

Weierstraß-Institut für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint

ISSN 0946 – 8633

Locally Time Homogeneous Time Series Modelling

Mstislav Elagin, Vladimir Spokoiny

submitted: 3rd December 2008

Mohrenstr. 39, 10117 Berlin
E-Mail: elagin@wias-berlin.de, spokoiny@wias-berlin.de

No. 1379
Berlin 2008



2000 *Mathematics Subject Classification.* 62M10.

Key words and phrases. Adaptive estimation, Local homogeneity, Model selection, Stagewise aggregation, Volatility model, Poisson model, Exponential model, Bernoulli model, Propagation, Oracle.

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

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

Abstract

In this paper three locally adaptive estimation methods are applied to the problems of variance forecasting, value-at-risk analysis and volatility estimation within the context of nonstationary financial time series. A general procedure for the computation of critical values is given. Numerical results exhibit a very reasonable performance of the methods.

Mstislav Elagin and Vladimir Spokoiny

1 Introduction

Modelling particular features (“stylized facts”) of financial time series such as volatility clustering, heavy tails, asymmetry, etc. is an important task arising in financial engineering. For instance, attempts to model volatility clustering, i.e. the tendency of volatility jumps to appear in groups followed by periods of stability, led to the development of conditional heteroskedastic (CH) models including ARCH by Engle (1982) and GARCH by Bollerslev (1986) as well as their derivatives. The main idea underlying the mentioned methods is that volatility clustering can be modelled globally by a stationary process.

However, the assumption of stationarity is often compromised by the shape of the autocorrelation function (ACF) of squared log returns that for a typical financial time series decays slower than exponentially. Furthermore, Mikosch & Stărică (2004) showed that long range memory effects in financial time series may be caused by structural breaks rather than constitute an essential feature of stationary processes to be modeled by global methods. Diebold & Inoue (2001) and Hillebrand (2005) argue that one can easily overlook structural breaks with negative impact on the quality of modelling, estimation and forecasting. This circumstance motivates the development of methods involving processes that are stationary only locally. Local methods consider just the most recent data and imply subsetting of data using some localization scheme that can itself be either global or local and adaptive. Methods of this kind have been presented e.g. in Fan & Gu (2003) for adaptive selection of the decay factor used to weight components of the pseudo-likelihood function, in Dahlhaus & Subba Rao (2006) for the formulation of the locally stationary ARCH(∞) processes, in Cheng, Fan & Spokoiny (2003) for locally choosing parameters of a filter. In a recent paper by Giacomini, Härdle & Spokoiny (2008) a local adaptive method has been applied to the problem of copulae estimation.

Below we compare three methods for estimation of parameters in the context of univariate time series: the local change point (LCP) procedure by Mercurio & Spokoiny (2004), the local model selection (LMS), also known as the intersection of confidence intervals (ICI) by Katkovnik & Spokoiny (2008), and the stagewise aggregation (SA) by Belomestny & Spokoiny (2007). A universal procedure for the choice of parameters (*critical values*) is given. The performance of the procedures is compared using genuine financial data. It is shown that adaptive methods often outperform the standard GARCH(1,1) method.

The chapter is organized as follows. Section 2 is devoted to the formulation of the problem and theoretical introduction. Section 3 describes the methods under comparison. In the Section 4 the procedure for obtaining *critical values*, essential parameters of the procedures, is given. Section 5 shows the application of the adaptive methods to the computation of the value-at-risk.

2 Model and setup

2.1 Conditional heteroskedastic model

Let U_t be a one-dimensional stochastic asset price process in discrete time $t \in \mathbb{N}$ and $R_t = \log \frac{U_t}{U_{t-1}}$ be the corresponding log returns process. The latter is typically described using the *conditional heteroskedastic* model

$$R_t = \sigma_t \varepsilon_t, \tag{1}$$

where ε_t are independent and identically (standard Gaussian) distributed innovations, and σ_t is the *volatility* process progressively measurable w.r.t. the filtration $(\mathcal{F}_{t-1}) = \mathcal{F}(R_1, \dots, R_{t-1})$ generated by past returns. Equivalently,

$$Y_t = \theta_t \varepsilon_t^2 \tag{2}$$

where $Y_t = R_t^2$ are the squared log returns and $\theta_t = \sigma_t^2$. We aim to estimate θ_t from the past observations Y_1, \dots, Y_{t-1} . This problem commonly arises in financial applications such as value-at-risk determination and portfolio optimisation.

2.2 Parametric and local parametric estimation and inference

If $\theta_t = \theta$ one can apply the method of maximum likelihood to obtain the estimate $\hat{\theta}$. The model (2) leads to the log-likelihood function

$$L(\theta) = \sum_t \ell(Y_t, \theta)$$

where $\ell(y, \theta) = -\frac{1}{2} \log(2\pi\theta) - y/(2\theta)$ is the log density of the normal distribution with zero mean. The estimate $\hat{\theta}$ is then obtained by maximizing the log-likelihood function w.r.t. to θ :

$$\hat{\theta} = \arg \max_{\theta} L(\theta) = \frac{\sum_t Y_t}{N},$$

where N is the sample size. When the volatility does depend on time, $\theta_t = \theta(t) \neq \text{const.}$, the method of maximum likelihood is not directly applicable, since the joint distribution of the observations and therefore the log likelihood function are not available. Hence, we take the *local parametric approach* by supposing that for the time point of estimation T there exists some interval $\mathcal{I} = [T - N_{\mathcal{I}}, T]$ of length $N_{\mathcal{I}}$, to be estimated from the data, within which the model (2) describes the process adequately. If the interval \mathcal{I} has been found, then the log likelihood function assumes the form

$$L_{\mathcal{I}}(\theta) = \sum_{t \in \mathcal{I}} \ell(Y_t, \theta)$$

and the maximum likelihood estimate corresponding to the interval \mathcal{I} is

$$\tilde{\theta}_{\mathcal{I}} = \arg \max_{\theta} L_{\mathcal{I}}(\theta) = \sum_{t \in \mathcal{I}} Y_t / N_{\mathcal{I}}.$$

For the purpose of describing the quality of estimation we use the fitted likelihood $L(\tilde{\theta}, \theta)$ defined as the difference between the likelihood corresponding to the ML estimate $\tilde{\theta}$ and the likelihood corresponding to a different parameter value:

$$L(\tilde{\theta}, \theta) = L(\tilde{\theta}) - L(\theta).$$

For the model considered here the fitted likelihood can be represented in the form

$$L(\tilde{\theta}_{\mathcal{I}}, \theta) = N_{\mathcal{I}} \mathcal{K}(\tilde{\theta}_{\mathcal{I}}, \theta), \quad (3)$$

where

$$\mathcal{K}(\theta_1, \theta_2) = \frac{1}{2}(\theta_1/\theta_2 - 1) - \frac{1}{2} \log(\theta_1/\theta_2)$$

denotes the Kullback – Leibler divergence that measures the “distance” between distributions indexed by θ_1 and θ_2 .

2.3 Nearly parametric case

In practice the parametric assumption may be overly stringent and not hold even within an arbitrarily small interval. We describe the deviation from the parametric situation within an interval \mathcal{I} by a random magnitude

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

that we shall call *divergence*. The following *small modelling bias* condition imposes a limit on the deviation from the parametric case providing for the applicability of the local parametric approach.

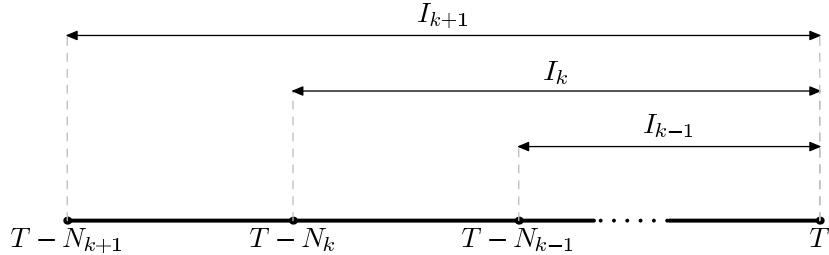


Figure 1: Nested intervals.

Condition 1 *There exists some parameter value $\theta \in \Theta$ and some interval \mathcal{I} such that the expectation under the true measure of the divergence $\Delta_{\mathcal{I}}(\theta)$ over the interval \mathcal{I} is bounded by some $\Delta \geq 0$:*

$$\mathbb{E} \Delta_{\mathcal{I}}(\theta) \leq \Delta.$$

If the small modelling bias condition holds, then for any $r > 0$ the risk of the local maximum likelihood estimate in the nearly parametric case satisfies

$$\mathbb{E} \log \left(1 + \frac{|N_{\mathcal{I}} \mathcal{K}(\tilde{\theta}_{\mathcal{I}}, \theta)|^r}{\mathcal{R}_{r,\theta}} \right) \leq \Delta + 1,$$

where

$$\mathcal{R}_{r,\theta} = \mathbb{E}_{\theta} \left| N_{\mathcal{I}} \mathcal{K}(\tilde{\theta}_{\mathcal{I}}, \theta) \right|^r \quad (4)$$

is the risk of the local maximum likelihood estimate in the parametric case. Here the logarithm under the expectation comes from the Cramér – Rao inequality, and the additional term Δ on the right-hand side can be interpreted as payment for the violation of the parametric assumption.

The last result leads to the notion of the *oracle* estimate as the “largest” one under the small modelling bias condition. In the next section we present three methods suitable for construction of estimates performing almost as well as the oracle estimate.

3 Methods for the estimation of parameters

3.1 Sequence of intervals

Local methods imply subsetting of data. A localization scheme that we use is a growing sequence of intervals.

Let T denote the time point at which the value of interest is to be estimated. We define an ordered sequence of intervals $\{\mathcal{I}_k\}_{k=1}^K$ of length N_k with the common right edge at T (Figure 1), so $\mathcal{I}_k = [T - N_k, T[$. We associate with each interval \mathcal{I}_k from this sequence the corresponding maximum likelihood estimate $\tilde{\theta}_k \equiv \tilde{\theta}_{\mathcal{I}_k}$, which we

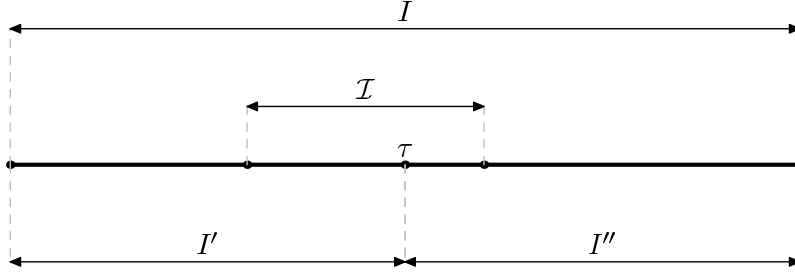


Figure 2: Intervals involved in the change point detection procedure.

shall call *weak estimate*. We aim to select or construct the “largest” one still satisfying the small modelling bias condition. The LCP and LMS procedures obtain the best estimate by choosing one from the sequence, whereas SSA builds the estimate by taking convex combinations of previously found estimates. Below we describe each of the methods.

3.2 Local change point selection

The method of local change point (LCP) selection introduced in (Mercurio & Spokoiny 2004) is a procedure that detects the largest interval of homogeneity and provides an adaptive estimate as the one associated with the interval found. The idea of the method consists in the testing of the null hypothesis of an interval containing no change points against the alternative hypothesis of a change point being present, whereas the interval under testing is taken from the growing sequence.

Consider a *tested interval* \mathcal{I} that possibly contains a change point, and an enclosing *testing interval* I (Figure 2). The statistic to test the hypothesis about the parameter change in some internal point τ of the candidate interval can be expressed as the difference between the sum of log likelihoods corresponding to the intervals I' , I'' into which the change point splits the testing interval, and the log likelihood corresponding to the testing interval containing no change points:

$$T_{\mathcal{I},\tau} = \max_{\theta',\theta''} \{L_{I''}(\theta'') + L_{I'}(\theta')\} - \max_{\theta} L_I(\theta) = L_{I'}(\tilde{\theta}_{I'}) + L_{I''}(\tilde{\theta}_{I''}) - L_I(\tilde{\theta}_I),$$

where $L(\cdot)$ denotes the log likelihood function. For the volatility distribution the test statistic can be represented in the form

$$T_{\mathcal{I},\tau} = \min_{\theta} \left(N_{I''} \mathcal{K}(\tilde{\theta}_{I''}, \theta) + N_{I'} \mathcal{K}(\tilde{\theta}_{I'}, \theta) \right) = N_{I''} \mathcal{K}(\tilde{\theta}_{I''}, \tilde{\theta}_I) + N_{I'} \mathcal{K}(\tilde{\theta}_{I'}, \tilde{\theta}_I) \quad (5)$$

due to (3). The test statistic for the whole candidate interval is the maximum of the pointwise statistics over all internal points:

$$T_{\mathcal{I}} = \max_{\tau \in \mathcal{I}} T_{\mathcal{I},\tau}$$

The hypothesis is rejected if the test statistic exceeds some *critical value* \mathfrak{z} , which is a parameter of the procedure specific to the problem design.

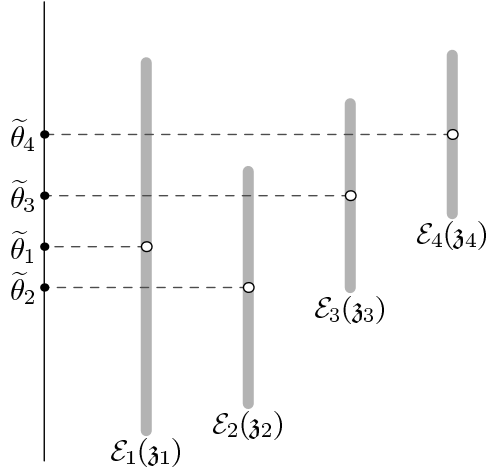


Figure 3: Principle of the local model selection. $\hat{\theta} = \tilde{\theta}_3$.

We let $\mathcal{I} = I_k \setminus I_{k-1}$ and $I = \mathcal{I}_{k+1}$ and take the adaptive estimate $\hat{\theta}$ to be equal to the \hat{k} -th weak estimate, where \hat{k} is the largest interval number such that all test statistics corresponding to the intervals $\mathcal{I}_1, \dots, \mathcal{I}_{\hat{k}}$ do not exceed their critical values with the opposite holding for $\hat{k} + 1$:

$$\hat{\theta} = \tilde{\theta}_{\hat{k}}, \text{ where } \hat{k} = \max k \text{ such that } T_l \leq \mathfrak{z}_l \text{ for all } l \leq \hat{k}.$$

The initial condition is that the smallest interval is always considered to be homogeneous. Since it is not feasible to test the largest interval, the greatest possible value of \hat{k} is $K - 1$.

3.3 Local model selection

The idea of the local model selection procedure introduced in (Katkovnik & Spokoiny 2008) consists in the choice of the “largest” weak estimate among $\tilde{\theta}_1 \dots \tilde{\theta}_K$ as the adaptive estimate $\hat{\theta}$ in such a way that the adaptive estimate belongs to the confidence interval \mathcal{E} of each of the previous weak estimates (cf. Figure 3). Formally, $\hat{\theta} = \tilde{\theta}_{\hat{k}}$, where

$$\hat{k} \text{ is such that } \begin{cases} \tilde{\theta}_{\hat{k}} \in \mathcal{E}_l & \text{for all } l < \hat{k} \\ \tilde{\theta}_{\hat{k}+1} \notin \mathcal{E}_l & \text{for some } l < \hat{k} + 1 \end{cases}$$

Confidence interval of level α for a weak estimate $\tilde{\theta}$ is provided by

$$\mathcal{E}(\mathfrak{z}_\alpha) = \left\{ \theta : L(\tilde{\theta}, \theta) \leq \mathfrak{z}_\alpha \right\}.$$

As with the LCP procedure, the first weak estimate is always accepted. However, the LMS procedure checks all estimates including the one corresponding to the last interval.

3.4 Stagewise aggregation

The stagewise aggregation (SA) procedure introduced in (Belomestny & Spokoiny 2007) differs from the two methods described above in that it does not choose the adaptive estimate $\hat{\theta}$ from the weak estimates $\tilde{\theta}_1 \dots \tilde{\theta}_K$. Instead, based on the weak estimates, it sequentially constructs *aggregated estimates* $\hat{\theta}_1 \dots \hat{\theta}_K$ possessing the property that any aggregated estimate $\hat{\theta}_k$ has smaller variance than the corresponding weak estimate $\tilde{\theta}_k$, while keeping “close” to it in terms of the statistical difference, the latter being measured through the likelihood ratio $L(\tilde{\theta}_k, \hat{\theta}_{k-1}) = L(\tilde{\theta}_k) - L(\hat{\theta}_{k-1})$. The adaptive estimate is finally taken equal to the last aggregated estimate: $\hat{\theta} = \hat{\theta}_K$ (unless the *early stopping* occurs, see below).

Formally, the first aggregated estimate is equal to the first weak estimate and every next aggregated estimate is a convex combination of the previous aggregated estimate and the current weak estimate:

$$\hat{\theta}_k = \begin{cases} \tilde{\theta}_1, & k = 1 \\ \gamma_k \tilde{\theta}_k + (1 - \gamma_k) \hat{\theta}_{k-1}, & k = 2, \dots, K \end{cases}$$

Here γ_k is the *mixing coefficient* that reflects the statistical difference between the previous aggregated estimate $\hat{\theta}_{k-1}$ and the current weak estimate $\tilde{\theta}_k$, and is obtained by applying an *aggregation kernel* K_{ag} to the likelihood ratio $L(\tilde{\theta}_k, \hat{\theta}_{k-1})$ scaled by the *critical value* \mathfrak{z}_k :

$$\gamma_k = K_{\text{ag}} \left(\frac{L(\tilde{\theta}_k, \hat{\theta}_{k-1})}{\mathfrak{z}_k} \right).$$

The aggregation kernel acts as a link between the likelihood ratio and the mixing coefficient. The principle behind its selection is that a smaller statistical difference between $\tilde{\theta}_k$ and $\hat{\theta}_{k-1}$ should lead to the mixing coefficient close to 1 and thus to the aggregated estimate $\hat{\theta}_k$ close to $\tilde{\theta}_k$, whereas a larger difference should provide the mixing coefficient close to zero and thus keep $\hat{\theta}_k$ close to $\hat{\theta}_{k-1}$. Whenever the difference is very large, the mixing coefficient is zero, and the procedure stops prematurely by setting $\hat{\theta} = \hat{\theta}_{k-1}$. We call this situation *early stopping*.

To satisfy the stated requirements, the kernel must be supported on the closed interval $[0, 1]$ and monotonously decrease from 1 on the left edge to 0 on the right edge. It is also recommended that the kernels have a plateau of size b starting with zero. Thus, the aggregation kernel assumes the form:

$$K_{\text{ag}}(u) = \begin{cases} 1, & 0 \leq u < b \\ 1 - \bar{K}_{\text{ag}}(u), & b \leq u \leq 1 \end{cases}$$

Examples of $\bar{K}_{\text{ag}}(u)$ include $\frac{u-b}{1-b}$ (triangular kernel), $(\frac{u-b}{1-b})^2$ (Epanechnikov kernel) etc.

4 Critical values and other parameters

All procedures described above depend on the set of parameters $\mathfrak{z}_1 \dots \mathfrak{z}_K$ known as critical values. The critical values reflect the problem design (interval length, model, method etc.). They are selected based on the following *propagation condition*:

Condition 2 (Propagation condition) For any $\theta^* \in \Theta$

$$\frac{\mathbb{E}_{\theta^*} |L(\tilde{\theta}_k, \hat{\theta}_k)|^r}{\mathcal{R}_{r, \theta^*}} \leq \alpha \frac{k}{K} \quad \text{for } k = 1, \dots, K, \quad (6)$$

where $\hat{\theta}_k$ is the adaptive estimate obtained on the k -th step and $\mathcal{R}_{r, \theta^*}$ is the risk delivered by the local maximum likelihood estimate in the parametric case (see (4)).

This condition means that in the homogeneous case the risk associated with the k -th adaptive estimate must not exceed a certain fraction of the risk in the parametric case.

Critical values constructed this way provide with high probability the prescribed performance of the procedures in the parametric situation (under the null hypothesis). Namely, under the parametric hypothesis on every step k the adaptive estimate $\hat{\theta}_k$ should be close enough to the oracle estimate $\tilde{\theta}_k$. However, the propagation condition is not explicit. For the computation of critical values we use the following sequential method based on Monte-Carlo simulations. Denote as $\hat{\theta}_l(\mathfrak{z}_k)$ for $l \geq k$ the adaptive estimate obtained after the l -th step of the procedure run with the critical values $\mathfrak{z}_1, \dots, \mathfrak{z}_{k-1}$ known and $\mathfrak{z}_{k+1}, \dots, \mathfrak{z}_K$ set to infinity:

$$\hat{\theta}_l(\mathfrak{z}_k) = \hat{\theta}_l(\mathfrak{z}_1, \dots, \mathfrak{z}_k, \mathfrak{z}_{k+1} = \infty, \dots, \mathfrak{z}_K = \infty).$$

The first critical value can be selected to satisfy the conditions

$$\frac{\mathbb{E}_{\theta^*} |L(\tilde{\theta}_l, \hat{\theta}_l(\mathfrak{z}_1))|^r}{\mathcal{R}_{r, \theta^*}} \leq \frac{\alpha}{K}, \quad l = 2, \dots, K.$$

Such a value exists, since for \mathfrak{z}_1 taken sufficiently large the weak and adaptive estimates coincide for any l and all Monte-Carlo paths, thus leading to the zero risk. With the first $k - 1$ critical values fixed the procedure is carried out sequentially for the remaining critical values. The k -th critical value is selected using the condition

$$\frac{\mathbb{E}_{\theta^*} |L(\tilde{\theta}_l, \hat{\theta}_l(\mathfrak{z}_k))|^r}{\mathcal{R}_{r, \theta^*}} \leq k \frac{\alpha}{K}, \quad l = k + 1, \dots, K.$$

Obviously, the critical values depend on the specific form of the likelihood function and hence of the Kullback-Leibler distance. Further, the critical values depend on the global parameters α and r .

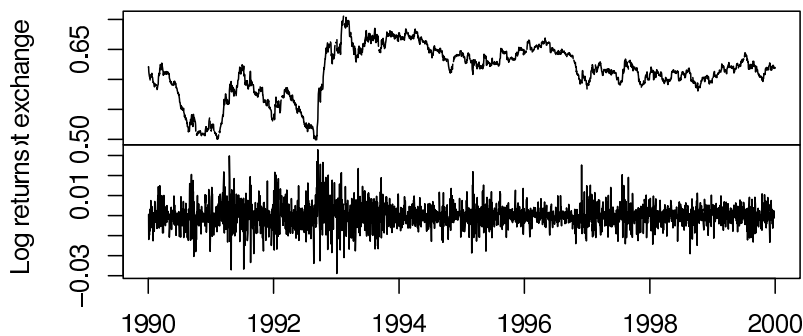


Figure 4: Exchange rate of the British pound to the US dollar and corresponding log returns.

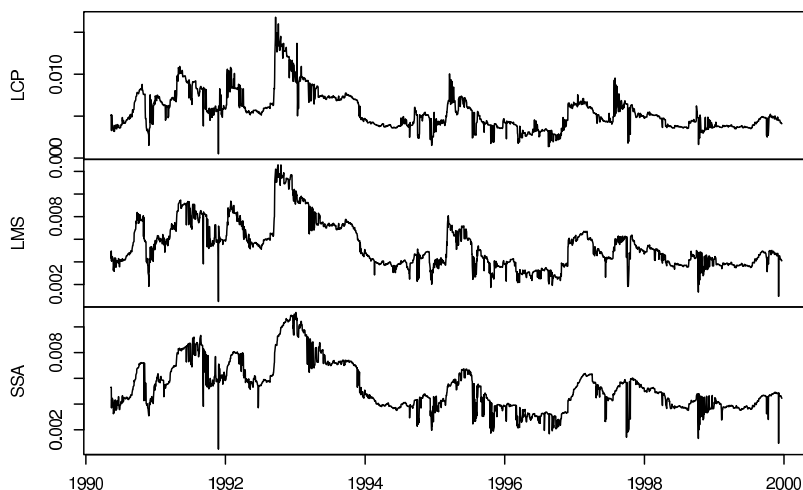


Figure 5: Volatility estimates for the British pound obtained by LCP, LMS and SA.

5 Applications

We illustrate the performance of the methods introduced in the section 3 by analyzing daily exchange rates of six currencies (GBP, AUD, NZD, JPY, CAD, DKR) to the US dollar available from the site of the US Federal Reserve. We use the data for the period from January 1, 1990 till December 31, 1999. Unless indicated otherwise, we use the GBP/USD exchange rate. Observed GBP/USD exchange rates along with the log returns are shown on the Figure 4, while Figure 5 presents the volatility estimates obtained by three adaptive methods.

A well known feature of financial time series is the uncorrelatedness of the log returns. However, in spite of the uncorrelatedness, the log returns are not independent, as one can see by plotting the autocorrelation of a non-linear transformation. For instance, absolute log returns show significant autocorrelation (Figure 6, upper plot). We obtain standardized absolute log returns by dividing the absolute log returns by the volatility estimated using the LCP method. The ACF plot (Figure 6, lower plot) shows that nearly all autocorrelation has been removed by standardizing. This result indicates the reasonable quality of volatility estimation.

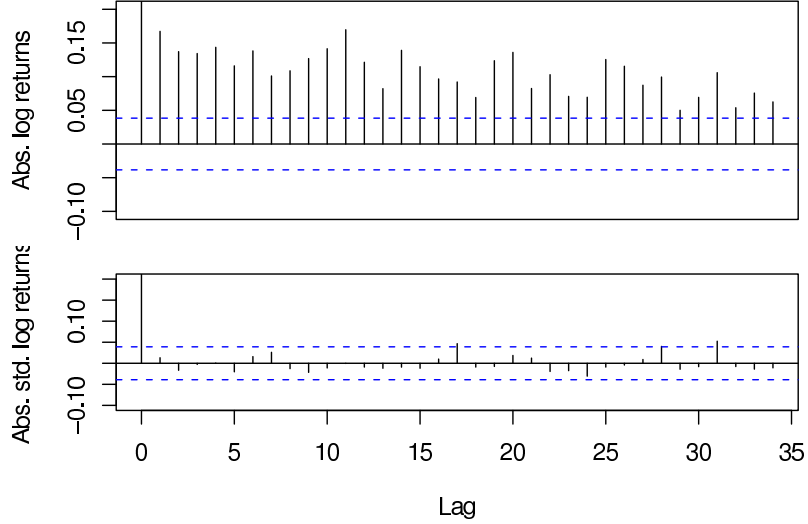


Figure 6: Autocorrelation functions of the absolute log returns (above) and of the absolute standardized log returns (below).

5.1 Variance forecasting

In order to assess the performance of the adaptive procedures we evaluate their ability to forecast the h -step ahead conditional variance of the aggregated returns originating at time t . We denote this magnitude as

$$V_{t,h} = \text{Var}(S_{t,h} | \mathcal{F}_t), \quad \text{where} \quad S_{t,h} = \sum_{k=1}^h R_{t+k}.$$

By definition of the conditional variance, employing that $\mathbb{E}(R_{t+k} | \mathcal{F}_t) = 0$ and in view of the conditional uncorrelatedness of the log returns as well as the linearity of the expectation, one can represent the conditional variance of the aggregated returns as the sum of conditional variances of log returns:

$$\begin{aligned} V_{t,h} &= \mathbb{E}([S_{t,h} - \mathbb{E}(S_{t,h} | \mathcal{F}_t)]^2 | \mathcal{F}_t) \\ &= \mathbb{E}\left(\left[S_{t,h} - \sum_{k=1}^h \mathbb{E}(R_{t+k} | \mathcal{F}_t)\right]^2 \middle| \mathcal{F}_t\right) = \mathbb{E}(S_{t,h}^2 | \mathcal{F}_t) \\ &= \mathbb{E}\left(\sum_{k=1}^h R_{t+k}^2 \middle| \mathcal{F}_t\right) = \sum_{k=1}^h \mathbb{E}(R_{t+k}^2 | \mathcal{F}_t) = \sum_{k=1}^h \theta_{t,k} \end{aligned} \quad (7)$$

By definition of the local constant approach the conditional variance of the log returns is constant for a certain horizon h :

$$\theta_{t,k}^{(A)} = \theta_t^{(A)}, \quad k = 1, \dots, h. \quad (8)$$

Therefore the forecasted conditional variance of the aggregated returns $S_{t,h}$ obtained by an adaptive method is simply

$$\widehat{V}_{t,h}^{(A)} = h\widehat{\theta}_t^{(A)}. \quad (9)$$

We compare the forecasts obtained by adaptive procedures with that of the GARCH(1,1) model, one of the most popular parameterizations of the volatility process of financial time series. The GARCH(1,1) model describes the volatility dynamics by the relation

$$\theta_t = \omega + \alpha R_{t-1}^2 + \beta \theta_{t-1},$$

where the requirement of the stationarity implies the following conditions on the coefficients:

$$\alpha > 0, \quad \beta > 0, \quad \alpha + \beta < 1.$$

The h -step ahead conditional variance forecast for the log returns is given by:

$$\theta_{t,h}^{(G)} = \mathbb{E} (R_{t+h}^2 | \mathcal{F}_t) = \bar{\sigma}^2 + (\alpha + \beta)^h (\theta_t - \bar{\sigma}^2),$$

where $\bar{\sigma}^2 = \omega / (1 - \alpha - \beta)$ is the unconditional variance. Finally, the realized variance of the aggregated returns is

$$\widehat{V}_{t,h}^{(G)} = \sum_{k=1}^h \widehat{\theta}_{t,k}^{(G)}. \quad (10)$$

As a criterion we use the mean square root error (MSqE) defined as

$$\text{MSqE}_{\mathcal{I},h} = \sum_{t \in \mathcal{I}} |\widehat{V}_{t,h} - V_{t,h}|^{1/2}, \quad (11)$$

where

$$V_{t,h} = \sum_{k=1}^h R_{t+k}^2$$

is the realized variance of h aggregated returns starting at time t , and $\widehat{V}_{t,h}$ is the conditional variance forecast of the aggregated returns. Substituting (9) or (10) for $\widehat{V}_{t,h}$ into (11), one obtains the performance data shown in the Figure 7. The results are presented for various years and forecasting horizons. As seen from the figure, adaptive methods outperform the GARCH(1,1) in many cases.

5.2 Value-at-risk

In the present section we apply the adaptive procedures to the computation of value at risk, an important problem in financial engineering. The value at risk (VaR) is defined as “the maximum loss not exceeded with a given probability defined as the confidence level, over a given period of time”. The problem of the VaR estimation can be represented as the problem of quantile estimation for the distribution of aggregated returns. We consider three distributions of innovations: standard Gaussian distributions, Student’s scaled distribution with 5 degrees of freedom and the empirical distribution:

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

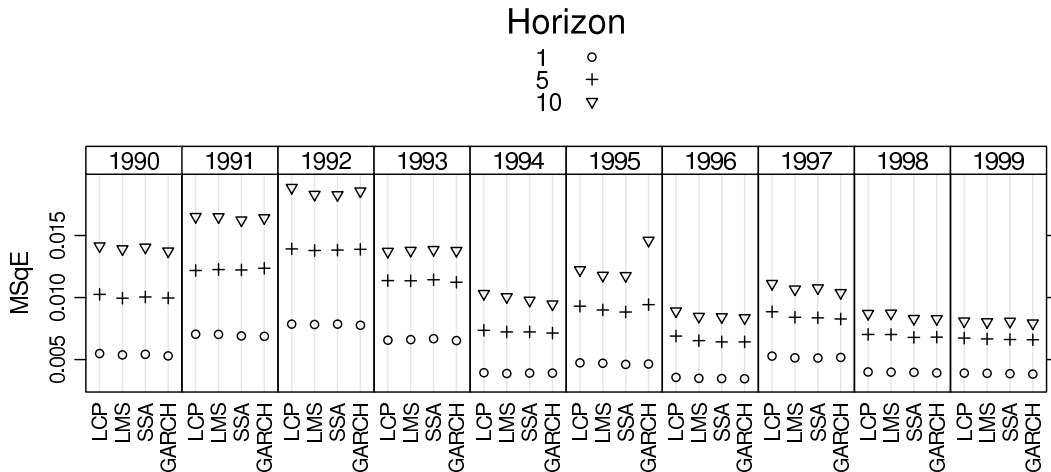


Figure 7: Performance of adaptive methods and GARCH(1,1) in terms of MSqE.

We aim to describe the quality of VaR computation in terms of the frequency of exceptions, where an “exception” is the event of the predicted value at risk exceeding the aggregated returns. According to the prescribed assessment rule, we examine the particular case of the value at risk predicted at 1% level for 10 steps ahead on 250 observations. Under the assumption that the exceptions follow the binomial distribution, we conduct a test with the null hypothesis about the probability of exception being equal to 0.01, and one-sided alternative hypothesis about the probability of exception exceeding 0.01. A procedure predicting the value at risk belongs in one of the three “zones”: “green” zone if the null hypothesis can not be rejected with 95% confidence (corresponding to not more than 5 exceptions on 250 observations, or 2% frequency), “yellow” zone if the null should be rejected with 95% confidence (from 6 to 10 exceptions, or not more than 4% frequency), and “red zone” if the null should be rejected with 99.99% confidence (11 or more exceptions, or more than 4% frequency).

Figure 8 shows the percentage of time points at which the loss within a certain horizon overshoots the value at risk predicted with the corresponding confidence level. The results were obtained for three distributions of innovations. One observes that none of the adaptive methods falls in the red zone. Stagewise aggregation always belongs to the green zone. LCP and LMS combined with the Gaussian innovations sometimes fall into the yellow zone. Use of Student’s innovations slightly, and of the empirically distributed innovations considerably improves the performance. Overall performance of the adaptive methods is rather good.

5.3 A multiple time series example

The local parametric approach can be extended to multiple time series. In this case one observes a vector of exchange rate processes $U_t \in \mathbb{R}^d$, $t = 1, 2, \dots$ and $R_{t,m}$ is the vector of the corresponding log returns:

$$R_{t,m} = \log(U_{t,m}/U_{t-1,m}), \quad m = 1, \dots, d.$$

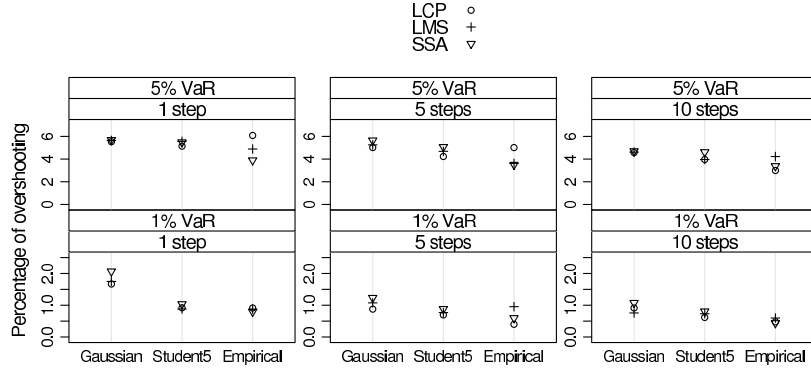


Figure 8: Percentage of overshooting the value-at-risk estimated by three methods for various distributions of innovations, number of forecasting steps and value-at-risk levels. Currency: Australian dollar.

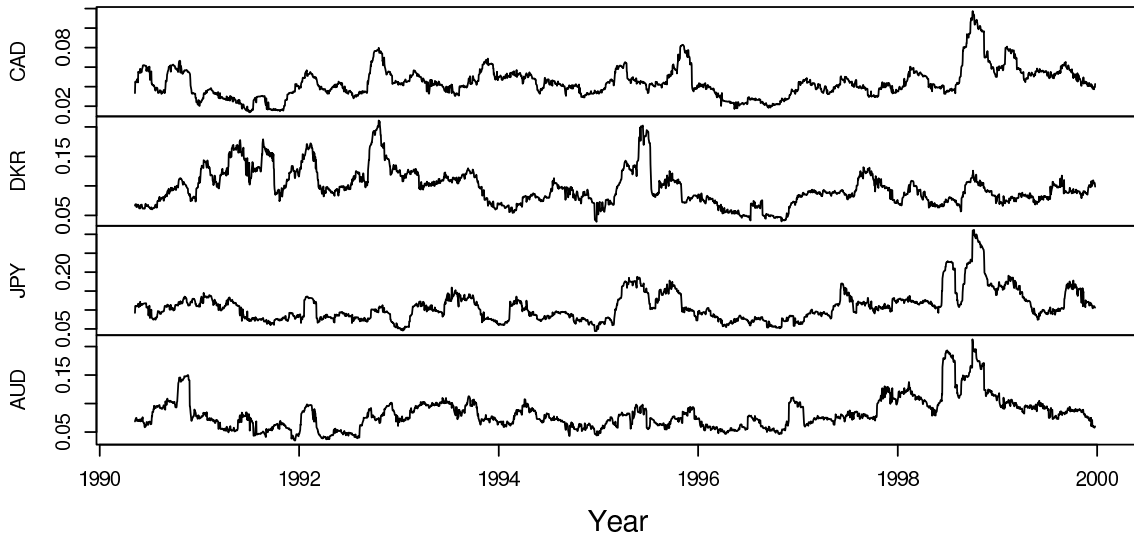


Figure 9: Adaptive estimation of the annualized volatility of four exchange rates.

The *conditional heteroskedasticity* model reads in this case as

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

where ε_t , $t \geq 1$, is a sequence of independent standard Gaussian random innovations and Σ_t is a symmetric $d \times d$ *volatility* matrix, which is to be estimated. As an example, Figure 9 shows annualized volatility estimated for exchange rates of several currencies to the US dollar. Annualized volatility is defined as $\sqrt{250 \widehat{\Sigma}_{ii}}$, where $\widehat{\Sigma}_{ii}$ represent diagonal elements of the volatility matrix, Similar evolution of the estimates indicates a possible common low-order component.

As in one-dimensional case, we observe significant correlation and autocorrelation of the absolute log returns (Figure 10, left) as a non-linear transformation of the log returns, indicating lack of independence in spite of the log returns being uncorrelated. We estimate the volatility matrix using the LCP method and obtain the

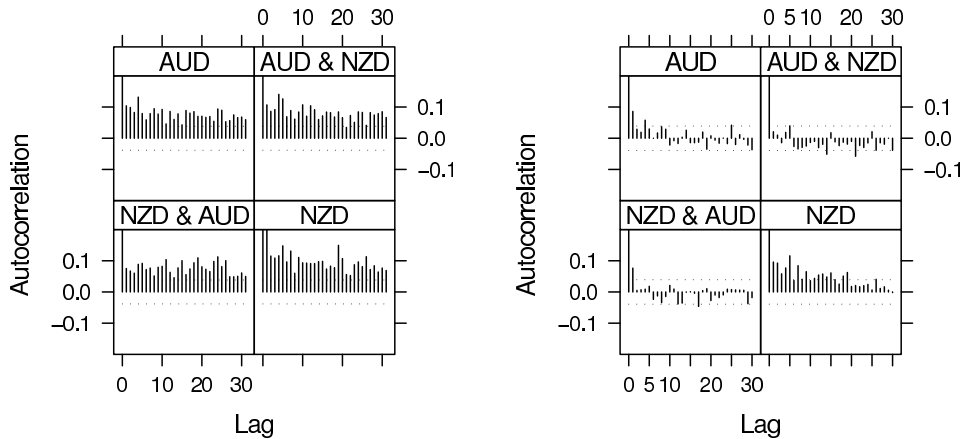


Figure 10: ACF for the NZD and AUD time series. Left: absolute log returns, right: absolute standardized log returns.

standardized absolute log returns by solving the equation

$$R_t = \hat{\Sigma}_t^{1/2} \xi_t$$

for ξ_t . The multivariate ACF plot of the standardized absolute log returns is shown in the right part of Figure 10. Although some autocorrelation still remains in the NZD series, the remaining three ACF plots show almost no significant correlation.

References

- Belomestny, D. & Spokoiny, V. (2007). Spatial aggregation of local likelihood estimates with applications to classification, *Ann. Statist.* **25**: 2287–2311.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* **31**(3): 307–327.
- Cheng, M.-Y., Fan, J. & Spokoiny, V. (2003). Dynamic nonparametric filtering with application to volatility estimation, in M. G. Akritas & D. N. Politis (eds), *Recent Advances and Trends in Nonparametric Statistics*, Elsevier.
- Dahlhaus, R. & Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes, *Ann. Statist.* **34**(3): 1075–1114.
- Diebold, F. X. & Inoue, A. (2001). Long memory and regime switching, *J. Econometrics* **105**(1): 131–159.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**(4): 987–1007.
- Fan, J. & Gu, J. (2003). Semiparametric estimation of value at risk, *Econom. J.* **6**(2): 261–290.

- Giacomini, E., Härdle, W. & Spokoiny, V. (2008). Inhomogeneous dependency modelling with time varying copulae, *Journal of Business and Economic Statistics*. Forthcoming.
- Hillebrand, E. (2005). Neglecting parameter changes in GARCH models, *J. Econometrics* **129**(1-2): 121–138.
- Katkovnik, V. & Spokoiny, V. (2008). Spatially adaptive estimation via fitted local likelihood techniques, *IEEE TRANSACTIONS ON SIGNAL PROCESSING* **56**(3): 873–886.
- Mercurio, D. & Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models, *Ann. Statist.* **32**(2): 577–602.
- Mikosch, T. & Stărică, C. (2004). Changes of structure in financial time series and the GARCH model, *REVSTAT* **2**(1): 41–73.