Long range dependence effects and ARCH modelling

Thomas Mikosch
University of Groningen

Cătălin Stărică
Chalmers University

Draft January 26, 2000

Short Title: Long range dependence and ARCH

Abstract

AMS 1991 Subject Classification: Primary: Secondary:
Key Words and Phrases.

1 Some preliminaries on ARCH and GARCH processes

Among the models for log-returns

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

of stock indices, foreign exchange rates, share prices, etc., P_t, $t = 0, 1, \dots$, the ARCH (*autoregressive conditionally heteroscedastic*) processes have gained particular popularity. Besides the *stochastic volatility models* (see for example Ghysels et al. [19] for a recent survey paper) they have become *the* standard models in the financial econometrics literature. In particular, they appear in many recent textbooks and monographs on time series analysis (see for example Brockwell and Davis [10] or Embrechts et al. [17]) or econometrics (see Campbell et al. [11]). Thus many students of statistics and virtually all students of econometrics have heard of them. The GARCH modules of various software packages have certainly contributed to the increasing popularity of this kind of econometric time series models as well.

The success story of the ARCH family started in 1982 when Engle introduced the ARCH(p) processes (ARCH of order p)

$$X_t = \sigma_t\, Z_t\,, \tag{1.1}$$

where σ_t (the so-called *stochastic volatility*) obeys the recurrence equation

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \cdots + \alpha_p X_{t-p}^2\,, \tag{1.2}$$

with α_i's non-negative parameters, while (Z_t) is a white noise process with variance 1, usually supposed to be iid. (In what follows we always assume (Z_t) to be iid.) This implies that, conditionally upon $X_{t-1}, \dots, X_{t-p}$ (the p past observations of the time series), X_t has variance σ_t^2.

The basic idea behind the construction (1.1) is quite intuitive: for a "forecast" of the distribution of X_t we only have to know two ingredients: σ_t and the distribution of Z_t. For example, if Z_t is normal $N(0,1)$, then, given the past observations of the time series, $X_t \sim N(0, \sigma_t^2)$. Hence, conditionally upon $X_{t-1}, \dots, X_{t-p}$, the present value X_t may assume values in $[-1.96\sigma_t, 1.96\sigma_t]$ with 95% probability. Similarly, there is a 5% chance for the log-return X_t to fall below the threshold $-1.64\sigma_t$. The 5%–quantile of the log-return distribution is considered as a measure of risk for the underlying asset. In the financial area, this quantile is known under the name of *Value at Risk* or *VaR*; see RiskMetrics [31].

These simple calculations show why models of type (1.1) have become so popular; in the presence of non-Gaussian distributions for log-returns (this is a fact no specialist would doubt!) mixture models such as (1.1) allow one to get updated (i.e. conditional) probability "forecasts" without too much sophistication.

Empirical work has shown that the simple ARCH(p) process given by (1.1) and (1.2) has a reasonable fit to real-life data *only if the number of the parameters* α_i *is rather large.* Since the rationale for the definition (1.2) is to take a time-changing weighted average of the past squared observations as an approximation to the conditional variance σ_t^2 (an "updated estimate of the variance", if you like), it is quite natural to define σ_t^2 not only as a weighted average of past X_j^2's but also of past σ_j^2's. This new idea resulted in Bollerslev's [2] and Taylor's [35] *generalised ARCH process of order* (p, q) (GARCH(p,q)): the process (X_t) is again given by (1.1), but now the squared stochastic volatility satisfies

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \, X_{t-j}^2 + \sum_{k=1}^{q} \beta_k \, \sigma_{t-k}^2 := \alpha_0 + \alpha(L) X_t^2 + \beta(L) \sigma_t^2, \tag{1.3}$$

where the α_j's and β_k's are non-negative parameters, L is the back-shift operator and $\alpha(L)$, $\beta(L)$ are the corresponding polynomials in L with coefficients α_j, β_k. Clearly, σ_t^2 could have been defined in many other reasonable ways, and therefore it is perhaps not totally surprising that a wave of different ARCH-type models has flooded the econometrics journals. Each of these models was introduced in order to improve upon (1.2) or (1.3) in some sense. Some of them have gained popularity such as Nelson's [30] EGARCH (exponential GARCH) model while most of them remained only of academic interest. From a mathematical point of view, not all of them are directly comparable with the GARCH processes. However, we do not have here the space to discuss these modifications in detail; see for example Bollerslev et al. [3] or Shephard [32] for review papers. In what follows, we mostly stick to the GARCH model and we do so for two reasons. The first is that, although apparently simple, its analytic study exemplifies the difficulties encountered when working with this class of models. Secondly, it is one of the models heavily used in practice.

Notice that we call (X_t) an ARCH or GARCH process and not the squared process (X_t^2). (Since both conventions have been used in the literature we want to make clear our preference.)

The connection with ARMA processes. It is straightforward that equation (1.3) can be rewritten as an ARMA equation with noise $\nu_t := X_t^2 - \sigma_t^2$:

$$(1 - \varphi(L)) X_t^2 = \alpha_0 + (1 - \beta(L)) \nu_t, \quad \varphi(L) := \alpha(L) + \beta(L) \, . \tag{1.4}$$

If (X_t) is strictly stationary and $EX^2 < \infty$, (ν_t) constitutes a strictly stationary martingale difference sequence. In the light of (1.4), the idea of viewing the GARCH process as an ARMA process for the squares X_t^2 was certainly a father of thought.

Representation (1.4) could give one the illusion that the theory for GARCH processes might be as easy as for ARMA processes. (Naturally, the notion of "simplicity" is a relative one; a thorough study of linear processes as provided by Brockwell and Davis [9] shows that simplicity can also have a high dimension of complexity.) However, (1.1) makes (X_t) a non-linear process. For this deviation from a linear (ARMA) process one has to pay a price. After 15 years of ARCH modelling we know it is a high one: we know very little about the theoretical (probabilistic) properties of ARCH and GARCH processes. The "pure theorist" and the "practical econometrician" will certainly deny this statement. However, just to give some examples of the difficulties one has to face: with a few exceptions (the ARCH(1) and GARCH(1,1) models) we can in general not decide whether a GARCH(p,q) process has a strictly stationary version, provided we know the distribution of the Z_t's and the parameters, we do not know much about the tails of the marginal distributions of (X_t), and very little about the finite-dimensional distributions, i.e. the dependence structure. We know almost nothing about the theoretical properties of multivariate extensions of ARCH-type models.

Surprisingly, the statistical estimation of the parameters α_j and β_k is not too difficult. (This fact is an essential argument in favour of GARCH processes.) Given the Z_t's are iid standard normal, the likelihood function of the vector $(X_1,\ldots,X_n)$ can be written down, and conditional maximum likelihood theory yields consistent and asymptotically normal estimates of the parameters. This theory remains valid even if one deviates from the Gaussianity of the Z_t's. The estimation theory for GARCH processes is provided for example in Gourieroux [23].

The stationarity issue. The GARCH(p,q) equations (1.1), (1.3) with iid innovations Z_t such that $EZ=0$ and $EZ^2=1$ have a strictly stationary finite first moment solution (σ_t^2) (and hence (X_t) is strictly stationary as well and has finite variance) if

$$\alpha_0>0 \quad \text{and} \quad \sum_{j=1}^{p}\alpha_j+\sum_{k=1}^{q}\beta_k<1\,. \tag{1.5}$$

See Nelson [29] for the GARCH(1,1) case and Bougerol and Picard [7] for the general case.

At this point it is worth to mention a possible source of confusion generated through borrowing terminology from ARMA models. In analogy with ARMA and integrated ARMA (ARIMA) processes, Engle and Bollerslev [18] coined the name *integrated* GARCH(p,q) (IGARCH(p,q)) process for the situation when

$$\alpha_0>0 \quad \text{and} \quad \sum_{j=1}^{p}\alpha_j+\sum_{k=1}^{q}\beta_k=1\,. \tag{1.6}$$

Although this name seems to be quite intuitive in the ARMA modeling framework, it is misleading from an ARCH perspective. Indeed, the name *integrated* suggests that there is a unit root problem, as for integrated ARMA processes, concerning the stationarity of such GARCH processes. However, this is not the case for the GARCH (1.1), (1.3) model. Bougerol and Picard [7] prove that, if all α_j's and β_k's are positive, (1.6) holds and the distribution of Z has unbounded support and no atom in zero, then (1.1) with (1.3) has a unique *strictly stationary* causal solution (σ_t^2) (although with infinite first moment) and hence (X_t) is strictly stationary as well (albeit with infinite second moment). *In the GARCH case* integrated *does not mean non-stationary.*

Though we do not know the most general conditions for strict stationarity of a GARCH(p,q) process we gain some information about this problem by considering the ARCH(1) (see Goldie

[22] and Section 8.4 in Embrechts et al. [17]) and GARCH(1,1) (see Nelson [30], Mikosch and Stărică [26]) cases. Given that Z is standard normal, a strictly stationary solution (X_t) exists in the ARCH(1) case if $\alpha_0 > 0$ and $\alpha_1 \in (0, 2e^\gamma)$, where γ is Euler's constant ($e^\gamma = 3.5621\ldots$). For a GARCH(1,1) process the conditions $\alpha_0 > 0$ and $E \ln(\alpha_1 Z^2 + \beta_1) < 0$ are necessary and sufficient for the existence of a strictly stationary version of (X_t). Hence $\alpha_1 + \beta_1 \geq 1$ is possible for certain choices of α_1 and β_1. The general GARCH(p,q) case is very complicated: a strictly stationary solution to (1.3) exists if the top Lyapunov exponent of the random matrices $\mathbf{A}_t$ given in (2.4) is negative; see Kesten [24], Bougerol and Picard [6, 7], Davis et al. [14] for more details on this difficult problem.

2 Measures of dependence

Before talking about the main core issue of the present paper — *long range dependence in volatility* — let us briefly remind the reader the more common ways of measuring the dependence of future observations on their past which one encounters in the econometrics literature. We focus on three such measures and their connections.

Persistence. This is a measure defined in the forecasting context. Formally, a process is said to be *persistent in variance* (see Bollerslev and Engle [4]) if

$$(2.1) \qquad \limsup_{t\to\infty} e_t := \limsup_{t\to\infty} |E(X_t^2|X_0, X_{-1}, \ldots) - E(X_t^2|X_1, X_0, \ldots)| \; > \; 0 \quad \text{a.s.}$$

In words, a process (X_t) is *persistent in variance* if the differences between the forecasts of the conditional variances at times 0 and 1 will never disappear, or if shocks to conditional variance persist indefinitely. In the case when the limit in (2.1) is 0, *the rate* at which shocks to conditional variance ultimately die out (i.e. *the rate* at which e_t goes to 0) is a measure of dependence in the forecasting framework. The notion of persistence in variance refers to *conditional* variances and not to the variance itself. Therefore it can still be defined in a meaningful way when the variance of X_t is infinite.

For example, consider a strictly stationary GARCH(1,1) process with iid noise sequence (Z_t) satisfying $EZ = 0$ and $EZ^2 = 1$. Using the recursion (1.4), direct calculation shows that

$$e_t = |\nu_1| \, \alpha_1 \, (\alpha_1 + \beta_1)^{t-1} \quad \text{a.s.}, \quad t > 1 \,.$$

Thus the stationary GARCH(1,1) process is *non-persistent in variance* if and only if $\alpha_1 + \beta_1 < 1$, i.e. when X_t has finite variance. Shocks to conditional variance disappear at the exponential rate $(\alpha_1 + \beta_1)^t$. The stationary IGARCH(1,1) model with $\alpha_1 + \beta_1 = 1$ is *persistent in variance*, i.e. shocks to conditional variance never die out.

Mixing. In order to measure dependence in the theoretical setting of strictly stationary sequences it is convenenient to use some kind of a *mixing condition.* Recall that a strictly stationary sequence of random vectors $\mathbf{Y}_t$ is *strongly mixing*, if there exist constants ϕ_k (the *mixing coefficients*) such that

$$\sup_{A\in\sigma(\mathbf{Y}_s\,,\,s\le 0)\,,\,B\in\sigma(\mathbf{Y}_s\,,\,s>k)} |P(A\cap B)-P(A)\,P(B)| \;=:\; \phi_k\to 0 \quad \text{as } k\to\infty. \tag{2.2}$$

The rate at which $\phi_k \to 0$ is a measure of the range of the memory of the time series . The slower ϕ_k goes to 0, the further in the past the memory of the process reaches. The sequence $(\mathbf{Y}_t)$ is *strongly mixing with geometric rate* if there exist constants $K>0$ and $a\in(0,1)$ such that $\phi_k \le K\,a^k$ for all $k\ge 1$. We refer to Doukhan [16] for a collection of facts on mixing properties. In words, *strongly mixing with geometric rate* stands for a process that forgets its past quickly.

The squared processes (X_t^2) and (σ_t^2) satisfy the following *stochastic recurrence* (or difference) *equation*:

$$\mathbf{X}_t = \mathbf{A}_t\mathbf{X}_{t-1}+\mathbf{B}_t\,, \tag{2.3}$$

where

$$\mathbf{X}_t = \left(X_t^2,\ldots,X_{t-p+1}^2,\sigma_t^2,\ldots,\sigma_{t-q+1}^2\right)'\,,$$

$$\mathbf{A}_t = \begin{pmatrix} \alpha_1 Z_t^2 & \cdots & \alpha_{p-1}Z_t^2 & \alpha_p Z_t^2 & \beta_1 Z_t^2 & \cdots & \beta_{q-1}Z_t^2 & \beta_q Z_t^2 \\ 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \\ \alpha_1 & \cdots & \alpha_{p-1} & \alpha_p & \beta_1 & \cdots & \beta_{q-1} & \beta_q \\ 0 & \cdots & 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \tag{2.4}$$

$$\mathbf{B}_t = (\alpha_0 Z_t^2,0,\ldots,0,\alpha_0,0\ldots,0)'\,.$$

Equation (2.3) allows one to embed the squared GARCH(p,q) process (X_t^2) and the squared volatility process (σ_t^2) in the Markov chain $(\mathbf{X}_t)$, and so standard theory on the mixing properties of such chains can be applied. This has been done in Davis et al. [14]. One conclusion of that paper is that, under mild conditions on the distribution of Z and negativity of the top Lyapunov exponent of the matrices $\mathbf{A}_t$ (which condition ensures stationarity), the Markov chain $(\mathbf{X}_t)$ is strongly mixing with geometric rate, which fact implies in turn that (X_t) and (σ_t) are *strongly mixing with geometric rates.*

The ACF. For any stationary sequence (Y_t) define the *autocorrelation function* (ACF) as

$$\gamma_Y(h)=\operatorname{corr}(Y_0,Y_h)\,,\quad h\in\mathbb{Z}\,.$$

We say that (Y_t) exhibits *long-range dependence* (*LRD*) if

$$\sum_{h=0}^{\infty} |\gamma_Y(h)| = \infty\,, \tag{2.5}$$

and *short-range dependence* otherwise.

A particular consequence of the strong mixing property with geometric rate of (X_t) is that

$$|\gamma_{f(X)}(h)| \le \text{constant } a^h \quad \text{for some } a \in (0,1)\,,$$

for any measurable function f, whenever these correlations are well defined. In particular, if $\gamma_{|X|}$ and γ_{X^2} are well defined these ACFs decay to zero at an exponential rate. Hence we may conclude:

GARCH models are not designed for modeling LRD.

Neither *persistence* nor *non-persistence in variance with a slow rate of decay of* e_t *to* 0 are directly comparable with the notions of *LRD* or *mixing*. Indeed, persistence in variance is defined via conditional moments whereas LRD and mixing are defined in terms of unconditional moments and probabilities. Also notice that LRD is defined only for finite variance stationary processes while mixing does not depend on the second order structure of the time series. For example, the IGARCH process is strongly mixing, i.e. "forgets" quickly in the sense of the unconditional probabilities. Nevertheless it is also persistent, i.e. from the conditional variance stand point, its memory extends indefinitely in the past. Moreover, it has infinite variance marginal distributions. Hence for such a model *it does not make sense* to talk about LRD. When one uses the terms "long memory" or "LRD", one first needs to make clear in which sense they are used.

3 A closer look at real-life data

3.1 The LRD effect

Long log-return series (X_t) of foreign exchange rates, stock indices and share prices have the following properties in common:

- The sample ACF $\widehat{\gamma}_X$ of the data is tiny for all lags, save possibly the first ones; the sample mean is not significantly different from zero. This indicates that (X_t) is a white noise process.
- The sample ACFs $\widehat{\gamma}_{|X|}$ and $\widehat{\gamma}_{X^2}$ of the absolute values and their squares
  - are all positive,
  - decay fast for the first few lags,
  - remain "almost constant" for larger lags.

  This is what we call the *LRD effect*.

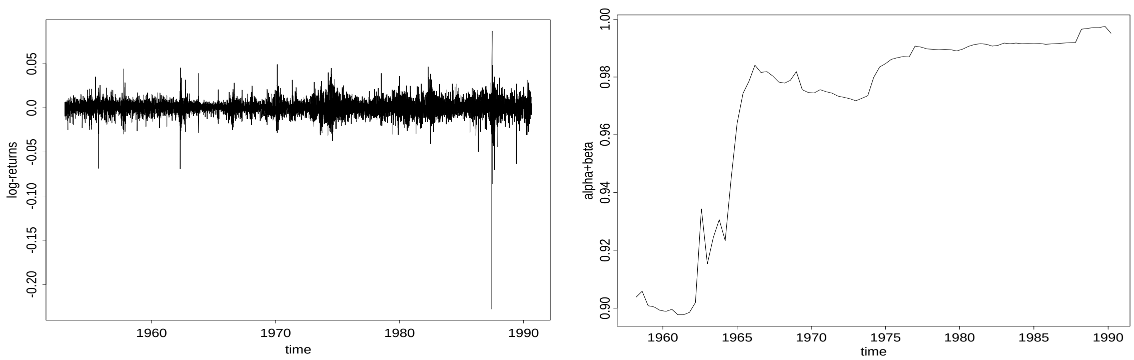

Figure 3.1 Left: *Plot of* 9558 S&P500 *log-returns. The year marks indicate the beginning of the calendar year.* Right: *The estimated values of* $\alpha_1 + \beta_1$ *for an increasing sample of* S&P500 *log-returns. An initial* GARCH(1,1) *model was estimated on the first* 1500 *observations. Then* α_1 *and* β_1 *were re-estimated on increasing samples of size* $1500 + k * 100$, $k > 0$. *The labels on the x-axis indicate the date of the latest observation used for the estimation procedure. The graph shows how the* IGARCH effect *builds up when the sample size increases.*

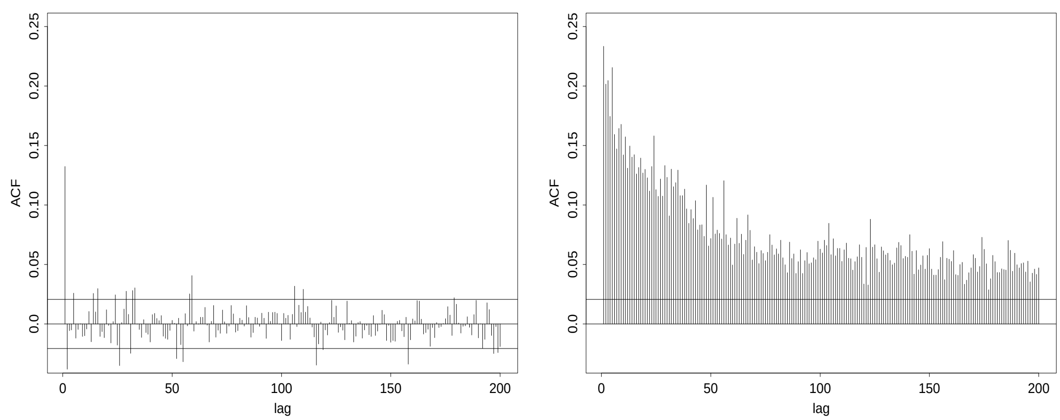

Figure 3.2 Left: *Sample ACF for the* S&P500 *log-returns. Here and in what follows, the horizontal lines in graphs displaying sample ACFs are set as the* 95% *confidence bands* ($\pm 1.96/\sqrt{n}$) *corresponding to the ACF of iid Gaussian white noise.* Right: *Sample ACF for the* S&P500 *absolute log-returns.*

The first mentioned empirical property of the sample ACF for the data fits nicely with the fact that the X_t's from a GARCH(p,q) process are uncorrelated (provided their second moment exists). Recalling that the GARCH(p,q) process actually has exponentially decaying autocorrelations, we may doubt that a GARCH process can capture the particular behaviour of the sample ACFs of the real-life $|X_t|$'s and X_t^2's described as the *LRD effect*.

In order to illustrate the mentioned "stylized sample ACF facts" we consider the daily log-returns of the Standard & Poor's 500 composite stock index from January 2, 1953, to December 31, 1990. The sample ACF of the log-returns and their absolute values (called *absolute log-returns* in what follows) are displayed in Figure 3.2. The same data set will be used in the sequel to substantiate most of our statements.

Since the GARCH(p,q) process cannot describe the observed sample ACF behaviour in an adequate way, we may ask for alternative explanations of this phenomenon. In the literature various answers have been given which we now want to mention.

3.2 The IGARCH effect

The estimation of GARCH processes on log-return data produces with regularity the following results:

- For longer samples, the estimated parameters $\alpha_1, \ldots, \alpha_p$ and $\beta_1, \ldots, \beta_q$ of the model (1.1), (1.3) sum up to values *close* to one.

- When shorter subsamples are used for estimation, the sum of the coefficients, although not small, *stays away* from 1.

We will refer to these two regularities as the *IGARCH effect*. Figure 3.1 illustrates the IGARCH(1,1) *effect* with the S&P500 data.

The first mentioned stylized fact motivated the introduction of the IGARCH(p,q) process (1.6) by Engle and Bollerslev [18] as a possible generating process for log-returns. Under the assumptions given above, in particular $EZ^2 = 1$, the IGARCH model has a strictly stationary solution, but the X_t's do not have a finite second moment. To see this take expectations in the defining equation (1.6) and note that $E\sigma^2 = EX^2$:

$$E\sigma^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \, EX^2 + \sum_{j=1}^{q} \beta_k \, E\sigma^2 = \alpha_0 + E\sigma^2 \,.$$

Since $\alpha_0 > 0$ is necessary for strict stationarity, $E\sigma^2 = \infty$ follows. For an IGARCH(1,1) process, if the distribution of Z satisfies some mild assumptions (such as the existence of a density with infinite support), it follows from a classical result of Kesten [24] (see also Goldie [22] for an alternative proof) that

$$P(X > x) \sim \text{constant}\; x^{-2}, \quad x \to \infty \,.$$

We refer to Mikosch and Stărică [26] for details and further references.

At this point it is important to notice that

the IGARCH model and the LRD notion are incompatible.

Indeed, our definition of *LRD* in terms of the ACF is not applicable since the ACF is not well defined. Thus, *if* the IGARCH model was correct, in particular the variance infinite, the sample ACFs of (X_t), $(|X_t|)$ and (X_t^2) would estimate nothing meaningful. A plausible explanation of the empirically observed *LRD effect* would then be:

If *the IGARCH model is the generating process of the log-returns, the* LRD effect *has nothing to do with* LRD*; it is simply an artifact since the sample ACFs do not measure anything.*

This point of view has been further explored in Mikosch and Stărică [26]. There we discuss the behaviour of the sample ACF for the less extreme case of an "almost integrated" GARCH(1,1) model when $\alpha_1 + \beta_1$ is less than, but close to 1. In this case, the ACFs for the log-returns and absolute log-returns are well defined and the corresponding sample ACFs have, possibly, a meaning. The outcome of this research is very much in line with the previous discussion and can be summarised as follows.

- X has power law behaviour in the tails:

  there exists $\kappa > 2$ but *close* to 2 such that $P(X > x) \sim \text{constant } x^{-\kappa}$ as $x \to \infty$.

- If the ACF is well defined, the rate of convergence of the sample autocorrelations (of the X_t's, $|X_t|$'s, X_t^2's) to their deterministic counterparts is *extremely slow*; it is the slower the smaller the value of κ.

- If the ACF is not defined (depending on κ, this can happen for certain powers of the absolute log-returns), the sample autocorrelations converge in distribution to a *non-degenerate limit.*

Thus, if one assumes an "almost integrated" GARCH model as log-return generating process, the sample ACF is again not very meaningful.

Although a possible explanation for the LRD effect, taking the IGARCH effect at face value contradicts other empirical findings. As we have already mentioned, assuming an estimated GARCH model (i.e. an integrated or "almost integrated" GARCH) as generating process for log-returns presupposes a tail index κ of (or close to) 2. However the existing statistical evidence shows quite convincingly that the tails of real log-returns are not so heavy; see for example Müller et al. [28] and Embrechts et al. [17].

We can offer two alternative explanations for the deviation of κ from 2.

- The statistical estimates of κ are poor.
- The *IGARCH effect* is spurious and occurs *because* the GARCH process is not a suitable model for the data.

The first fact has been discussed in detail (see Embrechts et al. [17] and the references therein; see also Stărică [34]) and cannot be neglected. However, despite the large variation of estimators for κ our experience shows:

The sum of estimated GARCH coefficients is close to 1 **always** *when one uses a sufficiently long log-return series, but it usually stays away from* 1 *when estimating on smaller subsamples.*

(See for example Figure 3.1.) This observation seems to indicate that a GARCH process is not a suitable model and that the *IGARCH effect* is just an artifact. Since changes of the structure in long log-return series are much more likely than in short ones, the following *hypothesis* (which goes back to Diebold [15] and Lamoureux and Lastrapes [25]) sounds plausible:

The IGARCH effect *is due to non-stationarity in log-return series.*

It is perhaps the time to soothe the worried reader who has not forgotten that our aim was to explain the *LRD effect* in the absolute log-returns. It seems that instead of doing that we ended up discussing possible connections between the *IGARCH effect* and non-stationarity. However, she should rest reassured, the detour was deliberate and brought us close to the question we believe to be central to the understanding of the issue at hand:

Is it possible that both, the LRD *and the* IGARCH effects, *are caused by the same simple reason:*
non-stationarity of the data?

A possible answer is given in the next section.

3.3 Checking the goodness of fit of GARCH processes

In order to verify in which period of time a GARCH(p,q) model gives a good fit to real-life data we constructed a goodness of fit test statistic in the spectral domain:

$$S_n := \sqrt{n} \sup_{\lambda\in[0,\pi]} \left| \sum_{h=1}^{n-1} \frac{\widehat{\gamma}_X(h)}{[\mathrm{var}(X_0X_h)]^{1/2}} \frac{\sin(\lambda t)}{t} \right| . \tag{3.1}$$

Under the null hypothesis that (X_t) comes from a GARCH(p,q) model with given parameters α_j and β_k, $\mathrm{var}(X_0X_h)$ can be calculated. Mikosch and Stărică [27] proved under the null hypothesis and assuming a finite 4th moment for X, that S_n converges in distribution to the supremum of a Brownian bridge on $[0,\pi]$. Thus the limit theory for S_n is very much the same as for the classical Kolmogorov–Smirnov goodness of fit test statistic; cf. Shorack and Wellner [33]. In Figure

3.3 we show how one can apply S_n in order to detect changes in the GARCH structure of the S&P500 log-return series. The graphs show that *the unconditional variance of the log-returns varies strongly through time.* A frequently re-estimated GARCH(1,1) model seems to capture the change in variance.

A glance at Figures 3.3 and 3.4 shows quite convincingly:

One particular GARCH process is a good model for the log-return time series only for a relatively short period of time, and therefore the underlying GARCH models have to be updated quite frequently.

Since the classical tools of time series analysis such as the *sample ACF* and the *periodogram* together with the results of *parametric model estimation* can be interpreted in a meaningful way *only if the underlying data can be thought of as coming from a strictly stationary process*, we may question everything we have done so far: the sample ACFs of Figure 3.2, in particular the *LRD effect*, and the parameter estimates for the α_j's and β_k's, in particular the *IGARCH effect*. However, is there a simple explanation for what we see in the data?

In [27] we considered a time series

$$Y_1^{(1)}, \ldots, Y_{[np]}^{(1)}, Y_{[np]+1}^{(2)}, \ldots, Y_n^{(2)},$$

where $p \in (0,1)$ is a fixed number. The two pieces of this time series come from distinct stationary ergodic models. (We focus here on two time series, the case of a finite number of such pieces can be treated completely analogously.) Simple calculation shows that as $n \to \infty$, the sample ACF at lag h converges:

$$\widehat{\gamma}_Y(h) \stackrel{P}{\to} p\,\gamma_{Y^{(1)}}(h) + (1-p)\,\gamma_{Y^{(2)}}(h) + p(1-p)\,(EY^{(1)} - EY^{(2)})^2\,. \tag{3.2}$$

Now assume that the two subsamples are also uncorrelated. Let $\lambda_j = 2\pi j/n$, $j = 1, 2, \ldots$, denote the Fourier frequencies. Then the periodogram $I_Y(\lambda_j)$ (the natural estimator of the spectral density f_Y of a *stationary* process (Y_t); see Brockwell and Davis [9]) at the Fourier frequencies satisfies as $n \to \infty$ and $\lambda_j \to 0$

$$\begin{aligned} &EI_Y(\lambda_j) \\ &\quad\sim\; p\,2\pi f_{Y^{(1)}}(\lambda_j) + (1-p)\,2\pi f_{Y^{(2)}}(\lambda_j) + \frac{2}{n\lambda_j^2}(EY^{(1)} - EY^{(2)})^2(1 - \cos(2\pi jp)). \end{aligned} \tag{3.3}$$

Let us apply our findings to a sample that consists of two subsamples from different GARCH(p,q) processes, $X_t^{(1)}, t = 1, \ldots, [np]$, and $X_t^{(2)}, t = [np]+1, \ldots, n$. Since all these variables have mean zero, we conclude from (3.2) and (3.3) that

$$\begin{aligned} \widehat{\gamma}_X &\stackrel{P}{\to} p\,\gamma_{X^{(1)}}(h) + (1-p)\,\gamma_{X^{(2)}}(h) = 0\,, \\ EI_Y(\lambda_j) &\sim p\,2\pi f_{X^{(1)}}(\lambda_j) + (1-p)\,2\pi f_{X^{(2)}}(\lambda_j) = p\,\mathrm{var}(X^{(1)}) + (1-p)\,\mathrm{var}(X^{(2)})\,. \end{aligned}$$

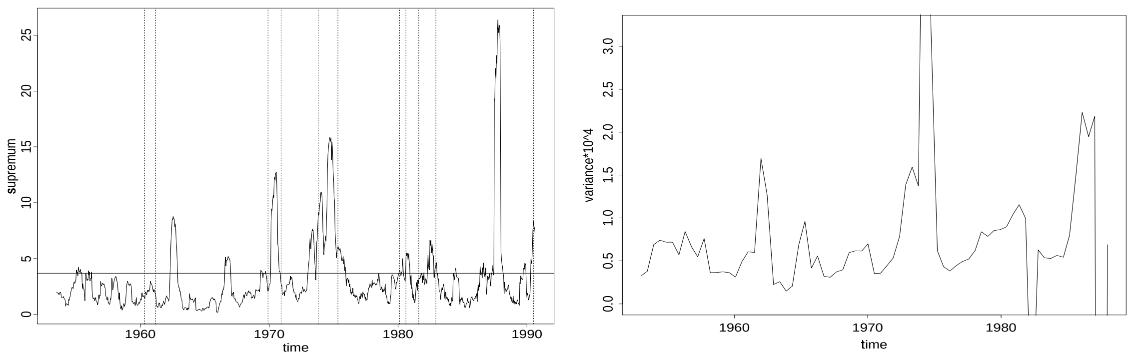

Figure 3.3 Left: *The goodness of fit test statistic* S_{125} (*see* (3.1)) *calculated on a weekly basis from previous* 125 *observations (approximately* 6 *months) of the* S&P500. *Based on estimation of the parameters from the first* 1500 *observations, we check the null hypothesis* $\sigma_{t+1}^2 = 8.58 \times 10^{-6} + 0.072X_t + 0.759\sigma_t^2$. *The horizontal line is set at* 3.6, *the* 99% *quantile of the limit distribution of* S_n. *Values above the confidence bound correspond to* 6 *months periods when the hypothesised model is inappropriate. Essentially, high values of the statistic* S_n *signal higher unconditional variance than that of the supposed model. The dotted vertical lines mark the beginning and end of economic recessions as determined by the National Bureau of Economic Research. They nicely show the coincidence between the recession periods and the intervals of higher unconditional variance detected by our tool.* 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. The graph displays the variances* $\sigma_X^2 = \alpha_0/(1 - \alpha_1 - \beta_1)$. *The similarities between the two graphs seem to show that a frequently re-estimated* GARCH(1,1) *model captures to a certain extent the changing unconditional variance of the log-returns.*

Figure 3.4 *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. Estimated* α_1 (left) *and* β_1 (right).

Thus we expect that *the sample ACF estimates zero at all lags and the periodogram estimates a constant;* see Figure 3.6. This is in agreement with the empirical findings for log-return series; see Figure 3.2.

Similar calculations for the absolute values and squares of the time series predict a totally different behavior. For example, assume that $E|X^{(1)}| \neq E|X^{(2)}|$ (the case of the squared time series is analogous). Then (3.2) implies that

$$\widehat{\gamma}_{|X|} \quad \stackrel{P}{\to} \quad p\,\gamma_{|X^{(1)}|}(h) + (1-p)\,\gamma_{|X^{(2)}|}(h) + p(1-p)(E|X^{(1)}| - E|X^{(2)}|)^2 \,.$$

Since the ACF of the absolute values of a GARCH process decays to zero exponentially the terms

$$p\,\gamma_{|X^{(1)}|}(h) + (1-p)\,\gamma_{|X^{(2)}|}(h)$$

decay to zero at an exponential rate, and so we may expect that we see a fast decay of the sample ACF at the first few lags. This is indeed in agreement with the sample ACF of the absolute values of various log-return series. The typical shape of the sample ACF at large lags of such a time series is however characterised by the constant term

$$p(1-p)\;(E|X^{(1)}| - E|X^{(2)}|)^2 \,,$$

which forces the sample ACF to stay positive and almost constant for a large number of lags and which produces the *LRD effect* in the absolute log-returns; see Figure 3.6.

Now consider the expected periodogram at small Fourier frequencies. From (3.3),

$$EI_{|X|}(\lambda_j) \quad \sim \quad p\;2\pi f_{|X^{(1)}|}(\lambda_j) + (1-p)\;2\pi f_{|X^{(2)}|}(\lambda_j) + \frac{2}{n\lambda_j^2}(E|X^{(1)}| - E|X^{(2)}|)^2(1-\cos(2\pi jp))\,.$$

Assume that $p = r_1/r_2$ for two relatively prime integers. Notice that the term with the cosine is either zero for some frequencies or bounded away from zero for the remaining Fourier frequencies. Since the ACF of the absolute values of a GARCH process decays exponentially, the corresponding spectral density is a continuous function on $[0,\pi]$. Hence

$$p\;2\pi f_{|X^{(1)}|}(\lambda_j) + (1-p)\;2\pi f_{|X^{(2)}|}(\lambda_j) \to p\;2\pi f_{|X^{(1)}|}(0) + (1-p)\;2\pi f_{|X^{(2)}|}(0) = \text{constant}\,.$$

On the other hand, for $n\lambda_j^2 \to 0$ the second term

$$\frac{2}{n\lambda_j^2}(E|X^{(1)}| - E|X^{(2)}|)^2(1-\cos(2\pi j\;r_1/r_2))$$

will give very large values for "almost all" such Fourier frequencies, and this will create the impression of a spectral density which has a singularity at zero.

The above discussion shows:

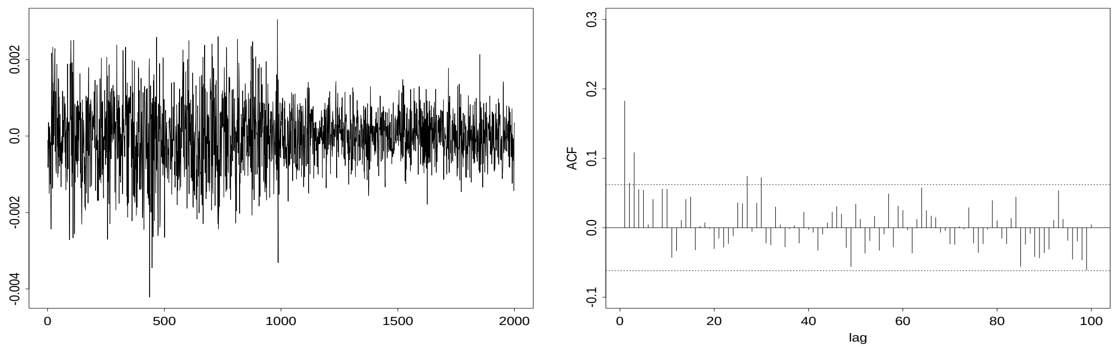

Figure 3.5 Left: *Two independent realizations of length* 1000 *of two* GARCH(1,1) *processes with parameters* $\alpha_0 = 0.13 \times 10^{-6}$, $\alpha_1 = 0.11$, $\beta_1 = 0.52$, *respectively* $\alpha_0 = 0.17 \times 10^{-6}$, $\alpha_1 = 0.20$, $\beta_1 = 0.65$, *are juxtaposed.* Right: *Sample ACF for* $|X_t|$, $t = 1, \ldots, 1000$. *The other sample ACF for* $|X_t|$, $t = 1001, \ldots, 2000$, *looks similar. The sample ACF quickly decreases to* 0.

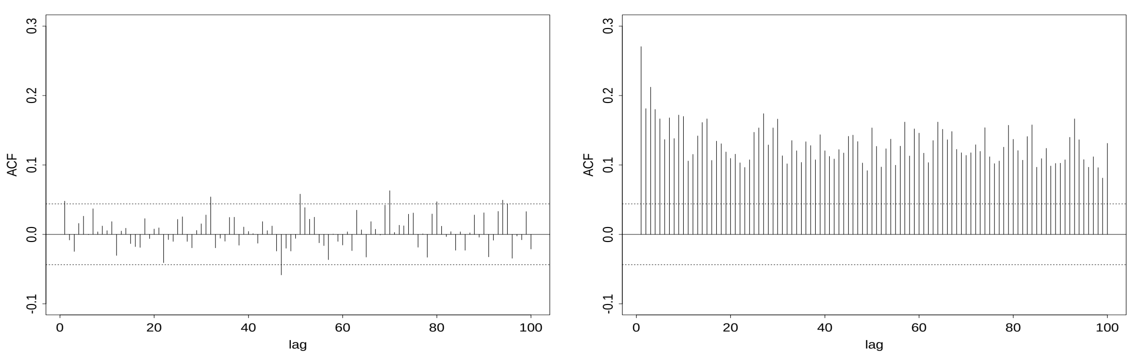

Figure 3.6 Left: *Sample ACF for* X_t, $t = 1, \ldots, 2000$. *The sample ACF is tiny.* Right: *Sample ACF for* $|X_t|$, $t = 1, \ldots, 2000$, *with the* LRD effect.

Non-stationarity of a time series could be responsible for the spurious LRD effect *in the behavior of the* sample ACF *and the* periodogram *of absolute log-returns. The* LRD effect *might have nothing to do with* LRD *as defined in* (2.5).

Finally, we also claim that the *IGARCH effect* might be due to non-stationarity as well. In [27] we showed that the Whittle estimate of α_1 and $\varphi_1 = \alpha_1 + \beta_1$, based on the squares of a GARCH(1,1) process, is consistent, provided X has a finite 4th moment. Moreover, if, as above, we assume that the sample consists of pieces from different GARCH(1,1) models, we showed that the Whittle estimate $\hat{\varphi}_1^W$ of φ_1 is of the order

$$\hat{\varphi}_1^W \sim 1 - \frac{c_1}{c_2 + [\mathrm{var}(X^{(1)}) - \mathrm{var}(X^{(2)})]^2},$$

where c_1 and c_2 are positive constants depending on the coefficients of both GARCH models. Notice that the estimate of φ_1 is the closer to one the larger the difference between the variances of the two models. This might explain the *IGARCH effect* since the longer the time series the larger the chance that strong non-stationarity will affect it and hence the closer to one the estimated value of φ_1 (see Figure 3.1 for an example of how the *IGARCH effect* builds up in longer time series).

It is not very realistic to assume that a real-life time series consists of disjoint pieces from distinct parametric models. It is more natural to think of a sample that, besides sharp switches from a certain regime to another, also contains periods described by models where the parameters change continuously. To understand this kind of behavior in log-returns more detailed research is needed. However, our simplistic model has already shown that *LRD*-type behavior of the sample ACF of absolute log-returns can be due to non-stationarity in the sample, and more sophisticated models will certainly support this hypothesis.

4 Description of LRD by using infinite ARCH models

¿From the theory discussed above we have learnt that ARCH(p) and GARCH(p,q) processes cannot explain *LRD*. If one wants to introduce *LRD* in the ARCH framework, one might be tempted to consider more general models of type (1.1), (1.2) with infinitely many parameters α_j, i.e.

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j X_{t-j}^2\,, \tag{4.1}$$

for non-negative parameters α_j. The ARCH(p) and GARCH(p,q) processes are particular cases.

Models of type (4.1) were introduced, among others, by Baillie et al. [1] and Ding and Granger [12]. Both references are frequently quoted in the LRD econometrics literature. The former authors proposed the fractionally integrated GARCH model (FIGARCH) in analogy to FARIMA processes. They focus on the issue of *persistence in variance* and aim at a model that is *non-persistent in* *variance* but for which shocks to conditional volatility vanish slower than in the case of the GARCH model, i.e at a polynomial rate. The latter authors define the Long Memory ARCH models (LM-ARCH) as the limiting case (as $N \to \infty$) of a model with N volatility components (which is a GARCH(N, N) model). Their aim is a model which displays *LRD* in powers of absolute log-returns.

A careful reading of the two papers reveals that both models were introduced with insufficient analytic description. Most noticeable, neither group of authors prove that a stationary version of their model exists. ([12] does not address this issue leaving open how they define LRD; [1] gives a vague argument for stationarity; see the discussion below.) This step, i.e. proving that the model has a stationary version, is the prerequisite for discussing any estimation and inference procedures, as well as for attempting any data analysis.

The authors of [1] attempt to define the FIGARCH model by the difference equation

$$\phi(L)(1-L)^d X_t^2 = \alpha_0 + (1-\beta(L))\nu_t\,, \quad \nu_t = X_t^2 - \sigma_t^2\,, \tag{4.2}$$

where $\phi(L)$, $1-\beta(L)$ are polynomials in the lag operator L with zeros outside the unit circle, $d \in [0,1]$ and $\alpha_0 > 0$. This definition is meant to remind one that of a FARIMA process (cf. Brockwell and Davis [9], Section 13.2) with X_t^2 on the left-hand side replacing a stationary sequence (Y_t) and (ν_t) on the right-hand side an iid noise sequence. We want to emphasize here that this connection is purely formal and it cannot be used as a waiver of rigourous proofs. Complications due to the formal nature of the relation between the two classes of processes surface in the very definition (4.2) of the FIGARCH process. While in the definition of the FARIMA process we may be assured that the iid noise sequence exists, in the definition (4.2) one constructs the noise sequence (ν_t) from the process (X_t) itself, i.e. from the process one tries to define! Next, it is claimed in [1] that the defining equation (4.2) can be rearranged into the following representation

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta(1)} + \left(1 - \frac{\phi(L)(1-L)^d}{1-\beta(L)}\right) X_t^2 =: \frac{\alpha_0}{1-\beta(1)} + \lambda(L) X_t^2\,. \tag{4.3}$$

In fact, the existence of a stationary solution for (1.1) together with (4.3) must be proved *first* in order to obtain the representation (4.2) of the FIGARCH process. In other words, representation (4.2) *cannot* serve as a definition. Moreover, it could be derived from (4.3) *only if* the time series defined by (1.1) together with (4.3) *has* a strictly stationary version (and, as we will see in the sequel, that it is still to be shown). Rearranging (4.3) to look like (4.2) would also need a formal proof (the same way as, for instance, the linear process representation with respect to the noise sequence of a FARIMA(p,d,q) process with $d \in (-0.5, 0.5)$ requires a formal proof; see [9], Theorem 13.2.1.)

Regarding the stationarity issue of FIGARCH, one reads on p. 158 of [5]: "Since the coefficients in the infinite lag polynomial, $\lambda(L)$, are dominated by the coefficients in the infinite ARCH representation of an appropriately defined high-order IGARCH model, it follows from Bougerol and

Picard [7] and Nelson [29] that the FIGARCH(p, d, q) model is strictly stationary and ergodic." We find this argument hard to follow : what is an "appropriately defined high-order IGARCH model" and how could one bound the coefficients in the infinite lag polynomial, $\lambda(L)$, which decay to 0 at a polynomial rate with the exponentially decaying coefficients in an infinite ARCH representation of an IGARCH model? To the best of our knowledge, a rigorous proof of the existence of a stationary version of the FIGARCH process is still not available. Recent efforts in this direction by Giraitis et al. [20] fell short of achieving it.

It is claimed on p. 8 of [1] and p. 158 of [5], and cited in various other papers, that a FIGARCH process has infinite variance marginals and, therefore, *cannot be covariance stationary.* If this was the case, one would immediately ask: which sense does it make to talk about LRD in the context of such a time series model and how should one interpret the sample ACF plots on pp. 153–155 in [5]? There we get presented the sample ACF plots of the absolute returns of the S&P500 data, their first and fractional differences with $d = 0.5$. In addition to that, the FIGARCH model would be at odds with the existing statistical evidence on tail estimation of log-returns which suggest that an infinite variance for log-returns is extremely unlikely. These questions would raise serious doubts about how appropriate it is to take the FIGARCH model as a log-return generating process.

Before turning to the LM-ARCH model of [12], we want to mention recent work by Giraitis et al. [20] which gave some needed theoretical insight into the class of ARCH(∞) models. They showed that the ARCH(∞) model (1.1), (4.1) with $EZ^2 = 1$, $EZ = 0$ has a strictly stationary non-degenerate version (σ_t^2) with finite first moment if

$$\alpha_0 > 0 \quad \text{and} \quad \sum_{j=1}^{\infty} \alpha_j < 1\,.$$

This condition looks very much the same as in the GARCH case, see (1.5). Giraitis et al. [20] also establish a link between the rate of convergence to 0 of the coefficients α_j and the rate at which the ACF vanishes.

As we have mentioned, Ding and Granger [12] introduced the LM-ARCH process with the aim to reproduce the *LRD effect* in the powers of absolute log-returns. One possible specification of their volatility process is the following:

$$\sigma_t^2 = \sigma^2(1-\mu) + \mu(1-(1-L)^d)X_t^2 \tag{4.4}$$

where $0 \le \mu \le 1$ and $0 < d < 1/2$. The authors claim in their equation (4.26) that $\text{corr}(X_t^2, X_{t-h}^2)$ is of the order h^{2d-1}, provided X has a finite 4th moment. In contrast to this statement, it is *proved* in Giraitis et al. [20] that, for $0 \le \mu < (EX^4)^{-1}$, *the ACF* $\gamma_{X^2}(h)$ cannot *decay at the rate* h^{2d-1}, *with* $d \in (0, 1/2)$. Hence for certain parameter choices which ensure stationarity, *the LM-ARCH model* (4.4) does not *exhibit* LRD. Giraitis et al. [20] mention that for other parameter choices in (4.4) it is not known whether the LM-ARCH model has a strictly stationary version and that "further research is needed". In the light of the discussion for the GARCH case this problem is a very delicate one since the parameter choice in the ARCH(∞) model heavily influences the tail behavior of X.

5 Some concluding remarks

So far we have argued that:

- Standard GARCH models cannot explain *LRD*.
- GARCH models give a reasonable fit to log-returns only for short time horizons. The models have to be updated frequently.
- *LRD* and *IGARCH effects* in log-returns series might be both due to non-stationarity.

This point of view is not shared by many authors who wrote papers on long memory in absolute log-return series. To the contrary, the empiricially observed *sample ACF* and *periodogram* behavior of such time series made absolute log-returns one of the warhorses of the *LRD* community.

We are aware that other models, for example stochastic volatility models (see for example Breidt et al. [8]) or the ARCH(∞) version of Giraitis et al. [21], *can* model the *LRD*-type behavior in the absolute log-returns and their squares.

However, common sense may allow one to ask at least two questions.

- Which economic reasons exist for *LRD* in absolute log-returns?
- What would we gain if we knew that there is *LRD* in absolute log-returns?

We have tried hard to find in the literature any *convincing* rational/economic argument in favor of *long-range dependent stationary log-returns*, but we did not find any, and so the above questions remain, to the best of our knowledge, unanswered. Since one cannot decide about the stationarity of a stochastic process on the basis of a finite segment (sample) the question as to whether there is *LRD* in the absolute log-returns or not will certainly keep a part of the academic community busy also in the future.

References

[1] Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996) Fractionally integrated generalized autoregressive conditional heteroskedasticity. *J. Econom.* **74**, 3–30.

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

[3] Bollerslev, T., Chou, R.Y. and Kroner, K.F. (1992) ARCH models in finance: A review of the theory and evidence. *J. Econometrics*, **52**, 5–59.

[4] BOLLERSLEV, T. AND ENGLE, T. (1993) Common persistence in conditional variances. *Econometrica* **61** (1993), 167–186.

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

[6] BOUGEROL, P. AND PICARD, N. (1992a) Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

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

[8] BREIDT, J.F., CRATO, N. AND LIMA, P. DE (1998) On the detection and estimation of long memory in stochastic volatility. *J. Econom.* **83**, 325–348.

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

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

[11] CAMPBELL, J.Y., LO, A.W. AND MACKINLEY, A.C. (1997) *The Econometrics of Financial Markets.* Princeton University Press, Princeton.

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

[13] DAVIS, R.A. AND MIKOSCH, T. (1998) The sample autocorrelations of heavy–tailed processes with applications to ARCH. *Ann. Statist.* **26**, 2049–2080.

[14] 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. Available at *www.math.rug.nl/∼mikosch*

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

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

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

[18] ENGLE, R.F. AND BOLLERSLEV, T. (1986) Modelling the persistence of conditional variances. *Econom. Rev.* **27**, 1–50.

[19] GHYSELS, E., HARVEY, A. AND RENAULT, E. (1997) Stochastic volatility. In: Madala, G.S. and Rao, C.R. (Eds.) *Statistical Methods of Finance. Handbook of Statistics,* vol. 14, pp. 119–191.

[20] GIRAITIS, L., KOKOSZKA, P. AND LEIPUS, R. (1998) Stationary ARCH models: dependence structure and Central Limit Theorem. *Econom. Theory*, to appear.

[21] GIRAITIS, L., ROBINSON, P. AND SURGAILIS, D. (1998) A model for long memory conditional heteroskedasticity. Preprint.

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

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

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

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

[26] 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. Available at *www.math.rug.nl/∼mikosch*

[27] Mikosch, T. and Stărică, C. (1999) Change of structure in financial time series, long–range dependence and the GARCH model. Technical Report. University of Groningen. Available at *www.math.rug.nl/∼mikosch*

[28] Müller, U.A., Dacorogna, M.M. and Pictet, O.V. (1998) Heavy tails in high-frequency financial data. In: Adler, R.J., Feldman, R.E. and Taqqu, M.S. (Eds.) *A Practical Guide to Heavy Tails*, Birkhauser, Boston, pp. 55–78.

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

[30] Nelson, D.B. (1991) Conditional heteroskedasticity in asset returns. A new approach. *Econometrica* **59**, 347–370.

[31] *RiskMetrics* (1995) Technical document. JP Morgan, New York.

[32] Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (Eds.) *Likelihood, Time Series with Econometric and other Applications.* Chapman and Hall, London.

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

[34] Stărică, C. (1998) On the tail empirical process for stochastic difference equations. Technical Report. Available at *www.math.chalmers.se/∼starica*

[35] Taylor, S. (1986) *Modelling Financial Time Series.* Wiley, New York.