Change of structure in financial time series, long range dependence and the GARCH model

Thomas Mikosch
University of Groningen

and

Cătălin Stărică
The Wharton School, Philadelphia
Chalmers University of Technology, Gothenburg

Draft December 7, 1998

Abstract

Functionals of a two–parameter integrated periodogram have been used for a long time for detecting changes in the spectral distribution of a stationary sequence. The bases for these results are functional central limit theorems for the integrated periodogram having as limit a Gaussian field. In the case of GARCH(p,q) processes a statistic closely related to the integrated periodogram can be used for the purpose of change detection in the model. We derive a central limit theorem for this statistic under the hypothesis of a GARCH(p,q) sequence with a finite 4th moment. Simulations show that our statistic is quite sensitive to changes of the parameters in the model and that it accurately detects the moment when a change of the model occurs.

When applied to real–life time series our method gives clear evidence of the fast pace of change in the data. One of the straightforward conclusions of our study is the infeasibility of modelling long return series with one GARCH model. The parameters of the model must be updated and we propose a method to detect when the update is needed.

Our study supports the hypothesis of global non–stationarity of the return time series. We bring forth both theoretical and empirical evidence that the long range dependence (LRD) type behaviour of the sample ACF and the periodogram of absolute return series documented in the econometrics literature could be due to the impact of non–stationarity on these statistical instruments.

Contrary to the common-hold belief that the LRD characteristic carries meaningful information about the price generating process, we show that the LRD behaviour could be just an artifact due to structural changes in the data. The effect that the switch to a different regime has on the sample ACF and the periodogram is theoretically explained and empirically documented using time series that were the object of LRD modelling efforts (S&P500, DEM/USD FX) in various publications.

AMS 1991 Subject Classification: Primary: 62P20 Secondary: 60G10 60F17 60B12 60G15 62M15 62P20

Key Words and Phrases. Integrated periodogram, spectral distribution, functional central limit theorem, Kiefer–Müller process, Brownian bridge sample autocorrelation, change point, GARCH process, long range dependence.

1 Introduction

In fields as diverse as time series analysis and extreme value theory it is generally assumed that the observations or a suitable transformation of them constitute a stationary sequence of random variables. In the context of this paper, stationarity is always understood as strict stationarity. It is the aim of this paper to provide a test for the change from one particular stationary model to another. To be precise, we assume that the data come from a *generalised autoregressive conditionally heteroscedastic process of order* (p, q), for short GARCH(p, q):

$$X_t = \sigma_t \, Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-i}^2 \,, \quad t \in \mathbb{Z} \,, \tag{1.1}$$

where (Z_t) is an iid sequence with $Z = Z_1$, $EZ = 0$, $EZ^2 = 1$ and the *stochastic volatility* σ_t is independent of Z_t for every fixed t. One often assumes Z to be standard normal. In what follows, we write σ for a generic random variable with the distribution of σ_1, X for a generic random variable with the distribution of X_1, etc.,

This kind of model is most popular in the econometrics literature for modelling the log–returns of stock indices, share prices, exchange rates, etc., and has found its way into financial practice for forecasting financial time series. See for example Engle [19] for a collection of papers on ARCH. We assume that, for a particular choice of parameters α_i and β_i, the sequence $((X_t, \sigma_t))$ constitutes a stationary sequence. Assumptions for stationarity of a GARCH(p, q) can be found for example in Bougerol and Picard [9] or Nelson [33].

Our analysis is based on the spectral porperties of the underlying time series. Recall that the *periodogram*

$$I_{n,X}(\lambda) = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathrm{e}^{-i\lambda t} X_t \right|^2 \,, \quad \lambda \in [0, \pi] \,,$$

is the natural (method of moment) estimator of the spectral density f_X of a stationary sequence (X_t); see Brockwell and Davis [10] or Priestley [37]. Under general conditions, the *integrated periodogram* or *empirical spectral distribution function*

$$\frac{1}{2\pi} J_{n,X}(\lambda) = \frac{1}{2\pi} \int_0^{\lambda} I_{n,X}(x) \, dx \,, \quad \lambda \in [0, \pi] \,, \tag{1.2}$$

is a consistent estimator of the *spectral distribution function* given by

$$F_X(\lambda) = \int_0^{\lambda} f_X(x) \, dx \,, \quad \lambda \in [0, \pi] \,,$$

provided the density f_X is well defined. Motivated by empirical process theory for iid sequences, various authors have built up a theory for the integrated periodogram and its modifications and ramifications. The main aim is to parallel the theory for the empirical process as much as possible.

For an introduction to the theory of empirical processes, see Pollard [36] or Shorack and Wellner [40]. The idea of treating the integrated periodogram as the empirical spectral distribution function has been used for a long time; see for example Grenander and Rosenblatt [23], Whittle [41] or Bartlett [3]. However, the theory for the integrated periodogram is not identical to the empirical process theory. For example, given a finite 4th moment for X, $\sqrt{n}(J_{n,X} - F_X)$ has a Gaussian process as weak limit in $\mathbb{C}[0,\pi]$, the space of continuous functions on $[0,\pi]$ endowed with the uniform topology, which can be a very unfamiliar process. Its ovariance structure depends on the spectral density f_X; see for example Anderson [2] or Mikosch [31]. Since one wants to use the distributional limit of $\sqrt{n}(J_{n,X} - F_X)$ for the construction of goodness–of–fit tests of the spectral distribution function, as proposed by Grenander and Rosenblatt [23], one needs to modify the integrated periodogram to get a more familiar Gaussian process, if possible a bridge–type process. Bartlett [3] (cf. Priestley [37]) had the idea to use a weighted form of the integrated periodogram:

$$J_{n,X,f_X}(\lambda) = \int_0^\lambda \frac{I_{n,X}(x)}{f_X(x)}\, dx\,, \quad \lambda \in [0,\pi]\,. \tag{1.3}$$

This process, when suitably normalised and centered, has a weak limit which, in the case of a linear process, is independent of the spectral density, but it is still dependent on the 4th moment of the underlying noise process; see for example Mikosch [31] for a discussion. There it is also mentioned that one can overcome the problem with the 4th moment by replacing the periodogram in (1.3) with a self–normalised (studentised) version, i.e.

$$\widetilde{I}_{n,X}(\lambda) = \frac{I_{n,X}(\lambda)}{n^{-1}\sum_{t=1}^n X_t^2}\,.$$

Then, under the assumption of a finite 2nd moment of X and mild conditions on the coefficients of the linear process (which include the long range dependence (LRD) case; see Kokoszka and Mikosch [30]) the self–normalised version $\widetilde{J}_{n,X,f_X}$ of J_{n,X,f_X} satisfies

$$\frac{1}{\sqrt{n}}\left(\widetilde{J}_{n,X,f_X}(\lambda) - \lambda)\right)_{\lambda\in[0,\pi]} \xrightarrow{d} (B(\lambda))_{\lambda\in[0,\pi]}\,,$$

where B denotes a Brownian bridge on $[0,\pi]$ and $\xrightarrow{d}$ stands for convergence in distribution in $\mathbb{C}[0,\pi]$.

The method of the integrated periodogram, as the empirical version of the spectral distribution function, can also be exploited for detecting changes in the spectral distribution function of a time series. This is analogous to the sequential empirical process known from empirical process theory; see for example Shorack and Wellner [40]. To the best of our knowledge, this idea was used first by Picard [35]. It was further developed for various linear processes under mild assumptions on the moments of X and the coefficients of the process; see Giraitis and Leipus [20] or Klüppelberg and Mikosch [29]. In the latter paper the method was applied to financial log–returns in order to test when a change from an iid sample to an ARMA process or vice versa occurred. We describe

the approach of that paper in more detail for one particular case. It will be further developed for constructing a change point test for GARCH processes. The test statistic is constructed from the two–parameter process

$$J_{n,X}(x,\lambda) = \int_0^\lambda \frac{I_{n,[nx],X}(y)}{f_X(y)}\, dy\,, \quad x \in [0,1]\,, \quad \lambda \in [0,\pi]\,, \tag{1.4}$$

where

$$I_{n,k,X}(\lambda) = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{k} \mathrm{e}^{-i\lambda t} X_t \right|^2\,, \quad k = 0,\ldots,n\,, \quad \lambda \in [0,\pi]\,.$$

Again, Bartlett's idea (divide the periodogram by the spectral density) was used to make the limit process independent of the form of the spectral density. Then, using a proper centring sequence for the process $J_{n,X}(x,\lambda)$, the normalised process converges in distribution in the Skorokhod space $\mathbb{D}([0,1]\times[0,\pi])$ to a two–parameter Gaussian process which, for fixed λ, is a Brownian motion and, for fixed x, a Brownian bridge. Such a process is known as *Kiefer–Müller process*; see Shorack and Wellner [40]. Functionals of $J_{n,X}(x,\lambda)$ can then be used to detect a change in the spectral distribution function. The basic idea is to calculate the empirical spectral distribution function for every subsample $(X_1,\ldots,X_{[nx]})$, $0 \le x \le 1$, and to compare it with the value of the true spectral distribution function. Under the null hypothesis that the whole sample $X_1,\ldots,X_n$ is taken from the same stationary sequence, the constructed change point statistic reacts quite sensitively to deviations from this hypothesis.

It is the aim of this paper to develop a similar change point analysis for GARCH processes. The situation is different from the linear process case insofar that any GARCH(p,q) process represents a white noise sequence and therefore its spectral density is a constant: $f_X \equiv \mathrm{var}(X)/(2\pi)$. If we use a statistic of type $J_{n,X}(x,\lambda)$ we thus test for a change of variance of the underlying time series which is determined by all parameters of the process. In contrast to the linear process case, where $I_{n,X}(\lambda)/f_X(\lambda)$ roughly behaves like $I_{n,Z}(\lambda)$ ((Z_t) is the underlying iid noise sequence), such a behaviour cannot be expected for the GARCH(p,q). In particular, the covariance structure of (X_t^2) determines the structure of the limit process. We explain this result in detail in Section 2, where we give the necessary limit theory for a modified two–parameter integrated periodogram.

In Section 5 we apply the theoretical results of Section 2 in a goodness–of–fir test for foreign exchange log–return data, the S&P500 series and simulated data consisting of subsamples from different GARCH$(1,1)$ models. We fit a GARCH$(1,1)$ to a subsample and compare it to to the overalll fit in the sample. It turns out that there is a very high probability for a change of the the GARCH$(1,1)$ models, in particular in periods of high volatility. Therefore it is reasonable to adjust the parameters to the new situation, and the goodness–of–fit analysis of Section 5 tells us when we have to re–estimate the new parameters.

In real–life financial data one often observes that the sample autocorrelation function of the absolute values and squares of log–returns decay to zero at a hyperbolical rate. Alternatively, the

estimated spectral density blows up at zero. We show in a simulation study in Section 5 that the same LRD effect can be achieved if the underlying sample consists of subsamples from different GARCH$(1,1)$ models, and in Section 3 we give theoretical reasons for this behaviour. We explain the LRD effect in non-stationary sequences by a change of expextation in the different subsamples. This in combination with the empirically observed changeability of real–life log–return models (see Section 5) provides a plausible explanation of the LRD effect in real–life data.

In real–life log–returns one often observes that the estimated parameters α_i and β_i of a fitted GARCH(p,q) model sum up to 1. In Section 4 we provide a theoretical explanation for this fact in the case of a non–stationary time series. In this case the differences between the variances in different subsamples cause this effect on the sum of the parameters. In Section 5 we give some empirical evidence on this observation. Larger values α_i and β_i imply that the tails of the marginal distributions become heavier. Our result on parameter estimation under non–stationarity implies that the tails of the marginal distributions of log–return series might be more light–tailed than one usually assumes.

2 Limit theory for the two–parameter integrated periodogram

In what follows, (X_t) is a GARCH(p,q) as defined in (1.1). Under general conditions, the tail of a GARCH(p,q) is Pareto–like:

$$P(X > x) \sim c\, x^{-\kappa} \quad \text{as } x \to \infty \text{ for some } \kappa > 0 \tag{2.1}$$

depending on the parameters α_i, β_i and the distribution of Z. The determination of κ is in general difficult and goes back to a classical paper by Kesten [27]; see Goldie [21] (cf. Embrechts et al. [17], Section 8.4) for the ARCH(1) case, Mikosch and Stărică [32] for the GARCH$(1,1)$ and Davis et al. [12] for the case of general stochastic recurrence equations including GARCH(p,q). In what follows, *we assume that the parameters α_i, β_i and the noise (Z_t) are chosen such that* (2.1) *holds for some $\kappa > 4$.* Moreover, *we suppose that Z is symmetric.*

As a motivation for the following, we start by considering the two–parameter process $J_{n,X}(x,\lambda)$ from (1.4):

$$\begin{aligned} J_{n,X}(x,\lambda) &= \int_0^\lambda \left(\gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,[nx],X}(h)\cos(yh) \right) dy \\ &= \lambda\gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,[nx],X}(h)\frac{\sin(\lambda h)}{h}\,, \end{aligned} \tag{2.2}$$

where

$$\gamma_{n,[nx],X}(h) = \frac{1}{n}\sum_{t=1}^{[nx]-h} X_t X_{t+h}\,, \quad h = 0,1,2,\dots\,, \quad x \in [0,1]\,.$$

Clearly,

$$\gamma_{n,X}(h) = \gamma_{n,n,X}(h)$$

denotes a version of the sample autocovariance at lag h; the standard version of the sample autocovariance is defined for the centered random variables $X_t - \overline{X}_n$, where $\overline{X}_n$ is the sample mean. We also write

$$\gamma_X(h) = \mathrm{cov}(X_0, X_h) \quad \text{and} \quad v_X(h) = \mathrm{var}(X_0 X_h) = E(X_0 X_h)^2\,, \quad h \in \mathbb{Z}\,.$$

The processes $\gamma_{n,[nx],X}(h)$ satisfy a fairly general functional central limit theorem (FCLT). Recall that $\mathbb{D}([0,1],\mathbb{R}^m)$ is the Skorokhod space of $\mathbb{R}^m$–valued cadlag functions on $[0,1]$ (continuous from the right in $[0,1)$, limits exist from the left in $(0,1]$) endowed with the J_1–topology and the corresponding Borel σ–field; see for example Jacod and Shiryaev [26] or Bickel and Wichura [6].

Lemma 2.1 *If* (2.1) *holds for some* $\kappa > 4$ *then for every* $m \geq 1$,

$$\sqrt{n}\left(\gamma_{n,[nx],X}(h), h = 1\ldots, m\right)_{x\in[0,1]} \stackrel{d}{\to} \left(v_X^{1/2}(h)\; W_h(x), h = 1, \ldots, m\right)_{x\in[0,1]}\,, \quad n \to \infty\,, \tag{2.3}$$

in $\mathbb{D}([0,1],\mathbb{R}^m)$, *where* $W_h(\cdot)$, $h = 0, 1, \ldots, m$, *are iid standard Brownian motions on* $[0,1]$.

Proof. We have to show the convergence of the finite–dimensional distributions and the tightness in $\mathbb{D}([0,1],\mathbb{R}^m)$. Notice first that for every fixed h,

$$\sqrt{n}(\gamma_{n,[nx],X}(h))_{x\in[0,1]} \stackrel{d}{\to} (v_X^{1/2}(h)\; W_h(x))_{x\in[0,1]}\,. \tag{2.4}$$

in $\mathbb{D}[0,1]$; see Oodaira and Yoshihara [34]; cf. Doukhan [16], Theorem 1 on p. 46. In the latter theorem one has to ensure that $E|X_0X_h|^{2+\epsilon} < \infty$ for some $\epsilon > 0$ (this follows from (2.1) for $\kappa > 4$) and that the sequence (X_tX_{t+h}) is α–mixing with a sufficiently fast rate for the mixing coefficients; see (2.5). However, the GARCH(p,q) is strongly mixing with geometric rate (see for example Davis et al. [12]), and so the mixing coefficients converge to zero at an exponential rate, which implies the conditions in the aforementioned theorem.

Thus each of the processes $\sqrt{n}\gamma_{n,[n\cdot],X}(h)$ is tight in $\mathbb{D}[0,1]$. Using a generalisation of the argument for Lemma 4.4 in Resnick [38], one obtains that the map from $(\mathbb{D}[0,1])^m$ into $\mathbb{D}([0,1],\mathbb{R}^m)$ defined by

$$(x_1, \ldots, x_m) \to (x_1(t), \ldots, x_m(t))_{t\geq 0}$$

is continuous at $(x_1, \ldots, x_m)$ in $(\mathbb{C}[0,1])^m$. This and the sample path continuity of the limit process ensure that the processes on the left–hand side of (2.3) are tight in $\mathbb{D}([0,1],\mathbb{R}^m)$.

Notice that the multivariate CLT

$$\sum_{t=1}^{[nx]} (X_tX_{t+1}, \ldots, X_tX_{t+h}) \stackrel{d}{\to} \left(v_X^{1/2}(1)W_1(x), \ldots, v_X^{1/2}(h)W_h(x)\right)$$

holds for every fixed x. This is again a consequence of the aforementioned CLT for α–mixing sequences in combination with the Cramér–Wold device. A similar argument for a finite number of x–values yields the convergence of the finite–dimensional distributions. This proves the lemma. □

Remark 2.2 It follows from the argument above that (2.3) remains valid for stationary strongly mixing sequences (X_t) with $EX = 0$, $E|X|^{4+\delta} < \infty$ for some $\delta > 0$ and such that $EX_0X_h = 0$ for $h \geq 1$, $\mathrm{cov}(X_0X_h, X_0X_l) = 0$ for all $h \neq l \geq 1$, and with α–mixing coefficients $\widetilde{\alpha}_i$ satisfying

$$\sum_{i=1}^{\infty} \widetilde{\alpha}_i^{\delta/(2+\delta)} < \infty \,. \tag{2.5}$$

The latter conditions are needed for the validity of the FCLT (2.4); see Oodaira and Yoshihara [34].

A naive argument, based on Lemma 2.1 and the decomposition (2.2), suggests that

$$\sqrt{n}(J_{n,X}(x,\lambda) - \lambda\gamma_{n,[nx],X}(0))_{x\in[0,1],\lambda\in[0,\pi]} \quad \xrightarrow{d} \quad 2\left(\sum_{h=1}^{\infty} v_X^{1/2}(h)\, W_h(x)\, \frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]},$$

in $\mathbb{D}([0,1]\times[0,\pi])$. This result can be shown to be true; one can follow the lines of the proof of Theorem 2.3 below. However, the two–parameter Gaussian limit field has a distribution that explicitly depends on the covariance structure of (X_t^2), which is not a very desirable property. Indeed, since we are interested in using functionals of the limit process, we would like to know the distribution of those functionals for the purposes of statistical inference. Therefore we want a "standard" Gaussian process in the limit; otherwise we would have to calculate the distributions of its functionals by Monte–Carlo simulations *for every choice of parameters of the* GARCH(p,q).

A glance at the right–hand side of (2.2) suggests another approach. The dependence of the limiting Gaussian field on the covariance structure of (X_t^2) comes in through the FCLT of Lemma 2.1. However, if we replaced in (2.2) the processes $\gamma_{n,[n\cdot],X}(h)$ with

$$\widetilde{\gamma}_{n,[nx],X}(h) = \frac{\gamma_{n,[n\cdot],X}(h)}{v_X^{1/2}(h)},$$

a naive argument would suggest that the limit process becomes independent of the covariance structure of (X_t^2).

Therefore we introduce the following two–parameter process which is a straightforward modification of $J_{n,X}(x,\lambda)$:

$$C_{n,X}(x,\lambda) = \sum_{h=1}^{n-1} \widetilde{\gamma}_{n,[nx],X}(h)\, \frac{\sin(\lambda h)}{h}, \quad x \in [0,1], \quad \lambda \in [0,\pi].$$

Our main result is a FCLT for $C_{n,X}$.

Theorem 2.3 *If* (2.1) *holds for some* $\kappa > 4$,

$$\sqrt{n}\,(C_{n,X}(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} \quad \stackrel{d}{\to} \quad (K(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} = \left(\sum_{h=1}^{\infty} W_h(x)\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]},$$

in $\mathbb{D}([0,1]\times[0,\pi])$ *where* $(W_h(\cdot))$ *is a sequence of iid standard Brownian motions on* $[0,1]$. *The infinite series on the right–hand side converges with probability* 1 *and represents a Kiefer–Müller process, i.e. a two–parameter Gaussian field with covariance structure*

$$\begin{aligned} E\left(K\left(x_1,\lambda_1\right)K\left(x_2,\lambda_2\right)\right) &= \min\left(x_1,x_2\right)\sum_{t=1}^{\infty}\frac{\sin\left(\lambda_1 t\right)\sin\left(\lambda_2 t\right)}{t^2} \\ &= 2^{-1}\pi^2\min\left(x_1,x_2\right)\left(\min\left(\frac{\lambda_1}{\pi},\frac{\lambda_2}{\pi}\right)-\frac{\lambda_1}{\pi}\frac{\lambda_2}{\pi}\right). \end{aligned} \tag{2.6}$$

Proof. We proceed analogously to Klüppelberg and Mikosch [29]. It follows from Lemma 2.1 and the continuous mapping theorem that, for every fixed $m \geq 1$, in $\mathbb{D}([0,1]\times[0,\pi])$

$$\sum_{h=1}^{m}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h} \;\stackrel{d}{\to}\; \sum_{h=1}^{m} W_h(x)\,\frac{\sin(\lambda h)}{h}.$$

According to Theorem 4.2 in Billingsley [7], it remains to show that for every $\epsilon > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left(\sup_{0\leq x\leq 1}\sup_{0\leq\lambda\leq\pi}\left|\sum_{h=m+1}^{[nx]}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right)=0. \tag{2.7}$$

Since Z is symmetric the sequences $(r_t)=(\mathrm{sign}(X_t))$ and $(|X_t|)$ are independent. Conditionally on $(|X_t|)$,

$$\sum_{h=m+1}^{k}\sqrt{n}\,\widetilde{\gamma}_{n,k,X}(h)\frac{\sin(\cdot h)}{h}, \quad k=m+1,\ldots,n-1,$$

is a sequence of quadratic forms in the iid Rademacher random variables r_t and with values in the Banach space $\mathbb{C}[0,\pi]$ endowed with the sup–norm. Now condition on $(|X_t|)$. Use first a decoupling inequality for Rademacher quadratic forms (e.g. de la Peña and Montgomery–Smith [13], Theorem 1) then the Lévy maximal inequality for sums of iid symmetric random variables, then again the decoupling inequality in reverse order, and finally take expectations with respect to $(|X_t|)$. Then we obtain the inequality

$$\begin{aligned} &P\left(\sup_{0\leq x\leq 1}\sup_{0\leq\lambda\leq\pi}\left|\sum_{h=m+1}^{[nx]}\sqrt{n}\,\widetilde{\gamma}_{n,[nx],X}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right) \\ &\leq\; c_1\,P\left(c_2\sup_{0\leq\lambda\leq\pi}\left|\sum_{h=m+1}^{n-1}\sqrt{n}\,\widetilde{\gamma}_{nX}(h)\,\frac{\sin(\lambda h)}{h}\right|>\epsilon\right) \end{aligned}$$

for certain positive constants c_1, c_2. The right–hand probability can be treated in the same way as the derivation of (6.3) in [28], pp. 1873–1876. Instead of Theorem 3.1 in Rosiński and Woyczyński [39] one can simply use the Cauchy–Schwarz inequality in the first display on p. 1876 in [28] with $\mu = 2$. Then all the calculations for (6.3) remain valid, implying that (2.7) holds. This concludes the proof of the theorem. □

Remark 2.4 The series representation of the Kiefer–Müller process can be found in Klüppelberg and Mikosch [29]. This process is known in empirical process theory as the limiting Gaussian field for the sequential empirical process; see Shorack and Wellner [40].

Remark 2.5 The statement of Theorem 2.3 remains valid for wider classes of stationary sequences. Assume the conditions in Remark 2.2 are satisfied. In addition, we need that (X_t) is symmetric and that $(|X_t|)$ and $(\mathrm{sign}(X_t))$ are independent. This condition is satisfied for any stochastic volatility model of form $X_t = \sigma_t Z_t$, where (Z_t) is a sequence of iid symmetric random variables and the random variables σ_t are adapted to the filtration $\sigma(Z_{t-1}, Z_{t-2}, \ldots)$.

Remark 2.6 The condition of symmetry of Z is needed only for the application of the Lévy maximal inequality for sums of independent random variables. Alternatively, one can proceed as in the proof of Theorem 3.1 in Klüppelberg and Mikosch [29], p. 980, last display, where instead of the Lévy maximal inequality Doob's 2nd moment maximal inequality for submartingales was used. Then one can follow the lines of the proof of Theorem 1 in Grenander and Rosenblatt [23], Chapter 6.4.

Remark 2.7 Theorem 2.3 can be used for testing the null hypothesis that the whole sample $X_1, \ldots, X_n$ comes from a particular GARCH(p, q) model with given parameters α_i and β_i. Now assume that $X_1, \ldots, X_n$ actually comes from a model with parameters $\widehat{\alpha}_i$ and $\widehat{\beta}_i$. Write $\widehat{v}_X(h) = E(X_0 X_h)^2$ for the corresponding variance of $X_0 X_h$. Then the same arguments as for the proof of Theorem 2.3 yield that, if (2.1) holds for some $\kappa > 4$, then

$$\sqrt{n}\,(C_{n,X}(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} \xrightarrow{d} \left(\sum_{h=1}^{\infty} \frac{\widehat{v}_X^{1/2}(h)}{v_X^{1/2}(h)}\, W_h(x)\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]},$$

in $\mathbb{D}([0,1]\times[0,\pi])$ where $(W_h(\cdot))$ is a sequence of iid standard Brownian motions on $[0,1]$.

Similar results can be derived under the alternative hypothesis that the sample $X_1, \ldots, X_n$ consists of subsamples from different GARCH(p, q) processes. However, the resulting Gaussian limit field is then dependent on the covariance structures of the X_ts in the different pieces of the sample. This fact makes an application of such a result extremely difficult; one would depend on Monte–Carlo based calculations for the quantiles of the functionals of those processes. In particular, these quantiles would depend on the parameters of the different GARCH(p, q) processes.

Immediate consequences of Theorem 2.3 are limit theorems for continuous functionals of the process $C_{n,X}$.

Corollary 2.8 *Under the assumptions of Theorem 2.3,*

$$\sqrt{n} \sup_{x\in[0,1],\lambda\in[0,\pi]} |C_{n,X}(x,\lambda)| \xrightarrow{d} \sup_{x\in[0,1],\lambda\in[0,\pi]} |K(x,\lambda)| .$$

$$n\int_0^1\int_0^\pi C_{n,X}^2(x,\lambda)\, dxd\lambda \xrightarrow{d} \int_0^1\int_0^\pi K^2(x,\lambda)\, dxd\lambda .$$

Setting $x = 1$, we also notice that

$$\sqrt{n}\,\widetilde{C}_{n,X}(\cdot) := \sqrt{n}\sum_{h=1}^{n-1} \frac{\gamma_{n,X}(h)}{v_X^{1/2}(h)} \frac{\sin(\cdot h)}{h} \xrightarrow{d} B(\cdot) := \sum_{h=1}^{\infty} W_h(1)\, \frac{\sin(\cdot h)}{h}, \tag{2.8}$$

where convergence happens in $\mathbb{C}[0,\pi]$. The series on the right–hand side represents a Brownian bridge on $[0,\pi]$; see (2.6) with $x = 1$. It is the so–called Paley–Wiener representation; see for example Hida [25]. The one–parameter process $\widetilde{C}_{n,X}$ is an alternative to $J_{n,X}$ in (1.2) and can be applied for testing the goodness of fit of the sample $X_1,\ldots,X_n$. The convergence of the following functionals can be used for constructing Kolmogorov–Smirnov– and Cramér–von Mises–type goodness–of–fit tests for a GARCH(p,q) process.

Corollary 2.9 *Under the assumptions of Theorem 2.3,*

$$\sqrt{n} \sup_{\lambda\in[0,\pi]} \left|\widetilde{C}_{n,X}(\lambda)\right| \xrightarrow{d} \sup_{\lambda\in[0,\pi]} |B(\lambda)| , \tag{2.9}$$

$$n\int_0^\pi \widetilde{C}_{n,X}^2(\lambda)\, d\lambda \xrightarrow{d} \int_0^\pi B^2(\lambda)\, d\lambda .$$

For comparison we also cite here a consequence of the results and proofs in Klüppelberg and Mikosch [29] for the case of iid Z_t. The following statement is a modification of their Proposition 2.2.

Proposition 2.10 *Let (Z_t) be an iid sequence with $EZ = 0$ and $\mathrm{var}(Z) < \infty$. Then*

$$\left(\sum_{h=1}^{n-1} \gamma_{n,[nx],Z}(h)\, \frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]} \xrightarrow{d} \mathrm{var}(Z)\ (K(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} ,$$

$$\left(\sum_{h=1}^{n-1} \frac{\gamma_{n,[nx],Z}(h)}{\gamma_{n,Z}(0)}\, \frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]} \xrightarrow{d} (K(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} ,$$

in $\mathbb{D}([0,1]\times[0,\pi])$ and

$$\left(\sum_{h=1}^{n-1} \gamma_{n,Z}(h)\, \frac{\sin(\lambda h)}{h}\right)_{\lambda\in[0,\pi]} \xrightarrow{d} \mathrm{var}(Z)\ (B(\lambda))_{\lambda\in[0,\pi]} ,$$

$$\left(\sum_{h=1}^{n-1} \frac{\gamma_{n,Z}(h)}{\gamma_{n,Z}(0)}\, \frac{\sin(\lambda h)}{h}\right)_{\lambda\in[0,\pi]} \xrightarrow{d} (B(\lambda))_{\lambda\in[0,\pi]} ,$$

in $\mathbb{C}[0,\pi]$.

These results allow one to compare goodness–of–fit tests and change point analyses for different kinds of white noise processes: iid sequences (Z_t) or GARCH(p,q) (X_t).

3 Long range dependence effects for non–stationary sequences

Financial log–return series such as the log–returns of foreign exchange (FX) rates often exhibit an interesting dependence structure which is described as follows: the sample autocorrelations of *long* log–return series are very small, whereas the sample autocorrelation functions (ACFs) of the absolute values and squares of the data usually decay to zero at a hyperbolic rate or even stay constant for a large number of lags. The slow decay of the sample ACF also manifests itself in large values of the estimated (via the periodogram) spectral density of the data for small frequencies.

This phenomenon is often interpreted as *long range dependence* (LRD). There does not exist a unique definition of LRD. One possible way to define it for a stationary sequence (Y_t) is via the condition $\sum_h |\gamma_Y(h)| = \infty$. Alternatively, one can require that the spectral density $f_Y(\lambda)$ of the sequence (Y_t) is asymptotically of the order $L(\lambda)\lambda^{-d}$ for some $d > 0$ and a slowly varying function L, as $\lambda \to 0$. See Beran [4] for various definitions of LRD. Since the detection of LRD effects is based on statistics of the underlying time series similar phenomena can also be observed for the sample ACF and the periodogram of data from non–stationary time series. For example, Bhattacharya et al. [5] detected such effects when the data contain a trend.

Similar effects can be observed for time series which consist of subsamples originating from different stationary models. In what follows we want to give some simple theoretical reasons for the appearance of LRD effects in the sample ACF and the estimated spectrum.

Let p_j, $j = 0, \ldots, r$, be positive numbers such that $p_1 + \cdots + p_r = 1$ and $p_0 = 0$. Define

$$q_j = p_0 + \cdots + p_j\,, \quad j = 0, \ldots, r\,.$$

Assume the sample $Y_1, \ldots, Y_n$ consists of r subsamples

$$Y_1^{(1)}, \ldots, Y_{[nq_1]}^{(1)}, \ldots, Y_{[nq_r]+1}^{(r)}, \ldots, Y_n^{(r)}\,. \tag{3.1}$$

The ith subsample comes from a stationary ergodic model with finite 2nd moment and spectral density $f_{Y^{(i)}}$. Define the sample autocovariances of the sequence (Y_t) as follows:

$$\widetilde{\gamma}_{n,Y}(h) = \frac{1}{n}\sum_{t=1}^{n-h} Y_t Y_{t+h} - (\overline{Y}_n)^2\,, \quad h \in \mathbb{Z}\,,$$

where $\overline{Y}_n$ denotes the sample mean. By the ergodic theorem it follows for fixed $h \geq 0$ as $n \to \infty$ that

$$\widetilde{\gamma}_{n,Y}(h) \;=\; \sum_{j=1}^{r} p_j \,\frac{1}{np_j} \sum_{t=[nq_{j-1}]+1}^{[nq_j]} Y_t^{(j)} Y_{t+h}^{(j)} - \left(\sum_{j=1}^{r} p_j \,\frac{1}{np_j} \sum_{t=[nq_{j-1}]+1}^{[nq_j]} Y_t^{(j)} \right)^2 + o(1)$$

$$\xrightarrow{P} \sum_{j=1}^{r} p_j E(Y_0^{(j)} Y_h^{(j)}) - \left(\sum_{j=1}^{r} p_j EY^{(j)} \right)^2$$

$$= \sum_{j=1}^{r} p_j \gamma_{Y^{(j)}}(h) + \sum_{1 \le i < j \le n} p_i p_j \left(EY^{(j)} - EY^{(i)} \right)^2 \quad \text{a.s.} \tag{3.2}$$

¿From (3.2) we can explain the LRD effect in the sample ACF: if the expectations differ in the subsequences $(Y_t^{(j)})$ and if the autocovariances $\gamma_{Y^{(j)}}(h)$ decay to zero exponentially as $h \to \infty$, the sample ACF $(\widetilde{\gamma}_{n,Y}(h))$ for sufficiently large h is close to a strictly positive constant given by the second term in (3.2). The overall picture should show a sample ACF $(\widetilde{\gamma}_{n,Y}(h))$ that decays exponentially for small lags and approaches a positive constant for larger lags. The presence of the positive constant in (3.2) forbids negative correlations for larger lags. For example, if (X_t) is a sequence which consists of pieces from different GARCH processes, the above argument applies to the sequences $(Y_t) = (|X_t|^l)$, $l = 1, 2$. See Section 5 for some evidence from simulated data.

Alternatively, one may consider estimates of the spectral density. The classical estimator in this case is the (smoothed) periodogram which is considered only at the Fourier frequencies

$$\lambda_j = \frac{2\pi j}{n} \in (-\pi, \pi]\,. \tag{3.3}$$

For convenience we exclude the Fourier frequencies at 0 and π. Since $\sum_{t=1}^{n} \mathrm{e}^{-i\lambda_j t} = 0$, the periodogram at the Fourier frequencies does not change its value if one replaces the Y_ts with the centered random variables $Y_t - c$, $t = 1, \ldots, n$, for any constant c, therefore centring is not necessary for the sample. We observe the following:

$$I_{n,Y}(\lambda_j) = \left| \frac{1}{\sqrt{n}} \sum_{l=1}^{r} \sum_{t=[nq_{l-1}]+1}^{[nq_l]} Y_t^{(l)} \mathrm{e}^{-i\lambda_j t} \right|^2$$

$$= \left| \frac{1}{\sqrt{n}} \sum_{l=1}^{r} \sum_{t=[nq_{l-1}]+1}^{[nq_l]} (Y_t^{(l)} - EY^{(l)}) \mathrm{e}^{-i\lambda_j t} + \frac{1}{\sqrt{n}} \sum_{l=1}^{r} EY^{(l)} \sum_{t=[nq_{l-1}]+1}^{[nq_l]} \mathrm{e}^{-i\lambda_j t} \right|^2 .$$

Notice that

$$\sum_{l=1}^{r} EY^{(l)} \mathrm{e}^{-i\lambda_j([nq_{l-1}]+1)} \sum_{t=0}^{[nq_l]-[nq_{l-1}]-1} \mathrm{e}^{-i\lambda_j t}$$

$$= \frac{\mathrm{e}^{-i\lambda_j}}{1 - \mathrm{e}^{-i\lambda_j}} \sum_{l=1}^{r} EY^{(l)} \left(\mathrm{e}^{-i\lambda_j [nq_{l-1}]} - \mathrm{e}^{-i\lambda_j [nq_l]} \right)$$

$$= \frac{\mathrm{e}^{-i\lambda_j}}{1 - \mathrm{e}^{-i\lambda_j}} \left[EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1} (EY^{(l)} - EY^{(l+1)}) \mathrm{e}^{-i\lambda_j [nq_l]} \right]$$

does not sum up to zero if the expectations $EY^{(j)}$s vary with j. Assuming uncorrelatedness between different subsamples, straightforward calculation yields for $\lambda_j \to 0$ that

$$
\begin{aligned}
& EI_{n,Y}(\lambda_j) \\
&= \sum_{l=1}^{r} p_l E \left| \frac{1}{\sqrt{np_l}} \sum_{t=1}^{[nq_l]-[nq_{l-1}]-1} (Y_t^{(l)} - EY^{(l)})\, \mathrm{e}^{-i\lambda_j t} \right|^2 + \left| \frac{1}{\sqrt{n}} \sum_{l=1}^{r} EY^{(l)} \sum_{t=[nq_{l-1}]+1}^{[nq_l]} \mathrm{e}^{-i\lambda_j t} \right|^2 \\
&= \sum_{l=1}^{r} p_l \left(\mathrm{var}(Y^{(l)}) + 2 \sum_{h=1}^{[np_l]-1} \left(1 - \frac{h}{[np_l]}\right) \gamma_{Y^{(l)}}(h)\ \cos(\lambda_j h) \right) \\
&\quad + \frac{1}{n} \frac{1}{\left|1 - \mathrm{e}^{-i\lambda_j}\right|^2} \left| EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1} (EY^{(l)} - EY^{(l+1)}) \mathrm{e}^{-i\lambda_j [nq_l]} \right|^2 + o(1) \\
&= \sum_{l=1}^{r} p_l [2\pi f_{Y^{(l)}}(\lambda_j)] \\
&\quad + \frac{1}{n} \frac{1}{\left|1 - \mathrm{e}^{-i\lambda_j}\right|^2} \left| EY^{(1)} - EY^{(r)} - (1 + o(1)) \sum_{l=1}^{r-1} (EY^{(l)} - EY^{(l+1)}) \mathrm{e}^{-i2\pi j q_l} \right|^2 + o(1) \\
&\sim \sum_{l=1}^{r} p_l [2\pi f_{Y^{(l)}}(\lambda_j)] + \frac{1}{n\lambda_j^2} \left| EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1} (EY^{(l)} - EY^{(l+1)}) \mathrm{e}^{-i2\pi j q_l} \right|^2 ,
\end{aligned}
\tag{3.4}
$$

where $f_{Y^{(l)}}$ denotes the spectral density of the sequence $(Y_t^{(l)})$.

Now assume that each of the subsequences $(Y_t^{(l)})$ has a continuous spectral density $f_{Y^{(l)}}$ on $[0, \pi]$. Then the first term in formula (3.4) is bounded for all frequencies λ_j, in particular for small ones. If $n\lambda_j^2 \to 0$ as $n \to \infty$, the order of magnitude of the second term in (3.4) is determined by $(n\lambda_j^2)^{-1}$. For the sake of illustration, assume $r = 2$. Then (3.4) turns into

$$
p_1 [2\pi f_{Y^{(1)}}(\lambda_j)] + p_2 [2\pi f_{Y^{(2)}}(\lambda_j)] + \frac{1}{n\lambda_j^2} \left| EY^{(1)} - EY^{(2)} \right|^2 2(1 - \cos(2\pi j p_1)\,. \tag{3.5}
$$

Under the assumption $EY^{(1)} \neq EY^{(2)}$, the right–hand probability for small $n\lambda_j^2$ is of the order

$$
(n\lambda_j^2)^{-1} (1 - \cos(2\pi \{j p_1\})\,, \tag{3.6}
$$

where $\{x\}$ denotes the fractional part of x. Now assume that p_1 is a rational number with representation $p_1 = r_1/r_2$ for relatively prime integers r_1 and r_2. Then $\{jp_1\}$ assumes values $0, r_2^{-1}, \ldots, (r_2 - 1)r_2^{-1}$. Thus, for j such that $n\lambda_j^2$ is small, the quantity (3.6) is either zero or bounded away from zero, uniformly for all j. The effect on (3.5) is that this quantity becomes arbitrarily large for various small values of j as $n \to \infty$ and is bounded from below by the weighted

sum of the spectral densities

$$p_1[2\pi f_{Y^{(1)}}(\lambda_j)] + p_2[2\pi f_{Y^{(2)}}(\lambda_j)]\,.$$

This fact, to some extent, explains the behaviour of the periodogram and smoothed versions of it for Fourier frequencies close to zero. In particular, if $(Y_t) = (|X_t|^l)$ for $l = 1$ or $l = 2$ and (X_t) consists of pieces of different GARCH processes, the above argument applies. Notice that the spectral densities corresponding to these subsequences are then continuous functions on $[0, \pi]$, and the blow up of the estimated spectrum at zero can be explained by the appearance of different expectations of the absolute values or squares in different subsamples. For some empirical evidence on this phenomenon, see Section 5.

4 The effect of non–stationarity on parameter estimation

The model estimation procedures for a GARCH(p, q) are also affected by non–stationarity of the data. In what follows, we are interested in assesing the impact of the presence of subsamples originating from different stationary GARCH$(1, 1)$ models on the Whittle estimates of the parameters of a global GARCH$(1, 1)$ model fit to the entire time series.

If one assumes that the whole sample comes from a GARCH$(1, 1)$ model with parameters α_1 and β_1, it follows from the calculations in the Appendix that $(U_t) = (X_t^2 - EX^2)$ can be rewritten as an ARMA(1,1) process with white noise innovations sequence $(\nu_t) = (X_t^2 - \sigma_t^2)$:

$$U_t - \varphi_1\, U_{t-1} = \nu_t - \beta_1 \nu_{t-j}\,, \quad t \in \mathbb{Z}\,, \tag{4.1}$$

where

$$\varphi_1 = \alpha_1 + \beta_1 \quad \text{and} \quad \Theta = (\varphi_1, \beta_1)\,.$$

One of the backbones of classical estimation theory in the analysis of causal invertible ARMA processes is the Whittle estimator. It is asymptotic equivalent to the least squares and Gaussian maximum likelihood estimates for causal and invertible ARMA processes with iid innovations in the sense that they yield consistent and asymptotic normal (with $\sqrt{n}$–rate) approximations to the true parameters of the model. We refer to Brockwell and Davis [10], Section 10.8, for the theoretical properties of this pseodo–maximum likelihood estimation procedure. The Whittle estimate $\Theta_n = (\overline{\varphi}_1, \overline{\beta}_1)$ of the ARMA$(1, 1)$ model (4.1) is obtained by minimizing the function

$$\overline{\sigma}_n^2(\Theta) = \frac{1}{n} \sum_j \frac{I_{n,U}(\lambda_j)}{f(\lambda_j, \Theta)} \tag{4.2}$$

with respect to Θ from the parameter domain

$$\mathcal{C} = \{(\varphi_1, \beta_1) :\ 0 \le \varphi_1, \beta_1 < 1\,, \quad \varphi_1, \beta_1 \ge 0\}\,.$$

The sum $\sum_j$ in (4.2) is taken over all Fourier frequencies $\lambda_j \in (-\pi, \pi]$ and $f(\lambda, \Theta)$ denotes the spectral density of the ARMA$(1,1)$ process (U_t) (see for example Brockwell and Davis [10], Chapter 4):

$$f(\lambda, \Theta) = \frac{\sigma_U^2}{2\pi} \frac{|1 - \beta_1 \mathrm{e}^{-i\lambda}|^2}{|1 - \varphi_1 \mathrm{e}^{-i\lambda}|^2} = \frac{\sigma_U^2}{2\pi} \frac{1 - 2\beta_1 \cos\lambda + \beta_1^2}{1 - 2\varphi_1 \cos\lambda + \varphi_1^2}. \tag{4.3}$$

Given $EX^4 < \infty$, the Whittle estimates of the parameters of a causal invertible stationary ergodic ARMA(p,q) process (U_t) with white noise innovation sequence (ν_t) are strongly consistent. This follows along the lines of the proof of Theorem 10.8.1 in Brockwell and Davis [10]. Therein, strong consistency has been proved for an ARMA(p,q) process with an iid white noise innovation sequence. However, a close inspection of pp. 378–385 in [10] shows that for the consistency of the Whittle estimates only the strict stationarity and ergodicity of the ARMA(p,q) process are required.

In practice, the GARCH$(1,1)$ model is one of the most frequently used models for fitting real–life financial time series. As a matter of fact, one often estimates a value $\varphi_1 \approx 1$ for long time series; see the empirical study in Section 5. This led to the introduction of the integrated GARCH models (IGARCH) corresponding to $\varphi_1 = 1$; see Engle and Bollerslev [18]. Under general conditions on Z, the stationary IGARCH process has infinite variance marginal distributions. This fact is in contrast to statistical evidence from tail estimation for log–return series. The case of an infinite variance is quite a rare event for financial time series; see for example the discussion on tail estimation in Chapter 6 of Embrechts et al. [17].

In what follows, we give reasons for the empirical fact $\varphi_1 \approx 1$. We will explain that this is an artifact which is due to non–stationarity in the data.

¿From now on, assume that the sample $X_1, \ldots, X_n$ consists of r subsamples as described in (3.1) from different GARCH$(1,1)$ models with corresponding parameters $\Theta^{(i)} = (\varphi_1^{(i)}, \beta_1^{(i)})$, $i = 1, \ldots, r$.

When (X_t^2) constitutes a stationary sequence centering in the sample $X_1^2, \ldots, X_n^2$ is not necessary in the definition of $\overline{\sigma}_n^2(\Theta)$ since $\sum_{t=1}^n \mathrm{e}^{-i\lambda_j t} = 0$ for $\lambda_j \neq 0$ and $\neq \pi$. (The two summands for $\lambda_j = 0$ and $\lambda_j = \pi$ in the definition of the Whittle likelihood $\overline{\sigma}_n^2(\Theta)$ can be neglected; they are of no importance for the asymptotic results considered. In what follows, we simply ignore these two special case.) Thus we have for the Fourier frequencies λ_j that

$$I_{n,X^2}(\lambda_j) = I_{n,U}(\lambda_j)\,, \quad \lambda_j \in (0, \pi)\,,$$

and therefore it is without loss of generality assumed in [10] that the sample is mean–corrected. In practice one does not know the value of EX_t^2. Therefore one has to mean–correct X_t^2, for example with the sample mean

$$\overline{X_n^2} = \frac{1}{n} \sum_{t=1}^n X_t^4\,.$$

It is natural to modify $\overline{\sigma}_n^2(\Theta)$ as follows:

$$\widehat{\sigma}_n^2(\Theta) = \frac{1}{n} \sum_j \frac{I_{n,X^2-\overline{X_n^2}}(\lambda_j)}{f(\lambda_j, \Theta)} .$$

By the above remark, $\overline{\sigma}_n^2(\Theta) = \widehat{\sigma}_n^2(\Theta)$.

We start with an analogue of Proposition 10.8.2 in [10].

Proposition 4.1 *Let $X_1, \ldots, X_n$ be a sample consisting of r subsamples as described in* (3.1). *Assume that the ith subsample comes from a* GARCH$(1,1)$ *model with parameters $\Theta^{(i)} = (\varphi_1^{(i)}, \beta_1^{(i)}) \in \mathcal{C}$ and that $E(X^{(i)})^4 < \infty$.. Then, for every $\Theta \in \mathcal{C}$ the following relation holds:*

$$\widehat{\sigma}_n^2(\Theta) \stackrel{\text{a.s.}}{\to}$$

$$\Delta(\Theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sum_{i=1}^r p_i \sigma^2_{(X^{(i)})^2} f(\lambda, \Theta^{(i)})}{f(\lambda, \Theta)} \, d\lambda + \frac{\sum_{1 \le i < j \le r} p_i p_j \left(\sigma^2_{X^{(i)}} - \sigma^2_{X^{(j)}} \right)^2}{f(0, \Theta)} , \tag{4.4}$$

where $\sigma_A^2 = \mathrm{var}(A)$. Moreover, for every $\delta > 0$, defining

$$f_\delta(\lambda, \Theta) = \frac{\left| 1 - \beta_1 \, \mathrm{e}^{-i\lambda} \right|^2 + \delta}{\left| 1 - \varphi_1 \, \mathrm{e}^{-i\lambda} \right|^2} ,$$

the following relation holds

$$\frac{1}{n} \sum_j \frac{I_{n,X^2-\overline{X_n^2}}(\lambda_j)}{f_\delta(\lambda_j, \Theta)}$$

$$\to \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sum_{i=1}^r p_i \sigma^2_{(X^{(i)})^2} f(\lambda, \Theta^{(i)})}{f_\delta(\lambda, \Theta)} \, d\lambda + \frac{\sum_{1 \le i < j \le r} p_i p_j \left(\sigma^2_{X^{(i)}} - \sigma^2_{X^{(j)}} \right)^2}{f_\delta(0, \Theta)} \tag{4.5}$$

uniformly in $\Theta \in \overline{\mathcal{C}}$, the closure of $\mathcal{C}$, almost surely.

Remark 4.2 The dependence structure between the different subsamples is inessential for the validity of the proposition.

Proof. For simplicity of presentation we restrict ourselves to the case of two subsamples. The general case is analogous. We follow the lines of proof of Proposition 10.8.2 in [10] specified to the ARMA$(1,1)$ process (X_t^2). Since each of the subsamples comes from a strictly stationary and ergodic model, both $((X_t^{(i)})^2)$ constitute stationary and ergodic sequences with $E(X^{(i)})^4 < \infty$, $i = 1, 2$. As in [10], we restrict ourselves to show that (4.5) is satisfied. The same arguments as on pp. 378–379 in [10] apply. The only fact one then has to check is the a.s. convergence of the sample autocovariances

$$\widetilde{\gamma}_{n,X^2}(h) = \frac{1}{n} \sum_{t=1}^{n-h} X_t^2 X_{t+h}^2 - (\overline{X_n^2})^2 .$$

The same arguments as for (3.2) show that

$$\widetilde{\gamma}_{n,X^2}(h) \stackrel{\text{a.s.}}{\to} \Gamma_{X^2}(h) := p_1\gamma_{(X^{(1)})^2}(h) + p_2\gamma_{(X^{(2)})^2}(h) + p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 .$$

Similarly to [10], p. 378, introduce the Cèsaro mean of the first m Fourier approximations to $f(\lambda,\Theta)$, given for every $m \geq 1$ by

$$q_m(\lambda,\Theta) = \sum_{|k|<m}\left(1 - \frac{|k|}{m}\right) b_k \mathrm{e}^{-ik\lambda},$$

where

$$b_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{ik\lambda}\frac{1}{f(\lambda,\Theta)}d\lambda .$$

Then the same arguments as for (10.8.11) in [10] and the display following it show that for every $m \geq 1$,

$$\frac{1}{n}\sum_j I_{n,X^2}(\lambda_j)q_m(\lambda_j,\Theta) \stackrel{\text{a.s.}}{\to} \sum_{|k|<m}\Gamma_{X^2}(h)\left(1-\frac{|k|}{m}\right)b_k$$

uniformly for $\Theta \in \overline{\mathcal{C}}$ and

$$\left|\sum_{|k|<m}\Gamma_{X^2}(h)\left(1-\frac{|k|}{m}\right)b_k \right.$$

$$\left. -\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{p_1\sigma^2_{(X^{(1)})^2}f(\lambda,\Theta_1) + p_2\sigma^2_{(X^{(2)})^2}f(\lambda,\Theta_2)}{f_\delta(\lambda,\Theta)}\,d\lambda - \frac{p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2}{f_\delta(0,\Theta)}\right|$$

$$\leq \quad const\ \epsilon$$

for every $\epsilon > 0$, uniformly in $\Theta \in \overline{\mathcal{C}'}$. The same arguments as in [10] conclude the proof. □

Next we formulate a result in the spirit of Theorem 10.8.1 of Brockwell and Davis [10].

Theorem 4.3 *Assume the conditions of Proposition 4.1 are satisfied. Let Θ_n be the minimizer of $\widehat{\sigma}^2_n(\Theta)$ for $\Theta \in \mathcal{C}$. Then $\Theta_n \stackrel{\text{a.s.}}{\to} \Theta_0$, where Θ_0 is the minimizer of the function $\Delta(\Theta)$ for $\Theta \in \mathcal{C}$ defined in* (4.4).

Remark 4.4 It also follows from Proposition 4.1 that $\widehat{\sigma}^2_n(\Theta_n) \stackrel{\text{a.s.}}{\to} \Delta(\Theta_0)$.

Proof. One can follow the arguments on p. 385 of [10]. We again assume for ease of presentation that $r = 2$. Assume that $\Theta_n \stackrel{\text{a.s.}}{\to} \Theta_0$ does not hold. Then by compactness there exists a subsequence

(depending on $\omega \in \Omega$) such that $\Theta_{n_k} \to \Theta$, where $\Theta \in \overline{\mathcal{C}}$ and $\Theta \neq \Theta_0$. By Proposition 4.1, for any rational $\delta > 0$,

$$\liminf_{k\to\infty} \widehat{\sigma}^2_{n_k}(\Theta_{n_k}) \geq \liminf_{k\to\infty} \frac{1}{n_k} \sum_j \frac{I_{n_k, X^2 - \overline{X^2_n}}(\lambda_j)}{f_\delta(\lambda_j, \Theta_{n_k})}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{p_1 \sigma^2_{(X^{(1)})^2} f(\lambda, \Theta^{(1)}) + p_2 \sigma^2_{(X^{(2)})^2} f(\lambda, \Theta^{(2)})}{f_\delta(\lambda, \Theta)} \, d\lambda + \frac{p_1 p_2 \left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2}{f_\delta(0, \Theta)} .$$

So by letting $\delta \to 0$ we have

$$\liminf_{k\to\infty} \widehat{\sigma}^2_{n_k}(\Theta_{n_k}) \geq \Delta(\Theta) > \Delta(\Theta_0) . \tag{4.6}$$

On the other hand, by definition of Θ_n as a minimizer, (4.4) implies that

$$\limsup_{n\to\infty} \widehat{\sigma}^2_n(\Theta_n) \leq \limsup_{n\to\infty} \widehat{\sigma}^2_n(\Theta_0) = \Delta(\Theta_0) .$$

This is a contradiction to (4.6). This concludes the proof. □

In what follows, we apply the above estimation theory to ARCH(1) processes, i.e. $\varphi_1 = \alpha_1$, in which case we can calculate the limit of the Whittle estimate explicitly. For illustrative purposes assume $r = 2$. Then

$$\mathcal{C} = \{\alpha_1 : 0 \leq \alpha_1 < 1\} .$$

The Whittle estimate is then the minimizer over $\alpha_1 \in \mathcal{C}$ of the function

$$\begin{aligned}
\Delta(\alpha_1) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\frac{p_1 \sigma^2_{(X^{(1)})^2}}{\left|1 - \alpha_1^{(1)} \mathrm{e}^{-i\lambda}\right|^2} + \frac{p_2 \sigma^2_{(X^{(2)})^2}}{\left|1 - \alpha_1^{(2)} \mathrm{e}^{-i\lambda}\right|^2} \right) \left|1 - \alpha_1 \mathrm{e}^{-i\lambda}\right|^2 \, d\lambda \\
&\quad + (1 - \alpha_1)^2 \, p_1 p_2 \left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 \\
&= (1 + \alpha_1^2) \left(p_1 \sigma^2_{(X^{(1)})^2} + p_2 \sigma^2_{(X^{(2)})^2}\right) - 2\alpha_1 \left(\alpha_1^{(1)} p_1 \sigma^2_{(X^{(1)})^2} + \alpha_1^{(2)} p_2 \sigma^2_{(X^{(2)})^2}\right) \\
&\quad + (1 - \alpha_1)^2 \, p_1 p_2 \left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 .
\end{aligned}$$

Minimization of $\Delta(\alpha_1)$ yields

$$\alpha_1 = 1 - \frac{(1 - \alpha_1^{(1)}) p_1 \sigma^2_{(X^{(1)})^2} + (1 - \alpha_1^{(2)}) p_2 \sigma^2_{(X^{(2)})^2}}{p_1 p_2 \left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 + p_1 \sigma^2_{(X^{(1)})^2} + p_2 \sigma^2_{(X^{(2)})^2}} .$$

The right–hand side becomes arbitrarily close to 1 if the differences in the variances of the subsamples increase to infinity.

In the general GARCH(1, 1) case a similar non–stationarity effect causes that φ_1 can be close to one. In that case, the minimizer Θ_0 cannot be expressed explicitly as in the ARCH(1) case.

However, one can exploit the following argument. The spectral density f_{X^2} of the ARMA$(1,1)$ process (X_t^2) is of the form

$$f_{X^2}(\lambda) = \frac{\sigma_{X^2}}{2\pi} \frac{\left|1-\beta_1 \mathrm{e}^{-i\lambda}\right|^2}{\left|1-\varphi_1 \mathrm{e}^{-i\lambda}\right|^2} = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{X^2}(h) \mathrm{e}^{-i\lambda h}, \quad \lambda \in [-\pi, \pi].$$

¿From the Appendix we know the explicit form of the ACF of an ARMA$(1,1)$ process. Denote by $(\widetilde{X}_t)$ an ARMA$(1,1)$ process with iid standard Gaussian innovations, AR–parameter β_1 and MA–parameter φ_1. Direct calculation shows that

$$\begin{aligned}
& \Delta(\Theta) - p_1 p_2 \left(\sigma^2_{X^{(2)}} - \sigma^2_{X^{(2)}}\right)^2 \frac{(1-\varphi_1)^2}{(1-\beta_1)^2} \\
= & \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(p_1 \sigma_{(X^{(1)})^2} \frac{\left|1-\beta_1^{(1)} \mathrm{e}^{-i\lambda}\right|^2}{\left|1-\varphi_1^{(1)} \mathrm{e}^{-i\lambda}\right|^2} + p_2 \sigma^2_{(X^{(2)})^2} \frac{\left|1-\beta_1^{(2)} \mathrm{e}^{-i\lambda}\right|^2}{\left|1-\varphi_1^{(2)} \mathrm{e}^{-i\lambda}\right|^2} \right) \frac{\left|1-\varphi_1 \mathrm{e}^{-i\lambda}\right|^2}{\left|1-\beta_1 \mathrm{e}^{-i\lambda}\right|^2} d\lambda \\
= & \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(p_1 \sum_{h=-\infty}^{\infty} \gamma_{(X^{(1)})^2}(h) \mathrm{e}^{-i\lambda h} + p_2 \sum_{h=-\infty}^{\infty} \gamma_{(X^{(2)})^2}(h) \mathrm{e}^{-i\lambda h} \right) \times \\
& \quad \times \sum_{h=-\infty}^{\infty} \gamma_{\widetilde{X}}(h) \mathrm{e}^{-\lambda h} \, d\lambda \\
= & \sum_{i=1}^{2} \left[p_i \left(\gamma_{(X^{(i)})^2}(0) \gamma_{\widetilde{X}}(0) + 2\gamma_{(X^{(i)})^2}(1) \gamma_{\widetilde{X}}(1) \sum_{h=0}^{\infty} (\varphi_1^{(i)} \beta_1)^h \right) \right] \\
= & \sum_{i=1}^{2} \left[p_i \left(\gamma_{(X^{(i)})^2}(0) \gamma_{\widetilde{X}}(0) + \gamma_{(X^{(i)})^2}(1) \gamma_{\widetilde{X}}(1) \frac{1}{1-\varphi_1^{(i)} \beta_1} \right) \right].
\end{aligned}$$

One obtains the minimizer of $\Delta(\Theta)$ by differentiating with respect to φ_1 and β_1, by setting the derivatives equal to zero and by solving the resulting system of two equations. By using the particular form of $\gamma_{\widetilde{X}}(i)$, $i = 1, 2$, we obtain for the derivative with respect to φ_1 the following identity:

$$2c_1 \frac{\varphi_1 - \beta_1}{1-\beta_1^2} + 2c_2 \left[-1 + 2\frac{(\varphi_1 - \beta_1)\beta_1}{1-\beta_1^2} \right] + 2c_3 \frac{\varphi_1}{(1-\beta_1)^2} = 0,$$

where

$$\begin{aligned}
c_1 &= \sum_{i=1}^{2} \left[p_i \gamma_{(X^{(i)})^2}(0) \right], \\
c_2 &= p_1 \frac{\gamma_{(X^{(1)})^2}(1)}{1-\varphi_1^{(1)} \beta_1} + p_2 \frac{\gamma_{(X^{(2)})^2}(1)}{1-\varphi_1^{(2)} \beta_1}, \\
c_3 &= p_1 p_2 \left(\sigma^2_{X^{(2)}} - \sigma^2_{X^{(2)}}\right)^2.
\end{aligned}$$

Hence the minimizing φ_1, as a function of β_1, can be written as follows:

$$\begin{aligned}\varphi_1 &= \frac{c_1\frac{\beta_1}{1-\beta_1^2}+c_2\left[1+2\frac{\beta_1^2}{1-\beta_1^2}\right]+c_3\frac{1}{(1-\beta_1)^2}}{c_1\frac{1}{1-\beta_1^2}+2c_2\frac{\beta_1}{1-\beta_1^2}+c_3\frac{1}{(1-\beta_1)^2}} \\ &= 1+\frac{1}{1+\beta_1}\frac{-c_1+c_2(1-\beta_1)}{c_1\frac{1}{1-\beta_1^2}+2c_2\frac{\beta_1}{1-\beta_1^2}+c_3\frac{1}{(1-\beta_1)^2}}\,.\end{aligned}$$

The second term on the right–hand side is negative for $\Theta \in \mathcal{C}$. Moreover, if the difference $|\sigma^2_{X^{(2)}} - \sigma^2_{X^{(2)}}|$ increases, the value of φ_1 increases to 1. This explains, to some extent, the behaviour of the estimates for α_1 and β_1 in real–life log–return data; see Section 5. We restricted ourselves to Whittle estimation because this kind of pseudo–maximum likelihood procedure is theoretically best understood. Its relation to quuasi–maximum likelihood methods as used for fitting GARCH$(1,1)$ models in Section 5 is not clear. However, the Whittle case indicates that a similar behaviour may be expected for the other estimation procedures as well.

5 An empirical study of the goodness of fit of GARCH models

5.1 A study of non–stationary simulated data

The first example deals with a set of simulated data. Two realisations of GARCH$(1,1)$ processes with parameters

$$\alpha_0 = 0.13 \times 10^{-6}, \quad \alpha_1 = 0.11, \quad \beta_1 = 0.52\,, \tag{5.1}$$

$$\alpha_0 = 0.17 \times 10^{-6}, \quad \alpha_1 = 0.20, \quad \beta_1 = 0.65\,, \tag{5.2}$$

are considered. The innovations are standard normal. We consider a simulation $X_1, \ldots, X_{3000}$. Both subsamples $(X_t)_{t=1,\ldots,500}$ and $(X_t)_{t=1501,\ldots,2500}$ stem from the GARCH$(1,1)$ with parameters (5.1), and the remaining values come from model (5.2).

5.1.1 Goodness–of–fit tests

In Figure 5.1 the resulting time series is displayed, together with the values of the goodness–of–fit test statistic

$$S_n = \sup_{\lambda\in[0,\pi]} |\widetilde{C}_{n,X}(\lambda)|\,, \tag{5.3}$$

defined in (2.8). Notice that S_n is some kind of Kolmogorov–Smirnov goodness–of–fit test statistic as used in empirical process theory. In Figure 5.1, S_n is calculated from $n = 250$ past values at every 5th instant of time. The statistic S_n is calculated under the hypothesis that the data come from a model with parameters (5.1). Notice that the values of $v_X(h)$ in S_n can be calculated from

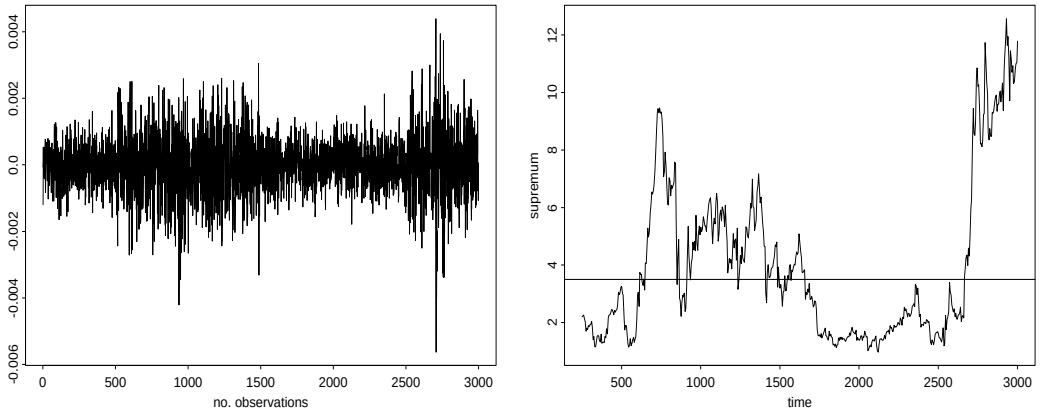

Figure 5.1 Left: 4 *pieces from* 2 *different* GARCH$(1,1)$ *models as explained after* (5.2). Right: *The goodness–of–fit test statistic* S_{250} *evaluated at every* 5*th instant of time from the past* 250 *values. The horizontal line is set at the* 99% *quantile of the limit distribution* $\sup_{\lambda\in[0,\pi]}|B(\lambda)|$.

the GARCH$(1,1)$ parameters; see (A.2) in the Apppendix. The horizontal line is set at $y = 3.6$, the 99% quantile of the limit distribution corresponding to $\sup_{\lambda\in[0,\pi]}|B(\lambda)|$ in (2.9). When all or most of the observations in the block over which the statistic S_{250} is computed come from the first process, the statistic should lie below the threshold most of the time. After the switch to the second process occurs, more and more values in the 250 long block (and finally all of them) will come from the second model and the statistic should pierce the threshold, staying above it most of the time until the switch back to the first model happens and the behaviour is reversed. In the right–hand graph in Figure 5.1 we see the statistic reacting to the first change of the model, by going above the 99% quantile of the limit distribution about day 580. It then stays above the threshold roughly until day 1600 when the change of model is again detected. It reacts once again to the last change of model by piercing the threshold about day 2600. Notice that the statistics S_{250} react to larger variance in the data by larger than expected values.

5.1.2 The sample ACF

Figure 5.2 displays the sample ACF of the absolute values $|X_t|$ for $t = 501,\ldots,1500$ and for $1501,\ldots,2501$. Under the choice of parameters (5.1) and (5.2), the 4th moment of X is finite. This follows for example from the theory developed in Mikosch and Stărică [32]. Indeed, $P(X > x) \sim c\,x^{-\kappa}$, where κ is determined as the unique solution to the non–linear equation $E|\alpha_1 Z^2+\beta_1|^{\kappa/2} = 1$. For Z standard normal, a numerical solution to this equation is given by $\kappa = 17.6$ for (5.1) and $\kappa = 17.9$ for (5.2). Moreover, GARCH$(1,1)$ processes are strongly mixing with geometric rate. Hence the theoretical ACF of the absolute values is well defined and decays to zero exponentially fast, while the sample ACF converges to the theoretical one at $\sqrt{n}$–rate and the asymptotic limits

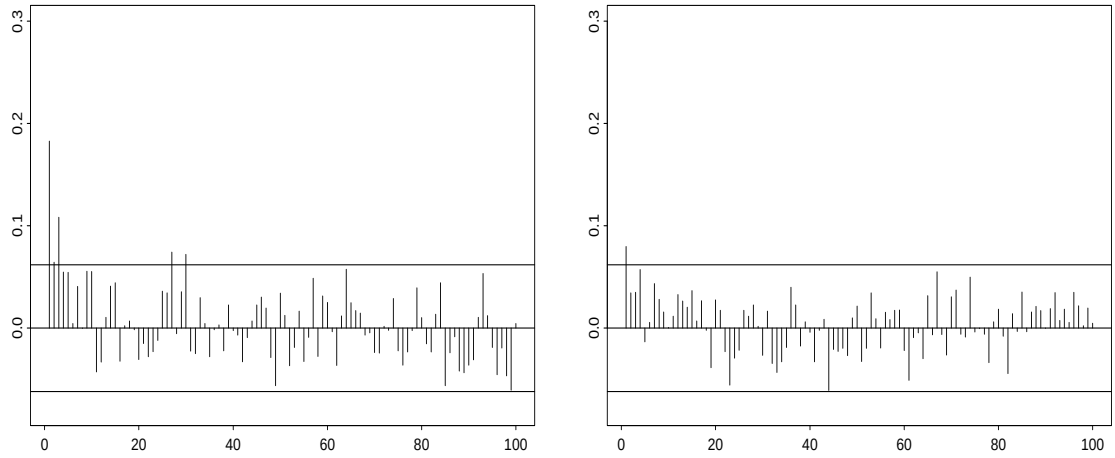

Figure 5.2 Left: *Sample ACF for* $|X_t|$, $t = 501, \ldots, 1500$. Right: *Sample ACF for* $|X_t|$, $t = 1501, \ldots, 2500$. *Here and in what follows, the horizontal lines are set as the* 95% *confidence bands* $(\pm 1.96/\sqrt{n})$ *corresponding to the ACF of iid noise with a finite* $2nd$ *moment.*

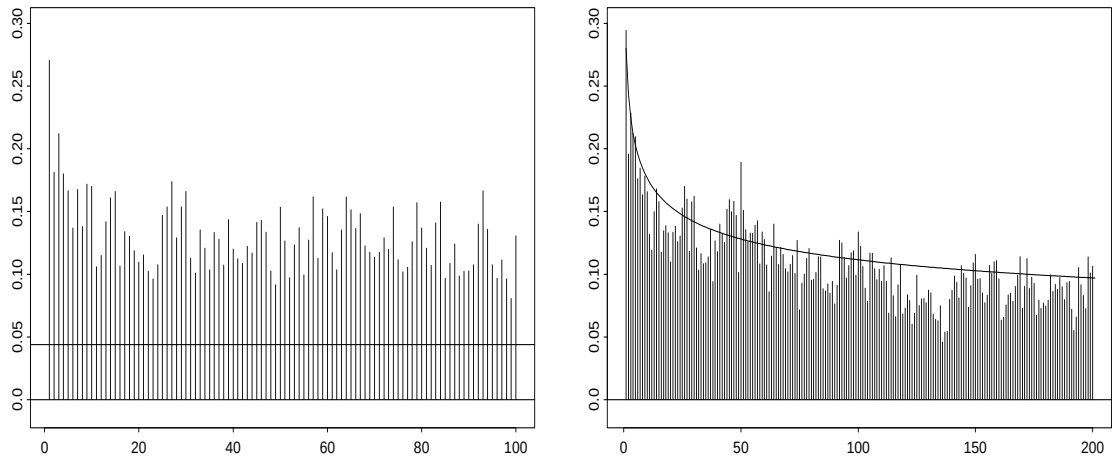

Figure 5.3 Left: *Sample ACF for* $|X_t|$, $t = 501, \ldots, 2500$. Right: *Sample ACF for* $|X_t|$, $t = 1, \ldots, 3000$ *with a fitted hyperbolic decay line.*

are normal; see Mikosch and Stărică [32].

Figure 5.3 displays the sample ACF for the juxtaposition of the two pieces, i.e. the absolute values $|X_t|$ for $t = 501, \ldots, 2500$ and shows how a resemblance of LRD behaviour of the sample ACF develops due to the non–stationarity of the data. Each of the pieces that compose the time series has exponentially decaying ACF structure and sample ACF that converges at rate $\sqrt{n}$ to the theoretical one, but the sample ACF of the series built from the two blocks seems to display LRD. The right–hand graph in Figure 5.3 shows an hyperbolic fit to the decay rate of the sample ACF of the absolute values for the whole time series. (We fitted the function $0.29h^{-0.23}$, $h = 1, 2, \ldots$.)

Figure 5.4 shows in more detail the misleading effect the non–stationarity paradigm can have on statistical estimation. It displays the results of two statistical estimation procedures for the so–called *Hurst exponent*. The underlying time series consists of the absolute values of the simulated data displayed in Figure 5.1. The quantity H has been proposed in the literature as a measure of LRD; see Beran [4] for details on the definition, properties and statistical estimation of H.

If one assumes that the theoretical ACF $\rho(h)$ of the time series has a hyperbolic decay rate, i.e. $\rho(h) \approx ch^{-\beta}$ for some positive β and c, the Hurst coefficient is usually determined as $H = 1 - \beta/2$. In particular, the presence of LRD in the time series is signalled if $H \in (0.5, 1)$. In this case, the sequence $(\rho(h))$ is not absolutely summable. The closer H to 1, the further the dependency reaches. An estimation procedure for H is then suggested by the following argument. Since

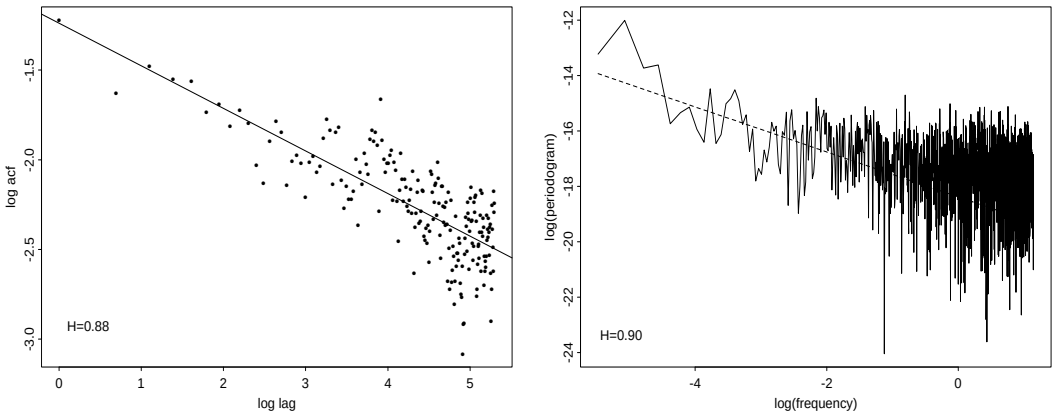

Figure 5.4 Left: *Log–log fit of the sample ACF for the absolute values of the simulated data displayed in Figure 5.1. Estimated Hurst coefficient* $H = 0.88$. Right: *Log–log fit of the periodogram for the same sample. Estimated* $H = 0.84$.

$\ln \rho(h) \approx \ln c - \beta \ln h$ and the sample ACF estimates the theoretical one, a log–log plot of the lags versus the sample ACF should be roughly linear, the slope of the regression line yielding an estimate of the quantity β, hence of H. The left–hand graph in Figure 5.4 displays the fit of a regression line through the plot of the log lags versus the log sample ACF. The slope is -0.23, the intercept

-1.23. Hence $\hat{\rho}(h) = 0.29h^{-0.23}$. The actual fit of this hyperbolically decaying function to the sample ACF is illustrated in Figure 5.3. The resulting Hurst coefficent is $H = 1 - 0.23/2 = 0.88$. This would imply a strong LRD effect if the data came from a stationary sequence.

5.1.3 The periodogram

It is a well–known fact from trigonometric function theory that power law decay of the ACF translates into power law behaviour of its Fourier transform (i.e. the spectral density) for small frequencies; see for example Zygmund [42]. (For a precise formulation of this statement the ACF has to satisfy some subtle conditions; we refrain from discussing them.) This fact suggests a second method for estimating the Hurst coefficient. Assume the spectral density f satisfies $f(\lambda) \approx c\lambda^{2H-1}$ for small $\lambda > 0$. Equivalently, $\ln f(\lambda)$ is linear in λ, $\ln f(\lambda) \approx \ln c + (2H-1)\ln\lambda$. Hence, on a log–log plot, the periodogram should roughly exhibit a linear behaviour, the slope of the line yielding an estimate of H. The right–hand graph in Figure 5.4 shows the plot of the log frequencies versus the log periodogram with a regression line fit to the first 10% of the frequencies yielding an estimate $H = 0.84$. Interestingly enough, both methods give similar values for the Hurst coefficients and suggest that there is a strong LRD effect in the data. The estimates for H are clearly subject to statistical uncertainty. It is not our primary purpose to give confidence bands for the estimation of H; all we intend to show here is that standard estimation procedures for stationary time series, when applied to a non–stationary sequence, may give misleading answers as to whether there is LRD in the data.

5.1.4 Parameter estimation

Finally, we check the impact which the change of regimes in the simulated data set has on the estimation of α_1 and β_1. We used quasi–maximum likelihood as for example proposed in Gourieroux [22]. GARCH$(1,1)$ models have been fitted for the increasing samples $X_{501}, \ldots, X_{650+t*50}$, $t = 1, \ldots, 48$. We decided to start from $t = 500$ in order to have a longer run of the same model before the switch to the second process occurs. For sample sizes less then 1000, the sum of the theoretical parameters is 0.85. The estimated sum varies for these sample sizes between 0.75 and 0.90. The graph in Figure 5.5 clearly shows how the switch of regimes (which happens at $t = 1000$) makes the sum increase to 1. This is in agreement with the theory in Section 4. There we explained that the Whittte estimate of $\alpha_1 + \beta_1$ increases the more the larger the difference in the variances in different subsamples. We expect that a similar effect occurs for the used quasi–maximum likelihood estimators.

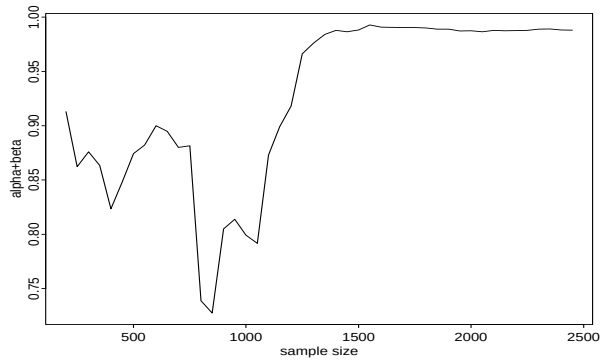

Figure 5.5 *The estimated values* $\alpha_1 + \beta_1$ *for a* GARCH$(1,1)$ *model fitted to an increasing sample from the simulated data in Figure 5.1. The labels on the x-axis indicate the size of the sample used in the quasi–maximum likelihood estimation.*

6 A study of the Standard & Poors series

Now we proceed by analysing a time series that has been previously used to exemplify the presence of LRD in financial log–return series: the Standard 90 and Standard and Poor's 500 composite stock index. This series, covering the period between January 3, 1928, to August 30, 1991, was used in Ding et al. [15], Granger et al. [24], Ding and Granger [14] for an analysis of the autocorrelation structure. It led the authors to the conclusion that the powers of the absolute values of the log–returns a re positively correlated over more than 2500 lags, i.e. 10 years. Hardly any proof is needed to convince one that this time series is likely to be non-stationary. It covers the Great Depression, a world war together with the most recent period, marked by major structural changes in the world's economy. Nevertheless, the authors performed their analysis under the strong assumption of strict stationarity. In addition, there was a compositional change in the S&P composite index that happened in January 1953 when the Standard 90 was replaced by the broader Standard and Poor's 500 index. Despite all these, Ding et al. [15] conclude the section which describes the data as follows (page 85): "During the Great Depression of 1929 and early 1930s, volatilities are much higher than any other period. There is a sudden drop in prices on Black Monday's stock market crash of 1987, but unlike the Great Depression, the high market volatility did not last very long. *Otherwise, the market is relatively stable.*"

Bollerslev and Mikkelsen [8] used the daily returns on the Standard and Poor's 500 composite stock index from January 2, 1953, to December 31, 1990 (a total of 9559 observations) to fit a FIGARCH model under the assumptions of stationarity and LRD.

In the sequel we perform a detailed analysis of the same data set covering the time span from January 2, 1953, to December 31, 1990. Contrary to the belief that the LRD characteristic carries meaningful information about the price generating process, we show that the LRD behaviour could

be just an artifact due to structural changes in the data. We use the sup-statistic (**??**) to detect the moment in time when a GARCH$(1,1)$ model estimated on past data stops describing the behaviour of the time series. Then we document the effect the switch to a different regime of variance has on the sample ACF. We find that the aspect of the sample ACF changes drastically after episodes of increased variance that cannot be properly described by the estimated model.

The S&P data are plotted in the left–hand graph of Figure 6.1. When fitting a GARCH$(1,1)$ model to the first 3 years of the data (750 observations), one obtains the following parameters by using quasi–maximum likelihood estimation (see for example Gourieroux [22]):

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759\,, \tag{6.4}$$

and an estimated 4th moment for the residuals of 3.72. These quantities are used for calculating $v_X(h) = \mathrm{var}(X_0 X_h)$ in (A.2).

6.1 Goodness–of–fit tests

The left–hand graph in Figure 6.2 shows the results of calculating the statistic S_n (see (5.3)) on a weekly basis (i.e. every 5th instant of time) with blocks of $n = 125$ past observations, corresponding to approximately 6 months. The horizontal line is again set at 3.6, the 99% quantile of the limit distribution of S_n.

A closer look at the graph of Figure 6.1 together with the left–hand graph in Figure 6.2 reveals an almost one–to–one correspondence between the periods of larger absolute log–returns (larger volatility) and the periods when the goodness–of–fit test statistic S_{125} falls outside the confidence region. This observation has a theoretical grounding in (A.2) which indicates that the statistics S_{125} will be sensitive to changes in the model mainly through changes in the variance σ_X^2 of the data. Therefore, we can identify the excursions of the statistics S_{125} above the 99% quantile threshold with bursts of volatility in the data that are beyond the explanatory capacity of the fitted GARCH$(1,1)$ model. So it seems that the model (6.4) fails to explain the more extreme bursts of volatility that occur every once in a while.

It is then interesting to verify if a periodically updated GARCH$(1,1)$ could account for the more pronounced volatility periods that are troublesome for the estimated GARCH$(1,1)$ model (6.4). One way is to calculate the implied unconditional GARCH$(1,1)$ variance for a periodically re–estimated GARCH$(1,1)$ model, i.e. one calculates the variance

$$\sigma_X^2 = \alpha_0/(1 - (\alpha_1 + \beta_1))$$

based on the estimated parameters α_1 and β_1; see (A.1).

More concretely, we fitted a GARCH$(1,1)$ model every 6 months, i.e. every 125 days, based on a moving window of 508 past observations, equivalent to roughly two years of daily returns. We

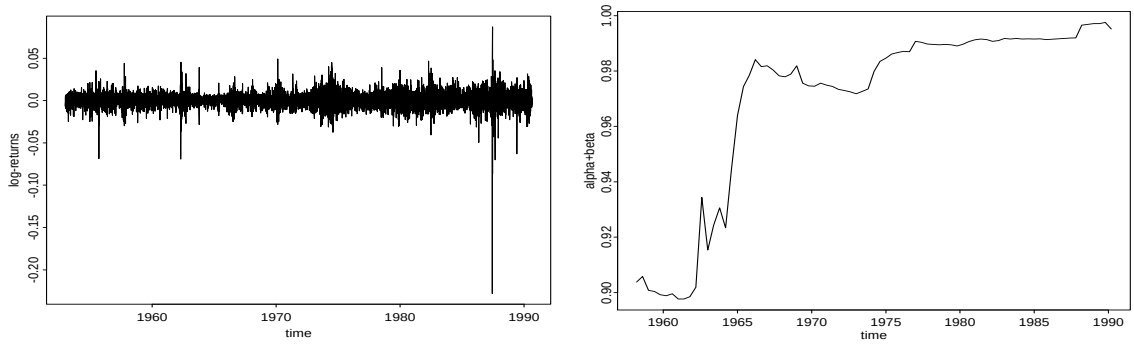

Figure 6.1 Left: *Plot of all* 9558 S&P500 *log–returns. The year marks indicate the beginning of the calendar year.* Right: *The estimated values* $\alpha_1 + \beta_1$ *for increasing samples of the* S&P500. *An initial* GARCH$(1,1)$ *model was estimated on* 1500 *observations, starting January 2, 1953. Then the parameters* α_1 *and* β_1 *were re–estimated by successively adding* 100 *new observations to the sample. The labels on the* x*–axis indicate the date of the latest observation in the sample.*

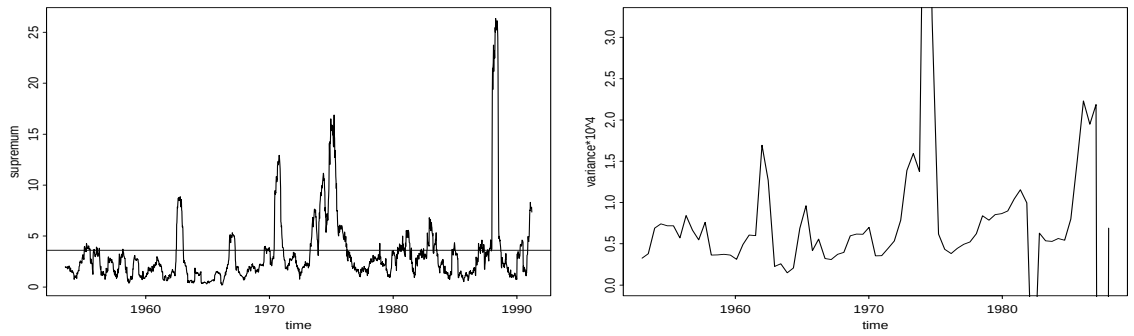

Figure 6.2 Left: *The goodness—of–fit test statistic* S_{125} *for the* S&P500 *data.* Right: *The implied* GARCH$(1,1)$ *unconditional variance of the* S&P500 *data. A* GARCH$(1,1)$ *model is estimated every* 6 *months using the previous* 2 *years of data* (*i.e.* 508 *observations*). *The graph displays the variances* $\sigma_X^2 = \alpha_0/(1-\alpha_1-\beta_1)$*; see* $(A.1)$.

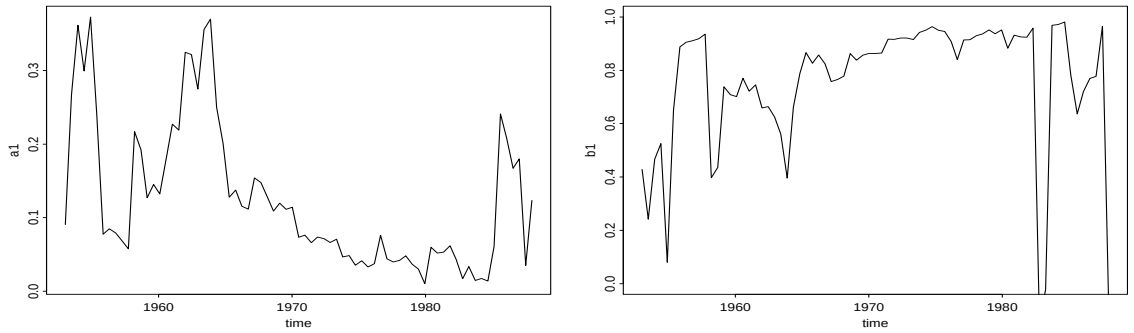

Figure 6.3 *A* GARCH$(1,1)$ *model is fitted to every block of* 6 *months data. The fit is based on the previous* 2 *years of data, i.e.* 508 *observations. Estimated* α_1 (left) *and* β_1 (right).

then plotted the implied variance σ_X^2 corresponding to every 6 months period. The results of this procedure are displayed in the right–hand graph of Figure 6.2. One notices that the pattern of increased implied variance is quite similar to the pattern of the excursions of the statistic S_{125} above the 99% quantile threshold. For us, this similarity has a double significance. First, it seems to imply that one can capture the changing patterns of volatility present in the data by periodically updating the GARCH$(1,1)$ model. Second, it confirms our observation that excursions of the goodness–of–fit test statistic above the threshold level are mainly caused by changes in the variance of the time series, more precisely by its increase. Figure 6.3 displays the changes in the estimated coefficients α_1 and β_1. The parameters vary widely, beyond the limits of the statistical uncertainty the assumption of a stationary time series would prescribe. For the sample sizes we use the statistical error is of an order lower than 10^{-1} for α_1 and of order 10^{-1} for b_1. The variation present in the parameters is much larger than these statistical error bounds. One also notices that the patterns of change in the coefficients α_1 and β_1 are quite different and do not resemble the ones in the graphs of Figure 6.2. This seems to support our observation that, for this data set, the statistic S_{125} detects structural changes in the model reacting mainly to variations in the variance. The right–hand graph of Figure 6.2 together with the ones in Figure 6.3 strengthen the impression, given by the statistical analysis displayed in the left–hand graph of Figure 6.2, of a relatively fast pace of change in the market.

In the end of this section, it is interesting to take a closer look at the behaviour of our statistic around the Black Monday crash in October 1987. Following on the observation that a frequently reestimated GARCH$(1,1)$ model seems to follow better the changing pattterns of volatility in the time series (see Figure 6.1), a GARCH$(1,1)$ model is estimated in the begining of June 1987 using the observations between January 1986 and June 1987 (375 observations). The estimated coefficients

$$\alpha_0 = 16 \times 10^{-6}, \quad \alpha_1 = 0.013, \quad \beta_1 = 0.812 \tag{6.5}$$

together with the fourth moment of the estimated residuals, $E\hat{Z}^4 = 3.4$ are used to build the S_{125} statistic. Figure 6.4 shows the behaviour of the statistic during the 100 days preceeding and the 25 following the crash. To allow for a better analysis, the log of the statistic is displayed. One sees that the statistic reaches above the 75% quantile (2.26) of the limiting distribution 6 weeks before the crash and never falls below. The days just before the crash mark an increase in the statistic. The day -6 up to -2 are above the 90% percentile of the limit distribution ($q_{0.90} = 2.71$), while day -1 is well above the 99% quantile ($q_{0.99} = 3.6$).

6.2 Parameter estimation

The right graph in Figure 6.1 displays the effect of the more extreme bursts of volatility on the quasi–maximum likelihood estimation procedure for the sum of the parameters α_1 and β_1. The graph clearly shows how episodes of higher volatility (that were detected through the behaviour of

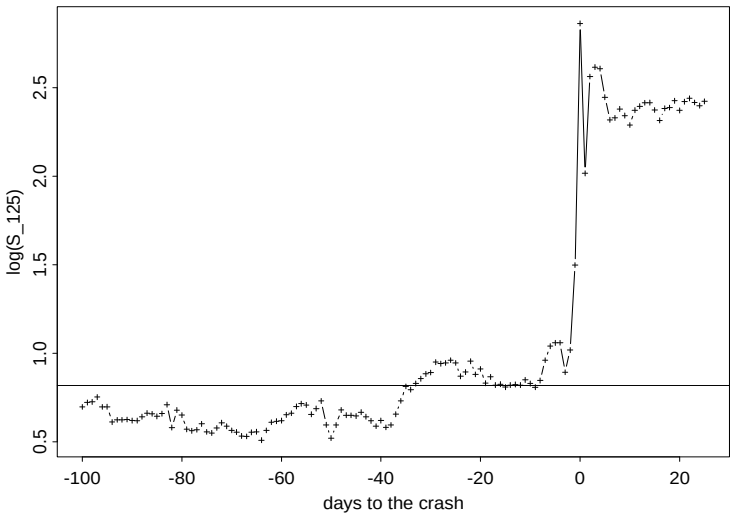

Figure 6.4 *The* $\log(S_{125})$ *statistic before and after the Black Monday 1987 crash. The horizontal line is the log of the 75% quantile (*$q_{0.75} = 2.26$*) of the limiting distribution of* S_{125}*.*

the S_{125} statistic and cannot be explained by the model (6.4)) increase the sum $\alpha_1 + \beta_1$, closer to 1. We see that the first burst of volatility recorded by the S_{125} statistic around 1963 pushes the sum from 0.90 up to almost 0.94, while the second burst about 1967 makes it climb from 0.92 to 0.98. The more prolonged period, where the model (6.4) does not fit the data (and which has such a strong impact on the behaviour of the sample ACF), causes an increase from 0.97 to 0.99, while the Black Monday 1987 stock market crash forces the sum to its highest value, 0.995. This is again in agreement with the explanation given for the Whittle estimate of $\alpha_1 + \beta_1$: The increased values of $\alpha_1 + \beta_1$ are due to an increase of differences between the variances in different subsamples; see Section 4.

6.3 The sample ACF

Let us now analyse the impact which these periods of different structural behaviour detected by the goodness–of–fit test statistic S_{125} have on the sample ACF of the time series. In the left bottom graph of Figure 6.1 one sees that, roughly for the first nine years, 1953 to 1962, the values of the statistic lie within the 99% confidence interval, with short exceptional periods that are not too extreme. Hence we do not have strong reasons to doubt the fit of the model for this part of the data. This calm period is followed by roughly a year during which S_{125} is well outside the confidence interval, indicating that the model ceased to describe the data. In Figure 6.5 the sample ACFs for the absolute values of the log–returns for the first 9–year and 11–year periods are compared. While the first period's autocorrelations seem to be insignificant after 50 lags, the autocorrelations

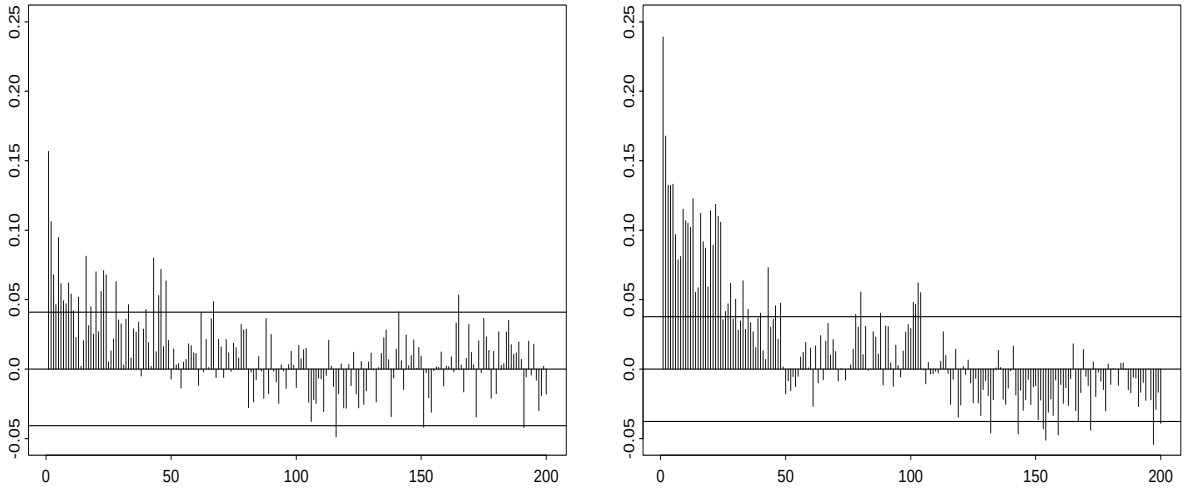

Figure 6.5 *The sample ACF for the absolute values of the log–returns for the first* 9 *years* (left) *and the first* 11 *years of the* S&P *data.*

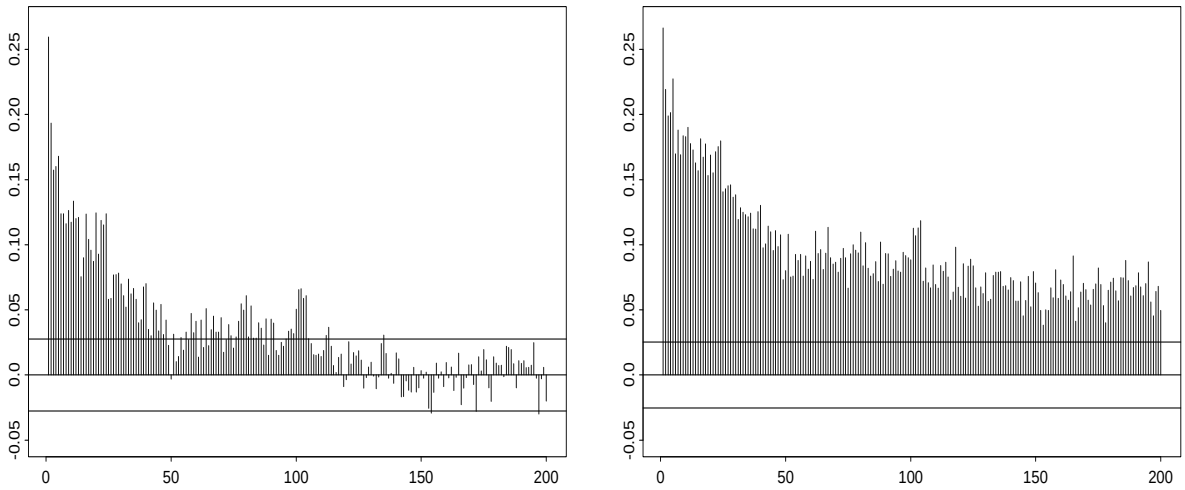

Figure 6.6 *The sample ACF for the absolute values of the log–returns of the first* 20 *years* (left) *and* 24 *years* (right) *of the S&P data.*

for the 11–year period are still significant at lag 100. We also note that the size of the significant autocorrelations increases together with the proportion of positively correlated lags. This is indeed consistent with the explanation of this phenomenon provided in Section 3.

After this extreme episode (which can also be identified in the plot of the data in Figure 6.1) a period of low volatility follows. It is interesting to notice that the statistic S_{125} now falls below the 1% quantile of the limit distribution. This could be interpreted either as an indication for the failure of the model or as a proof of extremely good fit. As the period the statistic spends below the 1% quantile is rather long (about 1.5 years), the first explanation is more plausible. Two more relatively short periods of increased volatility follow during which the model is again caught off–guard.

However, from the view point of the behaviour of the sample ACF, the most interesting part of the top graph in Figure 6.1 is the period beginning in 1973 and lasting for almost 4 years. The values of S_{125} are quite extreme, strongly indicating that the period between 1973 and 1977 is a long interval when the model (6.4) does not describe the data. Let us analyse the changes in the sample ACF caused by this long period of different behaviour. Figure 6.6 displays the sample ACF of the absolute values $|X_t|$ up to the moment when the change is detected, next to the sample ACF including the 4–year period that followed. We see that the sample ACF up to 1973 does not differ significantly from the one based only on the data up to 1964; see Figure 6.5. However the impact of the change in regime between 1973 and 1977 on the form of the sample ACF is extremely strong as one sees in the second graph of Figure 6.6. The graph clearly displays the LRD features given by the theoretical explanation of Section 3: exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations.

7 A study of exchange rates

As another example we consider 2032 daily log–returns from the DEM/USD foreign exchange (FX) rate between January 1975 and December 1982. The left–hand graph of Figure 7.1 displays the data. A GARCH$(1,1)$ fit to the first 2 years yields the following parameters:

$$\alpha_0 = 0.83 \times 10^{-6}\,, \quad \alpha_1 = 0.18\,, \quad \beta_1 = 0.77\,, \tag{7.6}$$

and an estimated 4th moment for the residuals of 4.6.

7.1 Goodness–of–fit tests

The right–hand graph of Figure 7.1 displays the values of the statistic S_n calculated on a weekly basis from the previous $n = 125$ observations which amount to roughly six months. The horizontal line is set at the a.s. 99% quantile. The graph shows the presence of two intervals where the estimated model clearly does not fit the data: a shorter period of less than a year, covering a part

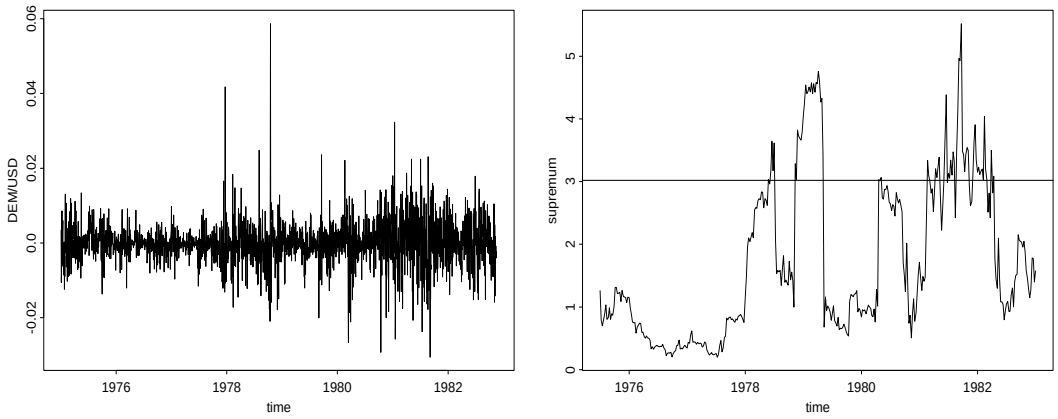

Figure 7.1 Left: *Plot of the* 2032 *DEM/USD log–returns. The year marks indicate the beginning of the calendar year.* Right: *The corresponding goodness–of–fit test statistics* S_{125}.

of 1978 and the beginning of 1979, and a longer period of about one year and a half, stretching from the end of 1980 to the beginning of 1982. Since the statistic S_{125} is calculated using observations from the previous 6 months and since the length of the first period, where the model does not seem to fit, is roughly of the order of 6 months, it is likely that the behaviour of the statistic during this period is caused by a short perturbance that keeps showing in the statistic for an entire half year after it occured. A first glance at Figure 7.1 allows one to identify as possible culprit the period around the large positive return at the end of 1978. A second glance at the sample ACF of the absolute values $|X_t|$ before and after the period in discussion reveals not too deep a change in the behaviour of this statistical instrument; see Figure 7.2. Here the first 100 lags of the sample ACF before and after the first episode are displayed.

In contrast to the first short episode, the second period, when the goodness–of–fit test statistic exceeds the threshold, lasts longer than half a year. A visual inspection of the graphs in Figure 7.1 gives the impression that the structure of the time series in the period between the middle of 1980 to the end of 1981 is quite different from the structure of the remaining observations. The statistic S_{125} confirms this fact by frequently switching sides of the threshold line during this period.

7.2 The sample ACF

The dramatic changes in the behaviour of the sample ACF are illustrated in Figure 7.3 where we can see the first 100 lags of the sample ACF before and after the second episode. Again, in accordance with our explanation, the sample ACF displays exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations.

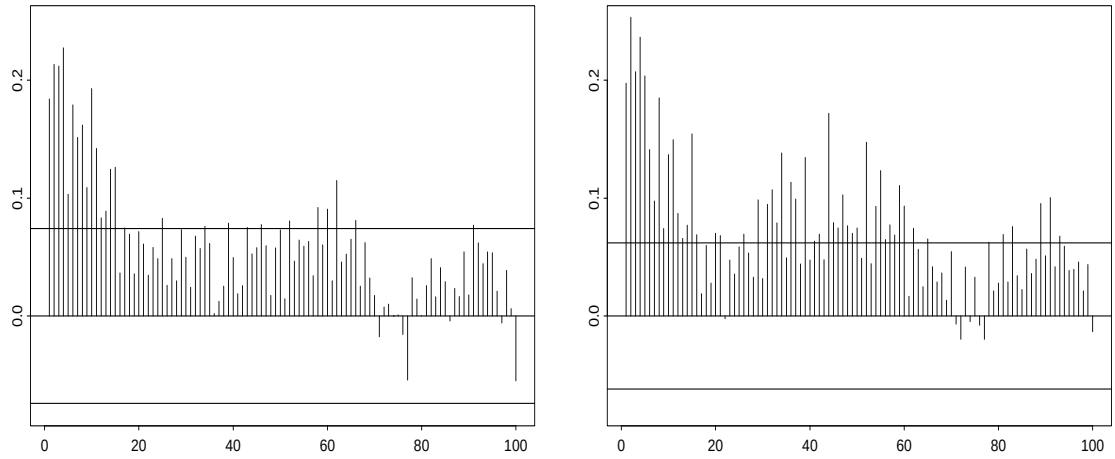

Figure 7.2 *The sample ACF for the absolute values of the log–returns of the DEM/USD FX data up to the beginning of* 1978 (left) *and up to June* 1979 (right).

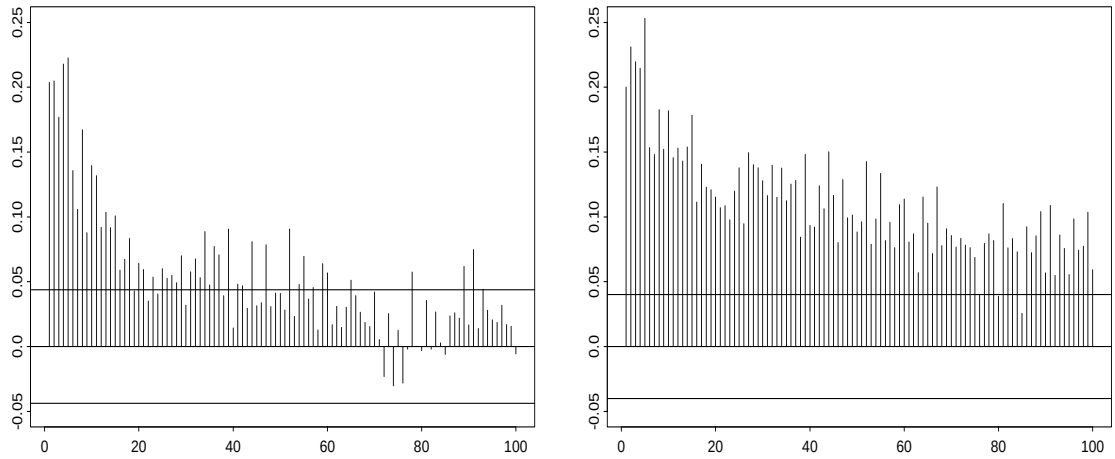

Figure 7.3 *The sample ACF for the absolute values of the log–returns of the DEM/USD FX data up to the beginning of* 1981 (left) *and up to June* 1982 (right).

Appendix

Consider a GARCH$(1,1)$ process (X_t) with parameters $\alpha_0, \alpha_1, \beta_1$. We write $\varphi_1 = \alpha_1 + \beta_1$ and assume $EX^4 < \infty$. ¿From the calculations below it follows that the condition

$$1 - (\alpha_1^2 EZ^4 + \beta_1^2 + 2\alpha_1\beta_1) > 0$$

must be satisfied. The squared GARCH$(1,1)$ process can be rewritten as an ARMA$(1,1)$ process by using the defining equation (1.1):

$$X_t^2 - \varphi_1 \, X_{t-1}^2 = \alpha_0 + \nu_t - \beta_1 \nu_{t-1} \,,$$

where $(\nu_t) = (X_t^2 - \sigma_t^2)$ is a white noise sequence. Thus, the covariance structure of

$$U_t = X_t^2 - EX^2 = X_2^2 - \frac{\alpha}{1-\varphi_1} \,, \quad t \in \mathbb{Z} \,,$$

is that of a mean–zero ARMA$(1,1)$ process. The values of $\gamma_U(h)$ are given on p. 87 in Brockwell and Davis [11]:

$$\begin{aligned}
\gamma_U(0) &= \sigma_\nu^2 \left[1 + \frac{(\varphi_1 - \beta_1)^2}{1 - \varphi_1^2} \right] , \\
\gamma_U(1) &= \sigma_\nu^2 \left[\varphi_1 - \beta_1 + \frac{(\varphi_1 - \beta_1)^2 \varphi_1}{1 - \varphi_1^2} \right] , \\
\gamma_U(h) &= \varphi_1^{h-1} \gamma_U(1) \,, \quad h \geq 2 \,.
\end{aligned}$$

Straightforward calculation yields

$$\begin{aligned}
\sigma_\nu^2 &= (EZ^4 - 1) \, E\sigma_1^4 = \frac{1+\varphi_1}{1-\varphi_1} \frac{\alpha_0^2 (EZ^4 - 1)}{1 - (\varphi_1^2 + \alpha_1^2 (EZ^4 - 1))} \,, \\
\sigma_X^2 &= \frac{\alpha_0}{1 - \varphi_1} \,. \qquad \text{(A.1)}
\end{aligned}$$

Thus we can calculate the quantities

$$v_X(h) = E(X_0^2 X_h^2) = \gamma_U(h) + \sigma_X^4 \,, \quad h \geq 1 \,,$$

which occur in the definition of the change point statistics and goodness–of–fit test statistics of Section 2. We obtain:

$$v_X(h) = \sigma_X^4 \left(\frac{(EZ^4 - 1)\alpha_1 \left(1 - \varphi_1^2 + \alpha_1 \varphi_1\right)}{1 - (\varphi_1^2 + \alpha_1^2 (EZ^4 - 1))} \, \varphi_1^{h-1} + 1 \right) \,, \quad h \geq 1 \,. \qquad \text{(A.2)}$$

Acknowledgement: Cătălin Stărică would like to thank the Department of Mathematics of the University of Groningen for financial support.

References

[1] ANDERSEN, T. ABD BOLLERSLEV T. (1998) Deutsche Mark–Dollar volatility: intraday activity patterns, macroeconomic announcements and longer run dependencies. *J. of Finance* **LIII**, 219–262.

[2] ANDERSON, T.W. (1993) Goodness of fit tests for spectral distributions. *Ann. Statist.* **21**, 830–847.

[3] BARTLETT, M.S. (1954). Problemes de l'analyse spectrale des séries temporelles stationnaires. *Publ. Inst. Statist. Univ. Paris.* **III-3**, 119–134.

[4] BERAN, J. (1994) *Statistics for Long–Memory Processes.* Monographs on Statistics and Applied Probability, No. 61. Chapman and Hall, New York.

[5] BHATTACHARYA, R.N., GUPTA, V.K. AND WAYMIRE, E. (1983) The Hurst effect under trends. *J. Appl. Probab.* **20**, 649–662.

[6] BICKEL, P.J. AND WICHURA, M.J. (1971). Convergence criteria for multiparameter stochastic processes and some applications. *Ann. Math. Statist.* **42**, 1656–1670.

[7] BILLINGSLEY, P. (1968) *Convergence of Probability Measures.* Wiley, New York.

[8] BOLLERSLEV, T. AND MIKKELSEN, H.O. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **74**, 3–30.

[9] BOUGEROL, P. AND PICARD, N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* **52**, 115–127.

[10] BROCKWELL, P.J. AND DAVIS, R.A. (1991) *Time Series: Theory and Methods,* 2nd edition. Springer, New York.

[11] BROCKWELL, P.J. AND DAVIS, R.A. (1996) *Introduction to Time Series and Forecasting.* Springer, New York.

[12] DAVIS, R.A., MIKOSCH, T. AND BASRAK, B. (1998) The sample ACF of solutions to multivariate stochastic recurrence equations. Technical Report, University of Groningen.

[13] DE LA PEÑA, V.H. AND MONTGOMERY–SMITH, S.J. (1995) Decoupling inequalities for the tail probabilities of multivariate U–statistics. *Ann. Probab.* **23**, 806–816.

[14] DING, Z. AND GRANGER, C.W.J. (1996) Modeling volatility persistence of speculative returns: A new approach. *J. Econometrics* **73**, 185–215.

[15] DING, Z., GRANGER, C.W.J. AND ENGLE, R. (1993) A long memory property of stock market returns and a new model. *J. Empirical Finance* **1**, 83–106.

[16] DOUKHAN, P. (1994) *Mixing. Properties and Examples.* Lecture Notes in Statistics **85**. Springer Verlag, New York.

[17] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[18] ENGLE, R.F. AND BOLLERSLEV, T. (1986) Modelling the persistence of conditional variances. With comments and a reply by the authors. *Econometric Rev.* **5**, 1–87.

[19] ENGLE, R.F. (ED.) (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (U.K.).

[20] GIRAITIS, L. AND LEIPUS, R. (1992) Testing and estimating in the change-point problem of the spectral function. *Lith. Math. Trans. (Lit. Mat. Sb.)* **32**, 20–38.

[21] GOLDIE, C.M. (1991) Implicit renewal theory and tails of solutions of random equations. *Ann. Appl. Probab.* **1**, 126–166.

[22] GOURIEROUX, C. (1997) *ARCH Models and Financial Applications.* Springer Series in Statistics. Springer, New York.

[23] GRENANDER, U. AND ROSENBLATT, M. (1984) *Statistical Analysis of Stationary Time Series,* 2nd edition. Chelsea Publishing Co., New York.

[24] GRANGER, C.W.J. AND DING, Z. (1996) Varieties of long memory models. *J. Econometrics* **73**, 61–77.

[25] HIDA, T. (1980) *Brownian Motion.* Springer, New York.

[26] JACOD, J. AND SHIRYAEV, A.N. (1987) *Limit Theorems for Stochastic Processes.* Springer, Berlin, New York.

[27] KESTEN, H. (1973) Random difference equations and renewal theory for products of random matrices. *Acta Math.* **131**, 207–248.

[28] KLÜPPELBERG, C. AND MIKOSCH, T. (1996) The integrated periodogram for stable processes. *Ann. Statist.* **24**, 1855–1879.

[29] KLÜPPELBERG, C. AND MIKOSCH, T. (1996) Gaussian limit fields for the integrated periodogram. *Ann. Appl. Probab.* **6**, 969–991.

[30] KOKOSZKA, P. AND MIKOSCH, T. (1997) The integrated periodogram for long–memory processes with finite or infinite variance. *Stoch. Proc. Appl.* **66**, 55–78.

[31] MIKOSCH, T. (1998) Periodogram estimates from heavy–tailed data. In: R. Adler, R. Feldman and M.S. Taqqu (eds.) *A Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy–Tailed Distributions,* pp. 241–258. Birkhäuser, Boston.

[32] MIKOSCH, T. AND STĂRICĂ, C. (1998) Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. Technical Report. University of Groningen.

[33] NELSON, D.B. (1990) Stationarity and persistence in the GARCH(1, 1) model. *Econometric Theory* **6**, 318–334.

[34] OODAIRA, H. AND YOSHIHARA, K. (1972) Functional central limit theorems for strictly stationary processes satisfying the strong mixing condition. *Kōdai Math. Sem. Rep.*, **24** 259–269.

[35] PICARD, D. (1985) Testing and estimating change-points in time series. *Adv. Appl. Probab.* **17**, 841–867.

[36] Pollard, D. (1984) *Convergence of Stochastic Processes.* Springer, Berlin.

[37] PRIESTLEY, M.B. (1981) *Spectral Analysis and Time Series, vols. I and II.* Academic Press, New York.

[38] RESNICK, S.I. (1986) Point processes, regular variation and weak convergence. *Adv. Appl. Probab.* **18**, 66–138.

[39] ROSINSKI, J. AND WOYCZYŃSKI, W.A. (1987) Multilinear forms in Pareto–like random variables and product random measures. *Coll. Math.* **51**, 303–313.

[40] SHORACK, G.R. AND WELLNER, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

[41] Whittle, P. (1951) *Hypothesis Testing in Time Series Analysis.* Almqvist och Wicksel, Uppsala.

[42] Zygmund, A. (1988) *Trigonometric Series.* First paperback edition. Cambridge University Press, Cambridge (UK).

Thomas Mikosch
Department of Mathematics
P.O. Box 800
University of Groningen
NL-9700 AV Groningen
THE NETHERLANDS

Cătălin Stărică
Department of Mathematics
Chalmers University of Technology
S–412 96 Gothenburg
SWEDEN