NON-STATIONARITIES IN FINANCIAL TIME SERIES, THE LONG RANGE DEPENDENCE AND THE IGARCH EFFECTS

THOMAS MIKOSCH AND CĂTĂLIN STĂRICĂ

ABSTRACT. In this paper we give the theoretical basis of a possible explanation for two stylized facts observed in long log-return series: the long range dependence (LRD) in volatility and the integrated GARCH (IGARCH). Both these effects can be theoretically explained if one assumes that the data is non-stationary.

1. Introduction

The long range dependence (LRD) in volatility and the integrated GARCH are common findings in the analysis of long series of log-returns, $X_t = \log P_t - \log P_{t-1}$, $t = 0, 1, \ldots$ of stock indices, foreign exchange rates, bond yields, etc. (with P_t we denote the prices of such instruments). More concretely, the sample ACFs of the absolute values and their squares have the following three features. First, they are all positive, second, they decay relatively fast at the first lags and third, they tend to stabilize around a positive value for larger lags. We will refer to this behavior as the LRD effect of absolute/squared log-return data. Concomitantly, the periodograms for the absolute values of the log-returns and their squares blow up at frequencies near zero.

The integrated GARCH finding can be observed when fitting a GARCH$(1,1)$ model

$$\left\{ \begin{array}{ll} X_t & = \sigma_t \, Z_t \,, \\ \sigma_t^2 & = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \alpha_1 X_{t-1}^2 \,, \end{array} \right. \qquad t \in \mathbb{Z}\,, \tag{1.1}$$

to the data. While for shorter samples, the estimated parameters α_1 and β_1 sum up to values significantly different from 1, for longer samples their sum becomes close to one. (This motivated the introduction of the integrated GARCH$(1,1)$ model (IGARCH(1,1)) with $\alpha_1 + \beta_1 = 1$ by Engle and Bollerslev [17] as a possible generating process for returns.) We refer to the ensemble of these two phenomena as the IGARCH effect of return data. Figure 1.1 illustrates both the LRD and the IGARCH effects on the daily log-returns of the Standard & Poor's 500 composite stock index from January 2, 1953, through December 31, 1990.

The main contribution of the paper is to explain by theoretical means how both mentioned effects could be due to a plausible type of non-stationarity of the data: changes in the unconditional variance. Some evidence for the presence of this type of non-stationarity in the daily log-returns is brought in a companion paper (Mikosch and Stărică [29]). There a goodness

1991 *Mathematics Subject Classification.* Primary: 62P20; Secondary: 60G10 60F17 60B12 60G15 62M15.

Key words and phrases. Sample autocorrelation, GARCH process, long range dependence, IGARCH, non-stationarity.

This research supported in part by a grant of the Dutch Science Foundation (NWO), by DYNSTOCH, a research training network under the programme Improving Human Potential financed by the The 5th Framework Programme of the European Commission, and the Danish Research Council Grant No 21-01-0546.

of fit test of GARCH$(1,1)$ model in the spectral domain is proposed and subsequently used to perform a thorough analysis of the Standard & Poor's 500 composite stock index.

The connection between non-stationarities and LRD has a long history in the applied probability literature (see Boes and Salas [5], Potter [30], Bhattacharya et al. [4], Anderson and Turkman [1], Teverovsky and Taqqu [31] just to mention a few contributions) and is present (to a much lesser extent) also in the econometrics literature (Hidalgo and Robinson [23], Lobato and Savin [26]). Contemporaneously with and independently of the present work, Granger and Hyung ([21]) and Diebold and Inoue ([13]) investigate in an econometric context the relationship between long memory and structural changes. Their studies focus on understanding this relationship in the concrete cases of a few simple econometric models with parameters that evolve in time. Our paper provides a general theoretical argument that explains, unhampered by particular model assumptions, how a very plausible type of non-stationarities in economic data, i.e. changes in the unconditional mean or variance, can

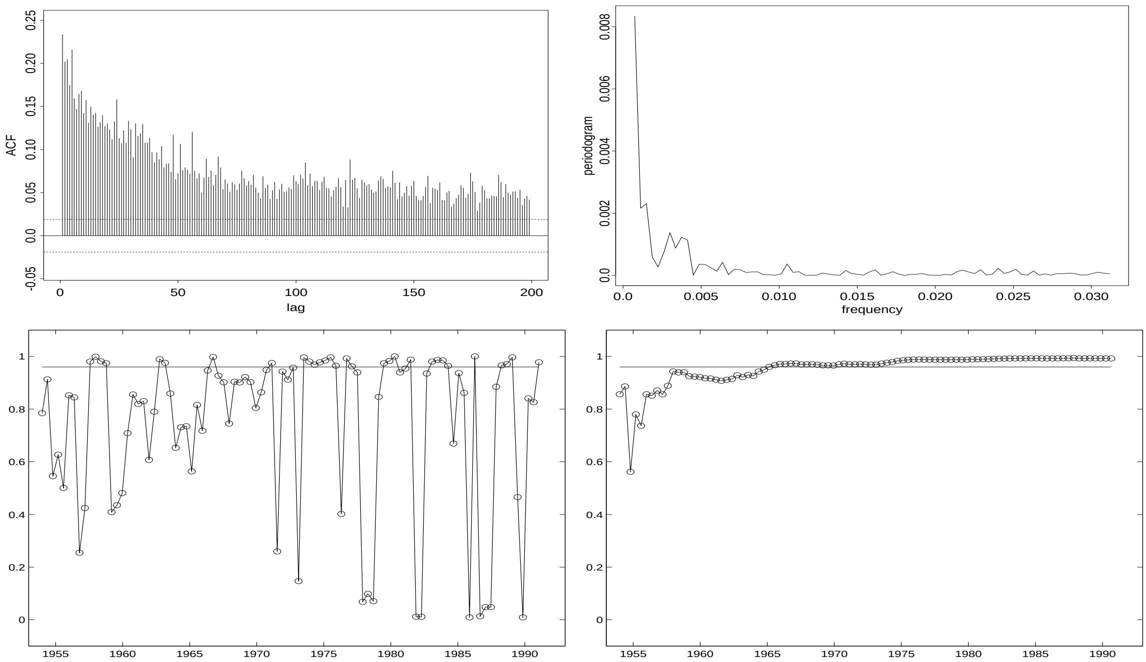

Figure 1.1. Top: *Sample ACFs* (left) *and the periodogram* (right) *of the absolute log-returns of the* S&P500. *Here and in what follows, the horizontal lines in graphs displaying sample ACFs are set as the* 95% *confidence bands* $(\pm 1.96/\sqrt{n})$ *corresponding to the ACF of iid Gaussian white noise.* Bottom: *Estimated values of* $\varphi = \alpha_1 + \beta_1$ *on a moving window with the length of a business-year* (250 *observations*) (left) *and for an increasing sample of* S&P500 *log-returns* (right). *In both graphs a* GARCH$(1,1)$ *model has been re-estimated every* 100 *days* (5 *business months*). *For the graph on the right an initial* GARCH(1,1) *model was estimated on the first* 500 *observations. Then* $k * 100$ *data were successively added to the sample and* α_1 *and* β_1 *were re-estimated on these samples. On the right, we notice the almost monotonous increase of the estimated* φ. *The labels on the* x*-axis indicate the date of the latest observation used for the estimation procedure. The estimation was done using the GARCH module of Splus.*

cause the statistical tools (sample ACF, periodogram) to behave the same way they would if used on stationary long-range dependent sequences.

The possible causal relation between non-stationarities and the integrated GARCH effect is a recurrent theme in the financial econometric literature (Lamoureux and Lastrapes [25], Hamilton and Susmel [22], Cai [11]) and can be traced back to Diebold [12]. As a common feature, all the references we are aware of make use of either simulations or indirect approaches to substantiate their claims. In the second half of the paper, we consider the Whittle estimation for the GARCH model (see Giraitis and Robinson [19] and Mikosch and Straumann [27]) and study the asymptotic behavior of the parameter estimators under structural breaks. We show theoretically that, at least in the frame of the Whittle estimation, the IGARCH effect could be due to the behavior of the estimators under misspecification. We do not have a similar theoretical result for the more common Gaussian quasi maximum likelihood estimator. Based on our results, both the LRD and the IGARCH effect could be spurious.

Our paper is organized as follows. In Section 2 we study the behavior of some statistical tools under non-stationarity. In Section 3 we show that the same type of non-stationarities can cause the IGARCH effect for the Whittle estimator of the parameters of a GARCH process. In Section 4 we substantiate by means of simulations the theoretically proven impact of non-stationarities on estimation of the long memory parameter. We also illustrate on the daily log-returns of the Standard & Poor's 500 composite stock the build-up of the LRD in volatility effect during the oil crisis of the 70's. Some concluding remarks are given in Section 5.

2. Long range dependence effects for non-stationary sequences

Before investigating the LRD effect of return data, a few remarks on the notion of long range dependence are needed. Various definitions of LRD exist in the literature; cf. Beran [2]. In the most general setup, a second order stationary sequence (Y_t) is said to exhibit LRD if the condition $\sum_h |\rho_Y(h)| = \infty$ holds, where $\rho_Y = \operatorname{corr}(Y_0, Y_h)$, $h \in \mathbb{Z}$, denotes the ACF of the Y-sequence. Most popular is the definition of LRD via power law decay of the ACF: assume there is a constant $c_\rho > 0$ such that

$$\rho_Y(h) \sim c_\rho \, h^{2d-1} \quad \text{for large } h \text{ and some } d \in (0, 0.5)\,. \tag{2.1}$$

In this case, the condition $\sum_h |\rho_Y(h)| = \infty$ is satisfied. 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 $\lambda \to 0$. Under some subtle conditions, (2.1) can be shown to be equivalent to the following: for some constant $c_f > 0$,

$$f_Y(\lambda) \sim c_f \, \lambda^{-2d} \quad \text{as } \lambda \downarrow 0. \tag{2.2}$$

The constant $d \in (0, 0.5)$ is called the long memory parameter.

In the econometrics literature, the study of LRD in log-return series is conducted on samples worth many years of data. When studying long time series, non-stationarities are quite likely. On the other hand, the statistical tools we are using are meaningful only under certain assumptions, the most crucial one being the stationarity. Hence the question arises as to what are the statistical tools telling us when used on non-stationary data. In particular, it is well known that various tools for detecting LRD are vulnerable to structural breaks. For example, Bhattacharya et al. [4] proved that the celebrated R/S statistic indicates LRD when the data contain a trend. Teverovsky and Taqqu [31] gave evidence that the sample variance of aggregated time series when applied to short memory data affected by trends or shifts in the mean exhibit the same behavior as long range dependent stationary sequences.

In what follows, we consider the sample ACF and the periodogram in situations when structural breaks occur. We assume that the sample $Y_1, \ldots, Y_n$ consists of different subsamples from distinct stationary models. To be precise, let p_j, $j = 0, \ldots, r$, be positive numbers such that $p_1 + \cdots + p_r = 1$ and $p_0 = 0$. Define

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

The sample $Y_1, \ldots, Y_n$ is written as

$$Y_1^{(1)}, \ldots, Y_{[nq_1]}^{(1)}, \ldots, Y_{[nq_r]+1}^{(r)}, \ldots, Y_n^{(r)}\,, \tag{2.3}$$

where the i subsamples come from distinct stationary ergodic models with finite second moment. The resulting sample is then not stationary.

2.1. **The sample ACF under non-stationarity.** Define the sample autocovariances of the sequence (Y_t) as follows:

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

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

$$\begin{aligned}
\widetilde{\gamma}_{n,Y}(h) &= \sum_{j=1}^{r} p_j \, \frac{1}{np_j} \sum_{t=[nq_{j-1}]+1}^{[nq_j]} Y_t^{(j)} Y_{t+h}^{(j)} - \left(\sum_{j=1}^{r} p_j \, \frac{1}{np_j} \sum_{t=[nq_{j-1}]+1}^{[nq_j]} Y_t^{(j)} \right)^2 + o(1) \\
&\to \sum_{j=1}^{r} p_j E(Y_0^{(j)} Y_h^{(j)}) - \left(\sum_{j=1}^{r} p_j E Y^{(j)} \right)^2 \\
&= \sum_{j=1}^{r} p_j \gamma_{Y^{(j)}}(h) + \sum_{1 \leq i < j \leq r} p_i p_j \left(EY^{(j)} - EY^{(i)} \right)^2 \quad \text{a.s.} \qquad (2.4)
\end{aligned}$$

From (2.4) we can explain the LRD effect in volatility. Suppose that $X_1, \ldots, X_n$ is a log-return series consisting of disjoint subsamples each one being a short memory (more precisely a strongly mixing with geometric rate) white noise with different variances. Then $EX_i = 0$ for all i, and setting $Y = X$ in (2.4) yields that the sample ACF would still estimate zero at all lags. This is in agreement with real-life data. Setting $Y = |X|$ or $Y = X^2$ in (2.4), the expectations of the subsequences $(Y_t^{(j)})$ differ and since the autocovariances $\gamma_{Y^{(j)}}(h)$ decay to zero exponentially as $h \to \infty$ (due to the short memory assumption), 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 (2.4). 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 (2.4) forbids negative correlations for larger lags. This is precisely the picture one sees in the sample ACFs of both simulated and real-life data; see also Section 4.

This is precisely the case if one assumes for example, that $X_1, \ldots, X_n$ is a log-return series consisting of disjoint subsamples which are modeled by distinct GARCH processes. We know that, under mild conditions on the distribution of the noise of a GARCH process, such as the existence of a Lebesgue density, a stationary GARCH process is strongly mixing with geometric rate; see Boussama [8]. This in turn implies exponential decay of the ACF of any function of the data; cf. Doukhan [16]. In particular, this argument applies to the ACF of the absolute values and squares of a GARCH process. Keeping this property in mind, we expect for the samples $|X_1|,\ldots,|X_n|$ and $X_1^2,\ldots,X_n^2$ that their sample ACFs decay quickly for the first lags and then they approach positive constants given by

$$\sum_{1\le i<j\le r} p_ip_j\left(E|X^{(j)}|-E|X^{(i)}|\right)^2 \quad\text{and}\quad \sum_{1\le i<j\le r} p_ip_j\left(E(X^{(j)})^2-E(X^{(i)})^2\right)^2, \tag{2.5}$$

respectively. This would explain the LRD effect we observe in log-return series.

2.2. **The periodogram under non-stationarity.** Alternatively, one may consider estimates of the spectral density. The classical estimator in this case is the *periodogram*

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

which is evaluated at the Fourier frequencies

$$\lambda_j=\frac{2\pi j}{n}\in(-\pi,\pi]. \tag{2.6}$$

The periodogram is the natural (method of moments) estimator of the spectral density of a second order stationary process; see Brockwell and Davis [9, 10].

It is our aim to show that the periodogram at the small Fourier frequencies can become arbitrarily large if the expectations $EY^{(j)}$ of the sequences $(Y_t^{(j)})$ differ. 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, for all t and any constant c, one replaces Y_t with the centered random variable Y_t-c. Therefore centering of Y_t is not necessary. 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

$$
\begin{aligned}
EI_{n,Y}&(\lambda_j)\\
&= \sum_{l=1}^{r} p_l E\left|\frac{1}{\sqrt{np_l}} \sum_{t=1}^{[nq_l]-[nq_{l-1}]-1} (Y_t^{(l)} - EY^{(l)})\,\mathrm{e}^{-i\lambda_j t}\right|^2 + \left|\frac{1}{\sqrt{n}}\sum_{l=1}^{r} EY^{(l)} \sum_{t=[nq_{l-1}]+1}^{[nq_l]} \mathrm{e}^{-i\lambda_j t}\right|^2\\
&= \sum_{l=1}^{r} p_l \left(\mathrm{var}(Y^{(l)}) + 2\sum_{h=1}^{[np_l]-1}\left(1-\frac{h}{[np_l]}\right)\gamma_{Y^{(l)}}(h)\,\cos(\lambda_j h)\right)\\
&\quad + \frac{1}{n}\frac{1}{|1-\mathrm{e}^{-i\lambda_j}|^2}\left|EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1}(EY^{(l)} - EY^{(l+1)})\mathrm{e}^{-i\lambda_j [nq_l]}\right|^2 + o(1)\\
&= \sum_{l=1}^{r} p_l [2\pi f_{Y^{(l)}}(\lambda_j)]\\
&\quad + \frac{1}{n}\frac{1}{|1-\mathrm{e}^{-i\lambda_j}|^2}\left|EY^{(1)} - EY^{(r)} - (1+o(1))\sum_{l=1}^{r-1}(EY^{(l)} - EY^{(l+1)})\mathrm{e}^{-i2\pi j q_l}\right|^2 + o(1)\\
&= \sum_{l=1}^{r} p_l [2\pi f_{Y^{(l)}}(\lambda_j)] + \frac{1}{n\lambda_j^2}\left|EY^{(1)} - EY^{(r)} - \sum_{l=1}^{r-1}(EY^{(l)} - EY^{(l+1)})\mathrm{e}^{-i2\pi j q_l}\right|^2 + o(1)\,,
\end{aligned}
\tag{2.7}
$$

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 (2.7) 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 (2.7) is determined by $(n\lambda_j^2)^{-1}$. For the sake of illustration, assume $r = 2$. Then (2.7) turns into

$$
p_1[2\pi f_{Y^{(1)}}(\lambda_j)] + p_2[2\pi f_{Y^{(2)}}(\lambda_j)] + \frac{1}{n\lambda_j^2}\left|EY^{(1)} - EY^{(2)}\right|^2 2(1-\cos(2\pi j p_1))\,. \tag{2.8}
$$

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

$$
(n\lambda_j^2)^{-1}(1-\cos(2\pi\{jp_1\}))\,, \tag{2.9}
$$

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 (2.9) is either zero or bounded away from zero, uniformly for all j. The effect on (2.8) 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

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

Assume now that a log-return series $X_1, \dots, X_n$ is modeled by disjoint subsamples from distinct GARCH models. Since $EX_i = 0$ we see that the second term in (2.7) disappears. Moreover, since a GARCH process is white noise, its spectral density is a constant. According to (2.7) we expect that the periodogram estimates are flat, i.e. approximate a constant. This is in agreement with periodogram estimates on log-return data. The situation changes when one considers the periodogram of the absolute values $|X_1|, \dots, |X_n|$ and the squares $X_1^2, \dots, X_n^2$. Since the ACFs of both series decay exponentially fast, the spectral densities $f_{|X^{(i)}|}$ and $f_{(X^{(i)})^2}$ corresponding to the ith GARCH model are continuous functions on $[0, \pi]$. Thus the apparent explosion of the spectral estimate for absolute and squares returns at small frequencies could be due to the second term in (2.7) which is non-negligible since the expectations $E|X_t^{(i)}|$ and $E(X_t^{(i)})^2$ differ in the different subsamples.

Let us conclude the section by summarizing our findings. Our theoretical explanations show how shifts in the variance of the data could explain the LRD effect both in the sample ACF and the spectral estimates we observe in real-life log-returns. Moreover we found theoretically that the stronger the non-stationarity (i.e. the bigger the differences $\left(E|X^{(1)}| - E|X^{(2)}|\right)^2$ and $\left(E(X^{(1)})^2 - E(X^{(2)})^2\right)^2$ in the case when $r = 2$), the more pronounced the LRD effect (2.5 and 2.8). See Section 4 for a simulation study on the impact the change of variance has on the estimation of d, the parameter of long memory and for some empirical evidence on the build-up of the LRD effect in the daily log-returns of the Standard & Poor's 500 composite stock.

3. The effect of non-stationarities on the Whittle parameter estimation of GARCH(1, 1) models

The model estimation procedures for a GARCH(1, 1) process (1.1) could also be affected by non-stationarity of the data. In this section we will show by theoretical means that, at least in the framework of Whittle estimation, the IGARCH effect explained in Section 1 appears when the sample displays non-stationarities of the type of changing unconditional variance. More concretely, we focus on the properties of the Whittle estimator for the parameters α_1 and β_1 of the model (1.1) when the sample consists of subsamples from distinct GARCH(1, 1) models.

Our main motivation for the choice of the Whittle estimator is that we can give a theoretical result for the asymptotic behavior of the parameter estimator under non-stationarity. Although the estimation procedure most often used in applications is Gaussian quasi maximum likelihood (see Berkes et al. [3] for some recent results including the consistency and asymptotic normality in a general GARCH(p, q) model) we cannot provide a similar result for this method.

The Whittle estimator is a well known classical pseudo likelihood estimator for ARMA processes. It is asymptotically equivalent to the Gaussian maximum likelihood and least squares estimators and yields consistent and asymptotically normal (with rate $\sqrt{n}$) estimators. Moreover, in the case of an autoregressive process it coincides with the Yule-Walker estimator. We refer to Brockwell and Davis [9], Section 10.8, for an encyclopedic treatment of the Whittle and related estimators.

The asymptotic behavior of the Whittle estimator for GARCH processes was studied in Giraitis and Robinson [19] under an 8th moment assumption and in the remaining cases in Mikosch and Straumann [27]. The convergence rates in the former case are comparable to those of the Gaussian quasi maximum likelihood estimator, but the asymptotic covariances are not comparable and, therefore, there is no obvious theoretical reason why one should prefer Gaussian quasi maximum likelihood. The discussion in [27] (see the simulation results in Section 4), which is also supported by simulations, shows that both estimators are poor for small and medium sample sizes up to 1000, say, but that the Gaussian quasi maximum likelihood estimator is superior for large sample sizes and also in the case when the 8th moment of the data does not exist.

If one assumes that the whole sample $X_1,\ldots,X_n$ comes from a GARCH$(1,1)$ model with parameters α_1 and β_1, it follows (see Appendix A2) that $(U_t)=(X_t^2-EX_t^2)$ can be written as an ARMA(1,1) process

$$U_t-\varphi_1\,U_{t-1}=\nu_t-\beta_1\nu_{t-1}\,,\quad t\in\mathbb{Z}\,, \tag{3.1}$$

with white noise innovations sequence $(\nu_t)=(X_t^2-\sigma_t^2)$, provided $EX_t^4<\infty$, and where

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

The Whittle estimate $\Theta_n=(\overline{\varphi}_1,\overline{\beta}_1)$ of the ARMA$(1,1)$ model (3.1) is obtained by minimizing the Whittle function

$$\overline{\sigma}_n^2(\Theta)=\frac{1}{n}\sum_j\frac{I_{n,U}(\lambda_j)}{g(\lambda_j,\Theta)} \tag{3.2}$$

with respect to Θ from the parameter domain

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

where

$$f(\lambda,\Theta)=\frac{\sigma_\nu^2}{2\pi}\,g(\lambda,\Theta)=\frac{\sigma_\nu^2}{2\pi}\,\frac{|1-\beta_1\mathrm{e}^{-i\lambda}|^2}{|1-\varphi_1\mathrm{e}^{-i\lambda}|^2}\quad\text{and }\sigma_\nu^2=\operatorname{var}(\nu_1^2). \tag{3.3}$$

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 (3.2) is taken over all Fourier frequencies $\lambda_j\in(-\pi,\pi)\backslash\{0\}$ and $f(\lambda,\Theta)$ denotes the spectral density of the ARMA$(1,1)$ process (U_t), see for example Brockwell and Davis [9], Chapter 4.

Given $EX_1^4<\infty$, 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 [9]. 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 [9] shows that for the consistency of the Whittle estimates only the strict stationarity and ergodicity of the ARMA(p,q) process are required, which follow from the corresponding properties of (X_t), see Bougerol and Picard [7].

Now we provide a possible explanation for the IGARCH effect. We show that this could be an artifact due to non-stationarity in the data. We assume that the sample $X_1,\ldots,X_n$ consists of r subsamples from different GARCH$(1,1)$ models (as described in (2.3)) with corresponding parameters $\Theta^{(i)}=(\varphi_1^{(i)},\beta_1^{(i)})$, $i=1,\ldots,r$.

If (X_t^2) constitutes a stationary sequence centering is not necessary in the definition of the Whittle likelihood $\overline{\sigma}_n^2(\Theta)$ since $\sum_{t=1}^n\mathrm{e}^{-i\lambda_jt}=0$ for $\lambda_j\neq0$. Thus, for the Fourier frequencies $\lambda_j\neq0$, $I_{n,X^2}(\lambda_j)=I_{n,U}(\lambda_j)$, and therefore it is without loss of generality assumed in [9] that the sample is mean-corrected. Hence the Whittle likelihood function $\overline{\sigma}_n^2(\Theta)$ can also be rewritten as:

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

where $\overline{X_n^2} = n^{-1}\sum_{t=1}^n X_t^2$.

We start with an analogue of Proposition 10.8.2 in [9].

Proposition 3.1. *Let $X_1,\ldots,X_n$ be a sample consisting of r subsamples as described in* (2.3). *Assume that the ith subsample comes from a* GARCH(1,1) *model with parameter $\Theta^{(i)} = (\varphi_1^{(i)},\beta_1^{(i)})$ in $\mathcal{C}$ and that $E(X^{(i)})^4 < \infty$. Then, for every $\Theta\in\mathcal{C}$ the following relation holds:*

$$\overline{\sigma}_n^2(\Theta) \stackrel{\text{a.s.}}{\to} \Delta(\Theta) := \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{i=1}^r p_i\sigma^2_{\nu^{(i)}}g(\lambda,\Theta^{(i)})}{g(\lambda,\Theta)}\,d\lambda + \frac{\sum_{1\le i<j\le r}p_ip_j\left(\sigma^2_{X^{(i)}}-\sigma^2_{X^{(j)}}\right)^2}{g(0,\Theta)}, \tag{3.4}$$

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

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

the following relation holds

$$\frac{1}{n}\sum_j\frac{I_{n,X^2-\overline{X_n^2}}(\lambda_j)}{g_\delta(\lambda_j,\Theta)} \quad\to\quad \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\sum_{i=1}^r p_i\sigma^2_{\nu^{(i)}}g(\lambda,\Theta^{(i)})}{g_\delta(\lambda,\Theta)}\,d\lambda + \frac{\sum_{1\le i<j\le r}p_ip_j\left(\sigma^2_{X^{(i)}}-\sigma^2_{X^{(j)}}\right)^2}{g_\delta(0,\Theta)} \tag{3.5}$$

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

The proof of the proposition is given in Appendix A1.

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 [9].

Theorem 3.2. *Assume the conditions of Proposition 3.1 are satisfied. Let Θ_n be the minimizer of $\overline{\sigma}_n^2(\Theta)$ 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* (3.4).

The proof of the theorem is given in Appendix A1. It also follows from Proposition 3.1 that $\overline{\sigma}_n^2(\Theta_n)\stackrel{\text{a.s.}}{\to}\Delta(\Theta_0)$.

We now specify the above results for the case of two subsamples, i.e., $p_1+p_2=1$. We exploit the following argument. The spectral density of the ARMA(1,1) process (U_t) is of the form (see (3.3)):

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

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 φ_1

and MA-parameter β_1. Direct calculation shows that

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

By using the particular form of $\gamma_{\widetilde{X}}(i)$ we obtain for $\Delta(\Theta)$, the function to be minimized, the following:

$$
\begin{aligned}
\Delta(\Theta) = \;& \sum_{i=1}^{2} \left[p_i \gamma_{(X^{(i)})^2}(0) \right] \left(1 + \frac{(\varphi_1-\beta_1)^2}{1-\beta_1^2} \right) + p_1 p_2 \left(\sigma^2_{X^{(1)}} - \sigma^2_{X^{(2)}}\right)^2 \frac{(1-\varphi_1)^2}{(1-\beta_1)^2} \\
&+ 2 \sum_{i=1}^{2} \left[p_i \frac{\gamma_{(X^{(i)})^2}(1)}{1-\varphi_1^{(i)}\beta_1} \right] (\varphi_1-\beta_1) \left(-1 + \frac{(\varphi_1-\beta_1)\beta_1}{1-\beta_1^2} \right) .
\end{aligned} \tag{3.6}
$$

It is not possible to obtain an explicit form of the minimizer of $\Delta(\Theta)$ over $\mathcal{C}$. However, minimizing it numerically gives a clear image of what goes on. As an example, we choose to investigate the behavior of the minimum of $\Delta(\Theta)$ when the sample consists of two subsamples of equal size, i.e., $p_1 = p_2 = 0.5$.

The choice of the GARCH$(1,1)$ parameters of one subsample is motivated by the data analysis in the companion paper Mikosch and Stărică [29]; see also Section 4. There a Student-t GARCH$(1,1)$ model was estimated on the first four years of the daily S& P500 log-returns sample covering the period from January 2, 1953, through December 31, 1990 producing the following coefficients

$$
\alpha_0 = 8.58 \times 10^{-6}, \quad \alpha_1 = 0.072, \quad \beta_1 = 0.759\,, \quad \nu = 5.24 \tag{3.7}
$$

where ν is the number of degrees of freedom of the noise sequence (Z_t) (the corresponding value of the fourth moment of the estimated residuals is $EZ^4 = 7.82$). The second subsample is a realization of a GARCH$(1,1)$ model with the same α_0 and ν as in (3.7). The other two parameters, α_1 and β_1 of the second model were chosen to vary around the values 0.072 and 0.759, respectively. The results are presented in Figure 3.3, illustrating the IGARCH effect. The two graphs establish a close connection between the size of the absolute differences in the variances of the two subsamples and φ_1's proximity to 1: the larger the absolute difference in variance, the closer to 1 the value of φ_1 that minimizes the function $\Delta(\Theta)$. This theoretical value is the limit of the Whittle estimate Θ_n.

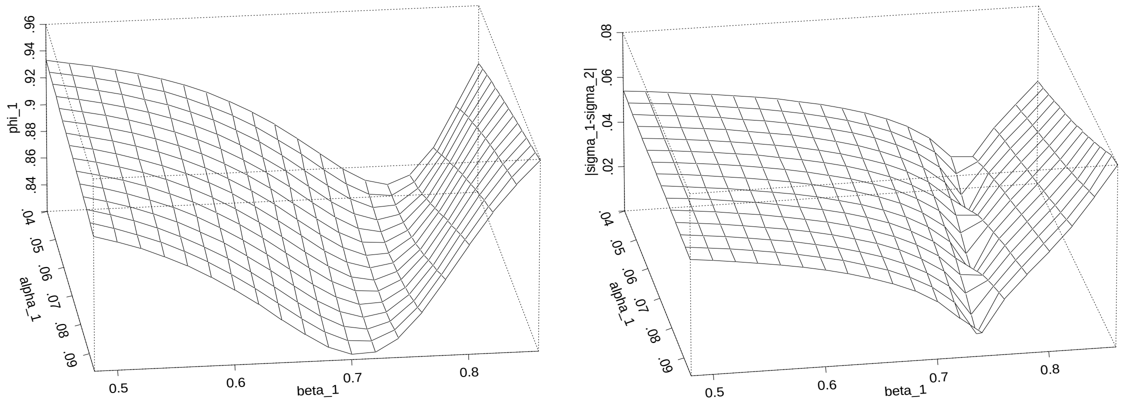

Figure 3.3. Left: *The minimizing φ_1–value of the function $\Delta(\Theta)$ when $p_1 = p_2$. The parameters $\alpha_0^{(1)}$, $\alpha_1^{(1)}$ and $\beta_1^{(1)}$ are fixed according to (3.7). The parameters of the second* GARCH(1, 1) *model $\alpha_1^{(2)}$ and $\beta_1^{(2)}$ vary, whereas the 4th moment of the noise (Z_t) and the parameter $\alpha_0^{(2)}$ are the same.* Right: *The absolute difference between the standard deviations of the two models generating the subsamples.*

The behavior observed in Figure 3.3 can, at least to some extent, explain the behavior of the estimates for α_1 and β_1 of real-life log-return data; see Section 4 for related empirical evidence.

4. Simulation studies and data analyzes

4.1. **The effect of non-stationarities on the Geweke–Porter–Hudak estimator.** The theory in Section 3 establishes a close connection between the explosion of the periodogram of non-stationary sequences around the origin and the difference between the variances of the subsamples. We further illustrate the consequences of non-stationarity for the estimation of the long memory parameter d, defined in (2.1) and (2.2). An estimated value significantly larger than 0 is often taken as evidence for the presence of LRD in the data; see for example Beran [2] for details on the statistical estimation of d. The closer the estimated d to 0.5, the further the dependency is thought to range.

An estimation procedure for d is suggested by the following argument. Assuming that (2.2) holds for $(|X_t|)$, d can be estimated via linear regression from a log-log plot of the periodogram versus the frequency λ, for small values λ

$$\log I_{n,|X|}(\lambda) \approx \log c_f - 2d \log \lambda.$$

This procedure yields the ubiquitous Geweke–Porter–Hudak estimator (GHP) [18] defined as:

$$\widehat{d} := -\frac{1}{2S_{xx}} \sum_{j=1}^{m} a_j I_{n,|X|}(\lambda_j) \tag{4.1}$$

where $a_j = U_j - \overline{U}$, $U_j = \log|2\sin(\lambda_j/2)|$, $\overline{U} = m^{-1}\sum_{j=1}^m U_j$ and $S_{xx} = \sum_{j=1}^m a_j^2$. The λ_j are the Fourier frequencies defined in (2.6) while m is the number of lower frequencies used in estimation. The choice of m is a delicate matter because a too small m causes the estimator to have a high variance while a too large m induces a high bias. Values of m of the order from $n^{0.3}$ to $n^{0.8}$ are common in the literature.

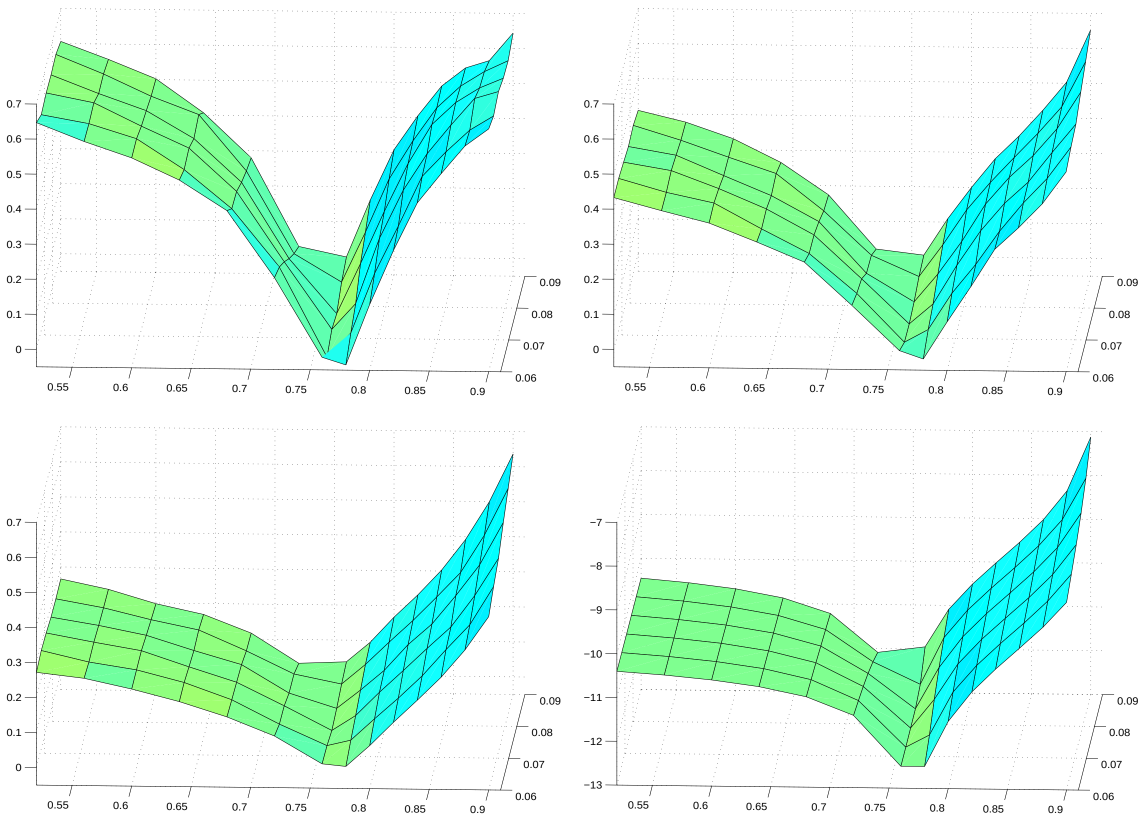

Figure 4.1. *Estimated long memory parameter d for various values of m. The simulated samples have length $n = 2048$. The first 1024 observations come from the* GARCH$(1,1)$ *model with parameters* (3.7)*. The other 1024 observations come from* GARCH$(1,1)$ *processes with parameters α_0 and ν fixed to the values in* (3.7)*, i.e.* 8.58×10^{-6} *and* 5.24 *respectively.* Top: $m = n^{0.4}$ (left) *and* $m = n^{0.5}$ (right). Bottom left: $m = n^{0.6}$. Bottom right: *The log of the absolute values of the differences between the variances of the first and second parts of the sample.*

In a simulation study the long memory parameter d is estimated on the absolute values of samples affected by non-stationarity of the unconditional variance. The simulated samples have length $n = 2048$. In the first set of simulations, with results presented in Figure 4.1, the first 1024 observations of every sample come from a GARCH$(1,1)$ model with Student-t innovations and parameters (3.7). The other 1024 observations were simulated using GARCH$(1,1)$ processes with parameters α_0 and ν fixed to the values in (3.7), 8.58×10^{-6} and 5.24 respectively, but varying α_1 and β_1 in the regions $\alpha_1 \in (0.06, 0.09)$ and $\beta_1 \in (0.52, 0.9)$. In the second set of simulations, with results presented in Figure 4.2, the second half of the sample was simulated using GARCH$(1,1)$ processes with parameters α_1 and β_1 fixed to the values in (3.7), 0.072 and 0.759 respectively, while the parameters α_0 and ν, the parameter of the Student-t distribution, varied between 0.15×10^{-5} and 4×10^{-5} and 4 and 16, respectively. The experiments were repeated 500 times; the estimated value d in the top and bottom left graphs in Figures 4.1 and 4.2 represents the average of the 500 estimates.

The calculations in Section 2.2 predict an explosion of the periodogram in a neighborhood of the origin for sequences affected by changes of variance. We found theoretically that the stronger the non-stationarity (i.e. the bigger the differences $\left(E|X^{(1)}| - E|X^{(2)}|\right)^2$ and $\left(E(X^{(1)})^2 - E(X^{(2)})^2\right)^2$ in the case when $r = 2$), the more pronounced the LRD effect (2.5) and (2.8). This connection can be clearly seen in the results of our simulations. The graphs in Figures 4.1 and 4.2 show that time series with changing unconditional variance produce estimates of the long memory parameter d that could erroneously be interpreted under the assumption of stationarity as evidence of long memory.

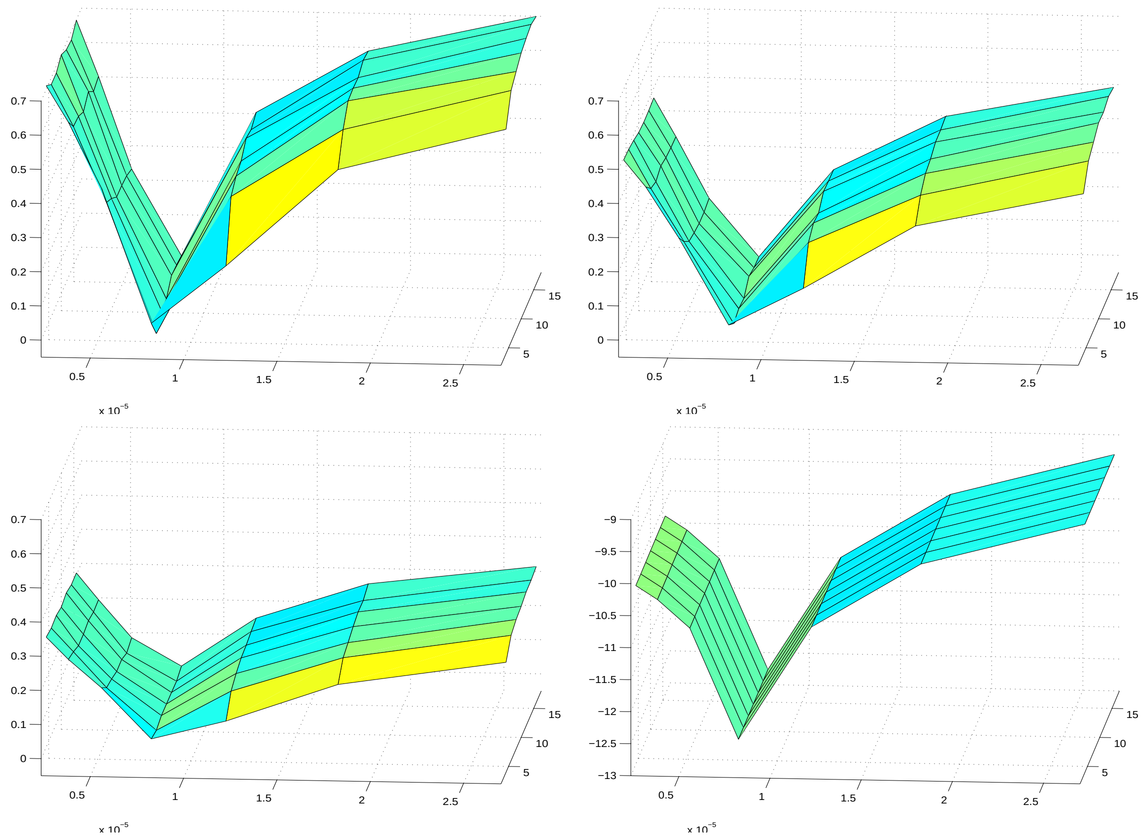

Figure 4.2. *Estimated long memory parameter d for various values of m. The simulated samples have length $n = 2048$. The first 1024 observations come from the* GARCH(1, 1) *model with parameters* (3.7)*. The other* 1024 *observations are simulated using* GARCH(1, 1) *processes with parameters α_1 and β_1 which are kept constant to the values in* (3.7)*,* 0.072 *and* 0.759 *respectively.* Top: $m = n^{0.5}$ *and* $m = n^{0.6}$, Bottom left: $m = n^{0.7}$. Bottom right: *The log of the absolute values of the differences between the variances of the models producing the second half of the sample and that of the model with parameters* (3.7).

4.2. **The Whittle estimator for** GARCH(1, 1) **models.** Since the emphasis in the literature on estimation of GARCH models is on the Gaussian quasi-MLE, other estimation techniques (as for example the Whittle estimation) have often been ignored. We are aware only of two references on this subject: Giraitis and Robinson [19] and Mikosch and Straumann [27]. In this section we present the results of two simulation studies designed to shed some light on the comparative behavior of the two methods of estimation. Figure 4.3 in-

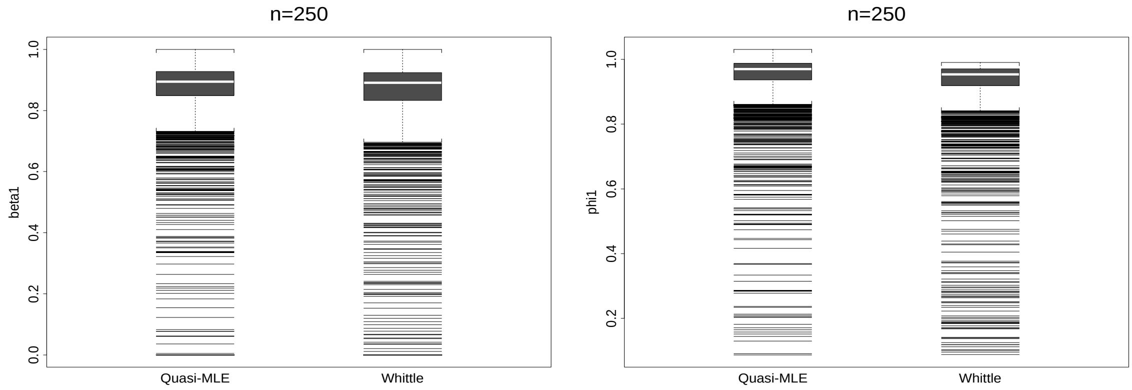

Figure 4.3. *A comparison of the Gaussian quasi-MLE and the Whittle estimator for a* GARCH$(1,1)$ *process with parameters* $\alpha_0 = 8.58 \times 10^{-6}$, $\alpha_1 = 0.072$, $\beta_1 = 0.92$, $\varphi_1 = \alpha_1 + \beta_1 = 0.992$ *and normal innovations. The sample size is* $n = 250$.

dicates that for small sample sizes such as the length of a business year (250 observations) both estimators perform equally (bad) for parameter values that ensure the existence of the 4th moment. For larger sample sizes the Whittle estimator is inferior to the MLE as Figure 4.4 clearly shows. We conclude by noting that in the context of efficiency it is almost impossible to make any theoretical statement concerning the two estimators. More concretely, it is difficult to directly compare the asymptotic covariance matrices of the Whittle estimator (which depend only on the variances of the ν_t's and the parameters) with those of the quasi-MLE (which depends on the distribution of the noise Z_t and the parameters, are rather unattractive and need to be evaluated through simulationss).

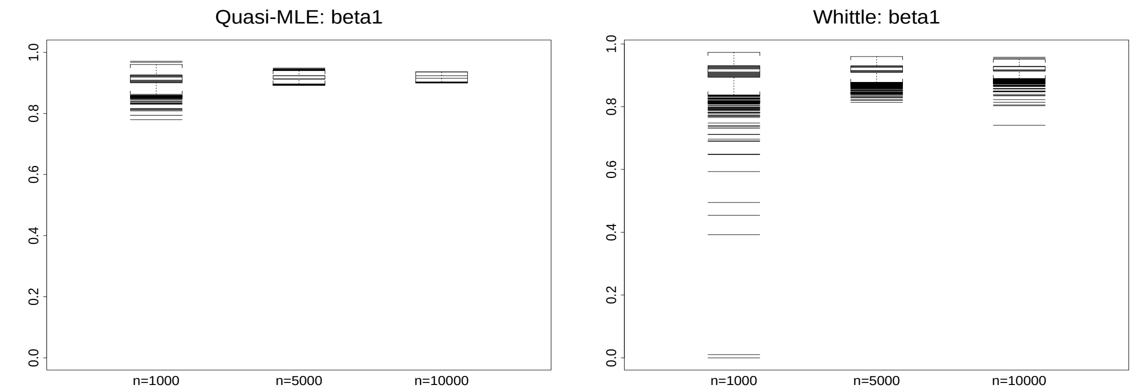

Figure 4.4. *A comparison of the Gaussian quasi-MLE and the Whittle estimators of* $\beta_1 = 0.92$ *in the* GARCH$(1,1)$ *process of Figure 4.3. Gaussian quasi-MLE outperforms the Whittle estimator for larger sample sizes. However, one needs samples of size about* $n = 5000$ *before the asymptotic normality results start working for the quasi-MLE. Both estimators are negatively biased.*

4.3. **LRD in the Standard & Poor's 500 index.** We conclude this section with an illustration of the build-up of the LRD effect based on a real data set. In a companion paper [29], our analysis of the Standard & Poor's 500 composite stock index from January

2, 1953, through December 31, 1990, identified most of the recessions of the period as being structurally different. More concretely, we found that most of the recessions coincide with an increase in the unconditional variance of the time series. We identified the period beginning in 1973 and lasting for almost 4 years as the longest and most significant deviation from the assumption of constant unconditional variance. This period is centered around the longest economic recession in the analyzed data. Figure 4.5 shows the impact which this structurally different period has on the sample ACF of the time series.

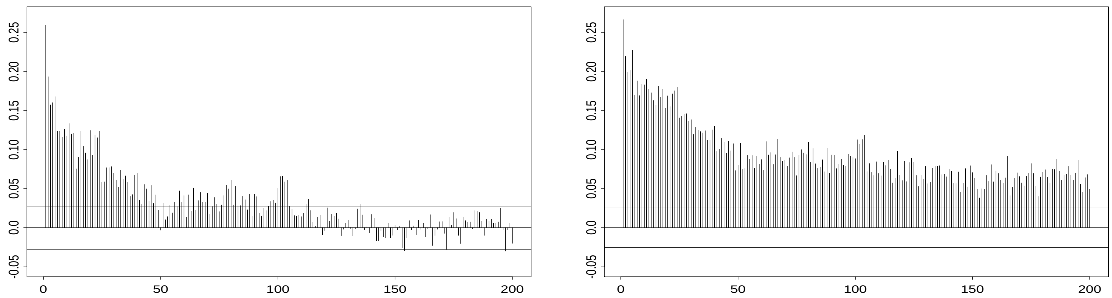

Figure 4.5. *The sample ACF for the absolute log-returns of the first* 20 *and* 24 *years* (left and right) *of the* S&P *data.*

It displays the sample ACF of the absolute values $|X_t|$ up to the moment when the change is detected, i.e., beginning of 1973, next to the sample ACF including the 4-year period that followed. The impact of the change in the structure of the time series between 1973 and 1977 on the sample ACF is extremely severe as one sees from the right graph of Figure 4.5. This graph clearly displays the LRD effect as explained in Section 2.1: exponential decay at small lags followed by almost constant plateau for larger lags together with strictly positive correlations. Contrary to the belief that the LRD characteristic carries meaningful information about the price generating process, these graphs suggest that the LRD behavior could be just an artifact due to very plausible structural changes in the log-return data: variations of the unconditional variance due to the business cycle.

5. Concluding remarks

In this paper we have argued that:

- The LRD effect in log-returns series might be due to non-stationarity. It could be spurious since the statistical tools used to detect it cannot discriminate between stationary, long memory and non-stationary time series.
- Modeling the changes in the conditional variance while assuming stationarity, i.e., constant unconditional variance (via GARCH-type models for example) leads possibly to spurious findings of integrated models (IGARCH).

As for the question as to whether there is LRD in the absolute log-returns or not, we believe that, since one cannot decide about the stationarity of a stochastic process on the basis of a finite sample, it will certainly keep the academic community busy also in the future.

A1. Appendix

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 [9] 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 [9], we restrict ourselves to show that (3.5) is satisfied. The same arguments as on pp. 378–379 in [9] apply. The only fact one then has to check is the a.s. convergence of the sample autocovariances

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

The same arguments as for (2.4) 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 [9], p. 378, introduce the Cèsaro mean of the first m Fourier approximations to $1/g_\delta(\lambda,\Theta)$, given for every $m\ge 1$ by

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

where

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

Then the same arguments as for (10.8.11) in [9] and the display following it show that for every $m\ge 1$,

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

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

$$\left|\sum_{|k|<m}\Gamma_{X^2}(h)\left(1-\frac{|k|}{m}\right)b_k-\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{p_1\sigma^2_{\nu^{(1)}}g(\lambda,\Theta_1)+p_2\sigma^2_{\nu^{(2)}}g(\lambda,\Theta_2)}{g_\delta(\lambda,\Theta)}\,d\lambda\right.$$
$$\left.-\frac{p_1p_2\left(\sigma^2_{X^{(1)}}-\sigma^2_{X^{(2)}}\right)^2}{g_\delta(0,\Theta)}\right|\le const\;\epsilon$$

for every $\epsilon>0$, uniformly in $\Theta\in\overline{\mathcal{C}}$. The same arguments as in [9] conclude the proof.

Proof of Theorem 3.2. One can follow the arguments on p. 385 of [9]. 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 $\delta > 0$,

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

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

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

$$\liminf_{k\to\infty} \widehat{\sigma}^2_{n_k}(\Theta_{n_k}) \geq \Delta(\Theta) > \Delta(\Theta_0) . \tag{A1.1}$$

On the other hand, by definition of Θ_n as a minimizer, (3.4) implies that

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

This is a contradiction to (A1.1). 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 (1.1):

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

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

$$U_t = X_t^2 - EX^2 , \quad t \in \mathbb{Z} ,$$

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 [10]:

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

Straightforward calculation yields

$$\sigma_\nu^2 = (EZ^4 - 1) \, E\sigma_1^4 = \frac{1 + \varphi_1}{1 - \varphi_1} \frac{\alpha_0^2 (EZ^4 - 1)}{1 - (\varphi_1^2 + \alpha_1^2(EZ^4 - 1))} ,$$

$$\sigma_X^2 = \frac{\alpha_0}{1 - \varphi_1} . \tag{A2.1}$$

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 3. We obtain:

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

Acknowledgment. Cătălin Stărică would like to thank the Department of Mathematics at 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. We are grateful to the Editor and the referee for various remarks which led to an improvement of the presentation of the paper.

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

[3] Berkes, I., Horváth, L. and Kokoszka, P. (2000) GARCH processes: structure and estimation. Technical Report. Available under www.math.usu.edu/~piotr/research.html.

[4] Bhattacharya, R.N., Gupta, V.K. and Waymire, E. (1983) The Hurst effect under trends. *J. Appl. Probab.* **20**, 649–662.

[5] 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.

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

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

[8] Boussama, F. (1998) *Ergodicité, mélange et estimation dans le modelès GARCH.* PhD Thesis, Université 7 Paris.

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

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

[13] Diebold, F.X. and Inoue, A. (2001) Long memory and regime switching. *J. Econometrics* **105**, 131–159.

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

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

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

[17] 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.

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

[19] Giraitis, L. and Robinson, P.M. (2000) Whittle estimation of ARCH models. *J. Econom. Th.*, to appear.

[20] Granger, C.W.J. and Ding, Z. (1996) Varieties of long memory models. *J. Econometrics* **73**, 61–77.
[21] Granger, C.W.J., Hyung, N. (1999) Occasional structural breaks and long-memory. Discussion paper 99-14, University of California, San Diego.
[22] Hamilton, J. and Susmel, R. (1994) Autoregresive conditional heteroskedasticity and changes in regime. *J. Econometrics* **64**, 307–333.
[23] Hidalgo, J. and Robinson, P.M. (1996) Testing for structural change in a long–memory environment. *J. Econometrics* **70**, 159 – 174.
[24] Horváth, L., Kokoszka, P. and Teyssiere, G. (2001) Empirical process of squared residuals of an ARCH sequence. *Ann. Statist.* **29**, 445–469.
[25] 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.
[26] Lobato, I. N. and Savin, N. E (1998). Real and spurious long-memory properties of stock-market data. *J. Business and Economic Statist.*, 261–268.
[27] Mikosch, T. and Straumann, D. (2002) Whittle estimation in a heavy-tailed GARCH(1,1) model. *Stoch. Proc. Appl.*, to appear.
[28] Mikosch, T. and Stărică, C. (2002) Long range dependence effects and ARCH modeling. In: Doukhan, P., Oppenheim, G. and Taqqu, M.S. (Eds.) *Long Range Dependence*. Birkhauser, Boston. To appear.
[29] Mikosch, T. and Stărică, C. (2002) Changes of structure in financial time series and the GARCH model *Preprint.* Available at www.math.chalmers.se/~starica
[30] Potter, K. (1976) Evidence for non stationarity as physical explanation of the Hurst phenomenon. *Water Resour. Res.* **12**, 1047–1052.
[31] 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.
[32] Whittle, P. (1951) *Hypothesis Testing in Time Series Analysis.* Almqvist och Wicksel, Uppsala.

Laboratory of Actuarial Mathematics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen, Denmark
E-mail address: mikosch@math.ku.dk

Department of Mathematical Statistics, Chalmers University of Technology, S–412 96 Gothenburg, Sweden
E-mail address: starica@math.chalmers.se