Giraitis, Liudas; Kokoszka, Piotr; Leipus, Remigijus; Teyssière, Gilles

Working Paper

Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity

SFB 373 Discussion Paper, No. 1999,81

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Giraitis, Liudas; Kokoszka, Piotr; Leipus, Remigijus; Teyssière, Gilles (1999) : Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity, SFB 373 Discussion Paper, No. 1999,81, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10046689

This Version is available at:
https://hdl.handle.net/10419/61738

Semiparametric estimation of the intensity of long memory in conditional heteroskedasticity *

Liudas Giraitis †
London School of Economics

Piotr Kokoszka
The University of Liverpool

Remigijus Leipus ‡
Vilnius University

Gilles Teyssière
Humboldt–Universität zu Berlin
& GREQAM, Marseille

May 31, 1999

Abstract. The paper is concerned with the estimation of the long memory parameter in a conditionally heteroskedastic model proposed by Giraitis, Robinson and Surgailis (1999). We consider methods based on the partial sums of the squared observations which are similar in spirit to the classical R/S analysis as well as spectral domain approximate maximum likelihood estimators. The finite sample performance of the estimators is examined by means of a Monte Carlo study.

Keywords: long memory, ARCH models, semiparametric estimation, modified R/S , KPSS and V/S statistics, periodogram

1. Introduction

Long memory, a term commonly used to describe persistent dependence between time series observations as the lag increases, has been shown to be present in geophysical and, more recently, in network traffic data. It is, however, still a matter of debate if market data also exhibit some form of long memory. Many earlier studies, focused on the returns themselves. Long memory in returns, or levels, as it is also commonly referred to, would, however, be a radical departure from the random walk hypothesis and the assumption of the unpredictability of asset returns which underlines the classical asset pricing theory. Empirical studies also suggest that the returns are essentially uncorrelated and the presence of a weak correlation can be to a large extent explained by factors like bid–ask spread and non–synchronous trading, see Campbell *et al.* (1997). However, the presence of strong dependence between the squares or absolute values of returns does not contradict the efficient market hypothesis and many empirical studies suggest that such transformations of returns exhibit some form of persistent dependence. The presence of long memory in the squares of returns may have profound implications. For example, the volatility estimators based on historical data can be affected, which may in turn impact pricing of derivative products.

In order to develop estimation procedures, a parametric or semiparametric model must be postulated in which the squares of returns form a long memory stationary sequence. Even though several attempts have been made to construct such models by modifying classical ARCH or

* Research partially supported by EPSRC grant GR/L/78222 at the University of Liverpool.
† Supported by ESRC grant R000235892.
‡ Partially supported by Lithuanian Science and Studies Foundation Grant K–014.

GARCH specifications, Giraitis, Kokoszka and Leipus (1999a) showed that some of these models have in fact short memory, see Section 2 for more details. Recall that in the context of covariance stationary linear time series, long memory is typically characterized by the requirement that the autocovariance function decays at the rate k^{2d-1}, $0 < d < 1/2$, and hence is not absolutely summable; a series is said to have short memory if the autocovariance function is absolutely summable. These definitions are applicable to *any* stationary sequences, and we adopt them in this paper to the sequences of squares r_t^2, where the r_t follow an ARCH type model developed by Giraitis, Robinson and Surgailis (1999). The new model is different from the traditional ARCH(∞) in that the parameter σ_t itself, not the conditional variance σ_t^2, is a linear function of the past returns. The construction implies that the autocovariance function $\mathrm{Cov}(r_t^2, r_{t+k}^2)$ decays at the rate k^{2d-1} for some $0 < d < 1/2$. We believe that it is not possible to modify the classical ARCH(∞) specification in such a way that the autocovariances of the r_t^2 decay like k^{2d-1}, see Proposition 2.1 and Giraitis *et al.* (1999a) for a more extensive discussion. The model of Giraitis *et al.* (1999) is described in detail in Section 2.

The paper examines two types of estimation procedures. The first class of estimators goes back to the pioneering work of Mandelbrot and his collaborators, see references in Section 3, who developed the *rescaled range*, or R/S , method of Hurst (1951) into a widely used tool for estimating the intensity of long memory. In addition to the R/S method, we also study estimators based on the KPSS statistic of Kwiatkowski *et al.* (1992) and the V/S statistic proposed by Giraitis *et al.* (1999b). In the latter two methods, the range of the partial sums appearing in the R/S statistic is replaced, respectively, by their "second moment" and "variance". Details are presented in Subsection 3.1. The above three methods are based on subdividing the sample into a number of blocks. The choice of the blocks is important as it affects the accuracy of the estimators. There is no theoretical guidance as to how to subdivide the sample, so Monte Carlo simulations must be employed. The second procedure is based on the spectral domain approximate maximum likelihood estimation developed by Robinson (1995) in the setting of linear long memory processes. In a practical implementation of this procedure, the choice of a bandwidth of Fourier frequencies around zero is crucial. Even though some theoretical results are available in the linear and Gaussian cases, see Subsection 3.2, Monte Carlo simulations offer a more detailed guidance.

The paper is organized as follows: Section 2 introduces the model of Giraitis *et al.* (1999). In section 3, we describe the estimators and develop the necessary theoretical background. Section 4 contains the results of an extensive simulation study and provides the technical details of the implementation of the estimation procedures presented in Section 3.

2. The model

We describe in this section the model of Giraitis *et al.* (1999) and discuss its main properties. The central feature of this model is that while the observations (returns) r_t are uncorrelated, their squares have autocovariance function which is not absolutely summable. This is in contrast to a classical ARCH(∞) sequence whose squares have an absolutely summable autocovariance function. To underline the differences between the two specifications, we begin by recalling some relevant properties of the classical ARCH(∞) model.

A random sequence $\{r_k,\ k \in \mathbf{Z}\}$ is said to satisfy ARCH(∞) equations if there exists a sequence of independent identically distributed zero mean random variables $\{\varepsilon_k,\ k \in \mathbf{Z}\}$ such that

$$r_k = \sigma_k \varepsilon_k, \quad \sigma_k^2 = a + \sum_{j=1}^{\infty} b_j r_{k-j}^2, \tag{2.1}$$

where $a \geq 0$, $b_j \geq 0$, $j = 1, 2, \ldots$. As mentioned in the introduction, in this paper we focus on the sequence of squares $X_k = r_k^2$. If the r_k obey (2.1), then the X_k satisfy the equations

$$X_k = \rho_k \xi_k, \quad \rho_k = a + \sum_{j=1}^{\infty} b_j X_{k-j}, \tag{2.2}$$

with $\xi_k = \varepsilon_k^2$ and $\rho_k = \sigma_k^2$. Using a Volterra-type representation

$$X_k = a + a \sum_{l=1}^{\infty} \sum_{j_1,\ldots,j_l=1}^{\infty} b_{j_1} \ldots b_{j_l} \xi_k \xi_{k-j_1} \cdots \xi_{k-j_1-\ldots-j_l},$$

Giraitis *et al.* (1999a) obtained a number of results which show that under mild assumptions, sequences X_k satisfying (2.2) cannot have long memory. These assumptions require essentially that $\sum_{j=1}^{\infty} b_j < \infty$, a condition imposed also in Ding and Granger (1996), Baillie *et al.* (1996) and related papers which aimed at constructing ARCH type models with long memory in squares. Kokoszka and Leipus (1999) showed that under the assumption

$$(E\xi_0^2)^{1/2} \sum_{j=1}^{\infty} b_j < 1 \tag{2.3}$$

there exists a unique weakly stationary solution to (2.2). Giraitis *et al.* (1999a), (1999b) established the following results which show that the classical ARCH model has short memory in squares.

PROPOSITION 2.1. *If assumption* (2.3) *is satisfied, then for any* $k \in \mathbf{Z}$

$$0 \leq \mathrm{Cov}(X_k, X_0) < \infty$$

and

$$\sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_k, X_0) < \infty. \tag{2.4}$$

THEOREM 2.1. *Suppose* $E\xi_0^4 < \infty$ *and*

$$(E\xi_0^4)^{1/4} \sum_{j=1}^{\infty} b_j < 1.$$

Then as $N \to \infty$

$$N^{-1/2} \sum_{j=1}^{[Nt]} (X_j - EX_j) \stackrel{D[0,1]}{\longrightarrow} \sigma W(t), \tag{2.5}$$

where $\{W(t), t \in [0,1]\}$ *is the standard Brownian motion,* $\stackrel{D[0,1]}{\longrightarrow}$ *means weak convergence in the space* $D[0,1]$ *endowed with the Skorokhod topology and* $\sigma^2 = \sum_{k=-\infty}^{\infty} \mathrm{Cov}(X_k, X_0)$.

In the model of Giraitis *et al.* (1999) relations (2.4) and (2.5) no longer hold; the covariances of the X_k decay at the rate k^{2d-1} for some $0 < d < 1/2$, and appropriately normalized partial sums converge to a fractional Brownian motion. The model is defined as follows. The r_k are assumed to satisfy

$$r_k = \sigma_k \varepsilon_k, \quad \sigma_k = \alpha + \sum_{j=1}^{\infty} \beta_j r_{k-j}, \tag{2.6}$$

where $\{\varepsilon_k,\ k \in \mathbf{Z}\}$ is a sequence of zero mean finite variance iid random variables, α is a real number and the weights β_j satisfy

$$\beta_j \sim cj^{d-1}, \quad 0 < d < 1/2, \tag{2.7}$$

for some $c > 0$.

Note that neither α nor the β_j are assumed positive and, unlike in (2.1), σ_k, not its square, is a linear combination of the past r_k, rather than their squares. Observe also that condition (2.7) implies only $\sum_j \beta_j^2 < \infty$ which contrasts with the assumption $\sum_j b_j < \infty$.

Giraitis *et al.* (1999) established the following results which show that the squares of the r_k satisfying (2.6) and (2.7) have two essential features of long memory: hyperbolically decaying non–summable covariances and attraction to a fractional Brownian motion.

THEOREM 2.2. *Suppose* $E\varepsilon_0^4 < \infty$ *and*

$$L(E\varepsilon_0^4)^{1/2} \sum_{j=1}^{\infty} \beta_j^2 < 1, \tag{2.8}$$

where $L = 7$ *if the* ε_k *are Gaussian and* $L = 11$ *in other cases. Then there is a stationary solution to equations* (2.6), (2.7) *given by orthogonal Volterra series*

$$r_k = \sigma_k \varepsilon_k, \quad \sigma_k = \alpha \sum_{l=0}^{\infty} \sum_{j_1,\ldots,j_l=1}^{\infty} \beta_{j_1} \ldots \beta_{j_l} \varepsilon_{k-j_1} \ldots \varepsilon_{k-j_1-\ldots-j_l}. \tag{2.9}$$

The sequence $X_k = r_k^2$ *is covariance stationary and as* $k \to \infty$

$$\mathrm{Cov}(X_k, X_0) \sim Ck^{2d-1}, \tag{2.10}$$

where C *is a positive constant.*

THEOREM 2.3. *If conditions of Theorem 2.2 are satisfied then as* $N \to \infty$

$$\frac{1}{N^{1/2+d}} \sum_{j=1}^{[Nt]} (X_j - EX_j) \stackrel{D[0,1]}{\longrightarrow} c_d W_{1/2+d}(t), \tag{2.11}$$

where c_d *is a positive constant.*

In (2.11), $W_{1/2+d}$ is a fractional Brownian motion with parameter $H = 1/2 + d$. Recall that a Gaussian process $\{W_H(t),\ t \geq 0\}$ is a fractional Brownian motion with parameter $H \in (0,1)$ if it has mean zero and covariances

$$E[W_H(t_1)W_H(t_2)] = \frac{1}{2}(t_1^{2H} + t_2^{2H} - |t_1 - t_2|^{2H}). \tag{2.12}$$

We conclude this section by noting that the smallest possible value of L in (2.8) is not known; this is a complex combinatorial problem. In the Gaussian case the third order cumulants in a diagram formula used in the proof vanish, so a smaller value of L can be taken.

In the simulations presented in Section 4 we also use coefficients β_j for which relation (2.8) fails to hold with $L = 7$, so, strictly speaking, there is no theoretical justification for the results obtained in such cases. The estimation procedures, however, continue to perform quite well, suggesting a need for further theoretical research in this direction.

3. The estimators

In this section, we describe the estimation procedures and provide some theoretical background. Throughout the present section $X_1, \ldots, X_N$ is the observed sample.

3.1. Estimators based on the partial sums

We present here a theoretical background for three estimation procedures based on Theorem 2.3. We begin with the rescaled range, or R/S analysis introduced by Hurst (1951) and subsequently refined by Mandelbrot and his collaborators, see Mandelbrot and Wallis (1969), Mandelbrot (1972, 1975) and Mandelbrot and Taqqu (1979).

The R/S statistic is defined as $\hat{R}_N/\hat{s}_N$ where

$$\hat{R}_N = \max_{1\le k\le N} \sum_{j=1}^{k} (X_j - \bar{X}_N) - \min_{1\le k\le N} \sum_{j=1}^{k} (X_j - \bar{X}_N) \tag{3.1}$$

is the *range* and

$$\hat{s}_N^2 = \frac{1}{N} \sum_{j=1}^{N} (X_j - \bar{X}_N)^2 \tag{3.2}$$

is a standard variance estimator. In (3.1) and (3.2), $\bar{X}_N$ is the sample mean $N^{-1}\sum_{j=1}^{N} X_j$. The identity

$$\sum_{j=1}^{k} (X_j - \bar{X}_N) = \sum_{j=1}^{k} (X_j - EX_j) - \frac{k}{N} \sum_{j=1}^{N} (X_j - EX_j)$$

and Theorem 2.3 imply that

$$\frac{\hat{R}_N}{N^{1/2+d}} \xrightarrow{d} c_d \left\{ \max_{0\le t\le 1} W^0_{1/2+d}(t) - \min_{0\le t\le 1} W^0_{1/2+d}(t) \right\}, \tag{3.3}$$

where

$$W^0_{1/2+d}(t) = W_{1/2+d}(t) - tW_{1/2+d}(1)$$

is a fractional Brownian bridge, cf. (2.12). It is equally easy to verify that

$$\hat{s}_N^2 \xrightarrow{P} \mathrm{Var} X_1. \tag{3.4}$$

Indeed,

$$\hat{s}_N^2 = \frac{1}{N} \sum_{j=1}^{N} \left(X_j^2 - EX_j^2\right) + \left(EX_1^2 - \left[\bar{X}_N\right]^2\right). \tag{3.5}$$

By the Volterra representation (2.9) X_j^2 can be written as $X_j^2 = f(\varepsilon_j, \varepsilon_{j-1}, ...)$ where f is a measurable function. Since $\{\varepsilon_j\}$ is an ergodic sequence this implies (cf. Theorem 3.5.8 of Stout (1974)) ergodicity of $\{X_j^2\}$. Under assumptions of Theorem 2.3 $EX_j^2 < \infty$. Therefore the first term in (3.5) tends to zero. By the same argument as above $\{X_j\}$ is ergodic as well, and therefore $\bar{X}_N \Rightarrow EX_1$. Hence the second term in (3.5) tends to $VarX_1$.

Combining (3.3) and (3.4), we see that as $N \to \infty$

$$\frac{1}{N^{1/2+d}} \frac{\hat{R}_N}{\hat{s}_N} \xrightarrow{d} \frac{c_d \left\{ \max_{0\le t\le 1} W^0_{1/2+d}(t) - \min_{0\le t\le 1} W^0_{1/2+d}(t) \right\}}{(\mathrm{Var} X_1)^{1/2}} =: R_d. \tag{3.6}$$

Relation (3.6) forms a theoretical foundation for the R/S method. Taking logarithms of both sides, we obtain a heuristic identity

$$\log\left(\hat{R}_n/\hat{s}_n\right) \approx \left(\frac{1}{2}+d\right)\log n + \log R_d, \quad \text{as } n\to\infty$$

which shows that $1/2+d$ can be interpreted as the slope of a regression line of $\log(\hat{R}_n/\hat{s}_n)$ on $\log n$ with random intercept $\log R_d$. The point of the R/S analysis is to consider many subsamples of varying size n from a given sample $X_1,\ldots,X_N$ in order to obtain many points which are used to estimate the slope of the regression line, see e.g. Mandelbrot and Taqqu (1979) or Beran (1994). The technical details of the implementation of this procedure are described in Section 4.

The above discussion shows that in place of the range (3.1), any other "simple" continuous functional of the partial sum process can form a basis for an estimation procedure of the type just described. We focus below on the KPSS and V/S statistics used by Giraitis *et al.* (1999b) to test for long memory in ARCH models.

The KPSS statistic was introduced by Kwiatkowski *et al.* (1992) to test trend stationarity against a unit root alternative. Lee and Schmidt (1996) used the KPSS statistic to test for the presence of long memory in a stationary linear time series and gave its asymptotic distribution under long memory alternatives, but provided only heuristic outlines of the proofs.

In the context of testing for long memory in a stationary sequence the KPSS statistic takes the form:

$$\hat{T}_N = \frac{\hat{M}_N}{N\hat{s}_N^2} \tag{3.7}$$

with $\hat{s}_N^2$ given by (3.2) and

$$\hat{M}_N = \frac{1}{N}\sum_{k=1}^{N}\left(\sum_{j=1}^{k}(X_j-\bar{X}_N)\right)^2.$$

We thus see that the range has been replaced by the second moment. We retained the N in the denominator of the RHS if (3.7) in order to conform to the original definition of Lee and Schmidt (1996); unlike the R/S statistic which must be divided by $\sqrt{N}$ in order to ensure convergence for weakly dependent X_j, the statistic $\hat{T}_N$ converges in this case without any normalization.

By Theorem 2.3,

$$\frac{\hat{M}_N}{N^{1+2d}} \xrightarrow{d} c_d^2\int_0^1\left[W^0_{1/2+d}(t)\right]^2 dt. \tag{3.8}$$

Combining relation (3.8) with (3.4), we see that the slope of the regression line of $\log\hat{T}_n$ on $\log n$ estimates $2d$, whereas the regression of $\log(\hat{M}_n^{1/2}/\hat{s}_n)$ on $\log n$ yields an estimate of $d+1/2$.

Giraitis *et al.* (1999b) considered the statistic

$$\hat{U}_N = \frac{\hat{V}_N}{\hat{s}_N^2 N^2}, \tag{3.9}$$

where

$$\hat{V}_N = \frac{1}{N}\left[\sum_{k=1}^{N}\left(\sum_{j=1}^{k}(X_j-\bar{X}_N)\right)^2 - \frac{1}{N}\left(\sum_{k=1}^{N}\sum_{j=1}^{k}(X_j-\bar{X}_N)\right)^2\right].$$

They called $\hat{U}_N$ the V/S statistic for "variance over S". This statistic is very similar to the KPSS statistic, the second sample moment $\hat{M}_N$ in (3.7) is replaced by the sample variance $\hat{V}_N$.

The statistic $\hat{U}_N$ contains a correction for a mean and is more sensitive to "shifts in variance" than $\hat{T}_N$, see Giraitis *et al.* (1999b) for further background and discussion.

Arguing as above, we conclude that the regressions of $\log \hat{U}_n$ and $\log(\hat{V}_n^{1/2}/\hat{s}_n)$ on $\log n$ will, respectively, yield estimates of $2d$ and $d + 1/2$.

In the context of testing for long memory, the estimator $\hat{s}_N^2$ defined by (3.2) is replaced by

$$\hat{s}_{N,q}^2 = \frac{1}{N}\sum_{j=1}^{N}(X_j - \bar{X}_N)^2 + 2\sum_{j=1}^{q}\left(1 - \frac{1}{q+1}\right)\hat{\gamma}_j, \tag{3.10}$$

where $\hat{\gamma}_j$ are the sample covariances

$$\hat{\gamma}_j = \frac{1}{N}\sum_{i=1}^{N-j}(X_i - \bar{X}_N)(X_{i+j} - \bar{X}_N), \quad 0 \le j < N.$$

The second term in (3.10) was introduced by Lo (1991) in order to construct a test for long memory based on the R/S statistic which is robust to many forms of weak dependence. He called the R/S statistic with $\hat{s}_N$ replaced by $\hat{s}_{N,q}$ a *modified* R/S *statistic*. The same modification can be made in (3.7) and (3.9).

3.2. Spectral domain estimation

We describe here the local Whittle estimator proposed by Künsch (1987) and developed by Robinson (1995) which is used to estimate the parameters $c > 0$ and $0 < d < 1/2$ assuming that the observed Gaussian or moving average series has spectral density $f(\lambda)$ which behaves at low frequencies like

$$f(\lambda) \sim c|\lambda|^{-2d} \quad (\lambda \to 0). \tag{3.11}$$

The estimator minimizes an approximate Gaussian maximum likelihood function:

$$\frac{1}{m}\sum_{j=1}^{m}\left\{\ln\left(C\lambda_j^{-2d}\right) + \frac{I(\lambda_j)}{C\lambda_j^{-2d}}\right\}$$

where

$$I(\lambda_j) = \frac{1}{2\pi N}\left|\sum_{k=1}^{N} X_j e^{ik\lambda_j}\right|^2$$

is the periodogram at the Fourier frequencies $\lambda_j = 2\pi j/N, j = 0, \ldots, m$. The bandwidth m increases more slowly than the sample size N:

$$\frac{1}{m} + \frac{m}{N} \to 0 \quad \text{as} \quad N \to \infty.$$

Robinson (1995) showed that under appropriate conditions, which include the existence of a linear moving average representation, the estimator of d is asymptotically normal and converges at the rate $\sqrt{m}$:

$$\sqrt{m}(\hat{d} - d) \sim \mathcal{N}\left(0, \frac{1}{4}\right).$$

In the case of long-memory ARCH sequences discussed in Theorem 2.2, no similar asymptotic theory is available at present. Note however that relation (2.10) implies that the spectral density f of the sequence $X_k = r_k^2$ satisfies (3.11). Thus, although the local Whittle estimator was

designed for Gaussian or moving average time series, we expect that it is applicable also to the ARCH(∞) series with the Volterra representation (2.9). This is because the weights $\beta_1, \beta_2, \ldots$ can be conveniently factorized and are square summable. We conjecture that, similarly as for moving averages, these properties effectively control the dependence structure of the X_k and allow to derive not only the CLT, Theorem 2.2, but also the asymptotic distribution of the local Whittle estimator.

In the Gaussian case, the problem of the choice of the bandwidth m is related to the smoothness of the short memory component $h(\lambda)$ appearing in the following factorization of the spectral density:

$$f(\lambda) = |1 - \exp(i\lambda)|^{-2d} h(\lambda).$$

Assuming that h is twice differentiable and $h(0) > 0$, Hurvich *et al.* (1998) and Delgado and Robinson (1996) proved that

$$m_{optimal} = c_{optimal} n^{4/5}, \tag{3.12}$$

where

$$c_{optimal} = \left(\frac{3}{4\pi}\right)^{4/5} \left(\frac{h''(0)}{2h(0)} + \frac{1}{12}d\right)^{-2/5}. \tag{3.13}$$

We conjecture that in the ARCH(∞) model $X_k = r_k^2$ with the r_k given by (2.9) the $m_{optimal}, c_{optimal}$ are also determined by (3.12), (3.13). We evaluate this optimal bandwidth from our data by using the iterative procedure proposed by Robinson and Henry (1996).

4. Simulations

We consider two sample sizes, $N = 3000$ and $N = 6000$. Once a sample of N observations has been generated, we subdivide it in B adjacent and non-overlapping blocks of observations of equal size $[N/B]$. We then obtain a grid $t_1 = 1, t_2 = [N/B]+1, \ldots, t_i = (i-1)[n/B]+1, \ldots, t_B = n-[N/B]+1$. For each point of the sequence $\{t_i\}_{i=1}^B$ we define a sequence of K increasing nested blocks with origin t_i, i.e., $\{[t_i, t_i+k_j]\}_{j=1}^K$, such that $t_i+k_j \leq N$, the sequence of K steps $\{k_j\}_{j=1}^K$ is given by a logarithmic grid. Given the existence of transcient effects,[1] the minimum value of k is set to 40. The number of blocks B is set to 40.

We calculate the R/S, V/S and KPSS statistics for each interval $\{\{[t_i, t_i + k_j]\}_{i=1}^B\}_{j=1}^K$ and obtain the sequences $\{\{R/S(t_i, k_j)\}_{i=1}^B\}_{j=1}^K$, $\{\{V/S(t_i, k_j)\}_{i=1}^B\}_{j=1}^K$, and $\{\{\mathrm{KPSS}(t_i, k_j)\}_{i=1}^B\}_{j=1}^K$. The denominator of these statistics is the variance the intervals $[t_i, t_i+k_j]$. We plot the logarithm of the statistics $\log(R/S(t_i, k_j)$, $\log(V/S(t_i, k_j)$, $\log(\mathrm{KPSS}(t_i, k_j)$, against $\log(k_j)$ and then obtain a "pox-plot". The estimates of $\hat{d}$, $\hat{d}_{R/S}$, $\hat{d}_{V/S}$, $\hat{d}_{\mathrm{KPSS}}$, are obtained from the OLS estimator. Let $\hat{b}$ be the estimated slope: $\hat{d}_{R/S} = \hat{b} - 1/2$, $\hat{d}_{V/S} = \hat{b}/2$, and $\hat{d}_{\mathrm{KPSS}} = \hat{b}/2$.

As we cannot use the Durbin-Levinson algorithm for generating the series of r_t, we generate each series with 12000 pre-sample values, the infinite order lag polynomial $\beta(L)$ being truncated at the order 5000.

We have considered three Data Generating Processes, which differ by the parameterization of the infinite order lag polynomial $\beta(L)$

- Model A: $\beta_j = b_j$, with $b_1 = d$, $b_j = b_{j-1}\frac{j-1+d}{j}$,
- Model B: $\beta_1 = b_1 + \phi$, $\beta_j = \sum_{k=1}^j \phi^k b_{j-k}$. The coefficients of this DGP are those of the MA form of a FARIMA(1,d,0) with AR coefficient equal to $1 - \phi$.

[1] See Beran (1994).

– Model C: $\beta_1 = b_1 - \theta$, $\beta_j = b_j - \theta b_{j-1}$. The coefficients of this DGP are those of a FARIMA(0,d,1), the MA coefficient being equal to $1 - \theta$.

For the Model A, condition (2.8) is satisfied if $d < 0.1865$. If this condition is not satisfied, there is a systematic bias for the "pox-plot" based estimators. For that reason, we do not report the estimates for $d > 0.225$. Condition (2.8) can be satisfied by multiplying all the β_j by a constant < 1. However, Monte Carlo simulation results show that this rescaling ends with a systematic bias.

For Models B and C, the coefficients β_j depends on d, but also on the parameters θ and ϕ. If the first elements of the sequence of the β_j are small, there is a systematic bias.

For Models B and C, we choose $\theta = 0.20$ and $\phi = -0.20$. The bias is quite large for small values of d, and becomes smaller when $d \in (0.20, 0.375)$, and increases for $d \geq 0.375$.[2]

For all the models, it apears that the Root Mean Squared Error of the R/S estimator is slightly smaller than the RMSE of the other estimators. Although these estimators are biased, they are good tools for an exploratory approach.

5. Conclusions

We have considered in this paper several methods for estimating the degree of long-memory for the long-memory conditional heteroskedastic model developed by Giraitis, Robinson and Surgailis (1999). Two of these estimators are similar to the "pox-plot" R/S estimator. Although Monte Carlo simulation results show that these estimators have a similar bias and variance, these estimators can be used as exploratory tools for detecting the presence of long-range dependence in the conditional variance of some time series.

[2] The whole set of results for $d \in [0.05, 0.5]$ are available upon request.

Table I. Estimation results for the GRS process, Model A: (Root Mean Squared Error between parentheses)

	3000 observations				6000 observations			
d	V/S	R/S	KPSS	Robinson	V/S	R/S	KPSS	Robinson
0.05	0.0092	0.0495	0.0114	0.0285	0.0122	0.0479	0.0137	0.0294
	(0.0622)	(0.0362)	(0.0662)	(0.0362)	(0.0527)	(0.0284)	(0.0557)	(0.0307)
0.075	0.0321	0.0677	0.0363	0.0553	0.0371	0.0682	0.0406	0.0560
	(0.0645)	(0.0378)	(0.0674)	(0.0378)	(0.0536)	(0.0301)	(0.0556)	(0.0324)
0.1	0.0800	0.1102	0.0835	0.0834	0.0826	0.1093	0.0848	0.0852
	(0.0528)	(0.0391)	(0.0585)	(0.0396)	(0.0425)	(0.0317)	(0.0470)	(0.0326)
0.125	0.0949	0.1190	0.1033	0.1125	0.1035	0.1235	0.1103	0.1145
	(0.0591)	(0.0398)	(0.0620)	(0.0413)	(0.0460)	(0.0319)	(0.0487)	(0.0331)
0.15	0.1284	0.1470	0.1383	0.1395	0.1375	0.1527	0.1454	0.1432
	(0.0562)	(0.0406)	(0.0604)	(0.0437)	(0.0438)	(0.0333)	(0.0480)	(0.0345)
0.175	0.1599	0.1738	0.1708	0.1656	0.1689	0.1801	0.1775	0.1708
	(0.0550)	(0.0415)	(0.0606)	(0.0471)	(0.0429)	(0.0343)	(0.0485)	(0.0370)
0.2	0.1776	0.1887	0.1894	0.1701	0.1872	0.1959	0.1965	0.1811
	(0.0578)	(0.0434)	(0.0619)	(0.0579)	(0.0448)	(0.0347)	(0.0490)	(0.0439)
0.225	0.1668	0.1786	0.1801	0.1778	0.1787	0.1874	0.1898	0.1902
	(0.0790)	(0.0625)	(0.0761)	(0.0688)	(0.0632)	(0.0510)	(0.0604)	(0.0531)

Table II. Estimation results for the GRS process, Model B: (Root Mean Squared Error between parentheses)

	3000 observations				6000 observations			
d	V/S	R/S	KPSS	Robinson	V/S	R/S	KPSS	Robinson
0.05	0.0048	0.0434	0.0069	0.0138	0.0075	0.0414	0.0085	0.0153
	(0.0643)	(0.0358)	(0.0678)	(0.0470)	(0.0556)	(0.0290)	(0.0584)	(0.0417)
0.075	0.0041	0.0453	0.0066	0.0069	0.0085	0.0449	0.0105	0.0078
	(0.0846)	(0.0462)	(0.0865)	(0.0742)	(0.0756)	(0.0410)	(0.0767)	(0.0708)
0.1	0.0227	0.0597	0.0276	0.0213	0.0297	0.0615	0.0342	0.0244
	(0.0907)	(0.0543)	(0.0907)	(0.0845)	(0.0797)	(0.0482)	(0.0788)	(0.0794)
0.125	0.0502	0.0815	0.0582	0.0446	0.0603	0.0861	0.0675	0.0516
	(0.0894)	(0.0576)	(0.0875)	(0.0877)	(0.0756)	(0.0493)	(0.0731)	(0.0788)
0.15	0.0841	0.1088	0.0949	0.0764	0.0963	0.1158	0.1059	0.0873
	(0.0833)	(0.0569)	(0.0804)	(0.0837)	(0.0673)	(0.0467)	(0.0644)	(0.0705)
0.175	0.1208	0.1390	0.1338	0.1133	0.1342	0.1478	0.1454	0.1268
	(0.0755)	(0.0543)	(0.0731)	(0.0755)	(0.0585)	(0.0429)	(0.0566)	(0.0593)
0.2	0.1573	0.1697	0.1718	0.1512	0.1709	0.1796	0.1831	0.1662
	(0.0686)	(0.0517)	(0.0680)	(0.0673)	(0.0518)	(0.0398)	(0.0520)	(0.0501)
0.225	0.1914	0.1989	0.2067	0.1875	0.2046	0.2093	0.2171	0.2032
	(0.0639)	(0.0501)	(0.0653)	(0.0620)	(0.0480)	(0.0383)	(0.0503)	(0.0450)
0.25	0.2215	0.2251	0.2370	0.2204	0.2340	0.2357	0.2464	0.2364
	(0.0615)	(0.0498)	(0.0642)	(0.0607)	(0.0465)	(0.0381)	(0.0501)	(0.0444)
0.275	0.2468	0.2474	0.2622	0.2490	0.2584	0.2579	0.2703	0.2647
	(0.0612)	(0.0512)	(0.0640)	(0.0629)	(0.0465)	(0.0393)	(0.0500)	(0.0472)
0.3	0.2676	0.2661	0.2828	0.2728	0.2766	0.2750	0.2876	0.2892
	(0.0624)	(0.0544)	(0.0641)	(0.0663)	(0.0491)	(0.0434)	(0.0510)	(0.0533)
0.325	0.2831	0.2799	0.2985	0.2909	0.2887	0.2864	0.2988	0.3048
	(0.0659)	(0.0609)	(0.0648)	(0.0731)	(0.0576)	(0.0533)	(0.0575)	(0.0611)
0.35	0.2918	0.2880	0.3057	0.3038	0.2968	0.2943	0.3061	0.3176
	(0.0774)	(0.0744)	(0.0742)	(0.0843)	(0.0687)	(0.0664)	(0.0664)	(0.0736)
0.375	0.2963	0.2922	0.3095	0.3115	0.2998	0.2976	0.3083	0.3238
	(0.0932)	(0.0923)	(0.0879)	(0.0992)	(0.0865)	(0.0853)	(0.0827)	(0.0896)

Table III. Estimation results for the GRS process, Model C: (Root Mean Squared Error between parentheses)

	3000 observations				6000 observations			
d	V/S	R/S	KPSS	Robinson	V/S	R/S	KPSS	Robinson
0.05	-0.0064 (0.0723)	0.0374 (0.0370)	-0.0058 (0.0760)	-0.0023 (0.0600)	-0.0053 (0.0656)	0.0344 (0.0316)	-0.0053 (0.0686)	-0.0023 (0.0568)
0.075	-0.0001 (0.0879)	0.0421 (0.0482)	0.0020 (0.0900)	0.0028 (0.0779)	0.0033 (0.0802)	0.0408 (0.0441)	0.0049 (0.0814)	0.0035 (0.0748)
0.1	0.0160 (0.0963)	0.0544 (0.0581)	0.0205 (0.0962)	0.0151 (0.0901)	0.0226 (0.0858)	0.0558 (0.0527)	0.0269 (0.0848)	0.0176 (0.0857)
0.125	0.0417 (0.0965)	0.0745 (0.0628)	0.0494 (0.0942)	0.0359 (0.0952)	0.0517 (0.0828)	0.0789 (0.0550)	0.0589 (0.0798)	0.0425 (0.0870)
0.15	0.0748 (0.0906)	0.1009 (0.0626)	0.0856 (0.0869)	0.0662 (0.0923)	0.0874 (0.0744)	0.1081 (0.0524)	0.0972 (0.0704)	0.0772 (0.0794)
0.175	0.1117 (0.0820)	0.1311 (0.0596)	0.1251 (0.0783)	0.1032 (0.0836)	0.1259 (0.0644)	0.1403 (0.0478)	0.1376 (0.0609)	0.1174 (0.0669)
0.2	0.1494 (0.0737)	0.1625 (0.0561)	0.1644 (0.0714)	0.1424 (0.0734)	0.1640 (0.0559)	0.1730 (0.0435)	0.1768 (0.0543)	0.1583 (0.0554)
0.225	0.1851 (0.0675)	0.1930 (0.0534)	0.2011 (0.0673)	0.1803 (0.0658)	0.1994 (0.0505)	0.2041 (0.0407)	0.2127 (0.0513)	0.1974 (0.0474)
0.25	0.2172 (0.0638)	0.2209 (0.0521)	0.2336 (0.0654)	0.2153 (0.0621)	0.2307 (0.0478)	0.2322 (0.0397)	0.2439 (0.0505)	0.2328 (0.0446)
0.275	0.2448 (0.0619)	0.2455 (0.0521)	0.2611 (0.0642)	0.2470 (0.0619)	0.2570 (0.0473)	0.2562 (0.0403)	0.2696 (0.0504)	0.2636 (0.0461)
0.30	0.2679 (0.0628)	0.2660 (0.0549)	0.2840 (0.0645)	0.2724 (0.0643)	0.2763 (0.0510)	0.2743 (0.0450)	0.2879 (0.0526)	0.2893 (0.0517)
0.325	0.2839 (0.0673)	0.2806 (0.0615)	0.2993 (0.0672)	0.2937 (0.0720)	0.2918 (0.0555)	0.2888 (0.0514)	0.3025 (0.0554)	0.3092 (0.0584)
0.35	0.2961 (0.0746)	0.2916 (0.0717)	0.3108 (0.0720)	0.3083 (0.0801)	0.3018 (0.0650)	0.2985 (0.0629)	0.3118 (0.0629)	0.3236 (0.0683)
0.375	0.3024 (0.0883)	0.2975 (0.0876)	0.3164 (0.0834)	0.3181 (0.0933)	0.3066 (0.0806)	0.3035 (0.0799)	0.3158 (0.0769)	0.3320 (0.0823)

References

Baillie, R. T., Bollerslev, T. and Mikkelsen, H. O.: Fractionally integrated generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **74** (1996), 3–30.

Beran, J.: *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.

Campbell, J. Y., Lo, A. W. and MacKinlay, A. C.: *The Econometrics of Financial Markets*, Princeton University Press, Princeton, 1997.

Delgado, M. A. and Robinson, P. M.: Optimal spectral bandwidth for long memory, *Statistica Sinica* **6** (1996), 97–112.

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

Giraitis, L., Kokoszka, P. and Leipus, R.: Stationary ARCH models: dependence structure and Central Limit Theorem, *Econometric Theory* **00** (1999a), 000–000.

Giraitis, L., Kokoszka, P. and Leipus, R.: Detection of long memory in ARCH models, Preprint, 1999b.

Giraitis, L., Robinson, P. and Surgailis, D.: A model for long memory conditional heteroskedasticity, Preprint, 1999.

Hurst, H.: Long term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers* **116** (1951), 770–799.

Hurvich, C. M., Deo, R. and Brodsky, J.: (1998). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series, *Journal of Time Series Analysis* **19** (1998), 19–46.

Kokoszka, P. and Leipus, R.: Change-point estimation in ARCH models, *Bernoulli* **00** (1999), 000–000.

Künsch, H.: Statistical aspects of self-similar processes, in Yu. A. Prohorov and V V. Sazonov (eds), *Proceedings of the 1st World Congress of the Bernoulli Society*, volume 1, VNU Science Press, Utrecht, 1987. pp. 67–74.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y.: Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics* **54** (1992), 159–178.

Lee, D. and Schmidt, P.: On the power of the KPSS test of stationarity against fractionally-integrated alternatives, *Journal of Econometrics* **73** (1996), 285–302.

Lo, A.: Long-term memory in stock market prices, *Econometrica* **59** (1991) 1279–1313.

Mandelbrot, B. B.: Statistical methodology for non-periodic cycles: from the covariance to R/S analysis, *Annals of Economic and Social Measurement* **1** (1972), 259–290.

Mandelbrot, B. B.: Limit theorems of the self-normalized range for weakly and strongly dependent processes, *Z. Wahrschein. verw. Gebiete* **31** (1975), 271–285.

Mandelbrot, B. B. and Taqqu, M. S.: Robust R/S analysis of long run serial correlation, in *42nd Session of the International Statistical Institute, Manila*, Book 2, 1979, pp. 69–99.

Mandelbrot, B. B. and Wallis, J. M.: Robustness of the rescaled range R/S in the measurement of noncyclic long run statistical dependence. *Water Resources Research* **5** (1969), 967–988.

Robinson, P. M.: Gaussian semiparametric estimation of long range dependence, *Annals of Statistics* **23** (1995), 1630–1661.

Robinson, P. M. and Henry, M.: Bandwidth choice in Gaussian semiparametric estimation of long-range dependence, in P. M. Robinson and M . Rosenblatt (eds), *Athens Conference on Applied Probability and Time Series Analysis*, volume II: Time Series Analysis. In memory of E. J. Hannan, Springer–Verlag, New York, 1996. pp. 220–232.

Stout, W. F.: *Almost Sure Convergence*, Academic Press, New York, 1974.

Address for Offprints: Liudas Giraitis, Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; l.giraitis@lse.ac.uk

Piotr Kokoszka, Department of Mathematical Sciences, University of Liverpool, Liverpool, L69 3BX, United Kingdom; P.S.Kokoszka@liverpool.ac.uk

Remigijus Leipus, Department of Mathematics and Informatics, Vilnius University, Naugarduko 24, Vilnius 2600, Lithuania; Remigijus.Leipus@Maf.Vu.lt

Gilles Teyssière, Humboldt–Universität zu Berlin & GREQAM, Institut für Statistik und Ökonometrie, Spandauer Str. 1, D–10178 Berlin, Germany; teyssier@wiwi.hu-berlin.de, gilles@ehess.cnrs-mrs.fr