\mathcal{A}}^{(k)}$ of $\mathcal{A}^k$. Define $\mathcal{B}_k = \mathcal{A}^{(k-1)} \setminus \mathcal{A}^{(k)}$. On the event $\mathcal{B}_k$, the index k is the first one for which the condition $T_{I_k} \le \mathfrak{z}_k$ is violated. It is obvious that $\overline{\mathcal{A}}^{(k)} = \bigcup_{l<k} \mathcal{B}_l$.

First we bound the probability $I\!P_{\theta^*}\bigl(\mathcal{B}_k\bigr)$. The definition of $T_{I_k,\tau}$ and (3.1) yield for every k and every $\tau \in J_k = \mathcal{I}_k \setminus \mathcal{I}_{k-1}$

$$T_{I_k,\tau} \le N_{I_k''}\mathcal{K}(\widetilde{\theta}_{I''}, \theta^*) + N_{I_k'}\mathcal{K}(\widetilde{\theta}_{I'}, \theta^*).$$

Therefore, by Theorem 2.2,

$$I\!P_{\theta^*}\bigl(T_{I_k,\tau} > \mathfrak{z}_k\bigr) \le 4\mathrm{e}^{-\mathfrak{z}_k/2},$$

and hence,

$$I\!P_{\theta^*}\bigl(\mathcal{B}_k\bigr) \le \sum_{\tau \in J_k} I\!P_{\theta^*}\bigl(T_{I_k,\tau} > \mathfrak{z}_k\bigr) \le 4|J_k|\mathrm{e}^{-\mathfrak{z}_k/2} \le 4m_k\mathrm{e}^{-\mathfrak{z}_k/2}.$$

Next, for every $l < k$ and any $r > 0$, by Lemma 6.1 and the elementary inequality $(a+b)^{2r} \le 2^{2r}(a \vee b)^{2r}$ with any $a, b \ge 0$

$$I\!E_{\theta^*}\mathcal{K}^r\bigl(\widetilde{\theta}_k, \widetilde{\theta}_l\bigr) \le (2\mathfrak{a})^{2r}\bigl\{I\!E_{\theta^*}\mathcal{K}^r\bigl(\widetilde{\theta}_k, \theta^*\bigr) + I\!E_{\theta^*}\mathcal{K}^r\bigl(\widetilde{\theta}_l, \theta^*\bigr)\bigr\} \le (2\mathfrak{a})^{2r}\mathfrak{r}_r m_l^{-r}.$$

This and Theorem 2.4 imply for every r and $l < k \le K$

$$\begin{aligned} m_k^r I\!E_{\theta^*}\mathcal{K}^r\bigl(\widetilde{\theta}_k, \widehat{\theta}_k\bigr) &\le m_k^r(2\mathfrak{a})^{2r} I\!E_{\theta^*} \sum_{l=1}^{k} \mathcal{K}^r(\widetilde{\theta}_k, \widetilde{\theta}_l)\mathbf{1}\bigl(\mathcal{B}_l\bigr) \\ &\le m_k^r(2\mathfrak{a})^{2r} \sum_{l=1}^{k} I\!E_{\theta^*}^{1/2}\mathcal{K}^{2r}(\widetilde{\theta}_k, \widetilde{\theta}_l) I\!P_{\theta^*}^{1/2}\bigl(\mathcal{B}_l\bigr) \\ &\le (2\mathfrak{a})^{2r}\mathfrak{r}_{2r}^{1/2} \sum_{l=1}^{k} \Bigl(\frac{m_k}{m_l}\Bigr)^r 4m_l^{1/2}\mathrm{e}^{-\mathfrak{z}_l/4}. \end{aligned}$$

It remains to check that the choice $\mathfrak{z}_k = a_0 + a_1\log\alpha^{-1} + a_2 r\log(m_K/m_k) + a_3\log m_k$ with properly selected a_0, a_1, a_2 and a_3 provide the required bound $I\!E_{\theta^*}\bigl|m_k\mathcal{K}\bigl(\widetilde{\theta}_k, \widehat{\theta}_k\bigr)\bigr|^r \le \alpha\mathfrak{r}_r$.

6.2 Proof of Theorem 4.7

Let ν be the location of the change within $\mathcal{I}_{k^*+1}$. It suffices to show that under the conditions of the theorem the corresponding test statistic $T_{I_{k^*+1},\nu}$ exceeds with a high probability the value $\mathfrak{z}_{k^*+1}$. This would ensure that the interval $\mathcal{I}_{k^*+1}$ is rejected. The point ν splits the testing interval $I_{k^*+1} = \mathcal{I}_{k^*+2}$ into two subintervals I' and I'', and within each of intervals I' and I'' the function $f(t)$ is constant: $f(t) \equiv \theta'$ for $t \in I'$ and $f(t) \equiv \theta''$ for $t \in I''$. Let a value $\mathfrak{z} > 0$ be fixed. Introduce the event

$$\mathcal{A}(\mathfrak{z}) = \mathbf{1}\bigl\{ N_{I'}\mathcal{K}\bigl(\widetilde{\theta}_{I'}, \theta'\bigr) \le \mathfrak{z}, \quad N_{I''}\mathcal{K}\bigl(\widetilde{\theta}_{I''}, \theta''\bigr) \le \mathfrak{z} \bigr\}.$$

By Theorem 2.2,

$$I\!\!P\bigl(\mathcal{A}(\mathfrak{z})\bigr) \ge 1 - 4\mathrm{e}^{-\mathfrak{z}}.$$

We now consider $\mathfrak{z}$ such that (4.5) holds and show that $T_{I_{k^*+1},\nu} > \mathfrak{z}_{k^*+1}$ on $\mathcal{A}(\mathfrak{z})$. By definition, it holds on the set $\mathcal{A}(\mathfrak{z})$ that $N_{I'}\mathcal{K}\bigl(\widetilde{\theta}_{I'}, \theta'\bigr) \le \mathfrak{z}$ and $N_{I''}\mathcal{K}\bigl(\widetilde{\theta}_{I''}, \theta''\bigr) \le \mathfrak{z}$.

By Lemma 6.1

$$\begin{aligned} \mathcal{K}^{1/2}\bigl(\theta', \widetilde{\theta}_{I_{k^*+1}}\bigr) &\le \mathfrak{a}\mathcal{K}^{1/2}\bigl(\widetilde{\theta}_{I'}, \theta'\bigr) + \mathfrak{a}\mathcal{K}^{1/2}\bigl(\widetilde{\theta}_{I'}, \widetilde{\theta}_{I_{k^*+1}}\bigr) \\ &\le \mathfrak{a}(\mathfrak{z}/N_{I'})^{1/2} + \mathfrak{a}\mathcal{K}^{1/2}\bigl(\widetilde{\theta}_{I'}, \widetilde{\theta}_{I_{k^*+1}}\bigr) \end{aligned}$$

Hence,

$$\mathcal{K}\bigl(\theta', \widetilde{\theta}_{I_{k^*+1}}\bigr) \le 2\mathfrak{a}^2\mathfrak{z}/N_{I'} + 2\mathfrak{a}^2\mathcal{K}\bigl(\widetilde{\theta}_{I'}, \widetilde{\theta}_{I_{k^*+1}}\bigr)$$

and

$$\mathcal{K}\bigl(\widetilde{\theta}_{I'}, \widetilde{\theta}_{I_{k^*+1}}\bigr) \ge (2\mathfrak{a}^2)^{-1}\mathcal{K}^{1/2}\bigl(\theta', \widetilde{\theta}_{I_{k^*+1}}\bigr) - \mathfrak{z}/N_{I'}.$$

Similarly

$$\mathcal{K}\bigl(\widetilde{\theta}_{I''}, \widetilde{\theta}_{I_{k^*+1}}\bigr) \ge (2\mathfrak{a}^2)^{-1}\mathcal{K}^{1/2}\bigl(\theta'', \widetilde{\theta}_{I_{k^*+1}}\bigr) - \mathfrak{z}/N_{I''}.$$

Now by definition of $T_{I_{k^*+1},\nu}$, see (3.1),

$$\begin{aligned} T_{I_{k^*+1},\nu} &= N_{I'}\mathcal{K}\bigl(\widetilde{\theta}_{I'}, \widetilde{\theta}_{I_{k^*+1}}\bigr) + N_{I''}\mathcal{K}\bigl(\widetilde{\theta}_{I''}, \widetilde{\theta}_{I_{k^*+1}}\bigr) \\ &\ge (2\mathfrak{a}^2)^{-1}\bigl\{ N_{I'}\mathcal{K}\bigl(\theta', \widetilde{\theta}_{I_{k^*+1}}\bigr) + N_{I''}\mathcal{K}\bigl(\theta'', \widetilde{\theta}_{I_{k^*+1}}\bigr)\bigr\} - \mathfrak{z} \\ &= (2\mathfrak{a}^2)^{-1} m_{k^*+2}\bigl\{ c\mathcal{K}\bigl(\theta', \widetilde{\theta}_{I_{k^*+1}}\bigr) + (1-c)\mathcal{K}\bigl(\theta'', \widetilde{\theta}_{I_{k^*+1}}\bigr)\bigr\} - \mathfrak{z} \end{aligned}$$

with $c = N_{I'}/m_{k^*+2}$. This and the definition of $d(\theta', \theta'')$, see (4.4), yields on $\mathcal{A}(\mathfrak{z})$

$$T_{I_{k^*+1},\nu} \ge \frac{m_{k^*+2}}{(2\mathfrak{a}^2)} d^2(\theta', \theta'') - \mathfrak{z}$$

and the theorem assertion follows.

Lemma 6.1 (Polzehl and Spokoiny, 2005, Lemma 5.2)**.** *Under condition* (Θ) *it holds for every sequence* $\theta_0, \theta_1, \ldots, \theta_m \in \Theta$ *that*

$$\begin{aligned} \mathcal{K}^{1/2}(\theta_1, \theta_2) &\leq \mathfrak{a}\big\{\mathcal{K}^{1/2}(\theta_1, \theta_0) + \mathcal{K}^{1/2}(\theta_2, \theta_0)\big\}, \\ \mathcal{K}^{1/2}(\theta_0, \theta_m) &\leq \mathfrak{a}\big\{\mathcal{K}^{1/2}(\theta_0, \theta_1) + \ldots + \mathcal{K}^{1/2}(\theta_{m-1}, \theta_m)\big\}. \end{aligned}$$

This lemma extends to the multivariate case

Lemma 6.2. *Under condition* (Θ) *it holds for every sequence* $\Sigma_0, \Sigma_1, \ldots, \Sigma_m \in \Theta$ *that*

$$\begin{aligned} \mathcal{K}^{1/2}(\Sigma_1, \Sigma_2) &\leq \mathfrak{a}\big\{\mathcal{K}^{1/2}(\Sigma_1, \Sigma_0) + \mathcal{K}^{1/2}(\Sigma_2, \Sigma_0)\big\}, \\ \mathcal{K}^{1/2}(\Sigma_0, \Sigma_m) &\leq \mathfrak{a}\big\{\mathcal{K}^{1/2}(\Sigma_0, \Sigma_1) + \ldots + \mathcal{K}^{1/2}(\Sigma_{m-1}, \Sigma_m)\big\}. \end{aligned}$$

The proof is quite straightforward is left to the reader.

References

Andersen, T. G., Benzoni, L., and Lund, J. (2002). An empirical investigation of continuous-time equity return models. *Journal of Finance*, 3:1239–1284.

Baillie, R., Bollerslev, T., and Mikkelsen, H. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74:3–30.

BIS (1996). Supervisory framework for the use of "backtesting" in conjunction with the internal models approach to market risk capital requirements. Basel Commitee on Banking Supervision.

Bollerslev, T. (1986). Generalised autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

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

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

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

Deutsch, H. P. (2001). *Derivate und Interne Modelle.* Schäffer-Poeschel Verlag, Stuttgart, 2 edition.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica*, 68:1343–1376.

Eberlein, E. and Prause, K. (2002). The generalized hyperbolic model: financial derivatives and risk measures. In Geman, H., Madan, D., Pliska, S., and Vorst, T., editors, *Mathematical Finance-Bachelier Congress 2000*, pages 245–267, Berlin. Springer Verlag.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 50:987–1008.

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

Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometrc Reviews*, 5:1–50, 81–87.

Fan, J. and 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., and Spokoiny, V. (2003). Time inhomogeous multiple volatility modelling. *Journal of Financial Econometrics*, 1:55–95.

Kitagawa, G. (1987). Non-Gaussian state-space modelling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032–1063.

Kullback, S. (1959). *Statistics and Information Theory*. Wiley and Sons, New York.

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

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, 32:577–602.

Mercurio, D. and Torricelli, C. (2003). Estimation and arbitrage opportunities for exchange rate baskets. *Applied Econometrics*, 35:1689–1698.

Mikosch, T. and Starica, C. (2000a). Change of structure in financial time series, long range dependence and the garch model. University of Aarhus, Aarhus school of Business, Discussion Paper Nr. 58.

Mikosch, T. and Starica, C. (2000b). Is it really long memory we see in financial returns? In Embrechts, P., editor, *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.