17 Locally Time Homogeneous Time Series Modelling

Mstislav Elagin and Vladimir Spokoiny

17.1 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 and Stărică (2004) showed that long range memory effects in financial time series may be caused by structural breaks rather than that constitutes an essential feature of stationary processes to be modeled by global methods. Diebold and 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 and Gu (2003) for adaptive selection of the decay factor used to weight components of the pseudo-likelihood function, in Dahlhaus and Subba Rao (2006) for the formulation of the locally stationary ARCH(∞) processes, in Cheng, Fan and Spokoiny (2003) for locally choosing parameters of a filter. In a recent paper

by Giacomini, Härdle and 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 and Spokoiny (2004), the local model selection (LMS), also known as the intersection of confidence intervals (ICI) by Katkovnik and Spokoiny (2008), and the stagewise aggregation (SA) by Belomestny and 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 17.2 is devoted to the formulation of the problem and theoretical introduction. Section 17.3 describes the methods under comparison. In Section 17.4 the procedure for obtaining *critical values*, essential parameters of the procedures, is given. Section 17.5 shows the application of the adaptive methods to the computation of the value-at-risk.

17.2 Model and Setup

17.2.1 Conditional Heteroskedastic Model

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

$$R_t = \sigma_t \varepsilon_t, \tag{17.1}$$

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 $(\mathcal{F}_{t-1}) = \mathcal{F}(R_1, \ldots, R_{t-1})$ generated by past returns. Equivalently,

$$Y_t = \theta_t \varepsilon_t^2 \tag{17.2}$$

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 $Y_1, \ldots, Y_{t-1}$. This problem commonly arises in financial applications such as value-at-risk determination and portfolio optimisation.

17.2.2 Parametric and Local Parametric Estimation and Inference

If $\theta_t = \theta$ one can apply the method of maximum likelihood to obtain the estimate $\widehat{\theta}$. The model (17.2) 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 $\widehat{\theta}$ is then obtained by maximizing the log-likelihood function w.r.t. to θ:

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

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 (17.2) 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

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

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

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

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

$$L(\widetilde{\theta}_\mathcal{I}, \theta) = N_\mathcal{I} \mathcal{K}(\widetilde{\theta}_\mathcal{I}, \theta), \tag{17.3}$$

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.

17.2.3 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 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 (SMB)* condition imposes a limit on the deviation from the parametric case which provides the applicability of the local parametric approach.

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 $\Delta \geq 0$:*

$$\mathsf{E}\, \Delta_{\mathcal{I}}(\theta) \leq \Delta. \tag{17.4}$$

If the SMB condition 17.4 holds, then for any $r > 0$ the risk of the local maximum likelihood estimate in the nearly parametric case satisfies:

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

where

$$\mathcal{R}_{r,\theta} = \mathsf{E}_\theta \left| N_{\mathcal{I}} \mathcal{K}(\widetilde{\theta}_{\mathcal{I}}, \theta) \right|^r \tag{17.5}$$

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.

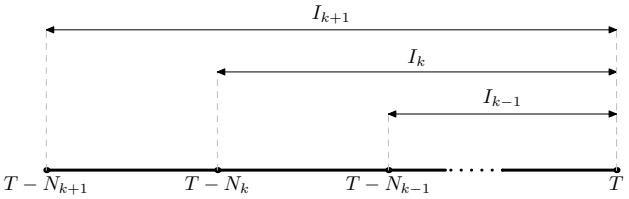

Figure 17.1. Nested intervals.

17.3 Methods for the Estimation of Parameters

17.3.1 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 17.1), so $\mathcal{I}_k = [T - N_k, T[$. We associate with each interval $\mathcal{I}_k$ from this sequence the corresponding maximum likelihood estimate $\widetilde{\theta}_k \equiv \widetilde{\theta}_{\mathcal{I}_k}$, which we 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.

17.3.2 Local Change Point Selection

The LCP method introduced in Mercurio and 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 17.2). 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

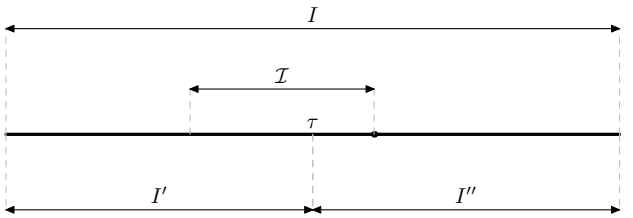

Figure 17.2. Intervals involved in the change point detection procedure.

containing no change points:

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

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

$$T_{\mathcal{I},\tau} = \min_{\theta} \left\{ N_{I''}\mathcal{K}(\widetilde{\theta}_{I''}, \theta) + N_{I'}\mathcal{K}(\widetilde{\theta}_{I'}, \theta) \right\} = N_{I''}\mathcal{K}(\widetilde{\theta}_{I''}, \widetilde{\theta}_I) + N_{I'}\mathcal{K}(\widetilde{\theta}_{I'}, \widetilde{\theta}_I) \tag{17.6}$$

due to (17.3). 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.

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

$$\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}, \text{ where } \widehat{k} = \max k \text{ such that } T_l \le \mathfrak{z}_l \text{ for all } l \le \widehat{k}.$$

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 $\widehat{k}$ is $K-1$.

17.3.3 Local Model Selection

The idea of the local model selection procedure introduced in Katkovnik and Spokoiny (2008) consists in the choice of the "largest" weak estimate among

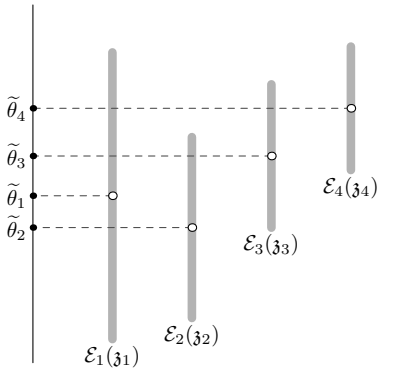

Figure 17.3. Principle of the local model selection. $\widehat{\theta} = \widetilde{\theta}_3$.

$\widetilde{\theta}_1 \dots \widetilde{\theta}_K$ as the adaptive estimate $\widehat{\theta}$ in such a way that the adaptive estimate belongs to the confidence interval $\mathcal{E}$ of each of the previous weak estimates (Figure 17.3). Formally, $\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}$, where

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

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

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

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.

17.3.4 Stagewise Aggregation

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

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:

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

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

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

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 $\widetilde{\theta}_k$ and $\widehat{\theta}_{k-1}$ should lead to the mixing coefficient close to 1 and thus to the aggregated estimate $\widehat{\theta}_k$ close to $\widetilde{\theta}_k$, whereas a larger difference should provide the mixing coefficient close to zero and thus keep $\widehat{\theta}_k$ close to $\widehat{\theta}_{k-1}$. Whenever the difference is very large, the mixing coefficient is zero, and the procedure stops prematurely by setting $\widehat{\theta} = \widehat{\theta}_{k-1}$. 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:

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

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

17.4 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* $\theta^* \in \Theta$

$$\frac{\mathsf{E}_{\theta^*} |L(\widetilde{\theta}_k, \widehat{\theta}_k)|^r}{\mathcal{R}_{r,\theta^*}} \le \alpha \frac{k}{K} \quad \textit{for } k = 1, \ldots, K, \tag{17.7}$$

where $\widehat{\theta}_k$ *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* (17.5)*).*

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 $\widehat{\theta}_k$ should be close enough to the oracle estimate $\widetilde{\theta}_k$. 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 $\widehat{\theta}_l(\mathfrak{z}_k)$ for $l \ge k$ the adaptive estimate obtained after the l-th step of the procedure run with he critical values $\mathfrak{z}_1, \ldots, \mathfrak{z}_{k-1}$ known and $\mathfrak{z}_{k+1}, \ldots, \mathfrak{z}_K$ set to infinity:

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

The first critical value can be selected to satisfy the conditions

$$\frac{\mathsf{E}_{\theta^*} \left| L(\widetilde{\theta}_l, \widehat{\theta}_l(\mathfrak{z}_1)) \right|^r}{\mathcal{R}_{r,\theta^*}} \le \frac{\alpha}{K}, \quad l = 2, \ldots, K.$$

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

$$\frac{\mathsf{E}_{\theta^*} \left| L(\widetilde{\theta}_l, \widehat{\theta}_l(\mathfrak{z}_k)) \right|^r}{\mathcal{R}_{r,\theta^*}} \le k \frac{\alpha}{K}, \quad l = k+1, \ldots, K.$$

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.

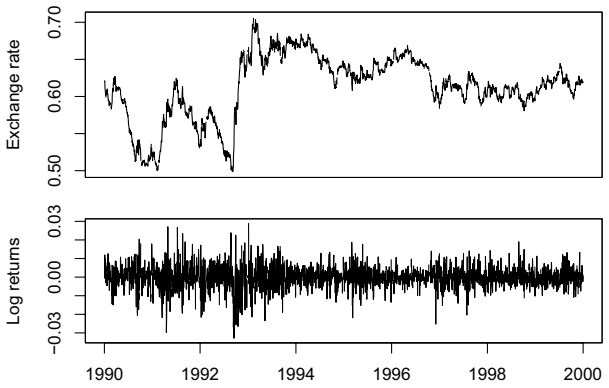

Figure 17.4. Exchange rate of the British pound to the US dollar 19900101-19991231 (above) and corresponding log returns (below).

17.5 Applications

We illustrate the performance of the methods introduced in the section 17.3 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 Januar 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 17.4, while Figure 17.5 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 nonlinear transformation. For instance, absolute log returns show significant autocorrelation (Figure 17.6, 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 17.6, lower plot) shows that nearly all autocorrelation has been removed by standardizing. This result indicates the reasonable quality of volatility estimation.

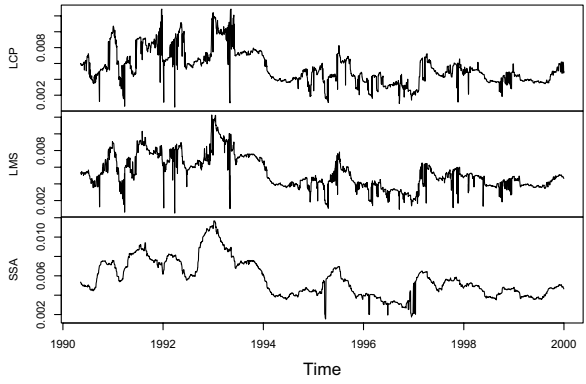

Figure 17.5. Volatility estimates for the British pound obtained by LCP, LMS and SA. XFGgbpvolaest

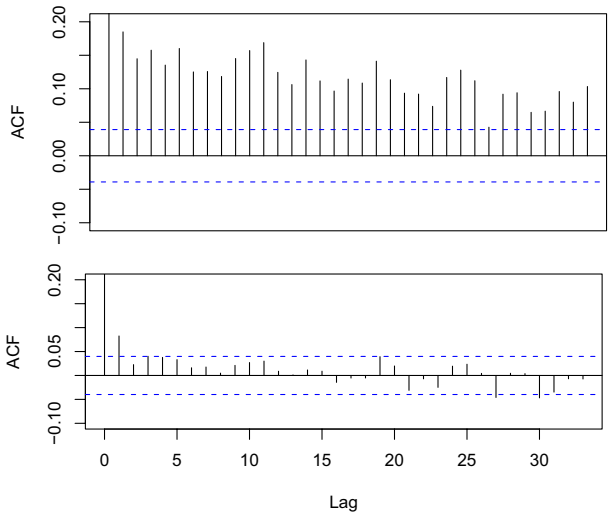

Figure 17.6. Autocorrelation functions of the absolute log returns (above) and of the absolute standardized log returns (below). XFGacfgbplog

17.5.1 Forecasting Performance for One and Multiple Steps

In order to assess the performance of the adaptive procedures we compare their ability to forecast the conditional variance of the aggregated returns

with that of the GARCH(1,1) model, one of the most popular parameterizations of the volatility process of financial time series. Namely, for a sequence of intervals and forecasting horizons we use the mean square root error (MSqE) criterion

$$\mathrm{MSqE}_{\mathcal{I}} = \sum_{t\in\mathcal{I}} |V^{\heartsuit}_{t,h} - V^{\circ}_{t,h}|^{1/2} \Big/ \sum_{t\in\mathcal{I}} |V^{\spadesuit}_{t,h} - V^{\circ}_{t,h}|^{1/2}, \tag{17.8}$$

where

$$V^{\circ}_{t,h} = R^2_{t+1} + \ldots + R^2_{t+h} \tag{17.9}$$

is the realized variance of h aggregated returns starting at time t, and $V^{\heartsuit}_{t,h}, V^{\spadesuit}_{t,h}$ denote the conditional variance forecast of the aggregated returns by an adaptive procedure and GARCH(1,1), respectively.

The h-step ahead conditional variance forecast originating at time t is defined as

$$V_{t,h} \stackrel{\mathrm{def}}{=} \mathsf{Var}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right).$$

By definition of the conditional variance

$$\mathsf{Var}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right) = \mathsf{E}\left[\left\{ \sum_{k=1}^{h} R_{t+k} - \mathsf{E}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right) \right\}^2 \middle| \mathcal{F}_t \right],$$

but since

$$\mathsf{E}\left(R_{t+k} \middle| \mathcal{F}_t \right) = 0 \tag{17.10}$$

the conditional variance simplifies to

$$\mathsf{Var}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right) = \mathsf{E}\left\{ \left(\sum_{k=1}^{h} R_{t+k} \right)^2 \middle| \mathcal{F}_t \right\}.$$

As the log returns are conditionally uncorrelated, conditional expectation of the squared sum is equal to the conditional expectation of the sum of squares:

$$\mathsf{Var}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right) = \mathsf{E}\left(\sum_{k=1}^{h} R^2_{t+k} \middle| \mathcal{F}_t \right).$$

Using the linearity of the expectation and equation (17.10), one finally obtains

$$\mathsf{Var}\left(\sum_{k=1}^{h} R_{t+k} \middle| \mathcal{F}_t \right) = \sum_{k=1}^{h} \mathsf{E}\left(R^2_{t+k} \middle| \mathcal{F}_t \right) = \sum_{k=1}^{h} \mathsf{Var}\left(R_{t+k} \middle| \mathcal{F}_t \right).$$

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

$$\mathsf{Var}\,(R_{t+k}|\,\mathcal{F}_t) = \widehat{\sigma}_t^2, \quad k = 1, \ldots, h. \tag{17.11}$$

Therefore the estimated conditional variance of the aggregated returns $R_t + R_{t+1} + \ldots + R_{t+h}$ is simply

$$V_{t,h}^{\heartsuit} = h\widehat{\sigma}_t^2. \tag{17.12}$$

The GARCH(1,1) model describes the volatility dynamics by the relation

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

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 variance forecast of the GARCH(1,1) model is given by:

$$\sigma_{t+h|t}^{2,\spadesuit} \stackrel{\text{def}}{=} \sum_{k=1}^{h} \mathsf{E}\left(R_{t+h}^2\middle|\,\mathcal{F}_t\right) = \bar{\sigma}^2 + (\alpha+\beta)^h(\sigma_t^2 - \bar{\sigma}^2),$$

where $\bar{\sigma}$ is the unconditional volatility. Thus, the conditional variance forecast of the aggregated returns is

$$V_{t,h}^{\spadesuit} = \sum_{k=1}^{h} \sigma_{t+k|t}^{2,\spadesuit}. \tag{17.13}$$

Substituting the expressions (17.9), (17.12) and (17.13) for $V_{t,h}^{\circ}$, $V_{t,h}^{\heartsuit}$ and $V_{t,h}^{\spadesuit}$ respectively in (17.8), one obtains the performance data shown in the Figure 17.7. 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.

17.5.2 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

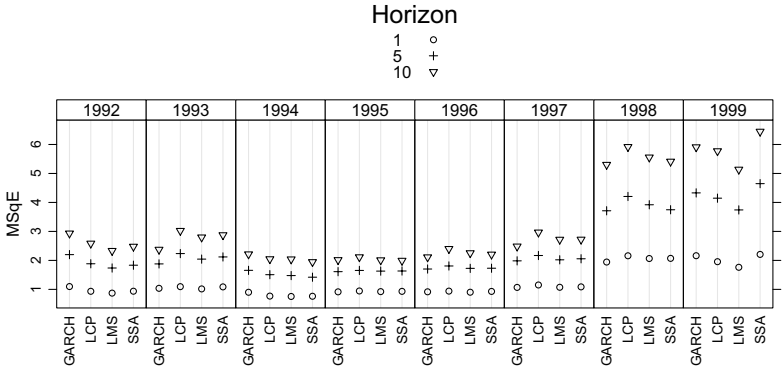

Figure 17.7. Peformance of adaptive methods and GARCH(1,1) in terms of MSqE. `XFGadamethperf`

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:

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

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 17.8 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

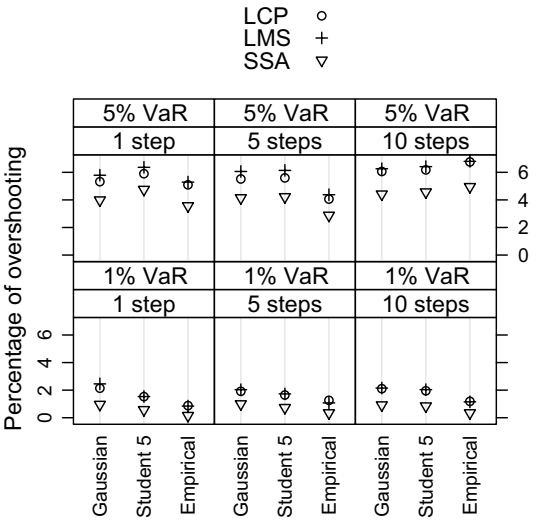

Figure 17.8. Percentage of overshooting the value-at-risk estimated by three methods for various distributions of innovations, number of forecasting steps and value-at-risk levels. Currency: Australian dollar. XFGvarAUD

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.

17.5.3 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 $S_t \in \mathbb{R}^d$, $t = 1, 2, \ldots$ and $R_{t,m}$ is the vector of the corresponding log returns:

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

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 17.9 shows annualized volatility estimated

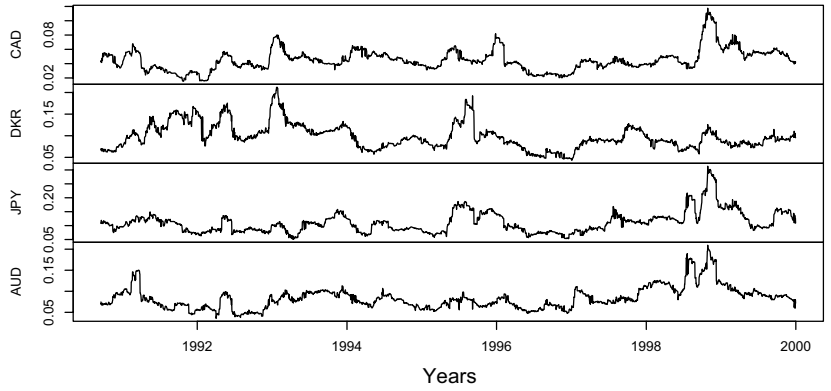

Figure 17.9. Adaptive estimation of the annualized volatility of four exchange rates. XFGcovmatexch

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.

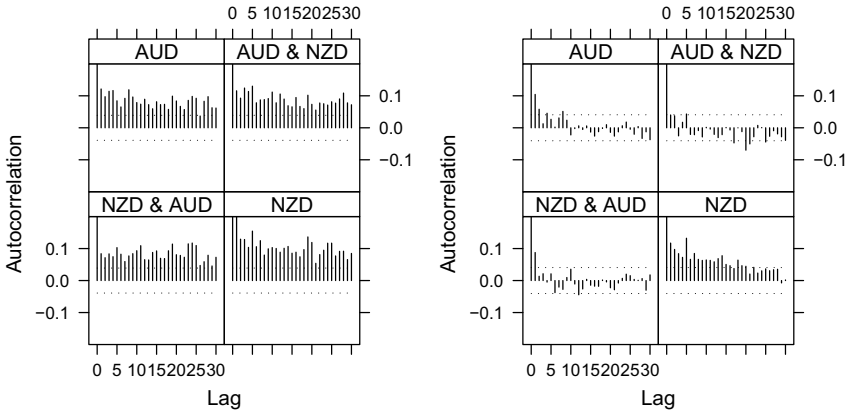

Figure 17.10. ACF for the NZD and AUD time series. Left: absolute log returns, right: absolute standardized log returns. XFGacfabsaud

As in one-dimensional case, we observe significant correlation and autocor-

relation of the absolute log returns (Figure 17.10, 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 standardized absolute log returns by solving the equation

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

for ξ_t. The multivariate ACF plot of the standardized absolute log returns is shown in the right part of Figure 17.10. Although some autocorrelation still remains in the NZD series, the remaining three ACF plots show almost no significant correlation.

Bibliography

Belomestny, D. and 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. and Spokoiny, V. (2003). Dynamic nonparametric filtering with application to volatility estimation, *in* M. G. Akritas and D. N. Politis (eds), *Recent Advances and Trends in Nonparametric Statistics*, Elsevier.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes, *Ann. Statist.* **34**(3): 1075–1114.

Diebold, F. X. and 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. and Gu, J. (2003). Semiparametric estimation of value at risk, *Econom. J.* **6**(2): 261–290.

Giacomini, E., Härdle, W. and 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. and Spokoiny, V. (2008). Spatially adaptive estimation via fitted local likelihood techniques, *ForthcomingIEEE Transactions on Signal Processing.* Forthcoming.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models, *Ann. Statist.* **32**(2): 577–602.

Mikosch, T. and Stărică, C. (2004). Changes of structure in financial time series and the GARCH model, *REVSTAT* **2**(1): 41–73.