CAF Centre for Analytical Finance
UNIVERSITY OF AARHUS • AARHUS SCHOOL OF BUSINESS

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

T. Mikosch and C. Starica

Working Paper Series No. 58 April 2000

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

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.

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 modeling 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 behavior 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. In other words, contrary to the common-hold belief that the LRD characteristic carries meaningful information about the price generating process, the LRD behavior 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 modeling efforts (S&P500, DEM/USD FX) in various publications.

*This research supported in part by a grant of the Dutch Science Foundation (NWO).
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 this paper we continue the discussion commenced in Mikosch and Stărică (1998) about various aspects of the goodness of fit of a *stationary* GARCH$(1,1)$ *process* to log-return data. This simple theoretical model is given by

$$
(1.1) \qquad \begin{cases} X_t &= \sigma_t \, Z_t \,, \\ \sigma_t^2 &= \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2 \,, \end{cases} \qquad t \in \mathbb{Z}\,,
$$

where $(Z_t)_{t\in\mathbb{Z}}$ is a sequence of iid random variables with $EZ = 0$, $EZ^2 = 1$. The parameters α_1 and β_1 are non-negative and α_0 is necessarily positive.

The GARCH$(1,1)$ process is an attractive modeling tool for several reasons. It is theoretically well understood (Bollerslev [13], Nelson [49], Bougerol and Picard [16]). It can be easily fitted to log-returns

$$
X_t = \log P_t - \log P_{t-1}\,, \quad t = 1, 2, \dots\,,
$$

of speculative prices P_t. Moreover, it is commonly believed that, despite its simplicity (3 parameters only), it captures some of the basic features of log-returns. In particular, this model adequately describes the *heavy-tailedness of the marginal distribution.* Indeed, under very general conditions on the noise sequence (Z_t), the GARCH$(1,1)$ in particular, and GARCH(p,q) processes in general, have Pareto-like marginal distributions, i.e.

$$
(1.2) \qquad P(X > x) \sim c_0 \, x^{-\kappa} \quad \text{as } x \to \infty \text{ for some } c_0, \kappa > 0.
$$

The GARCH$(1,1)$ model can also explain to some extent another empirical feature related to the extremal behavior of log-return data, i.e. *the exceedances of high and low thresholds do not occur separated over time but are heavily clustered.* This fact is usually referred to as *dependence in the tails.* For some recent research in this area we refer to Mikosch and Stărică [48] for the GARCH$(1,1)$ case and Davis et al. [22] in the general GARCH(p,q) case.

Although our discussion focuses on the GARCH$(1,1)$ model, the theoretical results of this paper are derived for the general GARCH(p,q) case. We want to emphasize that all the GARCH models, regardless their order, have the same basic properties. In particular, the extremal behavior encompassing heavy tails and dependence in the tails, the asymptotic behavior of the sample ACF, etc., are similar in nature, independent of the order of the process. For example, independently of the order, all GARCH(p,q) processes have heavy tails and display short memory. The only difference is the degree of analytical tractability of these properties. Therefore the conclusions we draw from examining the GARCH$(1,1)$ case are also valid for the wider class of GARCH(p,q) models.

The researcher browsing through the recent financial econometrics literature will find the following feature among the *stylized facts* that characterize the log-returns of foreign exchange rates, stock indices, share prices, bond yields, etc., P_t:

- Although the *sample autocorrelations* of the data are tiny (uncorrelated log-returns), the *sample autocorrelations* of the absolute and squared values are significantly different from zero even for large lags.

This empirical finding (which we will refer to as *long memory type of behavior of the sample ACF of absolute and squared log-returns*) is usually interpreted as evidence for *long memory in the volatility* of financial log-returns and seems to provide strong evidence against the GARCH models. Indeed, GARCH models have *short memory.* To be more precise, they are *strongly mixing with geometric rate*; see Davis et al. [22] for details. A particular consequence of this observation is that the theoretical autocorrelation functions (ACFs) of a GARCH process (X_t), and more generally, of the process $(f(X_t))$ for any measurable function f, decay to zero at an exponential rate. This fact implies *short memory* for the GARCH process and the corresponding processes of its absolute values and squares. The *short memory property* should be reflected in the behavior of the *sample ACF.* A hasty identification of the behaviors of the theoretical ACF and the sample ACF would imply that the sample ACF of a GARCH(p,q) process should decay to zero at an exponential rate and eventually vanish. This is in contradiction with the stylized fact mentioned above.

Before concluding that the GARCH(p,q) processes fail to capture the mentioned stylized feature, an attempt should be made to reconcile the empirical findings with the theoretical facts. A possible explanation (which we had considered as plausible for some time) for the mentioned contradiction could be the neglection of the statistical uncertainty present in the *sample ACF.* It is conceivable that the sample ACF at large lags could be non-zero and at the same time statistically insignificant. This argument seems especially plausible in the light of the mentioned heaviness of the tails of log-returns which implies a considerable variability in the estimated ACF. Since little was known about the behavior of the *sample ACF* of GARCH processes when the assumption of finite 4th moment for the marginal distribution is violated, investigations, aiming at assessing the truth behind the mentioned explanation, were conducted in Davis and Mikosch (1998) for the ARCH(1), Mikosch and Stărică [48] for the GARCH(1,1) and Davis et al. [22] in the general GARCH(p,q) case. This work can be summarized as follows. Log-returns could have infinite 3rd or 4th moments (see Embrechts et al. [29] for the statistical theory of tail and high quantile estimation, in particular Chapter 6 for methods to detect how heavy the tails of real-life data are). Therefore one expects that the rate of convergence of the sample autocorrelations to their theoretical counterparts is much slower than $\sqrt{n}$-rate encountered in classical time series theory and that the asymptotic normal limiting distributions in the classical large sample theory are replaced with much heavier tailed stable distributions. This results in extremely wide confidence bands that would render the non-zero values of the sample ACF of the absolute values and squares at large lags statistically unsignificant, even for huge sample sizes. In particular, the long memory type of behavior observed in the sample ACF of absolute values and squares would not be in contradiction with the short memory property of the GARCH processes.

The investigations conducted in the above mentioned papers show that the sample ACF is indeed a poor estimator of the theoretical ACF in the situations described. In particular, if the variance of the data is believed to be finite, but the 4th moment is not, it does not make sense to look at the sample ACF of the squared log-returns because the ACF is not defined and, even worse, the sample autocorrelations (which can be defined for any time series, independent of the existence of the moments) have non-degenerate limit distributions.

In the light of these theoretical findings, the heavy-tailedness of the GARCH model could be one possible explanation for the slow decay of the sample ACF of absolute and squared log-returns. This would render the long memory in the volatility spurious. In Mikosch and Stărică [48] this issue is investigated empirically by means of an analysis of a long high frequency time series of log-returns on the JPY/USD spot rates. Even when accounting for the mentioned larger-than-usual statistical uncertainty, by applying the wide confidence bands for the sample ACFs which are imposed by the estimated parameters α_0, α_1 and β_1 of a GARCH$(1,1)$ process, we did not find sufficient evidence that a GARCH process could explain the effect of almost constant sample ACF at large lags for absolute values and squares of long time series of log-returns.

In this paper we propose a simple explanation for the mentioned behavior of the sample ACFs of absolute and squared log-returns: a very plausible type of *non-stationarity*, i.e. *shifts in the unconditional variance* of the model underlying the log-returns.

Since the GARCH$(1,1)$ does not seem to provide a good description of long series of log-returns even when treated with the right amount of statistical care, and if one does not want to exclude GARCH processes as realistic models for log-returns, the modelling effort should be focused on shorter time series. We have noticed in our empirical work with log-returns that, at least when the sample ACF behavior is concerned, GARCH models fitted to daily time series covering one or two years of data are in better agreement with the theoretical results. More concretely, the sample ACFs of the data, their absolute values and squares behave very much in line with the theory on the sample ACF of GARCH processes: they vasnish at all lags or they decay exponentially to zero, respectively. Thus a possible strategy (popular among practitioners) is to fit the model to short time series and change it as soon as one realizes that something goes wrong with it (for example when one sees long memory developing in the sample ACF of the absolute values of the log-returns). This approach is clearly related to the question whether or not the real-life time series can be modeled as a sequence of stationary GARCH processes with varying parameters.

With this approach in mind, a question naturally arises:

How can we check whether a GARCH model gives a good fit to the data and when do we have to change it?

It is one of the aims of this paper to give a partial answer to this question.

Before discussing this in more detail, we would like to mention another less known anomaly that affects the modeling of long log-return series by one GARCH$(1,1)$ process. As we have already mentioned, it is known that GARCH processes have Pareto-like tails (1.2). However, only in a very few cases the value of κ, the tail index occurring in (1.2), can be obtained as closed form solution to an equation involving both the coefficients and the innovations in the model, allowing for a fully parametric estimation of the tail index. Hence, under most of the GARCH(p,q) models, one has to rely on semi-parametric statistical estimation procedures for the tail index. These methods can be applied whenever the tails of the marginal distribution have a behavior of the type (1.2); see Section 6.4 in Embrechts et al. [29]. *The* exception is the GARCH$(1,1)$ for which one can calculate κ as a function of the parameters α_1, β_1 and the distribution of the innovations Z. For this model we can compare the semi-parametric estimates of κ based only on the assumption (1.2) with the fully parametric ones based on the estimated parameters, α_1 and β_1, and the fitted innovations. The result of this comparison can be summarized as follows:

- The semi-parametric estimation techniques suggest values of κ which are significantly higher than the parametric estimates which are solutions to the equation for κ. In other words, the tails implied by the GARCH$(1,1)$ model fitted to long log-returns time series are significantly heavier than the tails of the data.

This is mainly due to the fact that the sum of the estimated coefficients of a GARCH$(1,1)$ model fit to longer time series, $\hat{\varphi}_1 = \hat{\alpha}_1 + \hat{\beta}_1$ is usually close to 1. Indeed, the theory for the GARCH$(1,1)$ in Bollerslev [13], cf. Mikosch and Stărică [48], explains that, in this case, one is close to the situation when the second moment of the theoretical model is infinite.

In the light of this fact and since there is little statistical evidence that log-returns have infinite variance (see Embrechts et al. [29], Section 6.4, for an extensive discussion and various examples) another natural question appears:

How does one explain that $\hat{\varphi}_1 \approx 1$?

Another aim of this paper is to show by theoretical means and empirical examples that the $\hat{\varphi}_1 \approx 1$ effect is spurious and might be due to non-stationarity in the time series.

We mention at this point that the occurence of almost integrated GARCH$(1,1)$ in the practice of fitting ARCH type models to log-returns, i.e. $\hat{\varphi}_1 \approx 1$, is another *stylized fact* of the empirical research in financial time series analysis . Besides its impact on the tail behavior of the estimated GARCH$(1,1)$ model, this empirical finding, if taken at face value, makes a strong statement about the forecasted volatility. This can be phrased as:

- The persistence in variance, i.e. the degree to which past volatility explains current volatility, as measured by ARCH models fitted to the data, is substantial.

Apparently, both the long memory type of behavior of the sample ACF of the absolute values and squares of the log-returns and the persistence in variance, seem to establish a strong connection between volatilities separated by large intervals of time. The first finding seems to say that significant correlation exists between the present volatility and remote past volatilities while an interpretation of the second fact can be given in terms of the forecast of the variance. The GARCH$(1,1)$ model describes the behavior of the m-period-ahead forecast of the variance

$$\sigma^2_{t+m|t} := E(X^2_{t+m} \mid X_t, X_{t-1}, \ldots)$$

through a first-order difference equation in the forecast horizon m:

$$\sigma^2_{t+m|t} = \alpha_0 + \varphi_1\, \sigma^2_{t+m-1|t}. \tag{1.3}$$

Since estimation produces values of φ_1 close to 1, the equation (1.3) implies a strong persistence of shocks to volatility. For example, a value of $\varphi_1 = 0.96$ estimated from weekly log-returns on the New York Stock Exchange *July 1962–December 1987* (Hamilton and Susmel [38]), would imply that any change in the stock market this week will continue to have non-negligible consequences a full year later: $0.96^{52} = 0.12$.

However, given the importance of measuring the degree to which past volatilities determine and explain the current volatility, no hasty conclusions should be drawn. The modeling of the prices of contingent claims, such as options, relies on the perception of how permanent shocks to variance are: a shock that is expected to vanish fast will have a smaller impact on the price of an option far from its expiration than a shock that is mostly permanent. The correct assessment of the relationship between past and present volatility is a key aspect of understanding such issues. Hence, a careful investigation of various possible explanations for the mentioned empirical facts should be carried out, with emphasis on the understanding of the statistical subtleties of this issue. Our paper is part of this effort.

Let us say a few words about the data which have been used to uncover the mentioned *stylized facts*, bearing in mind that the kind of data one analyzes can strongly determine the results of the analysis. One fact immediately draws even the attention of a not-so-careful reader of the econometrics literature on the long memory property of the volatility: an overwhelming proportion of the time series used to document and model long memory cover *extremely long time spans*, usually decades of economic activity. The latter are inevitably marked by turbulent years of crises and, possibly, structural shifts. Ding et al. [27] discuss the slow decay of the sample ACF of powers of absolute daily log-returns in the S&P *1928–1990*. Ding and Granger [26] found the same type of behavior for the sample ACF of the daily log-returns of the Nikkei index *1970–1992*, of foreign exchange (FX) rate log-returns Deutsche Mark versus U.S. Dollar *1971–1992* and of Chevron stock *1962–1991*. Bollerslev and Mikkelsen [15] use the S&P500 daily log-returns *1953–1990* to fit their

FIEGARCH model while Breidt et al. [17] fit a stochastic volatility model with long memory in the volatility to the daily log-returns of the value-weighted market index *1962–1987*.

There is also plenty of empirical evidence for integrated GARCH$(1,1)$ behavior, i.e. evidence for the parameter $\varphi_1 = \alpha_1 + \beta_1$ being close to 1; see Bollerslev, Chou and Kroner [14] and the references therein. The results reported in the two last mentioned studies are exemplary. For example, Bollerslev and Mikkelsen [15] fit a GARCH$(1,1)$ to the S&P500 daily log-returns *1953–1990* and get an estimate of $\hat{\varphi}_1 = 0.995$. For the same model, Breidt et al. [17] obtain an estimate of $\hat{\varphi}_1 = 0.999$ for the daily log-returns of the value-weighted market index *1962–1987*.

It is worth mentioning that, while studies of daily asset log-returns have frequently found integrated GARCH behavior, studies with higher-frequency data *over shorter time spans* have often uncovered weaker persistence. For example, Baillie and Bollerslev [5] report on the estimation of GARCH$(1,1)$ models for hourly log-returns on FX rates of British Pound (GBP), Deutsche Mark (DM), Swiss Franc (CHF) and Japanese Yen (JPY) versus U.S. Dollar (USD) *January 1986–July 1986* (this is only a 6.5 months period!). The estimated parameters $\hat{\varphi}_1^{GBP} = 0.606$, $\hat{\varphi}_1^{DM} = 0.568$, $\hat{\varphi}_1^{CHF} = 0.341$, $\hat{\varphi}_1^{JPY} = 0.717$ are in sharp contrast to the almost integrated GARCH$(1,1)$ models fitted to daily log-returns.

Much lower persistence in variance than suggested by almost integrated GARCH$(1,1)$ models can be detected if one allows for changes in the level of unconditional variance. Such models were considered, for example, by Hamilton and Susmel [38] and Cai [20]. Hamilton and Susmel fitted their SWARCH model to weekly log-returns from the New York Stock Exchange *July 1962–December 1987*. They reported a measure of persistence (related to $\hat{\varphi}_1$ of the GARCH$(1,1)$) of 0.4. This implies that shocks to volatility die out almost completely after a month $(0.4^4 = 0.05)$.

The message of this article can be formulated in one sentence: **it might be misleading to take the empirical evidence of long memory and strong persistence of the volatility in log-returns at face value, especially when it comes from the analysis of time series that cover long periods.**

In particular, we challenge the following two implications:

1. if *the sample autocorrelations of the absolute and squared values are significantly different from zero for large lags* then *there is long range dependence (LRD) or long memory in the data*,

2. if *the persistence in variance as measured by ARCH type processes is high* then *past volatility explains current volatility.*

In doing so we follow on the steps of Boes and Salas [12], Potter [53], Bhattacharya et al. [9], Anderson and Turkman [1], Teverovsky and Taqqu [59] and others with respect to the first statement (although these references are not directly related to finance) and Diebold [25], Lamoureux and Lastrapes [45], Hamilton and Susmel [38], Cai [20] with respect to the second one.

We will show that the type of behavior described by the mentioned stylized facts can be due simply to a very plausible type of *non-stationarity*: *shifts in the unconditional variance* of the model underlying the log-returns. Given the fast rate at which new technological and financial tools have been introduced in the financial markets, the case for the existence of structural changes (and thus for lack of stationarity) seems quite strong.

The goal of providing an alternative explanation for the observed empirical facts is achieved in a sequence of steps. *Firstly,* in Section 2 we build a statistical goodness of fit procedure (basically a Kolmogorov–Smirnov test in the frequency domain) which we use for detecting possible structural shifts in the log-return series. Our procedure is based on the belief that GARCH processes are reasonable models for log-returns over suitably short periods of time, and that changes from one GARCH model to another occur in the log-return time series.

Secondly, we prove that the mentioned type of non-stationarity

- causes the slow decay of the sample ACF of the absolute and squared log-returns (Section 3.1.1),
- forces the sum of the ARCH and GARCH parameters to 1, when such a non-stationary time series is modeled by fitting just one GARCH process (Section 3.1.2).

Thirdly, in Section 5 we use the goodness of fit procedure to detect possible structural changes in the unconditional variance in the S&P500 log-return series and show how a long memory type of behavior might be induced in the sample ACF by these shifts. Our procedure identifies the recession periods as being structurally different. The major structural change is detected between 1973 and 1975 corresponding to the oil crises.

A conclusion relevant to the methodology of GARCH modeling also follows: *long log-return series should not be modeled just by one GARCH process.* The parameters of the model must be periodically updated.

Our perspective on the time series analysis of long log-return data sets can be summarized as follows: statistical tools and procedures (such as the sample autocorrelations, parameter estimators, periodogram) gather meaningful information and perform the tasks for which they were designed only under certain assumptions (stationarity, light tails, ergodicity, etc.). Hence, if one or several of these assumptions are violated by the data at hand, the reading of these tools as well as the output of our procedures are rendered meaningless. More colorfully, when used inappropriately, the statistical tools and procedures could “see things that are not there”.

Let us briefly illustrate this point in relation to the firstly mentioned *stylized fact* that we reformulate in a more specific form: *power law decay of the sample ACF of the absolute values of*

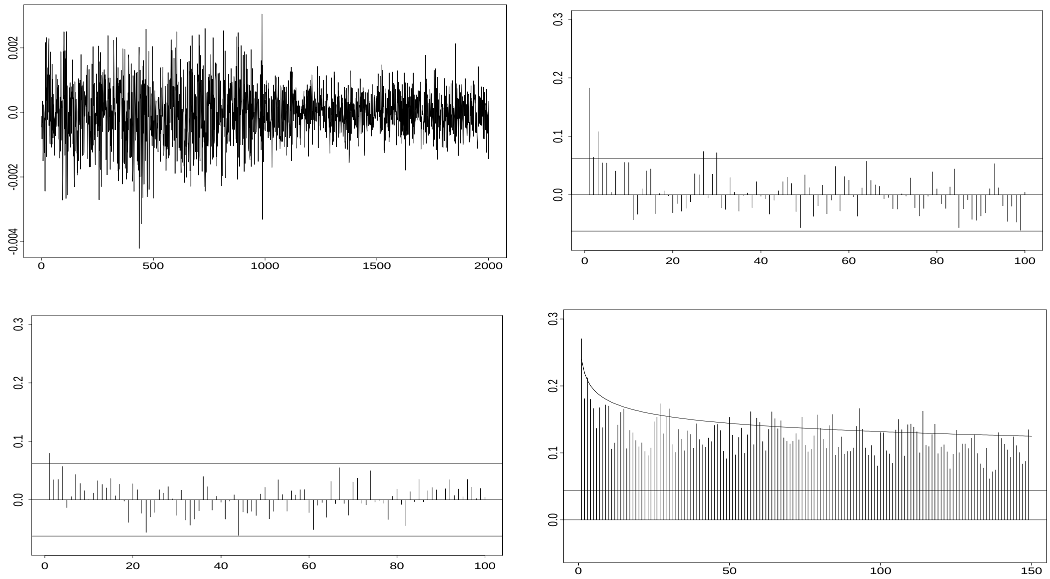


Figure 1.1 Top left: *Concatenation of two* GARCH(1, 1) *time series with parameters* (1.4) *and* (1.5). Top right: *Sample ACF of the absolute values of the first 1000 values.* Bottom left: *Sample ACF of the absolute values of the second 1000 values.* Bottom right: *Sample ACF for the absolute values of the whole time series. The hyperbolic line is given by the function* $\gamma(h) = 0.24h^{-0.13}$. *The horizontal lines in the ACF plots indicate the usual* 95% *confidence bands for the sample ACF of iid Gaussian noise.*

log-returns is an indication for long memory in the volatility of log-return process. We simulated realizations $(X_t)_{t=1,\dots,1000}$ and $(Y_t)_{t=1,\dots,1000}$ from two independent GARCH(1, 1) processes with corresponding parameters

$$\alpha_0 = 0.13 \times 10^{-6}, \quad \alpha_1 = 0.11, \quad \beta_1 = 0.52\,, \tag{1.4}$$

$$\alpha_0 = 0.17 \times 10^{-6}, \quad \alpha_1 = 0.20, \quad \beta_1 = 0.65. \tag{1.5}$$

The innovations are standard normal. The top left graph in Figure 1.1 displays the two time series concatenated. The difference in unconditional variance of the two pieces is clearly noticeable. The top right and bottom left graphs display the sample ACFs of the time series $|X_1|, \dots, |X_{1000}|$ and $|Y_1|, \dots, |Y_{1000}|$, respectively. GARCH(1, 1) processes are strongly mixing with geometric rate and, hence, the theoretical ACF of the absolute values is well defined (for our parameter choices) and decays to zero exponentially fast. Under the choice of parameters (1.4) and (1.5), the 4th moments of X and Y are finite. Therefore the sample ACFs converge to the theoretical ones at $\sqrt{n}$-rate and the asymptotic limits are normal; see Mikosch and Stărică [48]. Here we are using the sample ACF tool in a proper way and the readings are meaningful: the sample ACF decays to zero quickly, as expected.

The bottom right graph in Figure 1.1 displays the sample ACF for the juxtaposition of the two pieces, i.e. for $|X_1|, \ldots, |X_{1000}|, |Y_1|, \ldots, |Y_{1000}|$. Long range dependence type behavior of the sample ACF develops even though we know there is no long memory in the data. The explanation lies in the change of the unconditional variance of the time series as we will rigorously prove in Section 3.1. Inferring the presence of long memory from this sample ACF would be an instance where, through misuse, statistical tools could "see things that are not there". The sample ACF reading of LRD is rendered meaningless by the non-stationarity in the data. This example shows that LRD type behavior of the sample ACF can be caused either by stationary long memory time series or, equally well, by non-stationarity in the time series.

The field of long memory detection and estimation is particularly (in)famous for the numerous statistical instruments that behave similarly under the assumptions of long range dependence and stationarity or under weak dependence affected by some type of non-stationarity. Three examples of statistics that are frequently used in the detection and estimation of long memory and that are mired by the mentioned lack of power to discriminate between possible scenarios will set the frame for the developments in Section 3.1 and Section 4.

The first one is probably the most famous, as it is related to the very phenomenon that brought the issue of long range dependence to the forefront of statistical research, i.e the *Hurst effect.* This effect is defined in terms of the asymptotic behavior of the so-called R/S (range over standard deviation) statistic. This statistic was proved to have the same kind of asymptotic behavior when applied to a stationary long memory time series or to a short memory time series perturbed by a small monotonic trend that even converges to 0 as time goes to infinity (Bhattacharya et al. [9]). See Anderson and Turkman [1] for another instance where a long memory type behavior for the R/S statistic of a stationary 1-dependent sequence is proved.

The second tool deals with one of the traditional methods used in the detection of long memory. It examines the sample variance of the time series at various levels of aggregation. Under stationarity and long memory, the variance of the aggregated series

$$X^{(m)}(k) = \frac{1}{m} \sum_{i=(k-1)m+1}^{km} X_i, \quad m \geq 1, \quad k = 1, \ldots, [n/m]$$

behaves asymptotically like a power of the aggregation level m:

$$\text{var}(X^{(m)}) \sim cm^{\beta},$$

where $\beta = 2H - 2$ and H is the so-called *Hurst exponent.* This behavior suggests to take the slope of the regression line of the plot of the logarithm of the sample variance versus $\log m$ as an estimator of β. However, Teverovsky and Taqqu [59] showed that this estimator performs similarly when applied to a long memory stationary time series or to a stationary short memory one that was perturbed by shifts in the mean or small trends. This could make one believe that there is long range dependence in the data when in reality a jump or a trend is added to a series with no long range dependence.

In the more specialized context of long memory detection in the volatility of log-returns, the results of a simulation study in de Lima and Crato [24] give clear evidence of the unreliability of the methods for detecting long memory which are commonly used in the econometrics literature. In the first step of the study, the Geweke–Porter–Hudak procedure and the R/S statistic (adjusted in the spirit of Lo [46]) were used to document the presence of long memory in the volatility of five log-return time series. In the second step, GARCH$(1,1)$ models were fitted to these time series and a thousand simulated log-return series were generated for each set of parameters. The simulated time series clearly had short memory. The two mentioned methods were then used to test the short memory null hypothesis. *Both tests for long memory clearly reject the null of short memory in the data generated by the* GARCH$(1,1)$ *models.* (It is, however, surprising to see how determined the authors are to "find" long memory in the volatility. At odds with the evidence they provide against the methods under discussion, their conclusion reads: "Using both semi-parametric and non-parametric statistical tests, we found compelling evidence of persistent long run dependence in the squared returns series". However, this fact does not diminish the value of their evidence.)

In Section 3.1 we show, theoretically and through examples, that the basic statistical tools for gauging the long memory phenomenon: the sample ACF and the periodogram do also suffer from the mentioned lack of power to discriminate between long memory and non-stationarity. We prove that the sample ACF and the periodogram for a time series with deterministically changing mean exhibit behavior similar to that implied by the assumptions of stationarity and LRD. Our study shows how a change in the unconditional variance of a time series with short memory causes not only slow decay, but even almost constancy of the sample ACF of absolute and squared log-returns. This behavior translates into the periodogram of the absolute (squared) values: the periodogram has a significant peak at zero. In Section 5 we use the theoretical results of the paper to detect structural changes in the S&P500 log-return series. Then we show how a long memory type of behavior might be induced in the sample ACF by these changes.

The task of uncovering the phenomena behind the empirical findings of slowly decaying sample ACFs and high persistence of the volatility is rendered more difficult by the fine distinction (many times just a matter of belief) between non-stationarity and LRD stationarity. To illustrate this consider the time series in Figure 1.2. Here the observations X_t are 0 or 1 and the lengths of ON periods (spells of 1s) and OFF periods (spells of 0s) are iid Pareto distributed with tail index 1. It is easy to see that the behavior of the theoretical ACF is: $\mathrm{corr}(X_t, X_{t+h}) \approx ch^{-1}$, hence the time series exhibits long memory. When the time series in the left graph of Figure 1.2 is analyzed, the assumption of stationarity (and long memory) is plausible, while, when only in possession of

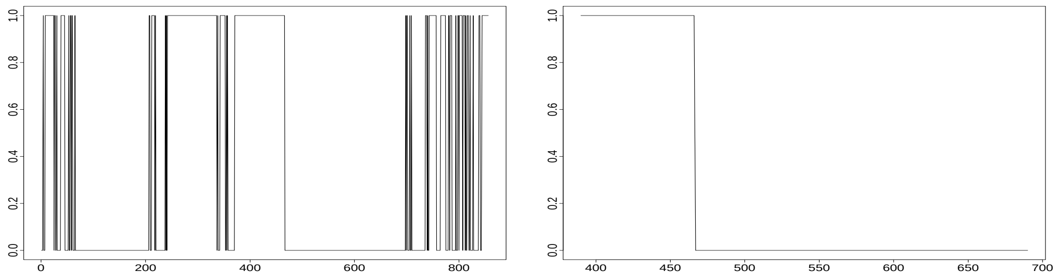


Figure 1.2 Left: $X_t,\ t = 1, \ldots, 856$. *The long memory stationary assumption is plausible.* Right: $X_t,\ t = 390, \ldots, 690$. *The hypothesis of structural change is plausible if only a part of the original time series is available.*

the observations displayed in the second graph, i.e. observation 390 up to 690, the hypothesis of a structural change is more plausible.

This fine distinction between long memory stationarity and non-stationarity is the source for various competing explanations for the empirical findings under discussion. Our paper postulates *non-stationarity of the unconditional variance* as a possible source of both the slow decay of the sample ACF and the high persistence of the volatility in long log-return time series as measured by ARCH type models. Hence we claim that both these findings in long time series could be spurious. Other studies (Baillie et al. [6]) have argued that the slow decay of the sample ACF correctly reflects the presence of a stationary long memory time series while the apparent integrated GARCH behavior is an artifact of a long memory process investigated through the estimation of a GARCH$(1,1)$ model. While the "correct" explanation behind the empirical facts is still elusive, it is worth mentioning that, while Baillie et al. [6] base their explanation on simulations, we prove our results analytically.

Part of this paper is in the spirit of work by Lamoureux and Lastrapes [45] which investigates the extent to which persistence in variance might be overstated in the ARCH framework because of the existence of and failure to take into account structural shifts in the model. The present paper extends their investigation in two directions. The first one addresses the issue of detection of the structural shifts in the model (and implicitly in the unconditional variance). While the mentioned authors acknowledge the difficulty of the task of detection and decide to allow for regular structural shifts that happen at ad-hoc chosen intervals of time, we propose a statistical procedure which we use to assess changes in the hypothesized model. Besides helping to investigate the possible causes of the empirical findings mentioned at the beginning, this part of the paper is particularly important as it proposes a quantitative method to assess the goodness of fit of a GARCH$(1,1)$ model.

The method, essentially a Kolmogorov–Smirnov goodness of fit procedure, is based on the spectral properties of the GARCH$(1,1)$ process. It is introduced and theoretically investigated in

Section 2. We assume that the data X_t come from a stationary GARCH$(1,1)$ process with spectral density f_X. Our test statistic, S is related to the integrated periodogram and can be used to detect a deviation of the empirical spectrum from the hypothesized one. The basic idea consists of calculating S on a moving window rolling over the time series $X_1, \ldots, X_n$ and checking whether the calculated values are explained by the distribution of the limit process of S under the hypothesized model. As long as the statistic lies inside the confidence region we conclude that the hypothesized model provides a satisfactory description to the data. The presence of the statistic outside the confidence region for consecutive windows signals that the model has stopped to describe the data. This procedure allows us to show that *one* GARCH$(1,1)$ *model cannot describe long log-returns time series* and gives evidence of fast pace of change in the log-return data. It can also be used as an auxiliary tool for deciding when the coefficients of the model should be re-estimated. Although only the implications of the goodness of fit test for the GARCH$(1,1)$ modeling methodology are discussed, we provide the theory for the goodness of fit test is the general case of the GARCH(p,q) model.

The second improvement upon the work of Lamoureux and Lastrapes [45] consists in providing an analytical treatment of the effect of misspecification of the GARCH$(1,1)$ model on the behavior of the parameter $\varphi_1 = \alpha_1 + \beta_1$. The mentioned authors use simulation studies to assess this impact. We show that change in the structure of a time series which consists of pieces from different GARCH models forces the sum of all ARCH and GARCH parameters close to one when these parameters are estimated from the whole time series. This will be made precise in Section 3.2.

We now proceed with the theory underlying the goodnes of fit test procedures used intensively in the rest of the paper. The limit theory for a certain two-parameter process which is provided in the next section is slightly more general than needed for the purposes of this paper. However, the theory for the corresponding one-parameter process is essentially the same as for the case of two parameters. The latter case can be used for change point detection in the spectral domain, the former one for goodness of fit tests in that domain. In the context of this paper, we are mainly interested in test statistics for the goodness of fit of GARCH processes.

2 Limit theory for the two-parameter integrated periodogram

2.1 A preliminary discussion

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. One of the aims of this paper is to provide a procedure for testing how good the fit of a GARCH(p,q) model is. This procedure will also allow us to single out the parts of the data which are not well described by the hypothesized model. To be precise, we assume that the data come from a stationary *generalised autoregressive conditionally heteroscedastic process of order* (p,q), for short GARCH(p,q):

$$X_t = \sigma_t\, Z_t, \quad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i X_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-i}^2\,, \quad t \in \mathbb{Z}\,, \tag{2.1}$$

where (Z_t) is an iid symmetric sequence with $EZ^2 = 1$, non-negative parameters α_i and β_j, and the *stochastic volatility* σ_t is independent of Z_t for every fixed t. 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 modeling the log-returns of stock indices, share prices, exchange rates, etc., and has found its way into practice for forecasting financial time series. See for example Engle [30] for a collection of papers on ARCH. We assume that, for a particular choice of parameters α_i and β_i, the sequence $((X_t,\sigma_t))$ is stationary. Assumptions for stationarity of a GARCH(p,q) can be found for example in Bougerol and Picard [16] or Nelson [49].

Our analysis is based on the spectral properties 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 [18] or Priestley [54]. Under general conditions, the *integrated periodogram* or *empirical spectral distribution function*

$$\frac{1}{2\pi} J_{n,X}(\lambda) = \frac{1}{2\pi} \int_0^\lambda I_{n,X}(x)\, dx\,, \quad \lambda \in [0,\pi]\,, \tag{2.2}$$

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. This approach is based on the idea of treating the integrated periodogram as the empirical spectral distribution function (see for example Grenander and Rosenblatt [37], Whittle [60] or Bartlett [7]) and its 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 [52] or Shorack and Wellner [57]. However, the theory for the integrated periodogram differs significantly from the empirical process theory. For example, given a finite 4th moment for X and supposing that X_t is a linear process, the limit of $\sqrt{n}(J_{n,X} - F_X)$ in $\mathbb{C}[0,\pi]$, the space of continuous functions on $[0,\pi]$ endowed with the uniform topology, is an unfamiliar Gaussian process, i.e. its covariance structure depends on the spectral density f_X; see for example Anderson [3] or Mikosch [47]. 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 [37], one needs to modify the integrated periodogram to get a more familiar Gaussian process, if possible a bridge type process. Bartlett [7] (cf. Priestley [54]) had the idea to use a weighted form of the integrated periodogram:

$$J_{n,X,f_X}(\lambda) = \int_0^\lambda \frac{I_{n,X}(x)}{f_X(x)}\, dx\,, \quad \lambda \in [0,\pi]\,. \tag{2.3}$$

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 of the process, but it is still dependent on the 4th moment of the underlying noise process; see for example Mikosch [47] for a discussion. There it is also mentioned that one can overcome the problem with the 4th moment by replacing the periodogram in (2.3) 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 [44]) 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 [57]. To the best of our knowledge, this idea was used first by Picard [51]. 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 [33] or Klüppelberg and Mikosch [43]. The basic idea that underlies the detection of changes in the spectral distribution function is as follows. The test statistic is constructed from the two-parameter process

$$J_{n,X}(x,\lambda) = \int_0^{\lambda} \frac{I_{n,[nx],X}(y)}{f_X(y)}\,dy\,,\quad x\in[0,1]\,,\quad \lambda\in[0,\pi]\,, \tag{2.4}$$

where

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

Again, Bartlett's idea (divide the periodogram by the spectral density) was used to make the limit process independent of the spectral density. The process $J_{n,X}(x,\lambda)$ (2.4), properly centered and scaled, 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 [57].

Functionals of $J_{n,X}(x,\lambda)$ of Kolmogorov–Smirnov type can be used to detect a deviation of the empirical spectrum from the hypothesized one as follows. The empirical spectral distribution function for every subsample $X_1,\ldots,X_{[nx]}$, $0\le x\le 1$, is calculated and compared with the value of the true spectral distribution function. The constructed change point statistic reacts to deviations from the null hypothesis that the whole sample $X_1,\ldots,X_n$ is taken from the same stationary sequence, with the prescribed spectrum.

2.2 Main results

In what follows, we develop a change point analysis for GARCH processes similar to the one for linear 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: $f_X \equiv \mathrm{var}(X)/(2\pi)$. By using a test statistic which is a functional of $J_{n,X}(x,\lambda)$ (or of a modied version of it), we can test for a change of variance of the underlying time series. That variance 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 behavior 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 below, where we give the necessary limit theory for a modified two-parameter integrated periodogram.

As a motivation for the following, we start by considering the two-parameter process $J_{n,X}(x,\lambda)$ related to (2.2) (see also Appendix A1):

$$\begin{aligned} J_{n,X}(x,\lambda) &= \int_0^\lambda \left(\gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,[nx],X}(h)\cos(yh) \right) dy \\ &= \lambda \gamma_{n,[nx],X}(0) + 2\sum_{h=1}^{n-1} \gamma_{n,[nx],X}(h)\frac{\sin(\lambda h)}{h}\,, \end{aligned} \tag{2.5}$$

where

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

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

$$\gamma_X(h) = \operatorname{cov}(X_0, X_h) \quad \text{and} \quad v_X(h) = \operatorname{var}(X_0X_h) = E(X_0X_h)^2\,, \quad h\in\mathbb{Z}\,.$$

The processes $\gamma_{n,[n\cdot],X}(h)$ 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 [40] or Bickel and Wichura [10].

Lemma 2.1 *Consider the* GARCH(p,q) *process* (X_t) *given by* (2.1). *Assume that*

$$E|X|^{4+\delta} < \infty \quad \textit{for some } \delta > 0. \tag{2.6}$$

Then for every $m\ge 1$, as $n\to\infty$

$$\sqrt{n}\left(\gamma_{n,[nx],X}(h), h=1,\ldots,m\right)_{x\in[0,1]} \stackrel{d}{\to} \left(v_X^{1/2}(h)\; W_h(x), h=1,\ldots,m\right)_{x\in[0,1]}\,, \tag{2.7}$$

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

The proof of the lemma is given in Appendix A1.

A naive argument, based on Lemma 2.1 and the decomposition (2.5), suggests that

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

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 2.2 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 for a goodness of fit procedure, we would like that the asymptotic distribution of those functionals is independent of the null hypothesis we test. In other words, we want a "standard" Gaussian process in the limit since otherwise we would have to evaluate the distributions of its functionals by Monte–Carlo simulations *for every choice of parameters of the* GARCH(p,q) we consider in the null hypothesis.

A glance at the right-hand side of (2.5) 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, it is intuitively clear that, if we replaced in (2.5) the processes $\gamma_{n,[n\cdot],X}(h)$ by

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

the limit process would become 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 2.2 *Let* (X_t) *be a stationary* GARCH(p,q) *process given by* (2.1). *Assume that* (2.6) *holds. Then*

$$\sqrt{n}\,(C_{n,X}(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} \quad \stackrel{d}{\to} \quad (K(x,\lambda))_{x\in[0,1],\lambda\in[0,\pi]} = \left(\sum_{h=1}^{\infty} W_h(x)\frac{\sin(\lambda h)}{h}\right)_{x\in[0,1],\lambda\in[0,\pi]}, \tag{2.8}$$

in $\mathbb{D}([0,1]\times[0,\pi])$ *where* $(W_h(\cdot))_{h=1,\ldots}$ *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*

$$\begin{aligned} E\left(K\left(x_1,\lambda_1\right)\ K\left(x_2,\lambda_2\right)\right) &= \min\left(x_1,x_2\right)\sum_{t=1}^{\infty}\frac{\sin\left(\lambda_1 t\right)\sin\left(\lambda_2 t\right)}{t^2} \\ &= 2^{-1}\pi^2\min\left(x_1,x_2\right)\left(\min\left(\frac{\lambda_1}{\pi},\frac{\lambda_2}{\pi}\right)-\frac{\lambda_1}{\pi}\frac{\lambda_2}{\pi}\right). \end{aligned} \tag{2.9}$$

The proof of the theorem is given in Appendix A1.

Remark 2.3 The series representation of the Kiefer–Müller process can be found in Klüppelberg and Mikosch [43]. This process is known in empirical process theory as the limiting Gaussian field for the sequential empirical process; see Shorack and Wellner [57].

Remark 2.4 The statement of Theorem 2.2 remains valid for wider classes of stationary sequences. In particular the result holds if the conditions in Remark A1.1 are satisfied and in addition, (X_t) is symmetric and $(|X_t|)$ and $(\mathrm{sign}(X_t))$ are independent. These later conditions are satisfied by any stochastic volatility model of the 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 $\sigma(Z_{t-1}, Z_{t-2}, \ldots)$, or alternatively, (σ_t) and (Z_t) are independent.

Remark 2.5 Theorem 2.2 can be used for testing the null hypothesis that the sample $X_1, \ldots, X_n$ comes from a GARCH(p,q) model with given parameters α_i and β_i against the alternative of a GARCH(p,q) model with parameters α_i^a and β_i^a. Write $v_X^a(h) = E(X_0 X_h)^2$ for the corresponding variance of $X_0 X_h$ under the alternative. Then the same arguments as for the proof of Theorem 2.2 yield under the alternative that

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

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

The dependence of the limiting field on the covariance structures of the X_t^2 makes an application of the latter result difficult. One would depend on Monte–Carlo based calculations for the quantiles of appropriate functionals of the limit processes. In particular, these quantiles would depend on the parameters under the alternative hypothesis. 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.

Immediate consequences of Theorem 2.2 and the continuous mapping theorem are limit theorems for continuous functionals of the process $C_{n,X}$ which can be used for the construction of goodness of fit tests and tests for detecting changes in the spectrum of the time series.

Corollary 2.6 *Under the assumptions of Theorem 2.2,*

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

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

For $x = 1$, convergence in (2.8) yields

$$\sqrt{n}\,\widetilde{C}_{n,X}(\cdot) := \sqrt{n}\sum_{h=1}^{n-1}\frac{\gamma_{n,X}(h)}{v_X^{1/2}(h)}\,\frac{\sin(\cdot h)}{h} \stackrel{d}{\to} B(\cdot) := \sum_{h=1}^{\infty} W_h(1)\,\frac{\sin(\cdot h)}{h}\,, \tag{2.10}$$

in $\mathbb{C}[0,\pi]$. The series on the right-hand side is the so-called Paley–Wiener representation of a Brownian bridge on $[0,\pi]$; see (2.9) with $x = 1$ (see for example Hida [39]).

The one-parameter process $\widetilde{C}_{n,X}$ will be our basic process for testing the goodness of fit of the sample $X_1,\ldots,X_n$ to a GARCH process. 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 2.7 *Under the assumptions of Theorem 2.2,*

$$\sqrt{n}\sup_{\lambda\in[0,\pi]}\left|\widetilde{C}_{n,X}(\lambda)\right| \stackrel{d}{\to} \sup_{\lambda\in[0,\pi]}|B(\lambda)|\,, \tag{2.11}$$

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

We are now ready to approach the study of the effects of non-stationary assumptions on the behavior of statistical tools and procedures that are designed to estimate characteristics of stationary time series.

3 Statistical instruments and procedures under non-stationarity

3.1 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 can be 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 shows up in large values of the estimated (via the periodogram) spectral density of the data for small frequencies.

This phenomenon is often interpreted as a manifestation of *long range dependence* (LRD) in the volatility of log-returns. 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 $\sum_h|\gamma_Y(h)| = \infty$, where γ_Y denotes the ACF of the sequence (Y_t). 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 [8] for various definitions of LRD.

The detection of LRD effects is based on statistics of the underlying time series, such as the sample ACF, the periodogram, the R/S statistic, the sample variances for aggregated time series, etc. The assumptions that the data are stationary and LRD imply a certain characteristic behavior of these statistics. The detection of LRD is reported when the statistics seem to behave in the way prescribed by the theory. However, as mentioned in the introduction, the field of LRD detection and estimation is well known for the great number of statistics that display the behavior prescribed by the theoretical assumptions of stationarity and LRD under alternative assumptions, most often short memory perturbed by some kind of non-stationarity. For example, Bhattacharya et al. [9] proved such a problematic behavior for the R/S statistc when the data contains a trend while Teverovsky and Taqqu [59] showed that the sample variances of aggregated time series develop LRD type of behavior when applied to short memory data affected by trends or shifts in the mean.

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

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 $f_{Y^{(i)}}$. Clearly, if the subsamples have different marginal distributions, the resulting sample $Y_1, \ldots, Y_n$ is non-stationary.

3.1.1 The sample ACF under non-stationarity

Define the sample autocovariances of the sequence (Y_t) as follows:

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

where $\overline{Y}_n$ denotes the sample mean.

In the sequel we show that if the expectations $EY^{(j)}$ of the subsequences $(Y_t^{(j)})$ are different, the sample ACF for sufficiently large h approaches a positive constant. We also show that the periodogram becomes arbitrarily large for Fourier frequencies close to 0. This behavior can mislead one in inferring LRD when in fact one analyses a non-stationary time series with subsamples that have different unconditional moments.

By the ergodic theorem it follows for fixed $h \geq 0$ as $n \to \infty$ that

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

$$\begin{aligned}
&\to \sum_{j=1}^{r} p_j E(Y_0^{(j)} Y_h^{(j)}) - \left(\sum_{j=1}^{r} p_j EY^{(j)}\right)^2 \\
&= \sum_{j=1}^{r} p_j \gamma_{Y^{(j)}}(h) + \sum_{1\le i<j\le r} p_i p_j \left(EY^{(j)} - EY^{(i)}\right)^2 \quad \text{a.s.} \qquad (3.2)
\end{aligned}$$

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. This is the picture one sees in the sample ACFs of both simulated and real-life data; see Sections 4–6.

The above argument applies, for example to the absolute values and squares of a sequence X_t which consists of subsamples from different GARCH processes with varying *unconditional variance.* If we apply the above argument to the non-stationary sequence (X_t) itself then (3.2) ensures that the sample ACF still estimates null at all lags as if (X_t) was a stationary GARCH process. We refer to Section 4 for empirical evidence from simulated data.

3.1.2 The periodogram under non-stationarity

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

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

In the sequel we show that if the expectations $EY^{(j)}$ of the subsequences $(Y_t^{(j)})$ are different, the periodogram becomes arbitrarily large for Fourier frequencies close to 0.

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 centering is not necessary for the sample. We observe the following:

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

Notice that

$$
\begin{aligned}
&\sum_{l=1}^{r} EY^{(l)}\, \mathrm{e}^{-i\lambda_j([nq_{l-1}]+1)} \sum_{t=0}^{[nq_l]-[nq_{l-1}]-1} \mathrm{e}^{-i\lambda_j t} \\
&= \frac{\mathrm{e}^{-i\lambda_j}}{1-\mathrm{e}^{-i\lambda_j}} \sum_{l=1}^{r} EY^{(l)} \left(\mathrm{e}^{-i\lambda_j[nq_{l-1}]} - \mathrm{e}^{-i\lambda_j[nq_l]}\right) \\
&= \frac{\mathrm{e}^{-i\lambda_j}}{1-\mathrm{e}^{-i\lambda_j}} \left[EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1} (EY^{(l)} - EY^{(l+1)})\, \mathrm{e}^{-i\lambda_j[nq_l]}\right]
\end{aligned}
$$

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

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

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

$$(n\lambda_j^2)^{-1}(1 - \cos(2\pi\{jp_1\}), \tag{3.6}$$

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 describes the behavior of the periodogram and smoothed versions of it for Fourier frequencies close to zero.

In particular, the above argument applies to a sequence (X_t) which consists of subsamples from GARCH processes with parameters that change from subsample to subsample and also to its sequence of absolute values and squares. Notice that the latter two processes consist of stationary subsamples with continous spectral densities on $[0, \pi]$. In particular, these densities do not have a singularity at zero. Thus, the apparent explosion of the estimated spectrum at zero is due to the different expectations of the absolute values or squares in various subsamples of the series and not to LRD. For some empirical evidence of this phenomenon, see Section 4.

3.2 The effect of non-stationarity on parameter estimation in GARCH$(1,1)$ models

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 Appendix A2 that $(U_t) = (X_t^2 - EX^2)$ can be rewritten as an ARMA(1,1) process with white noise innovations sequence $(\nu_t) = (X_t^2 - \sigma_t^2)$:

$$U_t - \varphi_1\, U_{t-1} = \nu_t - \beta_1 \nu_{t-1}, \quad t \in \mathbb{Z}, \tag{3.7}$$

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 asymptotically 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 asymptotically normal (with $\sqrt{n}$-rate) approximations to the true parameters of the model. We refer to Brockwell and Davis [18], Section 10.8, for the theoretical properties of this pseudo-maximum likelihood estimation procedure. The Whittle estimate $\Theta_n = (\overline{\varphi}_1, \overline{\beta}_1)$ of the ARMA$(1,1)$ model (3.7) is obtained by minimizing the Whittle likelihood function

$$\overline{\sigma}_n^2(\Theta) = \frac{1}{n} \sum_j \frac{I_{n,U}(\lambda_j)}{f(\lambda_j, \Theta)} \tag{3.8}$$

with respect to Θ from the parameter domain

$$\mathcal{C} = \{(\varphi_1, \beta_1) : \ -1 < \varphi_1, \beta_1 < 1\}.$$

(Clearly, both β_1 and φ_1 are non-negative. But for theoretical reasons we need $\mathcal{C}$ to be open, while for practical reasons we do not want to exclude $\beta_1 = 0$ or $\varphi_1 = 0$.) The sum $\sum_j$ in (3.8) 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 [18], Chapter 4):

$$f(\lambda, \Theta) = \frac{\sigma_U^2}{2\pi} \frac{|1 - \beta_1 \mathrm{e}^{-i\lambda}|^2}{|1 - \varphi_1 \mathrm{e}^{-i\lambda}|^2} = \frac{\sigma_U^2}{2\pi} \frac{1 - 2\beta_1 \cos \lambda + \beta_1^2}{1 - 2\varphi_1 \cos \lambda + \varphi_1^2}. \tag{3.9}$$

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 innovations sequence (ν_t) are strongly consistent. This follows along the lines of the proof of Theorem 10.8.1 in Brockwell and Davis [18]. Therein, strong consistency is proved for an ARMA(p,q) process with an iid white noise innovation sequence. However, a close inspection of pp. 378–385 in [18] 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 mentioned in the introduction, fitting of GARCH$(1,1)$ models to longer high frequency time series frequently yields estimates of φ_1 close to 1; see the empirical study in Section 4. This led to the introduction of the integrated GARCH models (IGARCH) corresponding to $\varphi_1 = 1$; see Engle and Bollerslev [31]. 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. [29]).

In what follows, we provide an explanation for the empirical finding of almost integrated GARCH ($\hat{\varphi}_1 \approx 1$). We show 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 from different GARCH$(1,1)$ models (as described in (3.1)) with corresponding parameters $\Theta^{(i)} = (\varphi_1^{(i)}, \beta_1^{(i)})$, $i = 1, \ldots, r$.

When (X_t^2) constitutes a stationary sequence , centering is not necessary in the definition of the Whittle likelihood, $\overline{\sigma}_n^2(\Theta)$ 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 $\overline{\sigma}_n^2(\Theta)$ can be neglected; they are of no importance for the asymptotic results considered. In what follows, we simply ignore these two special cases.) Thus, for the Fourier frequencies λ_j

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

and therefore it is without loss of generality assumed in [18] 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

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

It is natural to modify the Whittle likelihood function $\overline{\sigma}_n^2(\Theta)$ 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 neglected the Fourier frequencies at 0 and π, for ease of presentation).

We start with an analogue of Proposition 10.8.2 in [18].

Proposition 3.1 *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:*

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

$$\Delta(\Theta) := \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sum_{i=1}^r p_i \sigma^2_{(X^{(i)})^2} f(\lambda,\Theta^{(i)})}{f(\lambda,\Theta)}\,d\lambda + \frac{\sum_{1\le i<j\le r} p_i p_j \left(\sigma^2_{X^{(i)}} - \sigma^2_{X^{(j)}}\right)^2}{f(0,\Theta)}\,, \tag{3.10}$$

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

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

$$\to \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sum_{i=1}^r p_i \sigma^2_{(X^{(i)})^2} f(\lambda,\Theta^{(i)})}{f_\delta(\lambda,\Theta)}\,d\lambda + \frac{\sum_{1\le i<j\le r} p_i p_j \left(\sigma^2_{X^{(i)}} - \sigma^2_{X^{(j)}}\right)^2}{f_\delta(0,\Theta)} \tag{3.11}$$

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

The proof of the proposition is given in Appendix A1.

Remark 3.2 The dependence structure between the different subsamples is inessential for the validity of the proposition.

Next we formulate a result in the spirit of Theorem 10.8.1 of Brockwell and Davis [18].

Theorem 3.3 *Assume the conditions of Proposition 3.1 are satisfied. Let Θ_n be the minimizer of $\widehat{\sigma}_n^2(\Theta)$ 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* (3.10).

The proof of the proposition is given in Appendix A1.

Remark 3.4 It also follows from Proposition 3.1 that $\widehat{\sigma}_n^2(\Theta_n) \stackrel{\text{a.s.}}{\to} \Delta(\Theta_0)$.

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

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

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

$$\begin{aligned}
\Delta(\alpha_1) &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\frac{p_1\sigma^2_{(X^{(1)})^2}}{\left|1-\alpha_1^{(1)}\mathrm{e}^{-i\lambda}\right|^2} + \frac{p_2\sigma^2_{(X^{(2)})^2}}{\left|1-\alpha_1^{(2)}\mathrm{e}^{-i\lambda}\right|^2}\right)\left|1-\alpha_1\mathrm{e}^{-i\lambda}\right|^2\,d\lambda \\
&\quad + (1-\alpha_1)^2\,p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 \\
&= (1+\alpha_1^2)\left(p_1\sigma^2_{(X^{(1)})^2} + p_2\sigma^2_{(X^{(2)})^2}\right) - 2\alpha_1\left(\alpha_1^{(1)}p_1\sigma^2_{(X^{(1)})^2} + \alpha_1^{(2)}p_2\sigma^2_{(X^{(2)})^2}\right) \\
&\quad + (1-\alpha_1)^2\,p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2.
\end{aligned}$$

Minimization of $\Delta(\alpha_1)$ yields

$$\alpha_1 = 1 - \frac{(1-\alpha_1^{(1)})p_1\sigma^2_{(X^{(1)})^2} + (1-\alpha_1^{(2)})p_2\sigma^2_{(X^{(2)})^2}}{p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 + p_1\sigma^2_{(X^{(1)})^2} + p_2\sigma^2_{(X^{(2)})^2}}.$$

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 type of non-stationarity causes φ_1 to 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 Appendix A2 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

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

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

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

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

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

$$= \quad 1+\frac{1}{1+\beta_1}\frac{-c_1+c_2(1-\beta_1)}{c_1\dfrac{1}{1-\beta_1^2}+2c_2\dfrac{\beta_1}{1-\beta_1^2}+c_3\dfrac{1}{(1-\beta_1)^2}}.$$

The second term on the right-hand side is negative for $\Theta \in \mathcal{C}$. Moreover, if the difference $|\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}|$ increases, the value of φ_1 increases to 1. This explains, to some extent, the behavior of the estimates for α_1 and β_1 in real-life log-return data; see Section 4 for related empirical evidence. We restricted ourselves to Whittle estimation because this kind of pseudo-maximum likelihood procedure is theoretically well understood and easier to handle than Gaussian maximum-likelihood estimation. Its relation to quasi-maximum likelihood methods as used for fitting GARCH$(1,1)$ models in Section 4 is not clear. However, our findings for the Whittle case give an idea of what may be expected when other estimation procedures are used.

4 A study of non-stationary simulated data

This section deals with a set of simulated data and exemplifies the theoretical results of Sections 2, 3.1 and 3.2. Two realisations of GARCH$(1,1)$ processes with parameters

$$\alpha_0 = 0.13 \times 10^{-6}, \quad \alpha_1 = 0.11, \quad \beta_1 = 0.52\,, \tag{4.1}$$

$$\alpha_0 = 0.17 \times 10^{-6}, \quad \alpha_1 = 0.20, \quad \beta_1 = 0.65\,, \tag{4.2}$$

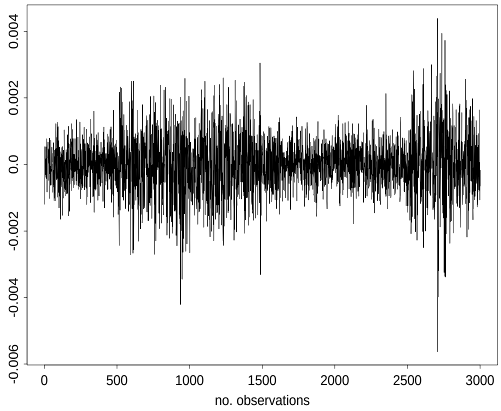
are considered. The innovations are standard normal. We consider a simulation $X_1, \ldots, X_{3000}$. Both subsamples $(X_t)_{t=1,\ldots,500}$ and $(X_t)_{t=1501,\ldots,2500}$ stem from the GARCH$(1,1)$ with parameters (4.1), and the remaining values come from model (4.2).

4.1 Goodness of fit tests

We are interested in checking the goodness of fit of the model (4.1) to various parts of the time series plotted in the left hand side graph of Figure 4.1. Towards this end the values of the goodness of fit test statistic

$$S_n = \sqrt{n} \sup_{\lambda \in [0,\pi]} |\widetilde{C}_{n,X}(\lambda)|\,, \tag{4.3}$$

defined in (2.10) are calculated over a window of width 250 that moves through the time series: $n = 250$ past values are used and the computation is performed every 5th instant of time. The statistic S_n is calculated under the hypothesis that the data come from a model with parameters (4.1). Notice that the values of $v_X(h)$ in S_n can be calculated from the GARCH$(1,1)$ parameters; see (A2.2) in the Appendix A2. The results are displayed in the right hand side graph of Figure 4.1. 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 (2.11); see Shorack and Wellner [57]. 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



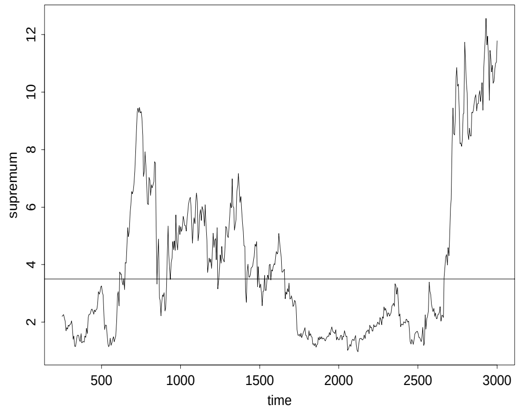


Figure 4.1 Left: 4 *pieces from* 2 *different* GARCH(1,1) *models as explained after* (4.2). Right: *The goodness of fit test statistic* S_{250} *evaluated at every* 5*th instant of time from the past* 250 *values. The horizontal line is set at the* 99% *quantile of the limit distribution* $\sup_{\lambda\in[0,\pi]}|B(\lambda)|$.

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 go through the threshold, staying above it most of the time until the switch back to the first model happens and the behavior is reversed. In the right-hand graph in Figure 4.1 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.

4.2 The sample ACF

In the sequel we illustrate the theoretical results of Section 3.1 by analysing the behavior of the sample ACF and of the periodogram of the simulated non-stationary time series. We will see how a a resemblance of LRD in the volatility can be caused by simple changes in the unconditional variance. Figure 4.2 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 (4.1) and (4.2), the 4th moment of X is finite. This follows for example from the theory developed in Mikosch and Stărică [48]. 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 (4.1) and $\kappa = 17.9$ for (4.2). 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

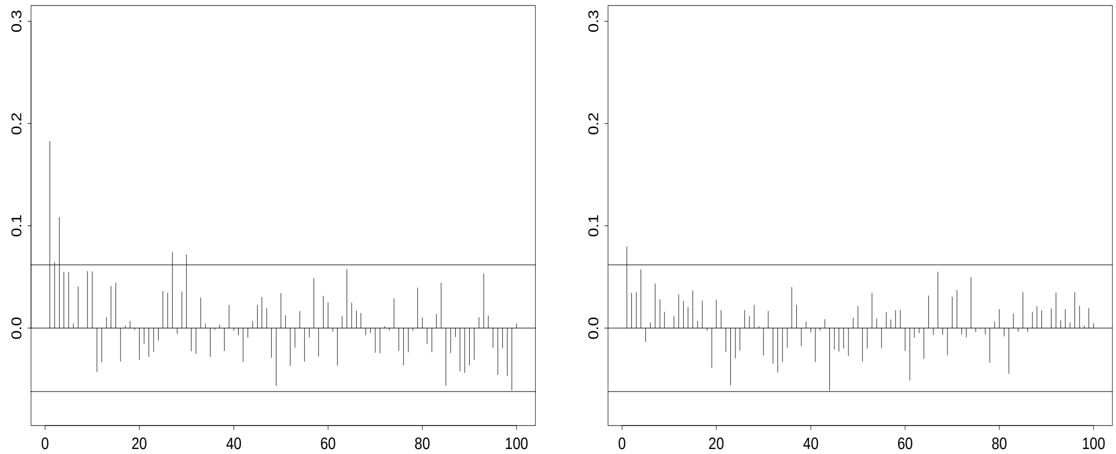


Figure 4.2 Left: *Sample ACF for* $|X_t|$, $t = 501, \ldots, 1500$. Right: *Sample ACF for* $|X_t|$, $t = 1501, \ldots, 2500$. *Here and in what follows, the horizontal lines are set as the* 95% *confidence bands* $(\pm 1.96/\sqrt{n})$ *corresponding to the ACF of iid noise with a finite* 2nd *moment.*

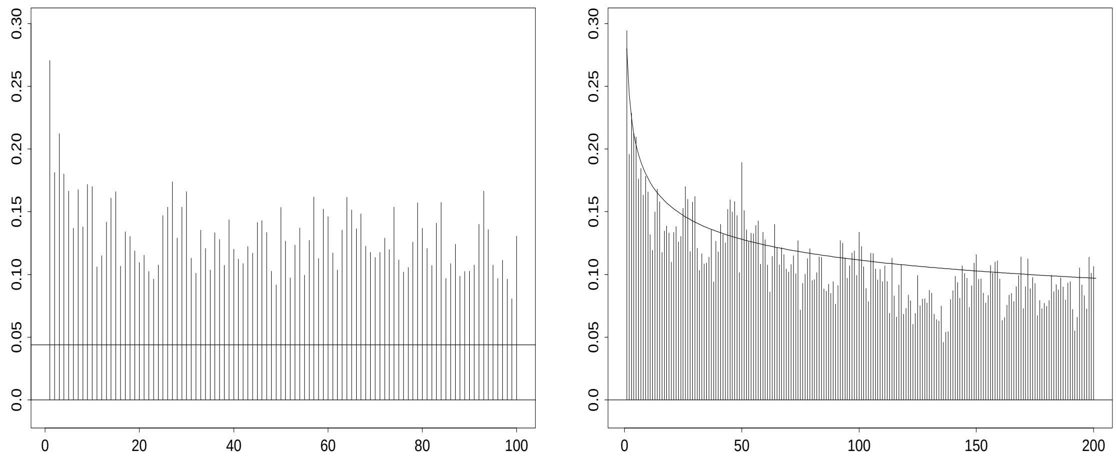


Figure 4.3 Left: *Sample ACF for* $|X_t|$, $t = 501, \ldots, 2500$. Right: *Sample ACF for* $|X_t|$, $t = 1, \ldots, 3000$ *with a fitted hyperbolic decay line.*

the asymptotic limits are normal; see Mikosch and Stărică [48].

Figure 4.3 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 behavior 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 4.3 shows a 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 4.4 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 4.1. The quantity H has been proposed in the literature as a measure of LRD; see Beran [8] 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.

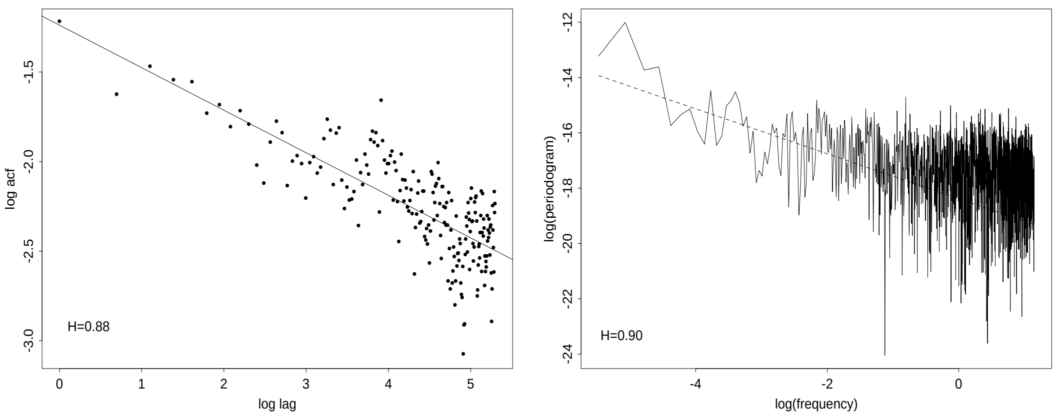


Figure 4.4 Left: *Log-log fit of the sample ACF for the absolute values of the simulated data displayed in Figure 4.1. Estimated Hurst coefficient* $H = 0.88$. Right: *Log-log fit of the periodogram for the same sample. Estimated* $H = 0.90$.

An estimation procedure for H is then suggested by the following argument. Since

$$\log \rho(h) \approx \log c - \beta \log h$$

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.

We apply this procedure to our simulated non-stationary time series. The left-hand graph in Figure 4.4 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 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. The actual fit of the hyperbolically decaying function to the sample ACF is illustrated in Figure 4.3 and is particularly noteworthy. A good fit of a hyperbolic decay line to the sample ACF is often used in the econometrics literature as an argument for the presence of LRD in the volatility process (see for example Anderson and Bollerslev [2]). Our example shows how misleading this approach could be. It says that a good hyperbolic fit might have nothing to do with the presence or absence of LRD in the volatility and it can be just a byproduct of changing unconditional variance.

4.3 The periodogram

It is a well known fact from trigonometric function theory that power law decay of the ACF translates into power law behavior of its Fourier transform (i.e. the spectral density) for small frequencies; see for example Zygmund [61]. (For a precise formulation of this statement, the ACF has to satisfy some subtle conditions; we refrain from discussing them.) Hence, frequently used statistical tests for detecting LRD and measuring its strength are based, one way or the other, on the following behavior of the spectral density:

$$f_{|X|}(\lambda) \approx c\lambda^{2H-1},$$

for small $\lambda > 0$. Equivalently, $\log f(\lambda)$ is linear in λ,

$$\log f_{|X|}(\lambda) \approx \log c + (2H - 1)\log\lambda.$$

Extrapolating this relationship to the sample, on a log-log plot, the periodogram should roughly exhibit a linear behavior, the slope of the line yielding an estimate of H. (For example, using a regression to capture the linear dependency yields the ubiquitous Geweke–Porter–Hudak estimator [32].)

We apply this procedure to our non-stationary simulated sequence. The right-hand graph in Figure 4.4 shows the plot of the log frequencies versus the log periodogram with a regression line fit to the first 10% of the frequencies (discarding the first $3000^{0.2} = 5$ lowest ones) yielding an estimate $H = 0.90$. Interestingly enough, both methods give similar values for the Hurst coefficients and suggest that there is a strong LRD effect in the data.

This concordance in the values yielded by the time and the frequency domain methods is sometimes used as an additional argument supporting the existence of LRD (see, for example

Anderson and Bollerslev [2]). Our example shows that this concordance might have nothing to say about the presence of LRD. It serves as a warning sign and shows that linear behavior in the log periodogram for low frequences can develop for other reasons than LRD.

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.*

4.4 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 [35]. GARCH$(1,1)$ models have been fitted to 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 4.5 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 3.2. There we explained that the Whittle 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 quasi-maximum likelihood estimators.

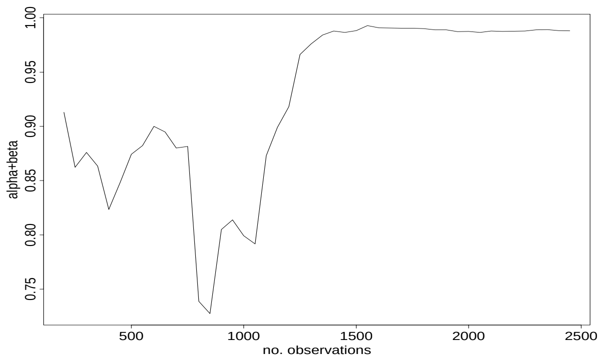


Figure 4.5 *The estimated values $\alpha_1 + \beta_1$ for a* GARCH$(1,1)$ *model fitted to an increasing sample from the simulated data in Figure 4.1. The labels on the x-axis indicate the size of the sample used in the quasi-maximum likelihood estimation.*

5 A study of the Standard & Poors 500 series

Now we proceed to analyse 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. [27], Granger et al. [36], Ding and Granger [26] for an analysis of its autocorrelation structure. It led the authors to the conclusion that the powers of the absolute values of the log-returns are 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. 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. [27] 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 [15] 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 behavior could be just an artifact due to structural changes in the data. We use the sup-statistic (4.3) to detect the moment in time when a GARCH$(1,1)$ model estimated on past data stops describing the behavior of the time series. Then we document the effect which 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 5.1. 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 [35]):

$$\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759\,, \tag{5.4}$$

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 (A2.2).

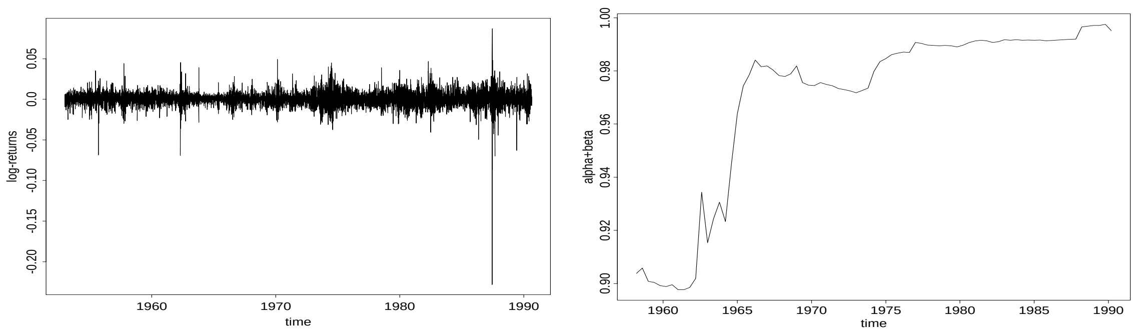


Figure 5.1 Left: *Plot of all* 9558 S&P500 *log-returns. The year marks indicate the beginning of the calendar year.* Right: *The estimated values* $\alpha_1 + \beta_1$ *for increasing samples of the* S&P500. *An initial* GARCH$(1,1)$ *model was estimated on* 1500 *observations, starting January 2, 1953. 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.*

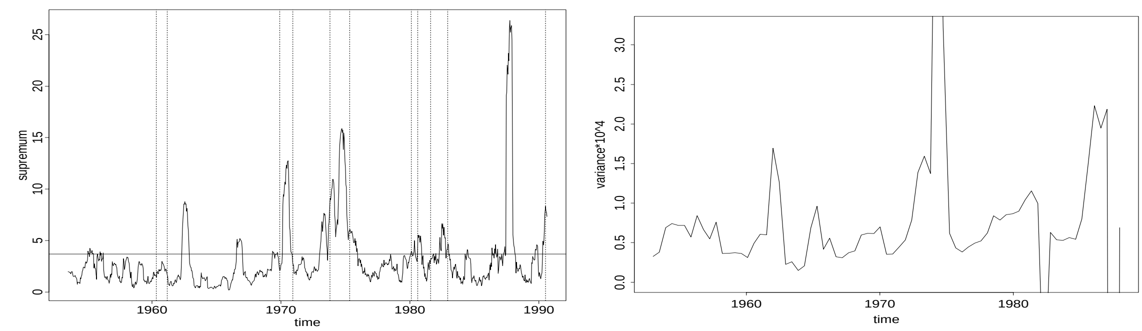


Figure 5.2 Left: *The goodness of fit test statistic* S_{125} *for the* S&P500 *data. The horizontal line is the* 99% *quantile of the limiting distribution of* S_{125}. *The dotted vertical vertical lines mark the start and the end of economic recessions as determined by the National Bureau of Economic Research.* 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*2.1).

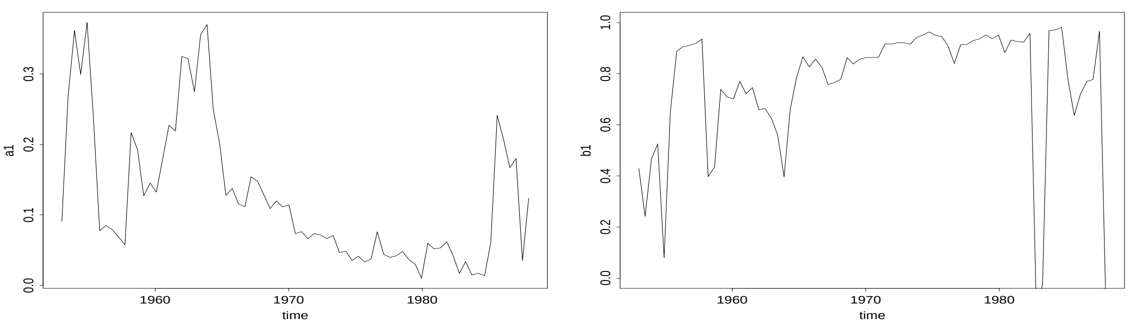


Figure 5.3 *A* GARCH$(1,1)$ *model is fitted to every block of* 6 *months data. The fit is based on the previous* 2 *years of data, i.e.* 508 *observations. Estimated* α_1 (left) *and* β_1 (right).

5.1 Goodness of fit tests

The left-hand graph in Figure 5.2 shows the results of calculating the statistic S_n (see (4.3)) 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. The dotted vertical lines mark the start and the end of economic recessions as determined by the National Bureau of Economic Research. The figure shows quite clearly that recessions are associated with larger volatilies and that the simple GARCH$(1,1)$ model cannot describe the complicated dynamics of longer, possibly non-stationary log-return time series.

A closer look at the graph of Figure 5.1 together with the left-hand graph in Figure 5.2 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 (A2.2) 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 (5.4) fails to explain the more extreme bursts of volatility that occur every once in a while.

It is then interesting to verify whether a periodically updated GARCH$(1,1)$ could account for the more pronounced volatility periods that are troublesome for the estimated GARCH$(1,1)$ model (5.4). 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 (A2.1).

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 log-returns. We 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 5.2. 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 5.3 displays the changes in the estimated coefficients α_1 and β_1. The parameters vary widely, beyond the limits of statistical uncertainty which 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 5.2. 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 5.2 together with the ones in Figure 5.3 strengthen the impression, given by the statistical analysis displayed in the left-hand graph of Figure 5.2, 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 behavior of our statistic around the Black Monday crash in October 1987. Following on the observation that a frequently re-estimated GARCH(1,1) model seems to follow better the changing patterns of volatility in the time series (see Figure 5.1), a GARCH(1,1) model is estimated in the beginning of June 1987 using the observations between January 1986 and June 1987 (375 observations). The estimated coefficients

$$\alpha_0 = 16 \times 10^{-6}, \quad \alpha_1 = 0.013, \quad \beta_1 = 0.812 \tag{5.5}$$

together with the 4th moment of the estimated residuals, $E\hat{Z}^4 = 3.4$ are used to build the S_{125} statistic. Figure 5.4 shows the behavior of the statistic during the 100 days preceding and the 25 following the crash. To allow for a better analysis, the log of the statistic is displayed. One sees

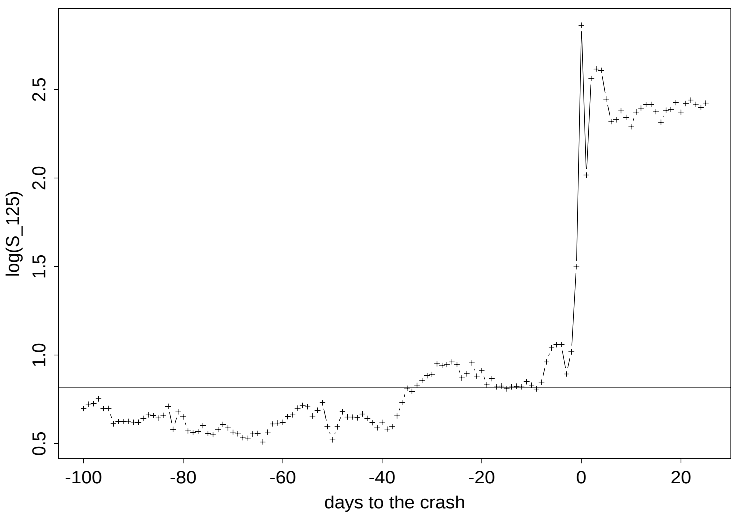


Figure 5.4 *The* $\log(S_{125})$ *statistic before and after the Black Monday 1987 crash. The horizontal line is the log of the* 75% *quantile* ($q_{0.75} = 2.26$) *of the limiting distribution of* S_{125}.

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 days -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 ($q_{0.99} = 3.6$).

5.2 Parameter estimation

The right graph in Figure 5.1 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 behavior of the S_{125} statistic and cannot be explained by the model (5.4)) 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 (5.4) does not fit the data (and which has such a strong impact on the behavior 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 3.2.

5.3 The sample ACF

Let us now analyze the impact which these periods of different structural behavior 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 5.1 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 5.5 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 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 3.1.

After this extreme episode (which can also be identified in the plot of the data in Figure 5.1) 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

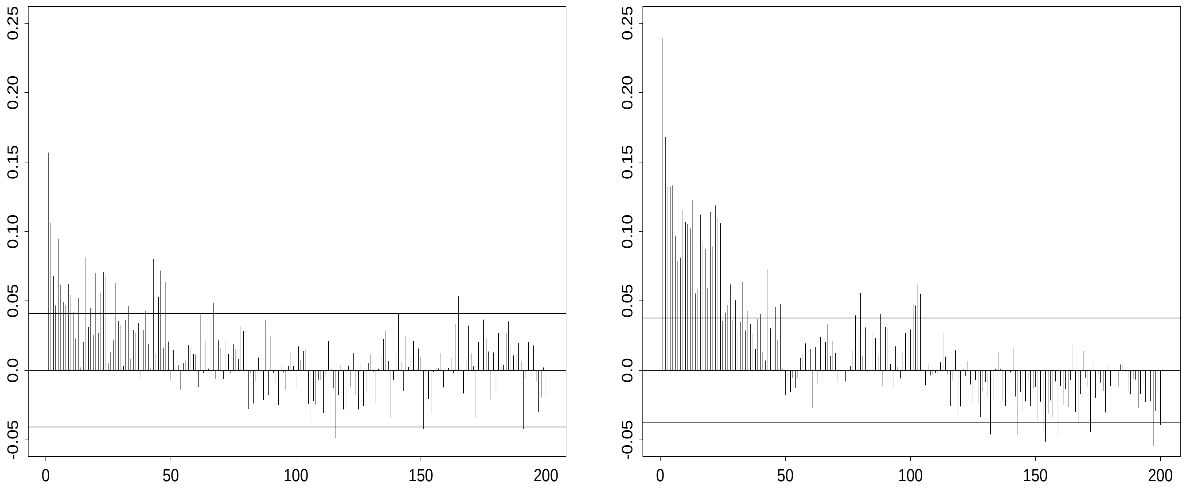


Figure 5.5 *The sample ACF for the absolute values of the log-returns for the first* 9 *years* (left) *and the first* 11 *years of the* S&P *data.*

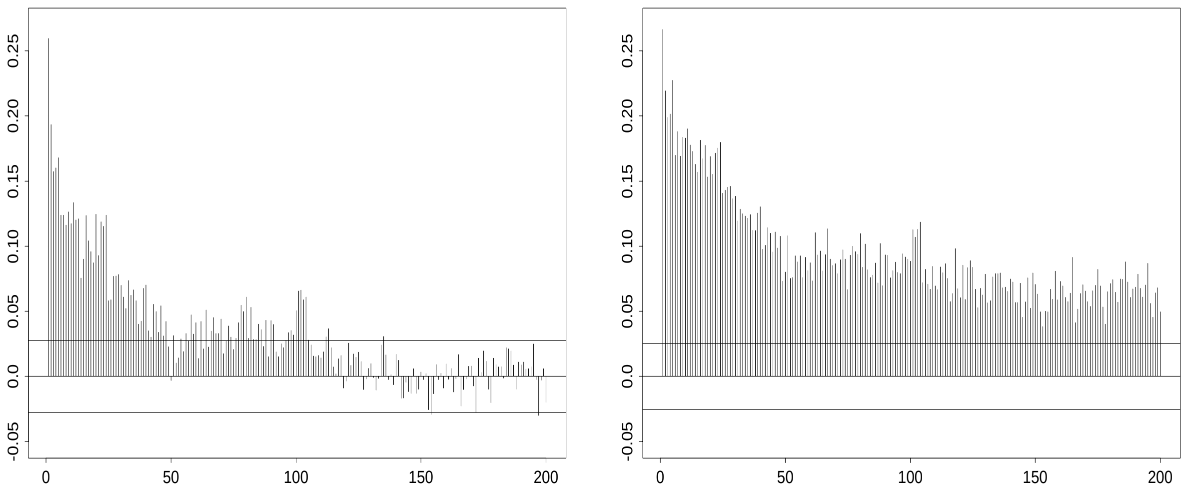


Figure 5.6 *The sample ACF for the absolute values of the log-returns of the first* 20 *years* (left) *and* 24 *years* (right) *of the* S&P *data.*

off-guard.

However, from the view point of the behavior of the sample ACF, the most interesting part of the top graph in Figure 5.1 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 (5.4) does not describe the data. Let us analyze the changes in the sample ACF caused by this long period of different behavior. Figure 5.6 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 5.5. 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 5.6. The graph clearly displays the LRD features given by the theoretical explanation of Section 3.1: exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations.

6 A study of exchange rates

In this section we focus on log-returns of foreign exchange (FX) data. We consider 2004 daily log-returns from the DEM/USD exchange rate between January 1975 and December 1982. We chose this period for a couple of reasons. This period is marked by two important events in the recent history of monetary policy. In Europe, several central banks improved upon the coordination of their monetary policy around March 1979. In the U.S., the Federal Reserve changed its instrument for controlling the money stock. It switched from using the federal funds rate to targeting the non-borrowed reserve. This change of policy lasted from *October 1979* until *October 1982* and, according to the economics literature (Spindt and Tarhan (1987), p. 107), it was "one of the more dramatic events in the recent history of monetary policy". Since monetary policy changes undoubtly have great impact on exchange rate dynamics, we expect to see the traces of these events in the data set.

Another reason is that data sets from this period which is marked by significant structural changes, have been included in studies conducted by various authors, both in the contexts of GARCH$(1,1)$ modeling and long memory analysis. Baillie and Bollerslev (1989) used daily spot exchange rates for the currencies of France, Italy, Japan, Switzerland, the United Kingdom and Germany against the USD between *March 1, 1980* and *Janury 28, 1985* to fit GARCH$(1,1)$ models with conditional t-distribution. They report estimated values of $\alpha_1 + \beta_1$ close to unity in all cases. Baillie et al. (1996) found evidence of long memory and fitted their FIGARCH model to a time series of daily DEM/USD spot exchange rates from *March 13, 1979* through *December 30, 1992*. Note that these periods contain (at least) the major event of the return of the Federal Reserve to the use of the federal funds rate to control the money stock.

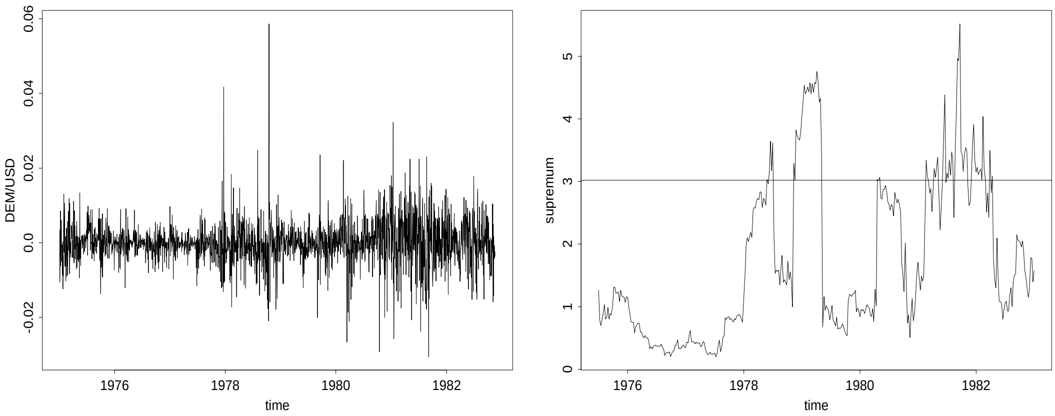


Figure 6.1 Left: *Plot of the* 2032 *DEM/USD log-returns. The year marks indicate the beginning of the calendar year.* Right: *The corresponding goodness of fit test statistics* S_{125}.

Our analysis shows that both events seem to have produced sensible changes in the dynamics of the exchange rates during this period. In the light of this analysis and of the theoretical developments of the paper, both findings of almost integrated GARCH$(1,1)$ and long memory in the mentioned studies are rendered questionable.

Further evidence for the fact that the mentioned structural shifts could be responsible for the spurious persistence in volatility documented in Baillie and Bollerslev (1989), comes from Cai (1994). This author finds that fitting a GARCH$(1,1)$ to the monthly excess returns of the three-month T-bill from *August 1964* to *November 1991* implies highly persistent volatility ($\hat{\phi}_1 = 0.98$). In contrast to that, fitting a model which allows for shifts in the unconditional variance yields significantly reduced ARCH parameters and, hence, a model with much less persistency in the volatility. The model clearly associates the periods of regime shifts with the oil shock and the Federal Reserve policy change.

Let us now commence our analysis. The left-hand graph of Figure 6.1 displays the data. Visible changes in the appearence of the data can be detected during 1978, begining of 1979 and end of 1980, 1981. A GARCH$(1,1)$ fit to the first 2 years yields the following parameters:

$$\alpha_0 = 8.3 \times 10^{-6}\,, \quad \alpha_1 = 0.18\,, \quad \beta_1 = 0.77\,, \tag{6.6}$$

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

6.1 Goodness of fit test and the sample ACF

The right-hand graph of Figure 6.1 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 asymptotic 95% 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 the beginning of 1979, and a longer period of about one year and a half, covering 1981 and the beginning of 1982. A look at the sample ACF of the absolute values $|X_t|$ before and after the period in the discussion reveals a minor change in the behavior of this statistical instrument; see Figure 6.2. Here the first 100 lags of the sample ACF before and after the first episode are displayed.

In contrast to the first episode that could possibly be associated with the mentioned increase in the policy coordination among several European countries, the second period during which the goodness of fit test statistic exceeds the threshold lasts longer. A visual inspection of the graphs in Figure 6.1 gives the impression that the structure of the time series in the period between spring 1980 and the end of 1981 is different from the remaining observations. The statistic S_{125} confirms this fact by frequently switching sides of the threshold line during this period. The dramatic changes in the behavior of the sample ACF are illustrated in Figure 6.3 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.

A1 Appendix

Proof of Lemma 2.1 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,

$$\sqrt{n}(\gamma_{n,[nx],X}(h))_{x\in[0,1]} \xrightarrow{d} (v_X^{1/2}(h)\, W_h(x))_{x\in[0,1]}\,. \tag{A1.1}$$

in $\mathbb{D}[0,1]$; see Oodaira and Yoshihara [50]; cf. Doukhan [28], 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 (2.6)) and that the sequence (X_tX_{t+h}) is α-mixing with a sufficiently fast rate for the mixing coefficients; see (A1.2). However, the GARCH(p,q) is strongly mixing with geometric rate (see for example Davis et al. [22]), 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 [55], one obtains that the map from $(\mathbb{D}[0,1])^m$ into $\mathbb{D}([0,1],\mathbb{R}^m)$ defined by

$$(x_1,\ldots,x_m)\to(x_1(t),\ldots,x_m(t))_{t\ge0}$$

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 (2.7) are tight in $\mathbb{D}([0,1],\mathbb{R}^m)$.

Notice that the multivariate CLT

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

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 A1.1 It follows from the argument in the proof of Lemma 2.1 that (2.7) 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

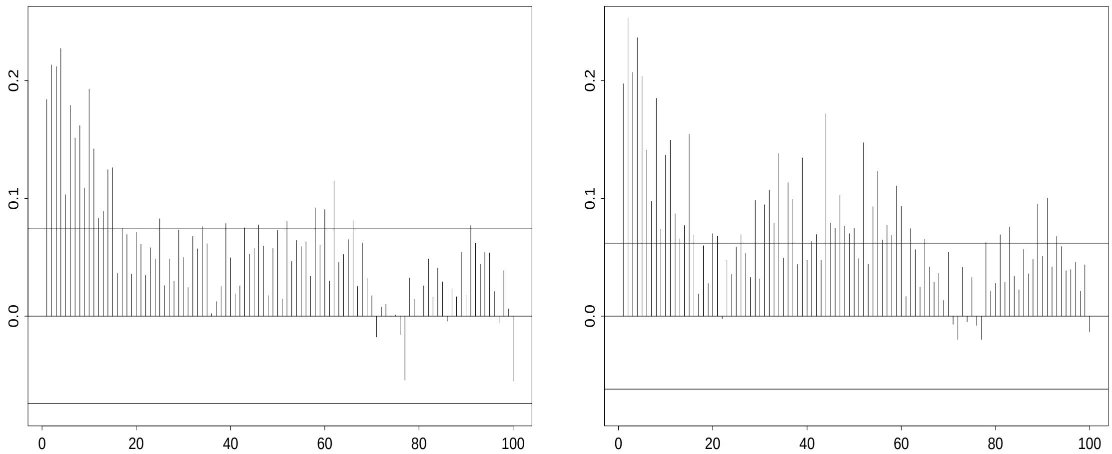


Figure 6.2 *The sample ACF for the absolute values of the log-returns of the DEM/USD FX data up to the beginning of* 1978 (left) *and up to June* 1979 (right).

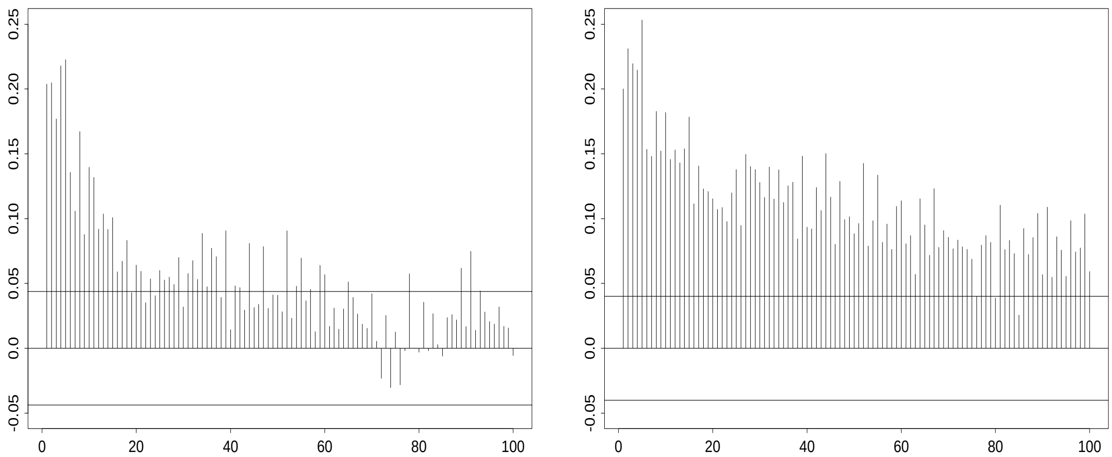


Figure 6.3 *The sample ACF for the absolute values of the log-returns of the DEM/USD FX data up to the beginning of* 1981 (left) *and up to June* 1982 (right).

$h \geq 1$, $\mathrm{cov}(X_0 X_h, X_0 X_l) = 0$ for all $h \neq l \geq 1$, and with α-mixing coefficients $\widetilde{\alpha}_i$ satisfying

$$\sum_{i=1}^{\infty} \widetilde{\alpha}_i^{\delta/(2+\delta)} < \infty \,. \tag{A1.2}$$

The latter conditions are needed for the validity of the FCLT in (A1.1); see Oodaira and Yoshihara [50].

Proof of Theorem 2.2. We proceed analogously to Klüppelberg and Mikosch [43]. 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 [11], it remains to show that for every $\epsilon > 0$,

$$\lim_{m \to \infty} \limsup_{n \to \infty} P\left(\sup_{0 \leq x \leq 1} \sup_{0 \leq \lambda \leq \pi} \left| \sum_{h=m+1}^{[nx]} \sqrt{n}\, \widetilde{\gamma}_{n,[nx],X}(h) \, \frac{\sin(\lambda h)}{h} \right| > \epsilon \right) = 0 \,. \tag{A1.3}$$

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 [23], 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 [42], pp. 1873–1876. Instead of Theorem 3.1 in Rosiński and Woyczyński [56] one can simply use the Cauchy–Schwarz inequality in the first display on p. 1876 in [42] with $\mu = 2$. Then all the calculations for (6.3) remain valid, implying that (A1.3) holds. This concludes the proof of the theorem. $\square$

Remark A1.2 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 [43], 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 [37], Chapter 6.4.

Proof of Proposition 3.1. 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 [18] 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 [18], we restrict ourselves to show that (3.11) is satisfied. The same arguments as on pp. 378–379 in [18] 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 [18], 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

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

Then the same arguments as for (10.8.11) in [18] 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

$$\begin{aligned} &\left| \sum_{|k|<m}\Gamma_{X^2}(h)\left(1 - \frac{|k|}{m}\right) b_k \right. \\ &\left. -\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{p_1\sigma^2_{(X^{(1)})^2}f(\lambda,\Theta_1) + p_2\sigma^2_{(X^{(2)})^2}f(\lambda,\Theta_2)}{f_\delta(\lambda,\Theta)}\,d\lambda - \frac{p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2}{f_\delta(0,\Theta)} \right| \\ \leq\ & \mathit{const}\ \epsilon \end{aligned}$$

for every $\epsilon > 0$, uniformly in $\Theta \in \overline{\mathcal{C}}$. The same arguments as in [18] conclude the proof. □

Proof of Theorem 3.3. One can follow the arguments on p. 385 of [18]. 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 3.1, for any rational $\delta > 0$,

$$\begin{aligned} \liminf_{k\to\infty}\widehat{\sigma}^2_{n_k}(\Theta_{n_k}) &\geq \liminf_{k\to\infty}\frac{1}{n_k}\sum_j \frac{I_{n_k,X^2-\overline{X^2_n}}(\lambda_j)}{f_\delta(\lambda_j,\Theta_{n_k})} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{p_1\sigma^2_{(X^{(1)})^2}f(\lambda,\Theta^{(1)}) + p_2\sigma^2_{(X^{(2)})^2}f(\lambda,\Theta^{(2)})}{f_\delta(\lambda,\Theta)}\,d\lambda + \frac{p_1p_2\left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2}{f_\delta(0,\Theta)} . \end{aligned}$$

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

$$\liminf_{k\to\infty}\widehat{\sigma}^2_{n_k}(\Theta_{n_k}) \geq \Delta(\Theta) > \Delta(\Theta_0) . \tag{A1.4}$$

On the other hand, by definition of Θ_n as a minimizer, (3.10) 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 (A1.4). This concludes the proof. □

A2 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 (2.1):

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

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 [19]:

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

Straightforward calculation yields

$$\begin{aligned}
\sigma_\nu^2 &= (EZ^4 - 1) \, E\sigma_1^4 = \frac{1+\varphi_1}{1-\varphi_1} \frac{\alpha_0^2 (EZ^4 - 1)}{1 - (\varphi_1^2 + \alpha_1^2 (EZ^4 - 1))} \, , \\
\text{(A2.1)} \qquad \sigma_X^2 &= \frac{\alpha_0}{1 - \varphi_1} \, .
\end{aligned}$$

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:

$$\text{(A2.2)} \qquad v_X(h) = \sigma_X^4 \left(\frac{(EZ^4 - 1)\alpha_1 \left(1 - \varphi_1^2 + \alpha_1 \varphi_1\right)}{1 - (\varphi_1^2 + \alpha_1^2 (EZ^4 - 1))} \, \varphi_1^{h-1} + 1 \right) , \quad h \geq 1 \, .$$

Acknowledgment: Cătălin Stărică would like to thank the Department of Mathematics of the University of Groningen and the Dutch Science Foundation (NWO) for financial support. Thomas Mikosch worked on this paper in January 1999 during a visit to the University of Aarhus. He would like to thank his colleagues at the Institute of Statistics, in particular Ole Barndorff-Nielsen, for their warm hospitality and generous support.

References

[1] ANDERSON, C.W. AND TURKMAN, K.F. (1995) Sums and maxima of stationary sequences with heavy tailed distributions. *Sankhya* **57** , 1–10.

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

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

[4] BAILLIE, R.T. AND BOLLERSLEV, T. (1989) The message in daily exchange rates: a conditional-variance tale. *J. Business and Economic Statist.* **7**, 297–305.

[5] BAILLIE, R.T. AND BOLLERSLEV, T. (1990) Intra-day and inter-market volatility in foreign exchange rates. *Review of Economic Studies* **58**, 565–585

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

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

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

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

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

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

[12] BOES, D. C. AND SALAS-LA CRUZ, J. D. (1978) Non stationarity of the mean and the Hurst phenomenon. *Water Resour. Res.* **14**, 135–143.

[13] BOLLERSLEV, T. (1986) Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* **31**, 307–327.

[14] BOLLERSLEV, T., CHOU, R.Y. AND KRONER, K.F. (1992) ARCH modeling in finance: a review of the theory and empirical evidence. *J. Econometrics* **52**, 5–59.

[15] BOLLERSLEV, T. AND MIKKELSEN, H.O. (1996) Modeling and pricing long memory in stock market volatility. *J. Econometrics* **73**, 151–184.

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

[17] BREIDT, F.J., CRATO, N. AND DE LIMA, P. (1996) The detection and estimation of long memory in stochastic volatility. *J. Econometrics* **83**, 325–348.

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

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

[20] Cai, J. (1994) A Markov model of unconditional variance in ARCH. *J. Business and Economic Statist.* **12**, 309–316.

[21] Davis, R.A. and Mikosch, T. (1999) The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049–2080.

[22] Davis, R.A., Mikosch, T. and Basrak, B. (1998) Sample ACF of multivariate stochastic recurrence equations with applications to GARCH. Technical Report, University of Groningen.

[23] 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.

[24] de Lima, P. and Crato, N. (1994) Long range dependence in the conditional variance of stock returns. *Economic Letters* **45**, 281–285.

[25] Diebold, F.X. (1986) Modeling the persistence of the conditional variances: a comment. *Econometric Reviews* **5**, 51–56.

[26] Ding, Z. and Granger, C.W.J. (1996) Modeling volatility persistence of speculative returns: A new approach. *J. Econometrics* **73**, 185–215.

[27] 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.

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

[29] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events for Insurance and Finance.* Springer, Berlin.

[30] Engle, R.F. (ed.) (1995) *ARCH Selected Readings.* Oxford University Press, Oxford (U.K.).

[31] 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.

[32] Geweke, J. and Porter–Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Series Analysis* **4**, 221–238.

[33] 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.

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

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

[36] Granger, C.W.J. and Ding, Z. (1996) Varieties of long memory models. *J. Econometrics* **73**, 61–77.

[37] Grenander, U. and Rosenblatt, M. (1984) *Statistical Analysis of Stationary Time Series,* 2nd edition. Chelsea Publishing Co., New York.

[38] Hamilton, J. and Susmel, R. (1994) Autoregresive conditional heteroskedasticity and changes in regime. *J. Econometrics* **64**, 307–333.

[39] Hida, T. (1980) *Brownian Motion.* Springer, New York.

[40] Jacod, J. and Shiryaev, A.N. (1987) *Limit Theorems for Stochastic Processes.* Springer, Berlin, New York.

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

[42] Klüppelberg, C. and Mikosch, T. (1996) The integrated periodogram for stable processes. *Ann. Statist.* **24**, 1855–1879.

[43] Klüppelberg, C. and Mikosch, T. (1996) Gaussian limit fields for the integrated periodogram. *Ann. Appl. Probab.* **6**, 969–991.

[44] Kokoszka, P. and Mikosch, T. (1997) The integrated periodogram for long–memory processes with finite or infinite variance. *Stoch. Proc. Appl.* **66**, 55–78.

[45] Lamoureux, C.G. and Lastrapes, W.D. (1990) Persistence in variance, structural change and the GARCH model. *J. Business and Economic Statist.* **8**, 225–234.

[46] Lo, A. (1991) Long memory in stock market prices. *Econometrica* **59**, 1279–1313.

[47] 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.

[48] 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.

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

[50] 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.

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

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

[53] Potter, K. (1976) Evidence for non stationarity as physical explanation of the Hurst phenomenon. *Water Resour. Res.* **12**, 1047–1052.

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

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

[56] Rosiński, J. and Woyczyński, W.A. (1987) Multilinear forms in Pareto–like random variables and product random measures. *Coll. Math.* **51**, 303–313.

[57] Shorack, G.R. and Wellner, J.A. (1986) *Empirical Processes with Applications to Statistics.* Wiley, New York.

[58] Spindt, P.A. and Tarhan, V. (1987) The Federal Reserve's new operating procedure: a postmortem. *J. Monetary Econom.* **19**, 107–123.

[59] Teverovsky, V. and Taqqu, M. (1997) Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *J. Time Ser. Anal.* **18**, 279–304.

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

[61] 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
mikosch@math.rug.nl

CĂTĂLIN STĂRICĂ
The Wharton School, Philadelphia, and
Department of Statistics
Chalmers University of Technology
S–412 96 Gothenburg
SWEDEN
starica@math.chalmers.se

List of CAF's Working Papers

1. O. E. Barndorff-Nielsen (November 1997), *Processes of Normal Inverse Gaussian Type*.

2. P. Honoré (November 1997), *Modelling Interest Rate Dynamics in a Corridor with Jump Processes*.

3. G. Peskir (November 1997), *The Concept of Risk in the Theory of Option Pricing*.

4. A. T. Hansen and P. L. Jørgensen (November 1997), *Analytical Valuation of American-style Asian Options*.

5. T. H. Rydberg (December 1997), *Why Financial Data are Interesting to Statisticians*.

6. G. Peskir (December 1997), *Designing Options Given the Risk: The Optimal Skorokhod-Embedding Problem*.

7. J. L. Jensen and J. Pedersen (December 1997), *A note on models for stock prices*.

8. M. Bladt and T. H. Rydberg (December 1997), *An actuarial approach to option pricing under the physical measure and without market assumptions*.

9. J. Aa. Nielsen and K. Sandmann (April 1998), *Asian exchange rate options under stochastic interest rates: pricing as a sum of delayed payment options.*

10. O. E. Barndorff-Nielsen and N. Shephard (May 1998), *Aggregation and model construction for volatility models.*

11. M. Sørensen (August 1998), *On asymptotics of estimating functions.*

12. A. T. Hansen and P. L. Jørgensen (August 1998), *Exact Analytical Valuation of Bonds when Spot Interest Rates are Log-Normal.*

13. T. Björk, B. J. Christensen and A. Gombani (September 1998), *Some Control Theoretic Aspects of Interest Rate Theory*. Published as: *Some System Theoretic Aspects of Interest Rate Theory* in: *Insurance, Mathematics & Economics* 1998, Vol. 22, pp. 17-23

14. B. J. Christensen and N. R. Prabhala (September 1998), *The Relation Between Implied and Realized Volatility*. Published in: *Journal of Financial Economics* 1998, Vol. 50, pp. 125-150.

15. O.E. Barndorff-Nielsen and W. Jiang (August 1998), *An initial analysis of some German stock price series.*

16. C. Vorm Christensen and H. Schmidli (September 1998), *Pricing catastrophe insurance products based on actually reported claims*.

17. M. Bibby and M. Sørensen (September 1998), *Simplified Estimating Functions for Diffusion Models with a High-dimensional Parameter*

18. P. Honoré (November 1998), *Pitfalls in Estimating Jump-Diffusion Models*

19. P. Honoré (November 1998), *Panel-Data Estimation of Non-Linear Term-Structure Models.*

20. T. Engsted and K. Nyholm (November 1998), *Regime shifts in the Danish term structure of interest rates.*

21. T. Hviid Rydberg and N. Shephard (December 1998), *Dynamics of trade- by-trade price movements: decomposition and models*.

22. E. Hansen and A. Rahbek (December 1998), *Stationarity and asymptotics of multivariate ARCH time series with an application to robustness of cointegration analysis.*

23. C. Strunk Hansen (January 1999), *The relationship between implied and realized volatility in the Danish option and equity markets*.

24. D. Duffie and D. Lando (January 1999), *Term structures of credit spreads with incomplete accounting information.*

25. B. Huge and D. Lando (January 1999), *Swap pricing with two-sided default risk in a rating-based model.*.

26. H. Sørensen (March 1999), *Approximation of the score function for diffusion processes.*

27. T. Björk and B.J. Christensen (March 1999), *Interest rate dynamics and consistent forward rate curves.*

28. C. Vorm Christensen (March 1999), *A new model for pricing catastrophe insurance derivatives.*

29. R. Poulsen (March 1999), *Approximate maximum likelihood estimation of discretely observed diffusion processes.*

30. A. Trolle Hansen and R. Poulsen (March 1999), *A simple regime switching term structure model.*

31. B.J. Christensen and N.M. Kiefer (March 1999), *Simulated moment methods for empirical equivalent martingale measures.*

32. A. Grosen and P. Løchte Jørgensen (April 1999), *Fair valuation of life insurance liabilities: The impact of interest rate guarantees, surrender options, and bonus policies.*

33. M. Sørensen (April 1999), *Prediction-based estimating functions.*

34. Lars Korsholm (April 1999), *The GMM estimator versus the semiparametric efficient score estimator under conditional moment restrictions.*

35. K. Nyholm (May 1999), *Inferring the Private Information Content of Trades: A Regime-Switching Approach.*

36. K. Nyholm (May 1999), *Analyzing Specialist's Quoting Behavior: A Trade-By-Trade Study on the NYSE.*

37. O.E. Barndorff-Nielsen and N. Shephard (May 1999), *Non-Gaussian OU based models and some of their uses in financial economics.*

38. B. Jensen and R. Poulsen (August 1999), *A comparison of approximation techniques for transition densities of diffusion processes.*

39. G. Peskir and J. Shorish (August 1999), *Market forces and Dynamic Asset Pricing*

40. T. Engsted and C. Tanggaard (August 1999), *The Danish stock and bond markets: Comovement, return predictability and variance decomposition.*

41. B.Jensen (August 1999), *Pricing in incomplete markets by fuzzy ranking.*

42. B. Jensen (August 1999), *Option pricing in the jump-diffusion model with a random jump amplitude: A complete market approach.*

43. C. Christiansen (September 1999), *Macroeconomic announcement effects on the covariance structure of bond returns.*

44. B.J. Christensen and N.M. Kiefer (November 1999), *Panel Data, Local Cuts and Orthogeodesic Models*.

45. O.E. Barndorff-Nielsen and K. Prause (November 1999), *Apparent Scaling*.

46. O. Linton, E. Mammen, J. Nielsen and C. Tanggaard (November 1999), *Yield Curve Estimation by Kernel Smoothing Methods.*

47. M. Kessler and A. Rahbek (November 1999), *Asymptotic likelihood based inference for cointegrated homogenous Gaussian diffusions.*

48. N. Væver Hartvig, J. Ledet Jensen and J. Pedersen (November 1999), *A class of risk neutral densities with heavy tails.*

49. H. Bunzel, B.J. Christensen, P. Jensen, N.M. Kiefer, L. Korsholm, L. Muus, G.R. Neumann, M. Rosholm (December 1999), *Investment in human capital versus differences in company productivity levels: Specification and estimation of equilibrium search models for Denmark.*

50. B.J. Christensen, P. Jensen, M. Svarer Nielsen, K. Poulsen, M. Rosholm (December 1999), *Public finance effects in an equilibrium search model with differences in company productivity levels: An application to Danish data.*

51. H. Bunzel, B.J. Christensen, N.M. Kiefer, L. Korsholm (December 1999), *The asset pricing approach to the rate of return to human capital: An equilibrium search framework for Denmark.*

52. M. Berg Jensen (January 2000), *Efficient method of moments estimation of the Longstaff and Schwartz interest rate model.*

53. C. Christiansen and C. Strunk Hansen (January 2000), *Implied volatility of interest rate options: An empirical investigation of the market model.*

54. N.M. Kiefer and T.J. Vogelsang (February 2000), *A new approach to the asymptotics of HAC robust testing in econometrics.*

55. B. Jensen, P. Løchte Jørgensen, A. Grosen (February 2000), *A finite difference approach to the valuation of path dependent life insurance liabilities.*

56. J. Lund Pedersen (March 2000), *Discounted optimal stopping problems for the maximum process.*

57. M. Sørensen (March 2000), *Small dispersion asymptotics for diffusion martingale estimating functions.*

58. T. Mikosch and C. Starica (April 2000), *Change of structure in financial time series, long range dependence and the GARCH model.*

59. T. Engsted (April 2000), *Measuring noise in the permanent income hypothesis.*

60. T. Engsted, E. Mammen and C. Tanggaard (April 2000), *Evaluating the C-CAPM and the equity premium puzzle at short and long horizons: A Markovian bootstrap approach.*

61. T. Engsted and C. Tanggaard (April 2000), *The relation between asset returns and inflation at short and long horizons.*

ISSN 1398-6163

Mailing address: University of Aarhus
Department of Economics
Building 350
DK - 8000 Aarhus C

Telephone: +45 8942 1580
Fax: +45 8613 6334
E-Mail: CAF@cls.dk
http://www.caf.dk