Weierstraß–Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Statistical inference for time-inhomogeneous volatility models

Danilo Mercurio[1], Vladimir Spokoiny[2]

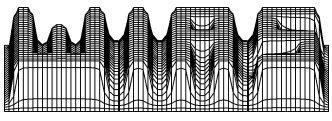
submitted: May,23, 2000

[1] Humboldt-Universität zu Berlin,
Spandauerstrasse 1,
D-10178 Berlin, Germany
E-Mail: isedoc02@wiwi.hu-berlin.de

[2] Weierstrass Institute for Applied Analysis
and Stochastics, Mohrenstraße 39
D – 10117 Berlin, Germany
E-Mail: spokoiny@wias-berlin.de
http://www.wias-berlin.de/~spokoiny

Preprint No. 583

Berlin 2000



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

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

Financial support by the Deutsche Forschungsgemeinschaft, Graduiertenkolleg für Angewandte Mikroökonomik at Humboldt University, is gratefully acknowledged.

The correspondence should be sent to Vladimir Spokoiny.

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

Fax: + 49 30 2044975
E-Mail (X.400): c=de;a=d400-gw;p=WIAS-BERLIN;s=preprint
E-Mail (Internet): preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

This paper offers a new approach for estimation and few-step ahead forecasting of the volatility of financial time series. No assumption is made about the parametric form of the processes, on the contrary we only suppose that the volatility can be approximated by a constant over some interval. In such a framework the main problem consists in filtering this *interval of time homogeneity*, then the estimate of the volatility can be simply obtained by local averaging. We construct an algorithm which can perform this task and investigate it both from the theoretical point of view and through Monte Carlo simulations. Finally the procedure is applied to some exchange rate data sets and a comparison with a standard GARCH model is also provided. Both models appear to be able of explaining many of the features of the data, nevertheless the new approach based on local constant approximation seems to be slightly superior as far as the out of sample results are taken into consideration.

1 Introduction

Stylized facts of financial asset returns such as stocks and exchange rates are: a leptokurtic density, variance clustering and highly persistent autocorrelation of square and absolute returns. A typical example can be seen in Figure 1, where the exchange rate returns for the period from 1 January 1990 to 7 April 2000 and the autocorrelation of their absolute values are plotted for the Canadian$/US$ exchange rate (first and second plot from the top) and for the Japanese Yen/US$ exchange rate (third and fourth plot). Further details and examples on this topic can be found in Taylor (1986).

Usually a white noise process with time varying variance is taken to model such features, so that the observed returns R_t follow the conditional heteroskedasticity model

$$R_t = \sigma_t \xi_t$$

where ξ_t are standard Gaussian independent innovations and σ_t is a time-varying *volatility* coefficient. In general situation σ_t is assumed to be a predictable random process. For modeling this volatility process one or another parametric assumption is usually used. The main model classes are the ARCH (Engle 1995a), GARCH (Bollerslev 1995) family, and the stochastic volatility (Harvey, Ruiz & Shephard 1995). A great amount of papers has followed the first publications on this topic, and the original models have been extended in order to provide better explanation. For example models which take into account asymmetries in volatility have been proposed, such as EGARCH (Nelson 1995),

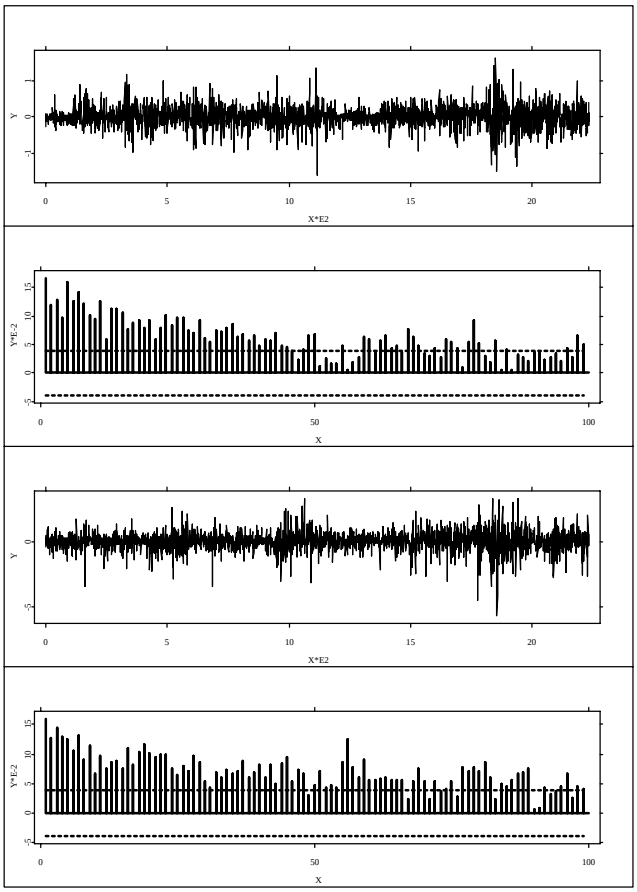


Figure 1: Exchange rate returns for the period from 1 January 1990 to 7 April 2000 and the autocorrelation of their absolute values for the Canadian$/US$ exchange rate (first and second plot from the top) and for the Japanese Yen/US$ exchange rate (third and fourth plot). The horizontal line indicates the 5% significance level of the estimated autocorrelation coefficients.

QGARCH (Sentana 1995) and GJR (Glosten, Jagannathan & Runkle 1992); furthermore the research on integrated processes has produced integrated (Engle & Bollerslev 1986) and fractal integrated versions of the GARCH model.

The availability of very large samples of financial data has given the possibility of constructing models which display quite complicated parameterizations in order to explain all the observed stylized facts. Obviously those models rely on the assumption that the parametric structure of the process remains constant through the whole sample. This is a nontrivial and possibly dangerous assumption in particular as far as forecasting is concerned (Clements & Hendry 1998). Furthermore checking for parameter instability becomes quite difficult if the model is nonlinear, and/or the number of parameters is

large. Whereby those characteristics of the returns which are often explained by the long memory and (fractal) integrated nature of the volatility process, could also depend on the parameters being time varying.

In this paper we propose another approach focusing on a very simple model but with a possibility for model parameters to depend on time. This means that the model is regularly checked and adapted to the data. No assumption is made about the parametric structure of the volatility process, we only suppose that it can be locally approximated by a constant, that is, for every time point τ there exists a past interval $[\tau - m, \tau]$ where the volatility σ_t did not vary much. This interval is referred to as *interval of time homogeneity*. An algorithm is proposed for data-driven estimation of the interval of time homogeneity, after which the estimate of the volatility can be simply obtained by averaging. It would be noted that the proposed approach attempts to describe the *local* dynamic of the volatility process. Such a strategy is particularly appealing for short term forecasting purposes which is an important building block e.g. in Value-at-Risk and portfolio hedging problems or backtesting (Härdle & Stahl 1999). At the same time, the underlying assumption of the local homogeneity can be hardly extended on a long forecasting horizon, so an application of the locally constant modelling to e.g. the problems of option price evaluation (where some *global* dynamic of the volatility process is important) is questionable.

The reminder paper is organized as follows. The next section introduces the adaptive modeling procedure, then some theoretical properties are discussed in the general situation and for a specific change-point model. A simulation study illustrates the performances of the new methodology with respect to different underling volatility processes. The question of selecting the smoothing parameters is also addressed and some solutions are proposed. Finally the procedure is applied to a set of nine exchange rates and it appears to be highly competitive with standard GARCH(1,1), which is used as a benchmark model. Mathematical proofs are given in the appendix.

2 Modeling volatility via power transformation

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

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

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

A *time-homogeneous* (*time-homoskedastic*) model means that σ_t is a constant. The process e^{S_t} is then a Geometric Brownian motion observed at discrete time moments. The assumption of time homogeneity is too restrictive in practical applications and it does not allow to fit well real data. In this paper we consider an approach based on the *local time-homogeneity* which means that for every time moment τ there exists a time interval $[\tau - m, \tau]$ where the volatility process σ_t is nearly constant. Under such a modeling, the main intention is both to describe the interval of homogeneity and to estimate the corresponding value σ_τ which can then be used for one-step forecasting etc.

2.1 Data transformation

The model equation (2.1) links the target volatility function σ_t with the observations R_t via the multiplicative errors ξ_t. The classical well developed regression approach relies on the assumption of additive errors which can be then smoothed out by some kind of averaging. A natural and widespread method of transforming the equation (2.1) into a regression like equation is to apply the log-function to both its sides squared:

$$\log R_t^2 = \log \sigma_t^2 + \log \xi_t^2 \tag{2.2}$$

which can be rewritten in the form

$$\log R_t^2 = \log \sigma_t^2 + C + v\zeta_t$$

with $C = \boldsymbol{E} \log \xi_t^2$, $v^2 = \operatorname{Var} \log \xi_t^2$ and $\zeta_t = v^{-1} \left(\log \xi_t^2 - C \right)$, see e.g. Gouriéroux (1997). This is a usual regression equation with the "response" $Y_t = \log R_t^2$, target regression function $f(t) = \log \sigma_t^2 + C$ and homogeneous "noise" $v\zeta_t$.

The main problem with this approach is due to the distribution of the errors ζ_t. It is log-normal and it has heavy tails. It is also highly skewed and it gives very high weights to the small values of the errors ξ_t, see Figure 3 for a typical sample from this distribution. Particularly this leads to a serious problem with missing data which are typically modeled equal to previous values providing $R_t = 0$.

Another possibility is based on power transformation, see Carroll & Ruppert (1988) which also leads to a regression with additive noise and this noise is much closer to a Gaussian one. Due to (2.1) the random variable R_t is conditionally on $\mathcal{F}_{t-1}$ Gaussian and it holds

$$\boldsymbol{E} \left(R_t^2 \,\middle|\, \mathcal{F}_{t-1} \right) = \sigma_t^2.$$

Similarly, for every $\gamma > 0$,

$$\boldsymbol{E}\left(|R_t|^\gamma \,|\, \mathcal{F}_{t-1}\right) = \sigma_t^\gamma \boldsymbol{E}\left(|\xi|^\gamma \,|\, \mathcal{F}_{t-1}\right) = C_\gamma \sigma_t^\gamma,$$
$$\boldsymbol{E}\left(|R_t|^\gamma - C_\gamma \sigma_t^\gamma \,|\, \mathcal{F}_{t-1}\right)^2 = \sigma_t^{2\gamma} \boldsymbol{E}\left(|\xi|^\gamma - C_\gamma\right)^2 = \sigma_t^{2\gamma} D_\gamma^2$$

where ξ denotes a standard Gaussian r.v., $C_\gamma = \boldsymbol{E}|\xi|^\gamma$ and $D_\gamma^2 = \mathrm{Var}\,|\xi|^\gamma$. Therefore, the process $|R_t|^\gamma$ allows for the representation

$$|R_t|^\gamma = C_\gamma \sigma_t^\gamma + D_\gamma \sigma_t^\gamma \zeta_t, \tag{2.3}$$

where ζ_t has conditionally on $\mathcal{F}_{t-1}$ the distribution $\left(|\xi|^\gamma - C_\gamma\right)/D_\gamma$. In the sequel we refer to the equation (2.3) as the *martingale representation* of the process $|R_t|^\gamma$. Indeed,

$$\boldsymbol{E}\left(|R_t|^\gamma - C_\gamma \sigma_t^\gamma \mid \mathcal{F}_{t-1}\right) = \boldsymbol{E}\left(D_\gamma \sigma_t^\gamma \zeta_t \mid \mathcal{F}_{t-1}\right) = D_\gamma \sigma_t^\gamma \boldsymbol{E}\left(\zeta_t \mid \mathcal{F}_{t-1}\right) = 0.$$

Note that the problem of estimating σ_t is in some sense equivalent to the problem of estimating $\theta_t = C_\gamma \sigma_t^\gamma$ which is the mean value of the transformed process R_t. This is already a kind of a heteroskedastic regression problem with additive errors $D_\gamma \sigma_t^\gamma \zeta_t$ satisfying

$$\begin{aligned} \boldsymbol{E}\left(D_\gamma \sigma_t^\gamma \zeta_t \mid \mathcal{F}_{t-1}\right) &= 0, \\ \boldsymbol{E}\left(D_\gamma^2 \sigma_t^{2\gamma} \zeta_t^2 \mid \mathcal{F}_{t-1}\right) &= D_\gamma^2 \sigma_t^{2\gamma}. \end{aligned}$$

A minimization of the skewness $\boldsymbol{E}\zeta_\gamma^3$ and the fat $\boldsymbol{E}\zeta_\gamma^4 - 3$ with respect to γ leads to the choice $\gamma \approx 1/2$. The corresponding density $p_{1/2}(x)$ of ζ_γ together with the standard normal density $\phi(x)$ is plotted in Figure 2.

3 Adaptive estimation under local time-homogeneity

Here we describe one approach to volatility modeling based on the assumption of local time homogeneity starting from the preliminary heuristic discussion. The assumption of local time homogeneity means that the function σ_t is nearly constant within an interval $I = [\tau - m, \tau[$, and the process R_t follows the regression-like equation (2.3) with the constant trend $\theta_I = C_\gamma \sigma_I^\gamma$ which can be estimated by averaging over this interval I:

$$\widetilde{\theta}_I = \frac{1}{|I|} \sum_{t \in I} |R_t|^\gamma. \tag{3.1}$$

By (2.3)

$$\widetilde{\theta}_I = \frac{C_\gamma}{|I|} \sum_{t \in I} \sigma_t^\gamma + \frac{D_\gamma}{|I|} \sum_{t \in I} \sigma_t^\gamma \zeta_t = \frac{1}{|I|} \sum_{t \in I} \theta_t + \frac{s_\gamma}{|I|} \sum_{t \in I} \theta_t \zeta_t \tag{3.2}$$

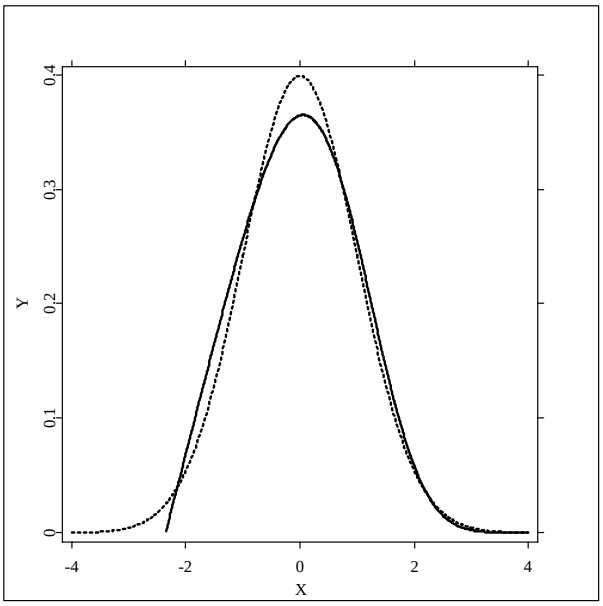


Figure 2: Density of $p_{1/2}(x)$ (straight line) and of a standard normal random variable (dotted line).

with $s_\gamma = D_\gamma / C_\gamma$ so that

$$\boldsymbol{E}\widetilde{\theta}_I = \boldsymbol{E}\frac{1}{|I|}\sum_{t\in I}\theta_t\,, \tag{3.3}$$

$$\frac{s_\gamma^2}{|I|^2}\boldsymbol{E}\left(\sum_{t\in I}\theta_t\zeta_t\right)^2 = \frac{s_\gamma^2}{|I|^2}\boldsymbol{E}\sum_{t\in I}\theta_t^2. \tag{3.4}$$

Define also

$$v_I^2 = \frac{s_\gamma^2}{|I|^2}\sum_{t\in I}\theta_t^2.$$

In view of (3.4) this value is called the *conditional variance* of $\widetilde{\theta}_I$.

Under local homogeneity it holds $\theta_t \equiv \theta_I$ for $t \in I$, and hence,

$$\begin{aligned}\boldsymbol{E}\widetilde{\theta}_I &= \theta_I\,,\\ v_I^2 &= \operatorname{Var}\widetilde{\theta}_I = \frac{s_\gamma^2\theta_I^2}{|I|}\,.\end{aligned}$$

3.1 Some properties of the estimate $\widetilde{\theta}_I$

Due to our assumption of local homogeneity, the value θ_t is close to θ_τ for all $t \in I$. This means that the value

$$\Delta_I^2 = |I|^{-1} \sum_{t \in I} (\theta_t - \theta_\tau)^2$$

is small.

Theorem 3.1 *Let the volatility coefficient σ_t satisfy the condition*

$$b \le \sigma_t^2 \le bB \tag{3.5}$$

with some positive constant b, B. Then there exists $a_\gamma > 0$ such that it holds for every $\lambda \ge 0$

$$\boldsymbol{P}\left(|\widetilde{\theta}_I - \theta_\tau| > \Delta_I + \lambda v_I\right) \le 4\sqrt{e}\lambda(1 + \log B) \exp\left(-\frac{\lambda^2}{2a_\gamma}\right).$$

Remark 3.1 This result can be slightly refined for the special case when the volatility function σ_t is deterministic:

$$\boldsymbol{P}\left(|\widetilde{\theta}_I - \theta_\tau| > \Delta_I + \lambda v_I\right) \le 2 \exp\left(-\frac{\lambda^2}{2a_\gamma}\right).$$

The result of this theorem bounds the loss of the estimate $\widetilde{\theta}_I$ via value Δ_I and the conditional standard deviation v_I. The latter term depends in its turn on the target function θ_t. One would be interested in another bound which does not involve the unknown function θ_t. Namely, basing on (3.4) and assuming Δ_I small, one may replace the conditional standard deviation v_I by its estimate

$$\widetilde{v}_I = s_\gamma \widetilde{\theta}_I |I|^{-1/2}.$$

Theorem 3.2 *Let $R_1, \ldots, R_\tau$ obey (2.1) and let (3.5) hold true. Then it holds for the estimate $\widetilde{\theta}_I$ of θ_τ :*

$$\begin{aligned}\boldsymbol{P}&\left(|\widetilde{\theta}_I - \theta_\tau| > \Delta_I(1 + \lambda s_\gamma |I|^{-1/2}) + \lambda \widetilde{v}_I\right) \\ &\le 4\sqrt{e}\lambda(1 + \log B) \exp\left(-\frac{\lambda^2}{2a_\gamma(1 + \lambda s_\gamma |I|^{-1/2})^2}\right).\end{aligned}$$

3.2 Adaptive choice of the interval of homogeneity

Given observations $R_1, \ldots, R_\tau$ following the time-inhomogeneous model (2.1), we aim to estimate the current value of the parameter θ_τ using the estimate $\widetilde{\theta}_I$ with a properly selected time interval I of the form $[\tau - m, \tau[$ to minimize the corresponding estimation error. Below we discuss one approach which goes back to the idea of pointwise adaptive estimation, see Lepski (1990), Lepski & Spokoiny (1997) and Spokoiny (1998). The idea of the method can be explained as follows. Suppose I is an interval-candidate, that is, we expect time-homogeneity in I and hence, in every subinterval of I. This particularly implies that the value Δ_I is negligible and similarly for all Δ_J, $J \subset I$ and that the mean values of the θ_t over I and over J nearly coincide. Our adaptive procedure roughly means a family of tests to check whether $\widetilde{\theta}_I$ does not differ significantly from $\widetilde{\theta}_J$ for any subinterval J of I. The latter is done on the base of Theorem 3.2 which allows under homogeneity within I to bound $|\widetilde{\theta}_I - \widetilde{\theta}_J|$ by $\lambda\widetilde{v}_I + \lambda\widetilde{v}_J$ provided that λ is sufficiently large. If there exists an interval $J \subset I$ such that the hypothesis $\widetilde{\theta}_I = \widetilde{\theta}_J$ cannot be accepted then we reject the hypothesis of homogeneity for the interval I. Finally, our adaptive estimate corresponds to the largest interval I such that the hypothesis of homogeneity is not rejected for I itself and all smaller intervals.

Now we present a formal description. Suppose a family $\mathcal{I}$ of interval-candidates I is fixed. Each of them is of the form $I = [\tau - m, \tau[$, $m \in \mathbb{N}$, so that the set $\mathcal{I}$ is ordered due to m. With every such interval we associate the estimate $\widetilde{\theta}_I$ of the parameter θ_τ due to (3.1) and the corresponding estimate $\widetilde{v}_I$ of the conditional standard deviations v_I.

Next, for every interval I from $\mathcal{I}$, we suppose to be given a set $\mathcal{J}(I)$ of testing subintervals J (one example of these sets $\mathcal{I}$ and $\mathcal{J}(I)$ is given in the next section). For every $J \in \mathcal{J}(I)$, we construct the corresponding estimate $\widetilde{\theta}_J$ from the observations Y_t for $t \in J$ according to (3.1) and compute $\widetilde{v}_J$.

Now, with two constants λ and μ, define the adaptive choice of the interval of homogeneity by the following iterative procedure:

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

Iteration Select the next interval I in $\mathcal{I}$ and calculate the corresponding estimate $\widetilde{\theta}_I$ and the estimated conditional standard deviation $\widetilde{v}_I$.

Testing homogeneity Reject I, if there exists one $J \in \mathcal{J}(I)$ such that

$$\left|\widetilde{\theta}_I - \widetilde{\theta}_J\right| > \lambda\, \widetilde{v}_J + \mu\, \widetilde{v}_I. \tag{3.6}$$

Loop If I is not rejected, then continue with the iteration step by choosing a larger interval. Otherwise, set $\widehat{I}$ = "the latest non rejected I".

The adaptive estimate $\widehat{\theta}_\tau$ of θ_τ is defined by applying this selected interval $\widehat{I}$:

$$\widehat{\theta}_\tau = \widetilde{\theta}_{\widehat{I}}\,.$$

It is supposed that the procedure is independently carried out at each time point τ. A possibility to reduce the computational effort of the selection rule is to make an adaptive choice of the interval of homogeneity only for some specific time points t_k and to keep the left end-point of the latest selected interval for all τ between two neighbor points t_k and t_{k+1}, see the next subsection for a proposal.

3.3 Choice of the sets $\mathcal{I}$, $\mathcal{J}(I)$ and the parameters λ and μ

The presented algorithm involves the sets $\mathcal{I}$ and $\mathcal{J}(I)$ of considered intervals and two numeric parameters λ and μ. We now discuss how these parameters can be selected starting from the set of intervals $\mathcal{I}$. The simplest proposal is to introduce a regular grid $\mathcal{G} = \{t_k\}$ with $t_k = m_0 k$ for some natural number m_0 and to consider the intervals $I_k = [t_k, \tau[$ for all $t_k < \tau$. It is also reasonable to carry out the adaptive procedure only for points τ from the same grid $\mathcal{G}$. The value m_0 can be selected, e.g., between 5 and 30.

If $\tau = t_{k^*}$ for some $k^* \geq 1$, then clearly every interval $I = [t_k, \tau[$ contains exactly $k^* - k$ smaller intervals $I' = [t_{k'}, \tau[$ for all $k < k' \leq k^*$. Next, for every such interval $I = [t_k, \tau[$, we define the set $\mathcal{J}(I)$ of testing intervals J by taking all smaller intervals $I' = [t_{k'}, \tau[$ with the right end-point τ and similarly all smaller intervals $[t_k, t_{k'}[$ with the left end-point t_k, $k < k' \leq k^*$:

$$\mathcal{J}(I_k) = \{J = [t_{k'}, \tau[\text{ or } J = [t_k, t_{k'}[\colon k < k' < k^*\}.$$

Let N_I stand for the number of subintervals J in $\mathcal{J}(I)$. Clearly, for $I = [t_k, t_{k^*}[$, the set $\mathcal{J}(I)$ contains at most $2(k^* - k)$ elements, that is, $N_I \leq 2(k^* - k)$.

3.4 Data-driven choice of parameters λ and μ

The behavior of the procedure critically depends on the parameters λ and μ. The simulation results from the next section indicate that there is no universal 'optimal'

choice. Below we discuss two possibilities: one is based on a more detailed consideration for a change-point model and another one is based on the data-driven selector based on the mean forecast error which we discuss right now.

First we have to make clear the difference between the data-driven choice of the interval of homogeneity I and data-driven choice of parameters λ and μ. The adaptive procedure proposed for selecting the interval of homogeneity is *local* in the sense that it is performed at every point τ independently. Such procedures are also called *pointwise* or *spatially* adaptive, among them: kernel smoothers with plug-in bandwidth selector (Brockmann, Gasser & Herrmann 1993) or pointwise adaptive bandwidth selector (Lepski, Mammen & Spokoiny 1997), nonlinear wavelet procedure (Donoho, Johnstone, Kerkyacharian & Picard 1994). All these procedures have been shown to possess some spatial adaptive properties. However, every such procedure contains some free parameter(s) which have strong influence on their behavior. The most well known example is given by the thresholding parameter for the wavelets method. The values λ and μ of the above procedure have the same flavor as the threshold for wavelets. These parameters are *global* in the sense that there is no way to select them optimally for one specific point but they determine the global performance of the procedure on a large observation interval. Namely, for every pair λ, μ we can build a corresponding procedure (estimator) $\widehat{\theta}_t^{(\lambda,\mu)}$ at every point t from the 'observations' $Y_1, \ldots, Y_{t-1}$ as described in Section 3.2. In view of the martingale representation (2.3), it holds $\boldsymbol{E}\,(Y_t \mid \mathcal{F}_{t-1}) = \theta_t$ and the estimate $\widetilde{\theta}_t$ constructed on the base of previous 'observations' $Y_1, \ldots, Y_{t-1}$ is a natural one-step forecast of the next 'observation' Y_t. Now we define the pair $(\widehat{\lambda}, \widehat{\mu})$ as the minimizer of the mean forecast error:

$$(\widehat{\lambda}, \widehat{\mu}) = \inf_{\lambda,\mu} \sum_{t=t_0}^{\tau} \left(|R_t|^\gamma - \widehat{\theta}_t^{(\lambda,\mu)} \right)^2 ,$$

where infimum is taken over all considered pairs λ, μ and t_0 is taken to provide enough data for the starting estimates. Similarly one can choose the grid step m_0 in a data-driven way.

In practical applications, especially for a short term forecasting of the volatility process, the parameters λ, μ can also be updated at every time points by application of a moving window of a fixed length. This would lead to the rule

$$(\widehat{\lambda}_\tau, \widehat{\mu}_\tau) = \inf_{\lambda,\mu} \sum_{t=\tau-M}^{\tau} \left(|R_t|^\gamma - \widehat{\theta}_t^{(\lambda,\mu)} \right)^2 .$$

Such a rolling estimator is also meaningful because it takes into consideration the requirement of financial regulators, such as the Bundesaufsichtamt für Kreditwesen in Germany,

which does not allow the use of very old data which leads to $M = 250$ corresponding to one calendar year.

The cardinality of the considered set of admissible pairs λ, μ can be drastically reduced. The latter is based on a more detailed study of a change-point model which is discussed in Section 5.

4 Theoretic properties

In this section we collect some results describing the quality of the proposed adaptive procedure.

4.1 Accuracy of the adaptive estimate

Let $\widehat{I}$ be the interval selected by our adaptive procedure. We also define the "ideal" choice

$$I\!I = \operatorname{argmax} \{|I| : I \in \mathcal{I},\, \Delta_I \le D v_I\} \tag{4.1}$$

where D is some fixed constant. For this 'ideal' choice $I = I\!I$ we have the balance between the accuracy of approximation (which is controlled by Δ_I) and the stochastic error characterized by the stochastic variance v_I. By definition $v_I = s_\gamma |I|^{-1} \left(\sum_{t \in I} \theta_t^2\right)^{1/2}$ so that v_I typically decreases when $|I|$ increases. For simplicity of notation, we shall suppose further that $v_I \le v_J$ for $J \subset I$.

The "ideal" choice $I\!I$ means that we select the largest interval I for which the variability of the function θ_t inside I is not too large compared to the conditional stochastic deviation v_I. This, due to Theorem 9.1, allows us to bound with a high probability the losses of the "ideal" estimate $\widetilde{\theta}_{I\!I}$ by $(D + \lambda) v_{I\!I}$ provided that λ is sufficiently large. The next assertion claims that the risk of the adaptive estimate is of the same order $v_{I\!I}$.

Theorem 4.1 *Let (3.5) hold true. Then it holds for the adaptive estimate $\widehat{\theta} = \widetilde{\theta}_{\widehat{I}}$ defined in Section 3.2 with $\mu \ge \lambda$:*

$$\boldsymbol{P}\left(|\widehat{\theta} - \theta_\tau| > 2(\lambda + \mu) v_{I\!I}\right) \le \sum_{J \in \mathcal{J}(I\!I)} 4\sqrt{e}\lambda_J (1 + \log B) \exp\left(-\frac{\lambda_J^2}{2a_\gamma}\right) \tag{4.2}$$

where

$$\lambda_J = \frac{\lambda}{1 + \lambda s_\gamma |J|^{-1/2}} - \delta_J\,, \qquad \delta_J = \frac{\Delta_J}{v_J}\,.$$

Remark 4.1 The result of Theorem 4.1 leads to the following interpretation of a "good" choice $I\!I$: an interval $I\!I$ is good if the value

$$\lambda^*_{I\!I} = \min_{J \in \mathcal{J}(I\!I)} \left(\frac{\lambda}{1 + \lambda s_\gamma |J|^{-1/2}} - \frac{\Delta_J}{v_J} \right)$$

is sufficiently large providing a small probability of the event $\{|\widehat{\theta} - \theta_\tau| > 2(\lambda + \mu) v_{I\!I}\}$. This particularly implies that the ratio $\Delta_{I\!I}/v_{I\!I}$ should not be too large: $\Delta_{I\!I}/v_{I\!I} < \lambda/(1 + \lambda s_\gamma |I\!I|^{-1/2})$.

5 Change-point model

An important special case of the model (2.1) is the so-called *change-point* model corresponding to the piecewise constant function θ_t. For this special case, the above procedure has a very natural interpretation: when estimating at the point τ we search for a largest interval of the form $[\tau - m, \tau[$ does not containing a change-point. This is doing via testing for a change-point within the interval-candidate $I = [\tau - m, \tau[$. It is worth mentioning that the classical maximum-likelihood test for no change-point in the regression case with Gaussian $\mathcal{N}(0, \sigma^2)$-errors is also based on comparison of the mean values of observations Y_t over the whole interval $I = [\tau - m, \tau[$ and every subinterval $J = [\tau - j, \tau[$ or $J' = [\tau - m, \tau - j[$ for different j, so that the proposed procedure has strong appeal in this situation. However, there is an essential difference between testing of a change-point and a testing homogeneity appearing as a building block of our adaptive procedure. Usually a test for a change-point is constructed in a way to provide the prescribed type I error (in the change-point framework such an error is called a "false alarm"). Our adaptive procedure involves a lot of such tests for every candidate I, which leads to a multiple testing problem. As a consequence, each particular test should be performed at a very small level, i.e., it should be rather conservative providing a joint error probability at a reasonable level.

5.1 Type I error

For the change-point model, the type I error would mean that the interval-candidate I is rejected although the hypothesis of homogeneity is still fulfilled. On the other hand, the type II error means that interval I is not rejected in spite of a violation from homogeneity, so that the type II error probability describes the sensitivity of the procedure to changes.

The arguments used in the proof of Theorem 4.1 lead to the following upper bound for the type I error probability:

Theorem 5.1 *Let $\widehat{I}$ be selected by the adaptive procedure with $\mu \geq \lambda$. If $I = [\tau - m, \tau[$ is an interval of homogeneity, that is $\theta_t = \theta_\tau$ for all $t \in I$, then*

$$\boldsymbol{P}\,(I \text{ is rejected}) \leq \sum_{J \in \mathcal{J}(I)} 2 \exp\left(- \frac{\lambda^2}{2 a_\gamma (1 + \lambda s_\gamma |J|^{-1/2})^2} \right).$$

This result is a special case of Theorem 4.1 with $\Delta_J \equiv 0$ and $B = 1$ when taking into account Remark 9.1.

As a consequence of this result one can immediately see that for every fixed value M there exists a fixed λ providing a prescribed upper bound α for the type one error probability for a homogeneous interval I of length M. Namely, the choice

$$\lambda \geq (1 + \epsilon) \sqrt{2 a_\gamma \log \frac{2M}{m_0 \alpha}}$$

leads for a proper small positive constant $\epsilon > 0$ to the inequality

$$\sum_{J \in \mathcal{J}(I)} 2 \exp\left(- \frac{\lambda^2}{2 a_\gamma (1 + \lambda s_\gamma |J|^{-1/2})^2} \right) \leq \alpha.$$

(Here $2M/m_0$ is approximately the number of intervals in $\mathcal{J}(I)$.) This bound is however, very rough and it is only of theoretical importance since we estimate the probability of the sum of dependent events by the sum of single probabilities. The problem of finding λ providing a prescribed type I error probability is discussed in the next section.

5.2 Type II error

Next we consider the case of estimation immediately after a change-point. Let a change occur at a moment T_{cp}. It is convenient to suppose that T_{cp} belongs to the grid $\mathcal{G}$ on which we carry out the adaptive choice of the interval of homogeneity. This assumption is not restrictive if the grid is 'dense', that is, if the grid step m_0 is not too large. In the case with $T_{\mathrm{cp}} \in \mathcal{G}$, the 'ideal' choice $I\!I$ is clearly $[T_{\mathrm{cp}}, \tau[$. Theorem 4.1 provides that the quality of estimation at τ is essentially the same as if we knew the latest change-point T_{cp} a priori. In fact, one can state a slightly stronger assertion: every interval I which is essentially larger than $I\!I$ will be rejected with a high probability provided that the magnitude of the change is large enough.

Denote $m' = |I\!I|$, that is, $m' = \tau - T_{\mathrm{cp}}$. Let also $I = [T_{\mathrm{cp}} - m, \tau[= [\tau - m' - m, \tau[$ for some m, so that $|I| = m + m'$, and let θ (resp. θ') denote the value of parameter θ_t before (resp. after) change-point T_{cp}. The magnitude of the change-point is measured by the relative change $b = 2|\theta' - \theta|/\theta$.

The interval I will be certainly rejected if, either $|\widetilde{\theta}_I - \widetilde{\theta}_J|$ or $|\widetilde{\theta}_I - \widetilde{\theta}_{I\!I}|$ is sufficiently large compared to the corresponding critical value.

Theorem 5.2 *Let* $\boldsymbol{E}Y_t = \theta$ *before the change-point at* $T_{\rm cp}$ *and* $\boldsymbol{E}Y_t = \theta'$ *after it, and let* $b = |\theta' - \theta|/\theta$ *. Let also* $m' = |I\!I| = \tau - T_{cp}$ *and* $I = [\tau - m' - m, \tau[$ *. Then*

$$\boldsymbol{P}\,(I \textit{ is not rejected}) \le 4e^{-\frac{\lambda^2}{2a_\gamma}}$$

provided that

$$1 - \delta - \frac{\mu}{\lambda\sqrt{2}}\delta(1+\delta) > 0 \quad \textit{and} \quad b \ge \frac{\delta + \delta(1+\delta) + \frac{\mu}{\lambda\sqrt{2}}\delta(1+\delta)}{1 - \delta - \frac{\mu}{\lambda\sqrt{2}}\delta(1+\delta)} \tag{5.1}$$

with $\delta = \frac{\lambda s_\gamma}{\sqrt{\min\{m,m'\}}}$ *.*

The result of Theorem 5.2 delivers some additional information about the sensitivity of the proposed procedure to change-points. One possible question is about the minimal delay m' between the change-point $T_{\rm cp}$ and the first moment τ when the procedure starts to indicate this change-point by selecting an interval of type $I\!I = [T_{\rm cp}, \tau[$. Due to Theorem 5.2, the change will be certainly 'detected' if the value $\delta = \lambda s_\gamma/\sqrt{m'}$ fulfills (5.1). With fixed $b > 0$, λ and μ, condition (5.1) leads to $\delta \le bC_0$ where C_0 depends on μ/λ only. The latter condition can be rewritten in the form

$$m' \ge \frac{b^{-2}\lambda^2 s_\gamma^2}{C_0^2}.$$

We see that the required delay m' depends quadratically on the change-point magnitude b and on the threshold λ. In its turn, for the prescribed type I error α of rejecting a homogeneous interval of length M, the threshold λ can be bounded by $C\sqrt{\log\frac{M}{m_0\alpha}}$. In particular, if we fix the length M and α, then $m' = O(b^{-2})$. If we keep fixed the values b and M but aim to provide a very small probability of a 'false alarm' by letting α go to zero, then $m' = O(\log\alpha^{-1})$. All these issues are in agreement with the theory of change-point detection, see, e.g. Pollak (1985) and Brodskij & Darkhovskij (1993).

6 Monte Carlo simulation

The aim of this section is to illustrate the performance of the proposed procedure on some simulated examples and to give some hints concerning the choice of the parameters λ, μ, m_0 and M. We first consider the simplest homogeneous model and study the stability of the procedure in such a situation. Then a change point model with two jumps is analyzed and the sensitivity with respect to the jump magnitude is measured. Finally we apply the procedure to a stochastic volatility model.

6.1 Type I error

It has already been mentioned that a reasonable approach for selecting m_0, λ, μ is by providing a prescribed level α for rejecting a homogeneous interval I of a given length M. This would clearly imply at most the same level α for rejecting a homogeneous interval of a smaller length. This can be made on the base of Theorem 5.1. However, the resulting upper bound for the error probability of the type I is rather conservative. More accurate choice of the parameters m_0, λ and μ can be made on the base of Monte Carlo simulation for the time homogeneous model. We examine the procedure described in Section 3.2 with the sets of intervals $\mathcal{I}$ and $\mathcal{J}(I)$ on the regular grid with the fixed step m_0. The time homogeneous model assumes that the parameter θ_t does not vary in time, i.e. $\theta_t \equiv \theta$. It can easily be seen that the value θ has no influence on the procedure under time homogeneity. One can therefore suppose that $\theta = 1$ and the original model (2.1) is transformed into the regression model $Y_t = 1 + s_\gamma \zeta_t$ with the constant trend and homogeneous variance s_γ. This model is completely described and therefore, one can define $r_1(m_0, \lambda, \mu)$ as the probability for this model to reject a homogeneous interval of length M if the parameters m_0, λ, μ are applied. To determine this parameters we simulate two set of white noise time series (i.i.d $N(0,1)$) of length $M = 40, 80$ respectively, each set containing 1000 realizations. We consider three testing steps: $m_0 = 5, 10, 20$ and for each testing step we compute the values of λ and μ for which the time homogeneous interval is not rejected with an "approximate" frequency of 95%.

The relationship among M, m_0, λ and μ can be summarized as follows.

- $M = 40, 80$ represents the length of the true time homogeneous interval, which is obviously known only for simulated data. For fixed m_0, λ and μ the probability of rejecting the true hypothesis of time homogeneity grows with M.

- $m_0 = 5, 10, 20$ is the grid step. If m_0 is small the test of homogeneity is performed very often and therefore, for fixed M, λ and μ the probability of rejecting the true hypothesis of time homogeneity decreases with increasing m_0.

- (λ, μ) are the smoothing parameters. For fixed M and m_0 the probability of rejecting the true hypothesis of time homogeneity decreases with increasing λ and/or μ. Furthermore if we also fix the rejection probability μ is decreasing in λ. Monte Carlo simulations displayed in Figure 3 show an approximately linear relationship between these two quantities.

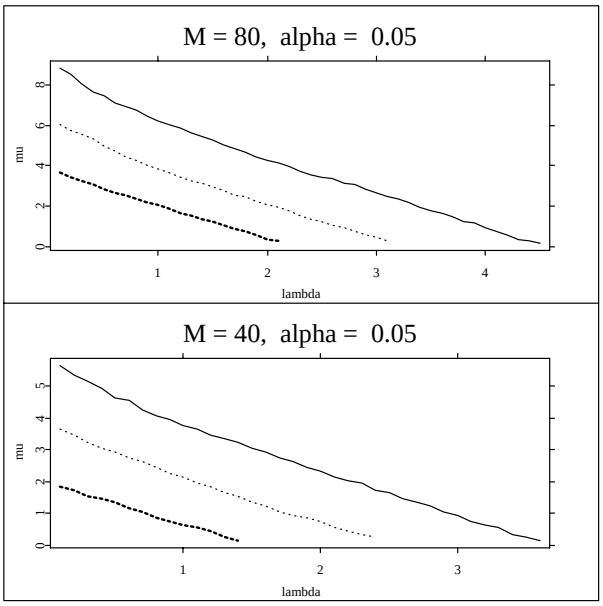


Figure 3: Pairs (λ, μ) which provide a type I error probability of 0.05 for different grid step: $m_0 = 5$ (solid line), $m_0 = 10$ (thin dotted line) and $m_0 = 20$ (thick dotted line).

6.2 Type II Error

To evaluate the type II error we consider the sets (M, m_0, λ, μ) which keep the frequency of the type I error at a 5%. Such sets guarantee that if the true interval of time homogeneity is M, then the algorithm will keep the whole interval on the average 95 times every 100. We evaluate the performances of the smoothing parameters with respect to different time inhomogeneous process. In particular we consider two deterministic piecewise constant processes with small, and large jumps respectively, and a stochastic volatility process.

Two jump processes of length $T = 240$ are considered with two jumps of the same magnitude in opposite directions, i.e.: $\sigma_t = \sigma$ for $t \in [1, 80]$ and $t \in [161, 240]$ and and $\sigma_t = \sigma'$ for $t \in [81, 160]$. Where $\sigma = 1$ and $\sigma' = 3$ and 5 respectively. For each model 500 realizations are generated, the estimation is performed at each time point $t \in [t_0, 240]$, where t_0 is set equal to 20.

The above model fulfills very well the assumptions of local homogeneity upon which our procedure relies. It is therefore interesting to see how the algorithm perform on a process, where a local constant estimation may not appear as the best strategy. The following example illustrates how the estimator behaves in a "worst case". We consider a stochastic volatility model, i.e. a white noise where the logs of the standard deviation follow an

AR(1) process:

$$\begin{aligned} R_t &= \sigma_t \xi_t \\ \log \sigma_t &= \omega + \phi \ln \sigma_{t-1} + \sigma_\zeta \zeta_t. \end{aligned}$$

The following parameters are chosen: $\omega = 0$, $\phi = 0.987$, and $\sigma_\zeta = 0.1$, ξ_t is standard normally distributed.

The performances of the different parameter sets are compared for two different criteria: averaged quadratic risk (MSE) and average absolute deviation risk (MAE), and also on their empirical counterparts based on forecast error (FE): mean squared forecast error (MSFE) and mean absolute forecast error (MAFE):

$$\begin{aligned} \text{MSE} &= \boldsymbol{E}^* \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} (\widehat{\theta}_t - \theta_t)^2, & \text{MSFE} &= \boldsymbol{E}^* \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} (\widehat{\theta}_t - Y_t)^2, \\ \text{MAE} &= \boldsymbol{E}^* \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} |\widehat{\theta}_t - \theta_t|, & \text{MAFE} &= \boldsymbol{E}^* \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} |\widehat{\theta}_t - Y_t|. \end{aligned}$$

Table 1 shows the selected parameters for both the theoretical and empirical criteria. Note that the choice among all criteria is not qualitatively different. Another interesting result is shown in Figure 4, where the result of the theoretical error measures are plotted against the empirical ones. A high positive correlation is observable. It must also be noted that the most conservative parameter sets, i.e. for $M = 80$, are never chosen. Furthermore the parameter sets with the smallest grid step $m_0 = 5$ are also never chosen. This result is very useful for practical application, because it allows us to restrict our attention to the values of λ and μ, for which $M = 40$, and $m_0 = 10$ and 20.

Table 1: The optimal results for all error measures for the three models.

M	m_0	λ	μ	MSE	MAE	MSFE	MAFE
jump process with ($\sigma = 1$, $\sigma' = 3$)							
40	10	0.4	3.05	0.035169*	0.1152*	0.24703*	0.38978
40	20	0.6	1.15	0.035345	0.11667	0.24706	0.38974*
jump process with ($\sigma = 1$, $\sigma' = 5$)							
40	10	2.2	0.45	0.06841*	0.13988	0.37332*	0.45954*
40	10	1.8	0.95	0.068481	0.13957*	0.37351	0.45966
stochastic volatility							
40	20	0.8	0.05	0.023386*	0.118*	0.14332*	0.29875*

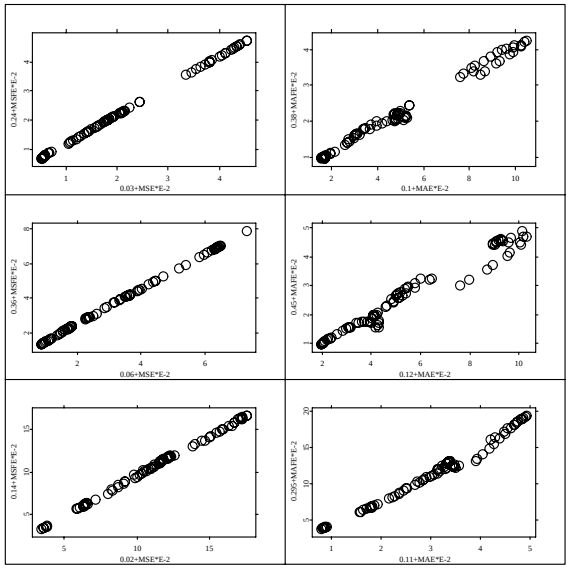


Figure 4: Scatterplot of the theoretical error measures against the empirical ones for the three models. Squared errors on the left, absolute errors on the right. From the top: ($\sigma = 1$, $\sigma' = 3$), ($\sigma = 1$, $\sigma' = 5$), SV .

Figure 5 shows the result of the estimation for the parameters selected with the MSFE. The first, third and fifth plots from the top indicate: the true process (straight line), the median among all estimates (dashed line) and the quartiles of the estimates (dotted line). The second, fourth and sixth plots from the top show the "ideal" interval of time homogeneity (straight line), the median among all estimates (dashed line) and the quartiles of the estimates (dotted line).

The results are very satisfactory in particular for the jump processes. As far as the SV process is concerned the plot shows some oversmoothing of the underlying volatility process. This can be explained by the fact that a local constant approximation is good, in this case, only for very small intervals. Nevertheless the behavior of the procedure under such very unfavorable conditions remain quite remarkable.

7 Empirical Evidence

We apply the local constant volatility estimation to a set of nine exchange rates, which are available from the web page of the Federal Reserve. The data sets represent the daily exchange rate of the US$ against the following currencies: Canadian$, Norwegian Krone, Swedish Krone, New Zeeland$, Japanese Yen, Danish Krone, Swiss Franc, Australian$

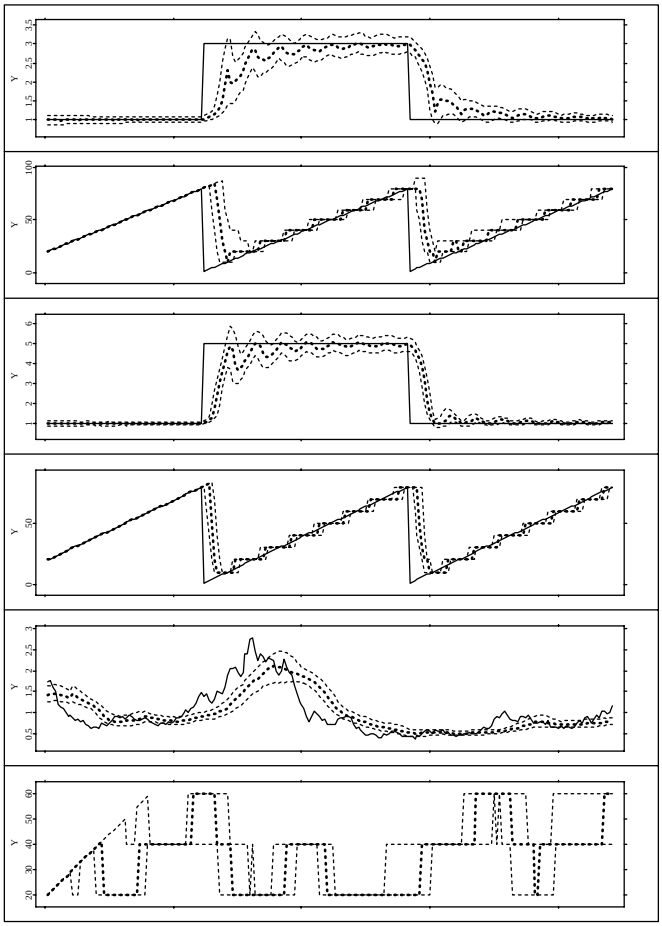


Figure 5: Estimation of the jump and SV processes and relative interval of homogeneity. From the top: ($\sigma = 1$, $\sigma' = 3$), ($\sigma = 1$, $\sigma' = 5$), SV. The smoothing parameters are the ones that minimize the MSFE.

and British Pound. The period under consideration goes from 1 January 1990 to 7 April 2000. The summary statistics of the exchange rate returns are shown in Table 2. All the data sets display a mean value which is very close to zero and an excess kurtosis. Furthermore the volatility clustering effect can be appreciated from Figure 7, while the first column of Figure 8 shows the typical persistent autocorrelation of the absolute returns.

7.1 Practical implementation of the local constant estimator

In Section 6 we have considered the question of selecting good smoothing parameters. Simulation results have shown that different kinds of processes may lead to different optimal smoothing parameters, depending on the magnitude and on the frequency of the

Table 2: Summary statistics.

currency	n	mean$\cdot 10^5$	variance$\cdot 10^5$	skewness	kurtosis
Canadian \$	2583	8.819	0.895	0.042	5.499
Norwegian Kr.	2583	9.493	4.251	0.313	8.630
Swedish Kr.	2583	12.660	4.615	0.372	9.660
New Zeeland \$	2583	-6.581	3.604	-0.356	49.178
Japanese Yen	2583	-12.700	5.486	-0.585	7.366
Danish Kr.	2583	6.097	4.201	-0.037	4.967
Swiss Fr.	2583	1.480	5.402	-0.186	4.526
Australian \$	2583	-10.410	3.191	-0.187	8.854
British Pd.	2583	-0.679	3.530	-0.279	5.792

jumps. Nevertheless the result of Section 6.2 allow us to exclude those parameters which make the procedure too much conservative, i.e. the case of $M = 80$, and those parameters which make the testing grid too dense, i.e. the case of $m_0 = 5$. All the combination of λ and μ, for which $M = 40$ and $m_0 = 5$ and 10 still remain. Nevertheless simulation results suggest that small changes in the parameters do not affect too much the performance of the estimator. Therefore we decide to keep a selection of the remaining smoothing parameters (Table 3), which should contain, according to the simulation result, a "good case" for any process similar to the ones that we have analyzed.

Table 3: Set of smoothing parameters which are used to perform the estimation.

M	m_0	λ	μ	M	m_0	λ	μ
40	10	0.2	3.45	40	10	2.2	0.45
40	10	0.6	2.75	40	20	0.2	1.75
40	10	1.0	2.15	40	20	0.6	1.15
40	10	1.4	1.55	40	20	1.0	0.65
40	10	1.0	0.95	40	20	1.4	0.15

For the parameter selection we apply the MSFE rule described in Section 3.4 with a moving window for the optimization of the smoothness parameters λ, μ. The MSFE has shown in the simulation a high correlation with the MSE (Figure 4), and a performance which is qualitative identical to the one of the MAE and MAFE.

7.2 A benchmark model

As a matter of comparison we also consider a model which is commonly used to estimate and forecast volatility processes: the GARCH(1,1) model, which was first proposed by Bollerslev (1995). Among all parametric volatility models it represents the most common specification: *"The GARCH(1,1) is the leading generic model for almost all asset classes of returns. ... it is quite robust and does most of the work in almost all cases."* (Engle 1995b).

$$\begin{aligned} R_t &= \sigma_t \xi_t \\ \sigma_t^2 &= \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

We do not require the parameters to be constant through the whole sample, but similarly to Franses & Dijk (1996) we consider a rolling estimator. We thus fit the model to a sample of 350 observations, generate the forecast, delete the first observation from the sample and add the next one. Such a procedure reduce the harmful effect of possible parameter shifts on the forecasting performances of the model, even if at the same time it may increase the estimation variability.

7.3 Forecast evaluation

The volatility is a hidden process which can be observed only together with a multiplicative error, therefore the evaluation of the forecasting performance of an algorithm is not straightforward. Due to the model (2.1), it holds $\boldsymbol{E}\left(R_{t+1}^2 \mid \mathcal{F}_t\right) = \sigma_{t+1}^2$ and $\boldsymbol{E}\left(|R_{t+1}| \mid \mathcal{F}_t\right) = c\sigma_{t+1}$ with $c = \sqrt{2/\pi}$. Therefore, given a forecast $\sigma_{t+1|t}$, the empirical mean value of the $|R_{t+1}^2 - \sigma_{t+1|t}^2|$'s resp. of the $\left(|R_{t+1}| - c\sigma_{t+1|t}\right)^2$'s can be used to measure the quality of the forecast $\sigma_{t+i|t}$. Similarly one can consider a larger forecasting horizon. The following different measures are computed in order to get a feeling of the robustness of the technique:

$$\begin{aligned} \mathrm{d}_1 &= \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} \left(|R_{t+i}| - c\sigma_{t+i|t}\right)^2, \\ \mathrm{d}_2 &= \frac{1}{T - t_0 - 1} \sum_{t=t_0}^{T} \left|R_{t+i}^2 - \sigma_{t+i|t}^2\right|. \end{aligned}$$

with $i = 1$ resp $i = 5$ corresponding to one step ahead resp. five step ahead forecast horizon. The relative performance of the local constant model and of the GARCH(1,1) model is displayed in Table 5.

7.4 Estimation results

The results of the estimation are quite satisfactory. Figure 7 shows two typical exchange return time series: Canadian$/US$ and Japanese Yen/US$, together with the local constant estimation of the volatility process, and of the interval of time homogeneity. The estimated standard deviation is nicely in accordance with the development of the volatility and in particular sharp changes in the volatility tend to be very quickly recognized. Note also that the variance of the estimated interval of time homogeneity appears to grow as the estimated interval becomes larger. This is a feature of the algorithm because the number of tests grows with the accepted interval, so that a rejection becomes more probable. Nevertheless this variability does not affect strongly the estimated volatility coefficient.

Figure 8 shows the first one hundred values of the autocorrelation of the absolute returns for the nine exchange rates in the first column. The second and the third columns show the autocorrelation of the absolute returns divided by the one step forecasted standard deviation estimated respectively with the local constant and with the GARCH $(1,1)$ model. The dashed line indicated the 5-percent level of significance. Both models seems to explain equally well the dynamics of the volatility. Particularly no significant positive correlation for the standardized returns is observed.

Both the local constant model and the GARCH model assume conditional normality of the returns. It is therefore interesting to test whether this hypothesis is supported by the empirical evidence. Figure 6 shows an example of the empirical densities for both models, plotted against a standard normal density. It can be seen that the three curves are mostly very close to each other and that the difference with the normal becomes larger as far as the tail behavior and the region about zero are concerned. One reasonable explanation of this fact is that both GARCH and local constant modelling fail to predict extreme events which typically occur in financial time series. A possible testing procedure compares the third and fourth empirical moments with skewness and kurtosis of the standard normal distribution. This approach mostly focuses on the tails of the empirical density and it usually leads to rejecting the hypothesis of conditional normality. We applied here the Kolmogorov/Smirnov test, which is based on the uniform distance between the empirical distribution and the theoretical one. The test statistic is:

$$KS_n = \sup_x |W_n(x) - \Phi(x)|,$$

where $W_n(x)$ is the empirical distribution and $\Phi(x)$ the standard normal distribution. The critical values (Bronstein & Semendjajew 1991) for an $\alpha = 0.05$ are between 1.35 and 1.36 and for an $\alpha = 0.01$ are between 1.62 and 1.63. Table 4 presents the results of

the test statistic for the nine exchange rates for both the local constant and the GARCH model. We remark that the returns are standardized by the one-step forecasted standard deviation. It appears that the local constant model at least partially succeeds in normalizing the returns, while the hypothesis of normality is always rejected for the GARCH model at high significance level. As far as the forecasting efficiency is concerned (Table 5), the GARCH $(1,1)$ is slightly outperformed by the local constant model. Both for the 1-step-ahead and 5-step-ahead forecast horizon the local constant model is preferred by all exchange rates for d_2-distance, while six exchange rates out of nine prefer the local constant model for d_1-distance.

Table 4: Results of the Kolmogorov/Smirnov test. (*) and (**) indicate the rejection of the hypothesis of normality at a 5%, 1% significance level respectively.

Currency	Loc. Const.	GARCH
Canadian \$	1.3990^{*}	1.7352^{**}
Norwegian Kr.	1.1014	1.7271^{**}
Swedish Kr.	1.1377	2.1376^{**}
New Zeeland \$	1.2432	2.7273^{**}
Japanese Yen	1.1577	2.1606^{**}
Danish Kr.	1.5544^{*}	1.6047^{*}
Swiss Fr.	1.2698	1.5872^{*}
Australian \$	1.3346	2.1127^{**}
British Pd.	1.3268	2.3691^{**}

Table 5: Relative forecasting performance: local constant versus GARCH modeling for different forecast horizons and different measures of the forecasting ability.

	1 step forecast		5 step forecast	
	d_1	d_2	d_1	d_2
Canadian \$	1.007	0.985	1.025	0.994
Norwegian Kr.	0.974	0.937	0.982	0.941
Swedish Kr.	0.983	0.942	0.997	0.950
New Zeeland \$	0.962	0.857	0.962	0.858
Japanese Yen	0.982	0.949	0.992	0.969
Danish Kr.	1.009	0.982	1.017	0.993
Swiss Fr.	0.990	0.979	0.992	0.984
Australian \$	0.975	0.919	0.982	0.921
British Pd.	1.010	0.955	1.029	0.968

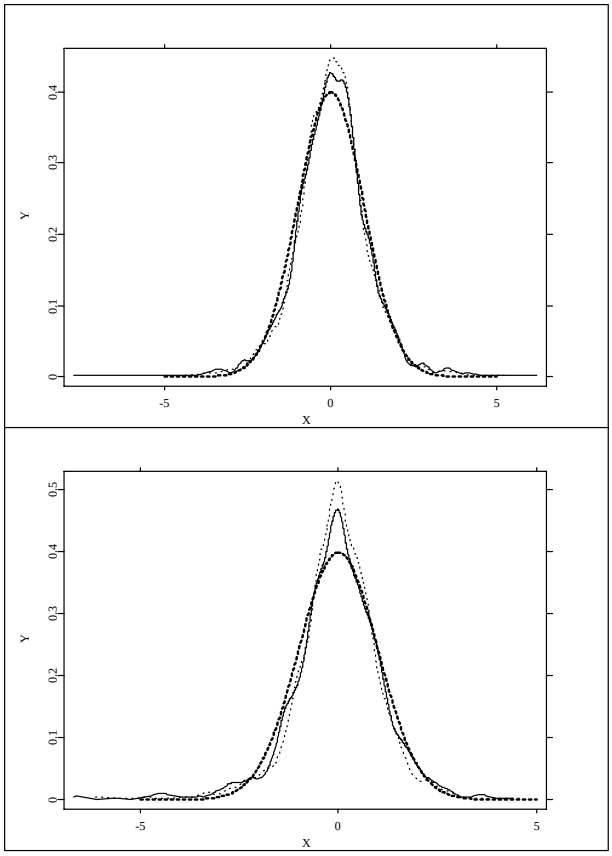


Figure 6: Empirical densities of the standardized returns of the Canadian$/US$ exchange rate (upper plot) Japanese Yen/US$ (lower plot). The three curves represent respectively: local constant model (straight line), GARCH (thin dotted line) and a standard normal density (thick dotted line).

8 Conclusions and outlook

A new algorithm for estimating and short term forecasting the volatility of financial returns is proposed. It is assumed that a local constant approximation of the volatility process holds over some unknown interval. The issue of filtering this interval of time homogeneity out of the return time series is considered, and a nonparametric approach is presented. The estimate of the volatility process is then found by averaging over the interval of time homogeneity.

The problem of selecting the smoothing parameters is analyzed through Monte Carlo simulation. First all the parameter sets are computed, which reject a true time homogeneous interval with an approximate frequency of 5%, then the set of smoothing parameters is

chosen that optimizes the forecasting performances.

Finally an empirical application to exchange rate returns and a comparison with a GARCH $(1,1)$ provides a good evidence that the new method is competitive and can even outperform the standard parametric models especially for forecasting with a short horizon.

An important feature of the proposed method is that it allows for a straightforward extension on the multivariate volatility estimation, see Härdle, Herwartz & Spokoiny (2000) for a detailed discussion.

Obviously, if the underlying conditional distribution is not normal, the estimated volatility can give only a partial information about the riskiness of the asset. Recent developments in the risk analysis tends to focus on the estimation of the quantiles of the distribution. In this direction the local constant estimator can be used as a convenient tool for pre-whitening the returns and obtain a sample of "almost" identical and independently distributed returns, which do not display any more variance clustering. So that the usual techniques of quantile estimation could be applied in a static framework. We regard such a development as a topic for future research.

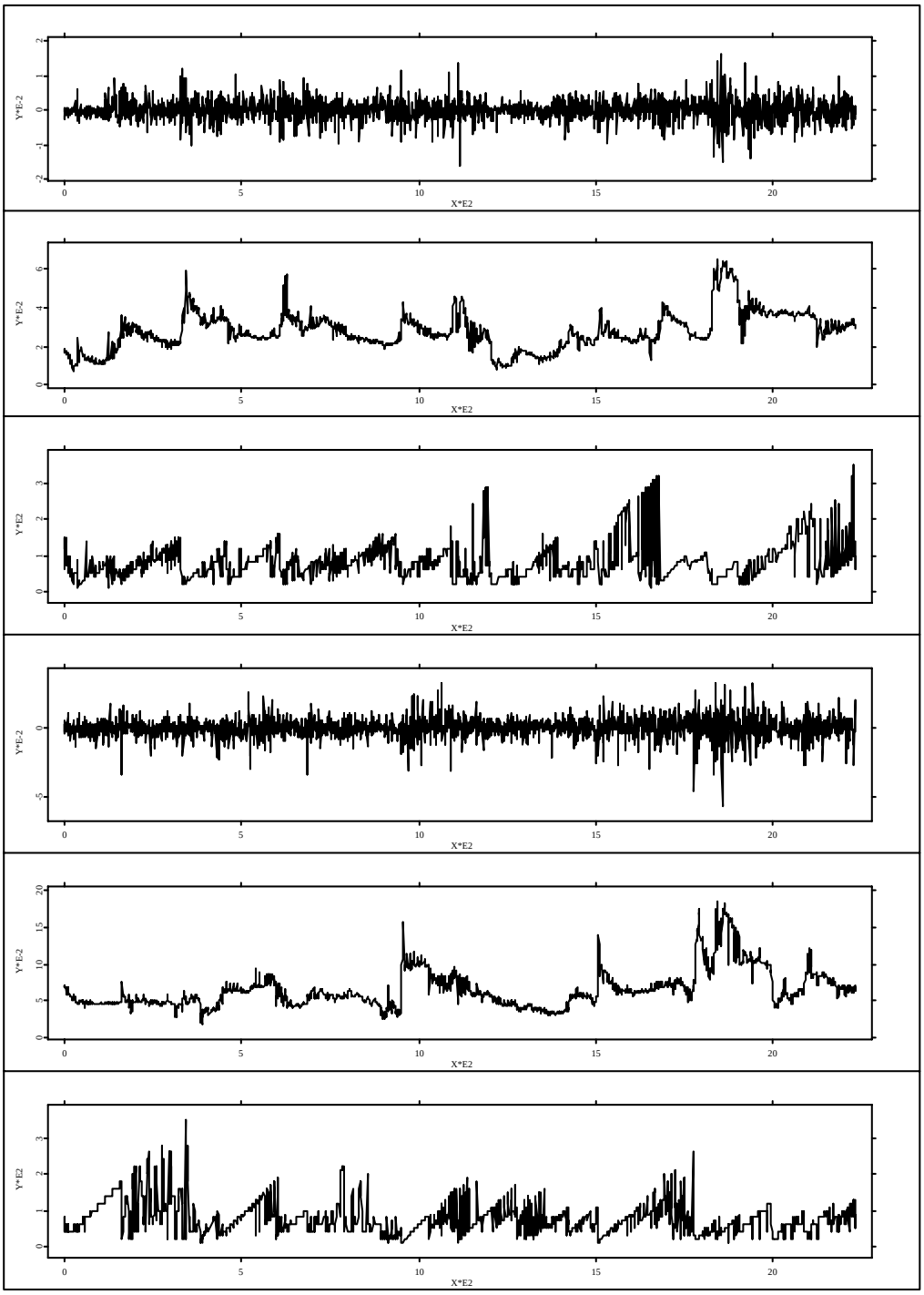


Figure 7: Exchange rate returns, estimated standard deviation and interval of homogeneity for Canadian$/US$ (the first three plots) and Japanese Yen/US$ (the last three plots).

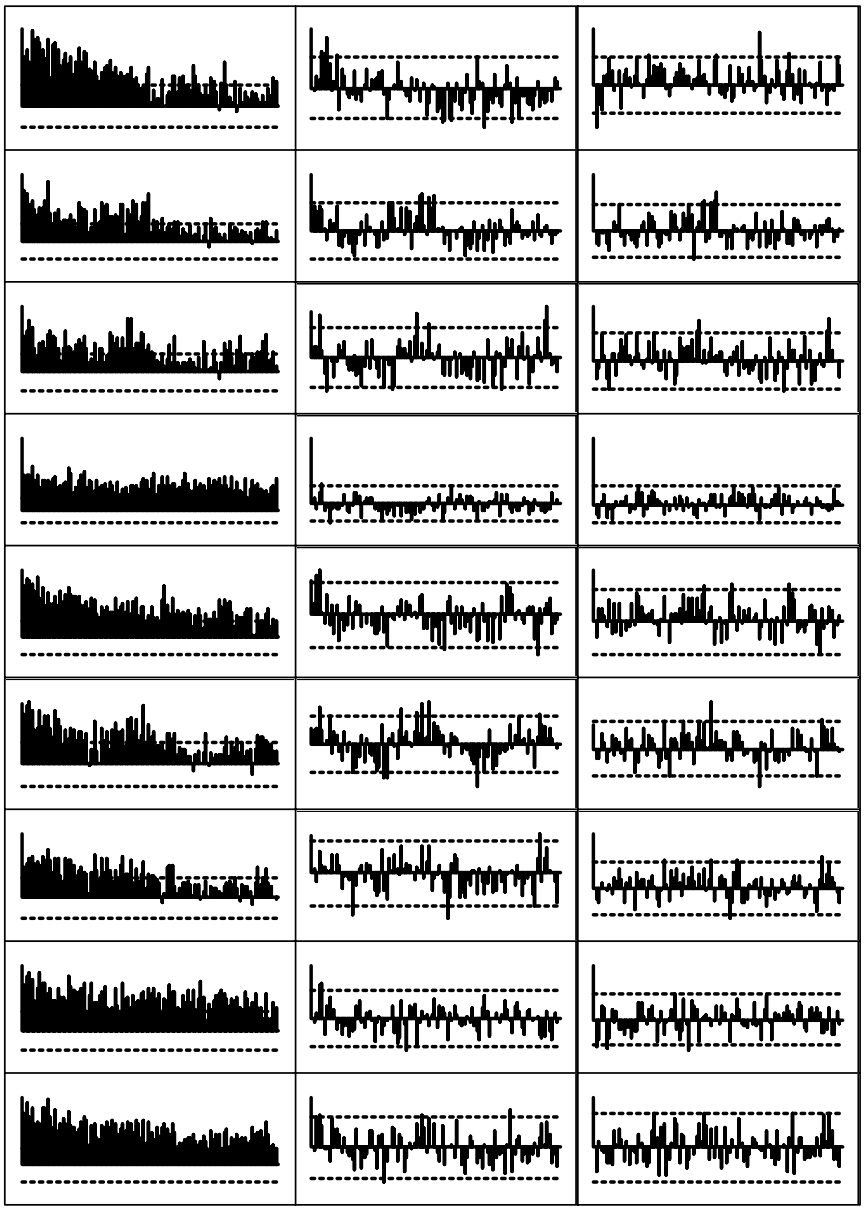


Figure 8: First one hundred lags of the autocorrelation function. From the left: ACF of the absolute exchange rate returns, ACF of the absolute returns divided by the standard deviation estimated with local constant, ACF of the absolute returns divided by the standard deviation estimated with GARCH $(1,1)$. From the top: Canadian\$, Norwegian Krone, Swedish Krone, New Zeeland\$, Japanese Yen, Danish Krone, Swiss Franc, Australian\$ and British Pound.

9 Proofs

In this section we collect the proofs of the results stated above. We begin by considering some useful properties of the power transformation introduced in Section 2.1.

Some properties of the power transformation

Let $g_\gamma(u)$ be the moment generating function of $\zeta_\gamma = D_\gamma^{-1}\left(|\xi|^\gamma - C_\gamma\right)$:

$$g_\gamma(u) = \boldsymbol{E} e^{u\zeta_\gamma}\,.$$

It is easy to see that this function is finite for $\gamma < 2$ and all u and for $\gamma = 2$ and $u < 1$. For $\gamma = 1/2$ the function $2u^{-2}\log g_\gamma(u)$ is plotted in Figure 9.

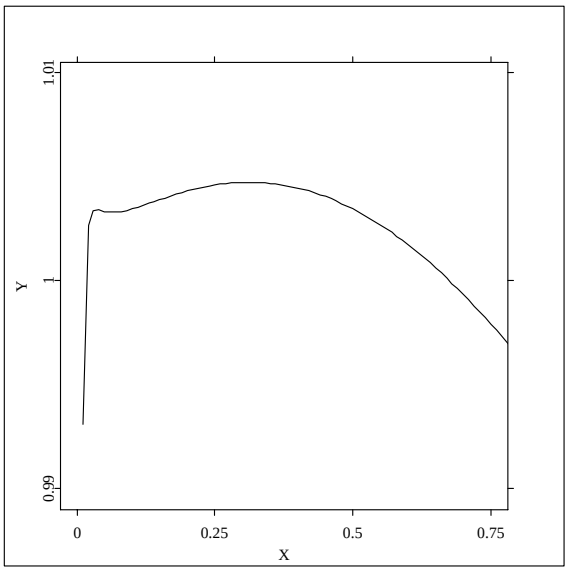


Figure 9: The log-Laplace transform of $\zeta_{1/2}$ divided by the log-Laplace transform of a standard normal r. v.

Lemma 9.1 *For every $\gamma \le 1$ there exists a constant $a_\gamma > 0$ such that*

$$\log \boldsymbol{E} e^{u\zeta_\gamma} \le \frac{a_\gamma u^2}{2}. \tag{9.1}$$

Proof. It is also easy to check that the function $g_\gamma(u)$ with $\gamma \le 1$ is positive and smooth (infinitely many times differentiable). Moreover, the function $h_\gamma(u) = \log g_\gamma(u)$ is also smooth and satisfies $h_\gamma(0) = h'_\gamma(0) = 0$, $h''_\gamma(0) = \boldsymbol{E}\zeta_\gamma^2 = 1$. This yields that

$u^{-2}h_\gamma(u) = u^{-2}\log g_\gamma(u)$ is bounded on every finite interval of the positive semiaxis $[0,\infty)$. It therefore remains to show that

$$\lim_{u\to\infty} u^{-2}\log \boldsymbol{E}e^{u\zeta_\gamma} < \infty.$$

Since $\zeta_\gamma(u) = D_\gamma^{-1}(|\xi|^\gamma - C_\gamma)$, it suffices to bound $u^{-2}\boldsymbol{E}e^{u|\xi|^\gamma/D_\gamma}$. It holds for every $t>0$

$$\begin{aligned}
\boldsymbol{E}e^{u|\xi|^\gamma D_\gamma^{-1}} &= \boldsymbol{E}e^{u|\xi|^\gamma D_\gamma^{-1}}\mathbf{1}(|\xi|\le t) + \boldsymbol{E}e^{u|\xi|^\gamma D_\gamma^{-1}}\mathbf{1}(|\xi|>t)\\
&\le e^{ut^\gamma D_\gamma^{-1}} + \boldsymbol{E}e^{u|\xi|t^{\gamma-1}D_\gamma^{-1}}\\
&\le e^{ut^\gamma D_\gamma^{-1}} + 2\boldsymbol{E}e^{u\xi t^{\gamma-1}D_\gamma^{-1}}\\
&= e^{ut^\gamma D_\gamma^{-1}} + 2e^{u^2t^{2\gamma-2}D_\gamma^{-2}}.
\end{aligned}$$

Next, with $t = u^{1/(2\gamma)}$ and $\gamma<1$, it holds for $u\to\infty$:

$$\begin{aligned}
u^{-2}\log e^{ut^\gamma D_\gamma^{-1}} &= u^{-1/2}D_\gamma^{-1}\to 0,\\
u^{-2}\log e^{u^2t^{2\gamma-2}D_\gamma^{-2}} &= u^{-(1-\gamma)/\gamma}D_\gamma^{-2}\to 0.
\end{aligned}$$

For $\gamma=1$, the last expression remains bounded and the assertion follows. ∎

For $\gamma=1/2$, condition (9.1) meets with $a_\gamma = 1.005$.

The next technical statement is a direct consequence of Lemma 9.1.

Lemma 9.2 *Let c_t be a predictable process w.r.t. the filtration $\mathcal{F}=(\mathcal{F}_t)$, i.e. every c_t is a function of previous observations $R_1,\dots,R_{t-1}$: $c_t = c_t(R_1,\dots,R_{t-1})$. Then the process*

$$\mathcal{E}_t = \exp\left(\sum_{s=1}^t c_s\zeta_s - \frac{a_\gamma}{2}\sum_{s=1}^t c_s^2\right)$$

is a supermartingale, that is,

$$\boldsymbol{E}\left(\mathcal{E}_t \mid \mathcal{F}_{t-1}\right) \le \mathcal{E}_{t-1}. \tag{9.2}$$

The next result has been stated in Lipster & Spokoiny (1999) for Gaussian martingales, however, the proof is based only on the property (9.2) and allows for a straightforward extension to the sums of the form $M_t = \sum_{s=1}^t c_s\zeta_s$.

Theorem 9.1 *Let $M_t = \sum_{s=1}^t c_s\zeta_s$ with predictable coefficients c_s. Let then T be fixed or a stopping time. For every $b>0$, $B\ge 1$ and $\lambda\ge 1$*

$$\boldsymbol{P}\left(|M_T| > \lambda\sqrt{\langle M\rangle_T},\ b\le\sqrt{\langle M\rangle_T}\le bB\right) \le 4\sqrt{e}\lambda\,(1+\log B)\,e^{-\frac{\lambda^2}{2a_\gamma}}$$

where

$$\langle M\rangle_T = \sum_{t=1}^{T} c_t^2.$$

Remark 9.1 If the coefficients c_t are deterministic then the quadratic characteristic $\langle M\rangle_T$ is also deterministic, and one derives directly from Lemma 9.1 using the Tschebyshev inequality:

$$\boldsymbol{P}\left(|M_T| > \lambda\sqrt{\langle M\rangle_T}\,\right) \le 2e^{-\frac{\lambda^2}{2a_\gamma}}.$$

Proof of Theorem 3.1

Due to (3.3) the bias $|\boldsymbol{E}\widetilde{\theta}_I - \theta_\tau|$ of the estimate $\widetilde{\theta}_I$ is bounded by Δ_I. Define

$$\bar{\theta}_I = \frac{1}{|I|}\sum_{t\in I}\theta_t.$$

Then by the Cauchy-Schwarz inequality

$$|\bar{\theta}_I - \theta_\tau| = |I|^{-1}\left|\sum_{t\in I}(\theta_t - \theta_\tau)\right| \le \left\{|I|^{-1}\sum_{t\in I}(\theta_t - \theta_\tau)^2\right\}^{1/2} \le \Delta_I \tag{9.3}$$

and, since $\bar{\theta}_I$ is the arithmetic mean of θ_t over I,

$$\sum_{t\in I}(\theta_t - \bar{\theta}_I)^2 \le \sum_{t\in I}(\theta_t - \theta_\tau)^2 \le |I|\Delta_I^2.$$

This yields

$$\sum_{t\in I}\theta_t^2 = |I|\bar{\theta}_I^2 + \sum_{t\in I}(\theta_t - \bar{\theta}_I)^2 \le |I|\left(\bar{\theta}_I^2 + \Delta_I^2\right) \le |I|\left(\bar{\theta}_I + \Delta_I\right)^2. \tag{9.4}$$

Next, by (3.2)

$$\widetilde{\theta}_I - \theta_\tau = |I|^{-1}\sum_{t\in I}(\theta_t - \theta_\tau) + s_\gamma|I|^{-1}\sum_{t\in I}\theta_t\zeta_t$$

and the use of (9.3) yields

$$\boldsymbol{P}\left(|\widetilde{\theta}_I - \theta_\tau| > \Delta_I + \lambda v_I\right) \le \boldsymbol{P}\left(\left|\sum_{t\in I}\theta_t\zeta_t\right| > \lambda\left(\sum_{t\in I}\theta_t^2\right)^{1/2}\right).$$

In addition, if the volatility coefficient σ_t satisfies $b \le \sigma_t^2 \le bB$ with some positive constant b, B, then the conditional variance $v_I^2 = s_\gamma^2|I|^{-2}\sum_{t\in I}\theta_t^2$ fulfills

$$b'|I|^{-1} \le v_I^2 \le b'|I|^{-1}B$$

with $b' = bs_\gamma^2$. Now the assertion follows from (3.5) and Theorem 9.1.

Proof of Theorem 3.2

Clearly

$$|\widetilde{\theta}_I - \theta_\tau| \le |\widetilde{\theta}_I - \bar{\theta}_I| + |\bar{\theta}_I - \theta_\tau| \le \Delta_I + |\widetilde{\theta}_I - \bar{\theta}_I|$$

and hence,

$$\begin{aligned}
&\boldsymbol{P}\left(|\widetilde{\theta}_I - \theta_\tau| > \Delta_I + \lambda s_\gamma(\widetilde{\theta}_I + \Delta_I)|I|^{-1/2}\right) \\
&\quad \le \boldsymbol{P}\left(|\widetilde{\theta}_I - \bar{\theta}_I| > \lambda s_\gamma(\bar{\theta}_I - |\widetilde{\theta}_I - \bar{\theta}_I| + \Delta_I)|I|^{-1/2}\right) \\
&\quad \le \boldsymbol{P}\left(|\widetilde{\theta}_I - \bar{\theta}_I| > \frac{\lambda s_\gamma}{1 + \lambda s_\gamma |I|^{-1/2}}(\bar{\theta}_I + \Delta_I)|I|^{-1/2}\right).
\end{aligned}$$

By (2.3)

$$\widetilde{\theta}_I - \bar{\theta}_\tau = |I|^{-1}\sum_{t\in I}(|R_t|^\gamma - \theta_t) = |I|^{-1}s_\gamma \sum_{t\in I}\theta_t\zeta_t$$

and the use of (9.4) implies

$$\begin{aligned}
&\boldsymbol{P}\left(|\widetilde{\theta}_I - \bar{\theta}_I| > \frac{\lambda s_\gamma}{1 + \lambda s_\gamma |I|^{-1/2}}(\bar{\theta}_I + \Delta_I)|I|^{-1/2}\right) \\
&\le \boldsymbol{P}\left(\left|\sum_{t\in I}\theta_t\zeta_t\right| > \frac{\lambda}{1 + \lambda s_\gamma |I|^{-1/2}}\left(\sum_{t\in I}\theta_t^2\right)^{1/2}\right).
\end{aligned}$$

Now the desirable result follows directly from Theorem 9.1.

Proof of Theorem 4.1

Let $I\!I$ be the "ideal" interval from (4.1). We intend to show that

$$\left\{\left|\widehat{\theta} - \theta_\tau\right| > 2(\lambda + \mu)v_{I\!I}\right\} \subseteq \bigcup_{J\in\mathcal{J}(I\!I)}\left\{|\widetilde{\theta}_J - \bar{\theta}_J| > \lambda_J v_J\right\}$$

which would imply the assertion in view of Theorem 9.1, cf. the proof of Theorem 3.1. This statement is equivalent to saying that the inequality $|\widehat{\theta} - \theta_\tau| > 2(\lambda + \mu)v_{I\!I}$ is impossible if

$$|\widetilde{\theta}_J - \bar{\theta}_J| \le \lambda_J v_J\,, \qquad \forall J \in \mathcal{J}(I\!I). \tag{9.5}$$

Obviously

$$\left\{|\widehat{\theta} - \theta_\tau| > 2(\lambda + \mu)v_{I\!I}\right\} \subseteq \left\{|\widehat{\theta} - \theta_\tau| > 2(\lambda + \mu)v_{I\!I}\,,\ I\!I \subseteq \widehat{I}\right\} + \{I\!I \text{ is rejected}\}\,.$$

We consider separately each event in the right side of this inequality using the following

Lemma 9.3 *Let (9.5) hold true. Then, for every* $\forall J \in \mathcal{J}(I\!I)$,

$$\begin{aligned} \widetilde{v}_J &\geq \frac{v_J}{1+\lambda s_\gamma |J|^{-1/2}}, \\ \widetilde{v}_J &\leq v_J \left(2 - \frac{1}{1+\lambda s_\gamma |J|^{-1/2}}\right) \end{aligned}$$

Proof. Define $\theta'_J = \left(|J|^{-1}\sum_{t\in J}\theta_t^2\right)^{1/2}$. Then $v_J = s_\gamma |J|^{-1/2}\theta'_J$ and $\widetilde{v}_J = s_\gamma |J|^{-1/2}\widetilde{\theta}_J$. The definition of Δ_I implies

$$|\theta'_J - \bar{\theta}_J| = \left(\bar{\theta}_J^2 + \frac{1}{|J|}\sum_{t\in J}(\theta_t - \bar{\theta}_J)^2\right)^{1/2} - \bar{\theta}_J \leq \left(\bar{\theta}_J^2 + \Delta_J^2\right)^{1/2} - \bar{\theta}_J \leq \Delta_J .$$

Along with (9.5) this implies

$$\begin{aligned} \widetilde{v}_J &= s_\gamma |J|^{-1/2}\widetilde{\theta}_J \\ &\geq s_\gamma |J|^{-1/2}\left(\theta'_J - |\widetilde{\theta}_J - \bar{\theta}_J| - |\theta'_J - \bar{\theta}_J|\right) \\ &\geq v_J - s_\gamma |J|^{-1/2}(\lambda_J v_J + \Delta_J) \\ &= v_J \left(1 - \frac{\lambda s_\gamma |J|^{-1/2}}{1+\lambda s_\gamma |J|^{-1/2}}\right) \end{aligned}$$

and the first assertion of the lemma follows. The second one is proved similarly. ∎

It holds on the event $\{I\!I \subseteq \widehat{I}\}$ in view of the definition of $\widehat{I}$

$$|\widetilde{\theta}_{\widehat{I}} - \widetilde{\theta}_{I\!I}| \leq \lambda \widetilde{v}_{I\!I} + \mu \widetilde{v}_{\widehat{I}} \leq (\lambda+\mu)\widetilde{v}_{I\!I}$$

and by Lemma 9.3

$$|\widetilde{\theta}_{\widehat{I}} - \widetilde{\theta}_{I\!I}| \leq (\lambda+\mu) v_{I\!I} \left(2 - \frac{1}{1+\lambda s_\gamma |I\!I|^{-1/2}}\right)$$

Next, by (9.5)

$$\begin{aligned} |\widetilde{\theta}_{I\!I} - \theta_\tau| &\leq |\widetilde{\theta}_{I\!I} - \bar{\theta}_{I\!I}| + |\bar{\theta}_{I\!I} - \theta_\tau| \leq |\widetilde{\theta}_{I\!I} - \bar{\theta}_{I\!I}| + \Delta_{I\!I} \\ &\leq \lambda_{I\!I} v_{I\!I} + \Delta_{I\!I} = \frac{\lambda v_{I\!I}}{1+\lambda s_\gamma |I\!I|^{-1/2}}. \end{aligned}$$

Hence, $\{I\!I \subseteq \widehat{I}\}$ implies

$$\begin{aligned} |\widehat{\theta} - \theta_\tau| &\leq |\widetilde{\theta}_{\widehat{I}} - \widetilde{\theta}_{I\!I}| + |\widetilde{\theta}_{I\!I} - \theta_\tau| \\ &\leq 2\lambda v_{I\!I} + \mu v_{I\!I}\left(2 - \frac{1}{1+\lambda s_\gamma |I\!I|^{-1/2}}\right) \\ &\leq 2(\lambda+\mu) v_{I\!I} . \end{aligned}$$

Now we study the event $\{I\!I \text{ is rejected}\}$. By definition

$$\{I\!I \text{ is rejected}\} = \bigcup_{I \in \mathcal{J}(I\!I)} \bigcup_{J \in \mathcal{J}(I)} \left\{ |\widetilde{\theta}_I - \widetilde{\theta}_J| > \mu \widetilde{v}_I + \lambda \widetilde{v}_J \right\}.$$

Since $|\bar{\theta}_J - \theta_\tau| \le \Delta_J = \delta_J v_J$ for all $J \in \mathcal{J}(I\!I)$, condition (9.5) yields for every pair $J \subset I \in \mathcal{J}(I\!I)$

$$\begin{aligned} |\widetilde{\theta}_I - \widetilde{\theta}_J| &\le |\widetilde{\theta}_I - \bar{\theta}_I| + |\bar{\theta}_I - \theta_\tau| + |\widetilde{\theta}_J - \bar{\theta}_J| + |\bar{\theta}_J - \theta_\tau| \\ &\le (\lambda_I + \delta_I) v_I + (\lambda_J + \delta_J) v_J \\ &= \frac{\lambda v_I}{1 + \lambda s_\gamma |I|^{-1/2}} + \frac{\lambda v_J}{1 + \lambda s_\gamma |J|^{-1/2}}. \end{aligned}$$

By Lemma 9.3

$$\mu \widetilde{v}_I + \lambda \widetilde{v}_J \ge \frac{\mu v_I}{1 + \lambda s_\gamma |I|^{-1/2}} + \frac{\lambda v_J}{1 + \lambda s_\gamma |J|^{-1/2}}$$

so that the event $\{I\!I \text{ is rejected}\}$ is impossible under (9.5) in view of $\mu \ge \lambda$.

Proof of Theorem 5.2

To simplify the exposition, we suppose that $\theta = 1$. (This does not restrict generality since one can always normalize each 'observation' Y_t by θ.) We also suppose that $\theta' > 1$ and $b = 2(\theta' - 1)$. (The case when $\theta' < \theta$ can be considered similarly.) Finally we assume that $m' = m$ (One can easily see that this case is the most difficult one.) Under the change-point model, the 'observations' $Y_t = |R_t|^\gamma$ are independent for all t and identically distributed within each interval of homogeneity. In particular, it holds for $\widetilde{\theta}_J$ with $J = [T_{\mathrm{cp}} - m, T_{\mathrm{cp}})$:

$$\widetilde{\theta}_J = \frac{1}{m} \sum_{t \in J} Y_t = 1 + \frac{s_\gamma}{\sqrt{m}} \xi,$$

with $\xi = m^{-1/2} \sum_{t \in J} \zeta_t$. Similarly, for $I = [\tau - 2m, \tau[$,

$$\begin{aligned} \widetilde{\theta}_I &= \frac{1}{2m} \sum_{t \in I} Y_t = \frac{1 + \theta'}{2} + \frac{s_\gamma}{2m} \sum_{t \in J} \zeta_t + \frac{s_\gamma \theta'}{2m} \sum_{t \in I\!I} \zeta_t \\ &= \frac{1 + \theta'}{2} + \frac{s_\gamma}{2\sqrt{m}} \xi + \frac{s_\gamma \theta'}{2\sqrt{m}} \xi' \end{aligned}$$

with $\xi' = m^{-1/2} \sum_{t \in I\!I} \zeta_t$, and hence,

$$\widetilde{\theta}_I - \widetilde{\theta}_J = b - \frac{s_\gamma}{2\sqrt{m}} \xi + \frac{s_\gamma \theta'}{2\sqrt{m}} \xi'.$$

Since $\boldsymbol{E}|\xi|^2 = \boldsymbol{E}|\xi'|^2 = 1$, by Lemma 9.1 (see also Remark 9.1)

$$\boldsymbol{P}\left(|\xi| > \lambda\right) + \boldsymbol{P}\left(|\xi'| > \lambda\right) \leq 4e^{-\frac{\lambda^2}{2a\gamma}}$$

and it suffices to check that the inequalities $|\xi| \leq \lambda$, $|\xi'| \leq \lambda$ and (5.1) imply

$$|\widetilde{\theta}_J - \widetilde{\theta}_I| \geq \lambda\widetilde{v}_J + \mu\widetilde{v}_I\,.$$

Since $1 + \theta' = 2b$ and since $\widetilde{v}_J = s_\gamma |J|^{-1/2}\widetilde{\theta}_J$ and similarly for $\widetilde{v}_I$, it holds under the assumptions made:

$$\begin{aligned}
|\widetilde{\theta}_J - \widetilde{\theta}_I| &\geq b - \frac{\lambda s_\gamma}{2\sqrt{m}}(1+\theta') = b(1-\delta) - \delta, \\
\widetilde{v}_J &= \frac{s_\gamma}{\sqrt{m}}\left(1 + \frac{s_\gamma}{\sqrt{m}}\xi\right) \leq \lambda^{-1}\delta\,(1+\delta)\,, \\
\widetilde{v}_I &= \frac{s_\gamma}{\sqrt{2m}}\left(\frac{1+\theta'}{2} + \frac{s_\gamma(\xi + \theta'\xi')}{2\sqrt{m}}\right) \leq \frac{s_\gamma}{\sqrt{2m}}\frac{1+\theta'}{2}(1+\delta) = \frac{(1+b)\delta(1+\delta)}{\lambda\sqrt{2}}\,.
\end{aligned}$$

Therefore

$$\begin{aligned}
&|\widetilde{\theta}_J - \widetilde{\theta}_I| - \lambda\widetilde{v}_J - \mu\widetilde{v}_I \\
&\quad \geq b(1-\delta) - \delta - \delta(1+\delta) - \frac{\mu}{\lambda\sqrt{2}}(1+b)\delta(1+\delta) \\
&\quad = b\left(1 - \delta - \frac{\mu}{\lambda\sqrt{2}}\delta(1+\delta)\right) - \delta - \delta(1+\delta) - \frac{\mu}{\lambda\sqrt{2}}\delta(1+\delta) > 0
\end{aligned}$$

in view of (5.1) and the assertion follows.

References

Bollerslev, T. (1995). Generalised autoregressive conditional heteroskedasticity, *in* Engle (1995b).

Brockmann, M., Gasser, T. and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators, *Journal of the American Statistical Association* **88**: 1302–1309.

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

Bronstein, I. and Semendjajew, K. (1991). *Taschenbuch der Mathematik*, B.G. Teubner Verlagsgesellschaft, Stuttgart.

Carroll, R. and Ruppert, D. (1988). *Transformation and Weighting in Regression*, Chapman and Hall, New York. Chapter 4.

Clements, M. P. and Hendry, D. F. (1998). *Forecasting Economic Time Series*, Cambridge University Press, Cambridge.

Donoho, D., Johnstone, I., Kerkyacharian, G. and Picard, D. (1994). Wavelet shrinkage: asymptopia ?, *Journal of the Royal statistical Society* **57**: 301–369. Series B.

Engle, R. F. (1995a). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, in *ARCH, selected readings* (Engle, 1995b).

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

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

Franses, P. and Dijk, D. V. (1996). Forecasting stock market volatility using (non-linear) GARCH models, *Journal of Forecasting* **15**: 229–235.

Glosten, L., Jagannathan, R. and Runkle, D. (1992). On the relation between the expected value and the volatility of nominal excess return on on stocks, *Journal of Finance* **46**: 1779–1801.

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

Härdle, W., Herwartz, H. and Spokoiny, V. (2000). Time inhomogeneous multiple volatility modelling. mimeo.

Härdle, W. and Stahl, G. (1999). Backtesting beyond var, *Technical Report 105*, Sonderforschungsbereich 373, Humboldt University, Berlin.

Harvey, A., Ruiz, E. and Shephard, N. (1995). Multivariate stochastic variance models, *in* Engle (1995b).

Lepski, O. (1990). One problem of adaptive estimation in Gaussian white noise, *Theory Probab. Appl.* **35**: 459–470.

Lepski, O., Mammen, E. and Spokoiny, V. (1997). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection, *Annals of Statistics* **25**: 929–947.

Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation, *Annals of Statistics* **25**: 2512–2546.

Lipster, R. and Spokoiny, V. (1999). Deviation probability bound for martingales with applications to statistical estimation, *Stat. & Prob Letter* **46**: 347–357.

Nelson, D. B. (1995). Conditional heteroscedasticity in asset returns: A new approach, *in* Engle (1995b).

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

Sentana, E. (1995). Quadratic ARCH models, *Review of Economic Studies* **62**: 639–661.

Spokoiny, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice, *Annals of Statistics* **26**: 1356–1378.

Taylor, S. J. (1986). *Modelling Financial Time Series*, Wiley, Chichester.