ROBUST ESTIMATION OF LONG MEMORY

Lopes, S.R.C.[a1] **and Mendes, B.V.M.**[b]

[a] Instituto de Matemática - UFRGS, Porto Alegre, RS, Brazil
[b] Instituto de Matemática - UFRJ, Rio de Janeiro, RJ, Brazil

Abstract

The problem of tuning an estimator by selecting bandwidth or truncation values is at the core of most semiparametric estimation procedures. This paper investigates the trade-off bias-variance implied by the tuning constant α, which governs the number of frequencies m used by the regression based estimates of the fractional parameter d. We apply classical least squares and robust methodologies to well known regression type estimators and assess their performance as α ranges in $[0.50, 0.86]$. We consider models with long-range dependence in mean and in volatility, and show that short-range dependence structure may affect the estimates and thus the optimal value for the bandwidth m.

Key words: Long memory; FIGARCH models; Stochastic Volatility models; Semiparametric estimation; Robust estimation.

1 Introduction

Models for long memory in mean were first introduced by Granger and Joyeux (1980) and Hosking (1981), following the seminal work of Hurst (1951). The important characteristic of an Autoregressive Fractionally Integrated (ARFIMA) process is its autocorrelation function decay rate. In an ARFIMA process, the autocorrelation function exhibits a hyperbolic decay rate, differently from an ARMA model which presents a geometric rate. Long memory in mean has been observed in data from areas such as meteorology, astronomy, hydrology, and economics, as reported in Beran (1994).

The ARFIMA framework was naturally extended towards volatility models. The Fractionally Integrated Generalized Autoregressive Conditionally Heteroskedastic (FIGARCH) models were introduced by Baillie, Bollerslev and Mikkelsen (1996) and Bollerslev and Mikkelsen (1996), motivated by the fact that autocorrelation function of the squared, log-squared, or the absolute value series of an asset return decays slowly, even when the return series has no serial correlation. Also aiming to model long memory in the second moment, Breidt et al. (1998) introduced the Fractionally Integrated Stochastic Volatility (FISV) model.

[1]Corresponding author: S. R. C. Lopes, Instituto de Matemática - UFRGS, Av. Bento Gonçalves, 9500. 91509-900 Porto Alegre, RS, Brasil. E-mail: slopes@mat.ufrgs.br

Models for heteroskedastic time series with long memory are of great interest in econometrics and finance, where empirical facts about asset returns have motivated the several extensions of GARCH type models (FIEGARCH, TGARCH, SW-ARCH, LM-ARCH, and so on; for a review, see Lopes and Mendes (2005)). Many empirical papers have detected the presence of long memory in the volatility of risky assets, market indexes, exchange rates. As the number of models available increases, it becomes of interest a simple, fast, and accurate estimation procedure for the fractional parameter d, independent of the specification of a parametric model. The regression based semiparametric (semiparametric in the sense that a full parametric model is not specified for the spectral density of the process) estimators seem to be the natural candidates. However, their asymptotic statistical properties, besides depending on their definition and estimation method, are also heavily dependent on the number of frequencies m used for the regression. In addition, their performances are also affected by other structures in the data. In this paper we put some light on this issue, by considering several long memory models and 300 regression type estimators. To specify the bandwidth m we consider the tuning constant α, by setting $m = n^{\alpha}$, where n is the sample size.

The regression method was introduced in the pioneer work of Geweke and Porter-Hudak (1983), giving rise to several other proposals. Hurvich and Ray (1995) introduced a cosine-bell function as a spectral window, to reduce bias in the periodogram function. They found that data tapering and the elimination of the first periodogram ordinate in the regression equation, could increase the estimator accuracy. However, smaller bias was obtained at the cost of a larger variance. Reisen (1994) and Velasco (1999a) considered smoothed versions of the periodogram function. Velasco (1999b) proved consistency and asymptotic normality of the regression estimators for any d, considering non-stationary and non-invertible processes. Reisen et al. (2001) carried out an extensive simulation study comparing both the semiparametric and parametric approaches in ARFIMA processes. Monte Carlo methods were also used by Lopes et al. (2004) in the case of non-stationary ARFIMA processes.

Despite the large number of regression type estimators available, a comprehensive evaluation of their performances in models for long memory in volatility, addressing the trade off bias-variance resulting from the choices of the tuning constant α is still missing. By considering 20 values for α in the range $[0.50, 0.86]$, in this paper we evaluate the performance of 5 semiparametric regression estimates of the fractional parameter in ARFIMA, FIGARCH, and FISV models. Besides the classical least squares method, robust estimation procedures are applied and also tuned with the constant α. We use the efficient 0.50 breakdown point robust estimates Least Trimmed Squares (LTS, Rousseeuw, 1984) and the MM-estimates (Yohai, 1987). Including the Whittle estimator, a total of 301 estimates are implemented in a Monte Carlo study.

Two related works are Taqqu and Teverovsky (1996) and Henry (2001). By noting that high frequencies tend to bias the estimates, and using only low frequencies eliminates the bias but increases the variance, Taqqu and Teverovsky (1996) suggest plotting the estimates as a function of m and the series length n, which would balance bias versus variance. Henry (2001) develops formulae and approximations for an (mean squared error) optimal bandwidth m when estimating long memory in the series level, considering conditionally heteroskedastic errors specifications.

Applications where the only parameter of interest is d may be found in many areas. In finance, for example, where a huge variety of conditionally heteroskedastic models are available, one may first remove the long-range dependence of return series, and then fit to the residuals some GARCH type model accounting for leverage terms, regime switching, different conditional distributions, and so on.

The remainder of this paper is as follows. In Section 2 we define the ARFIMA, FIGARCH and FISV models. In Section 3 we briefly review the semiparametric estimators used and give their robust versions. In Section 4 we carry on several simulation experiments according to 31 different data generating processes, and evaluate the performance of the estimators considering the trade off bias-variance implied by the choice of α. In Section 5 we illustrate using a real data set and in Section 6 we summarize the results.

2 Long-Memory Models

In this section we define the ARFIMA, FIGARCH and FISV models.

2.1 ARFIMA Models

Let $\{X_t\}_{t\in\mathbb{Z}}$ be an ARFIMA(p,d,q) process given by

$$\Phi(\mathcal{L})(1-\mathcal{L})^d X_t = \Theta(\mathcal{L})\epsilon_t, \;\; d\in\mathbb{R}, \tag{2.1}$$

where $\mathcal{L}$ is the backward-shift operator, that is, $\mathcal{L}^k X_t = X_{t-k}$. The polynomials $\Phi(\mathcal{L}) = \sum_{i=0}^{p}(-\phi_i)\,\mathcal{L}^i$ and $\Theta(\mathcal{L}) = \sum_{j=0}^{q}(-\theta_j)\,\mathcal{L}^j$ have degree p and q, respectively, with $\phi_0 = -1 = \theta_0$. The process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is white noise with zero mean and finite variance σ_ϵ^2. The term $(1-\mathcal{L})^d$ is the binomial, or Maclaurin, series expansion in $\mathcal{L}$.

The process $\{X_t\}_{t\in\mathbb{Z}}$, given by expression (2.1), is called a *general fractional differenced zero mean process*, where d is the *fractional differencing parameter*. This process is both stationary and invertible if the roots of $\Phi(\cdot)$ and $\Theta(\cdot)$ are outside of the unit circle and $|d| < 0.5$. Its spectral density function, $f_X(\cdot)$, is given by

$$f_X(w) = f_U(w)\left[2\,\sin(\frac{w}{2})\right]^{-2d}, \; w\in[-\pi,\pi], \tag{2.2}$$

where $f_U(\cdot)$ is the spectral density function of an ARMA(p,q) process. One observes that $f_X(w) \simeq w^{-2d}$, when $w\to 0$.

The ARFIMA(p,d,q) process exhibits *long memory* when $d\in(0.0,0.5)$, *intermediate memory* when $d\in(-0.5,0.0)$ and *short memory* when $d=0$.

2.2 FIGARCH Models

Denote by $\mathcal{F}_t$ the σ-field of events generated by $\{X_s; s\le t\}$ and assume that $\mathbb{E}(X_t|\mathcal{F}_{t-1}) = 0$ *a.s.*. Following Engle (1982), and Bollerslev (1986) we specify a GARCH(r,s) model by

$$X_t = \sigma_t Z_t, \tag{2.3}$$

where Z_t is an independent identically distributed ($i.i.d.$) random variable with zero mean and unit variance such that $X_t|\mathcal{F}_{t-1} \sim i.i.d.(0, \sigma_t^2)$, and $\sigma_t^2 = Var(X_t|\mathcal{F}_{t-1})$ is defined by

$$\sigma_t^2 = \omega + \alpha(\mathcal{L})X_t^2 + \beta(\mathcal{L})\sigma_t^2, \tag{2.4}$$

where $\omega > 0$ is a real constant, $\alpha(\mathcal{L}) = \sum_{i=1}^{r} \alpha_i \mathcal{L}^i$ and $\beta(\mathcal{L}) = \sum_{j=1}^{s} \beta_j \mathcal{L}^j$. For a FIGARCH process (see Baillie et el., 1996, and Bollerslev and Mikkelsen, 1996) the σ_t, in expression (2.3), is defined as

$$\begin{aligned} \sigma_t^2 &= \omega\,(1-\beta(\mathcal{L}))^{-1} + \{1-(1-\beta(\mathcal{L}))^{-1}[1-\alpha(\mathcal{L})-\beta(\mathcal{L})](1-\mathcal{L})^d\}X_t^2 \\ &= \omega\,(1-\beta(\mathcal{L}))^{-1} + \{1-(1-\beta(\mathcal{L}))^{-1}\phi(\mathcal{L})(1-\mathcal{L})^d\}X_t^2 \\ &= \omega\,(1-\beta(\mathcal{L}))^{-1} + \lambda(\mathcal{L})X_t^2, \end{aligned} \tag{2.5}$$

where

$$\lambda(\mathcal{L}) = \sum_{k=0}^{\infty} \lambda_k \mathcal{L}^k = 1 - (1-\beta(\mathcal{L}))^{-1}\phi(\mathcal{L})(1-\mathcal{L})^d, \tag{2.6}$$

$\phi(\mathcal{L}) = 1 - \alpha(\mathcal{L}) - \beta(\mathcal{L}))$, and the binomial series expansion in $\mathcal{L}$ is given by

$$\begin{aligned} (1-\mathcal{L})^d &= 1 + \sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}\mathcal{L}^k = 1 - d\sum_{k=1}^{\infty} \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(1-d)}\mathcal{L}^k \\ &= 1 - d\mathcal{L} - \frac{d}{2!}(1-d)\mathcal{L}^2 - \frac{d}{3!}(1-d)(2-d)\mathcal{L}^3 - \cdots \\ &= 1 - \sum_{k=1}^{\infty} \delta_{d,k}\,\mathcal{L}^k = 1 - \delta_d(\mathcal{L}). \end{aligned} \tag{2.7}$$

The coefficients $\delta_{d,k} = d\frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(1-d)}$, in expression (2.7), are such that

$$\delta_{d,k} = \delta_{d,k-1}\left(\frac{k-1-d}{k}\right), \tag{2.8}$$

for all $k \geq 1$, where $\delta_{d,0} \equiv 1$.

The following proposition totally characterizes any FIGARCH(r, d, s) process and also gives a recurrent formula for the coefficients λ_k's given in expression (2.6).

Proposition 2.1: *Let* $\{X_t\}_{t\in\mathbb{Z}}$ *be any FIGARCH*(r, d, s) *process, for* $d \in [0, 1]$, *defined by expressions (2.3) and (2.5). Then, the coefficients* λ_k, *for* $k \in \mathbb{N}$, *in expression (2.6), are given by*

$$
\begin{aligned}
\lambda_0 &= 0 \\
\lambda_n &= \sum_{i=1}^{r} \beta_i \lambda_{n-i} + \alpha_n + \delta_{d,n} - \sum_{j=1}^{\max\{r,s\}} \gamma_j \delta_{d,n-j}, \quad \textit{if} \quad 1 \le n \le r \\
\lambda_n &= \sum_{i=1}^{s} \beta_i \lambda_{n-i} + \delta_{d,n} - \sum_{j=1}^{\max\{r,s\}} \gamma_j \delta_{d,n-j}, \quad \textit{if} \quad n > r, \qquad (2.9)
\end{aligned}
$$

where

$$
\gamma_j = \begin{cases} \alpha_j, & \textit{if} \quad r > s, \\ \alpha_j + \beta_j, & \textit{if} \quad r = s, \\ \beta_j, & \textit{if} \quad r < s. \end{cases} \qquad (2.10)
$$

Proof: The proof is straightforward if one compares the coefficients of $\mathcal{L}^n$ in both sides of the following expression

$$
\begin{aligned}
[1 - \beta(\mathcal{L})]\, \lambda(\mathcal{L}) &= 1 - \beta(\mathcal{L}) - \phi(\mathcal{L})(1 - \mathcal{L})^d \\
&= 1 - \beta(\mathcal{L}) - [1 - \alpha(\mathcal{L}) - \beta(\mathcal{L})]\,(1 - \delta_d(\mathcal{L})) \\
&= \alpha(\mathcal{L}) + \phi(\mathcal{L})\delta_d(\mathcal{L}). \qquad (2.11)
\end{aligned}
$$

□

For any FIGARCH$(1, d, 1)$ process the parameters have to fulfill some restrictions to ensure positivity of the conditional variance σ_t^2. Besides of ω, α_1 and β_1 being non-negative numbers, these inequalities are as follows

- $\beta_1 - d \le \phi_1 \le \frac{2-d}{3}$
- $d(\phi_1 - \frac{1-d}{2}) \le \beta_1(d + \alpha_1)$, where $\phi_1 = \alpha_1 + \beta_1$.

In a FIGARCH$(1, d, 0)$ process, $\beta_1 = 0$, and in a FIGARCH$(0, d, 1)$, $\alpha_1 = 0$. For any FIGARCH$(0, d, 0)$ there are no further restrictions besides ω being non-negative.

2.3 FISV Models

Let $\{Y_t\}_{t=1}^{n}$ be such that

$$
Y_t = g(X_t)\sigma_\varepsilon \varepsilon_t, \qquad (2.12)
$$

where X_t is a long-memory in mean time series, $g(\cdot)$ is a continuous function and ε_t is an *i.i.d.* time series with zero mean and unit variance. Since $Var(Y_t|X_t) = g(X_t)^2\sigma_\varepsilon^2$, for certain functions $g(\cdot)$ model (2.12) may be described as a long-memory stochastic volatility process (see Robinson, 1999). This large class of volatility models include the long-memory nonlinear moving average models of

Robinson and Zaffaroni (1998) and Zaffaroni (1999), and the FISV process introduced by Breidt et al. (1998).

In a FISV$(p, d, q, \sigma_\varepsilon)$ process $\{Y_t\}_{t\in\mathbb{Z}}$, the function $g(\cdot)$ in (2.12) is given by

$$g(X_t) = \exp\left(\frac{X_t}{2}\right), \tag{2.13}$$

where $\{X_t\}_{t\in\mathbb{Z}}$ is an ARFIMA(p, d, q) process given by (2.1), and ε_t and ϵ_t are $i.i.d.$ standard normal, and mutually independent. One observes that $Var(Y_t|X_t) = \exp(X_t)\sigma_\varepsilon^2$. In particular, squaring both sides of equation (2.13) and taking logarithms,

$$\ln(Y_t^2) = \mu_\xi + X_t + \xi_t, \tag{2.14}$$

where $\mu_\xi = \ln(\sigma_\varepsilon^2) + \mathbb{E}[\ln(\varepsilon_t^2)]$, and $\xi_t = \ln(\varepsilon_t^2) - \mathbb{E}[\ln(\varepsilon_t^2)]$. Hence, $\ln(Y_t^2)$ is the sum of a Gaussian ARFIMA process and independent non-Gaussian noise with zero mean. Consequently, the autocovariance function of the process $\ln(Y_t^2)$, when $d \in (-0.5, 0.5)$, is such that

$$\gamma_{\ln(Y_t^2)}(k) \sim k^{2d-1}, \tag{2.15}$$

when $k \to \infty$, while its spectral density function has the property that

$$f_{\ln(Y_t^2)}(\lambda) \sim \lambda^{-2d}, \tag{2.16}$$

when the frequency $\lambda \to 0$. For $d \in (0.0, 0.5)$, the spectral density function in expression (2.16) is unbounded when $\lambda \to 0$. This forms the basis for the application of the traditional log-periodogram estimation procedures, given in the next section.

3 Classical and Robust Estimation Procedures

In the literature of the stochastic ARFIMA processes, there exist several estimation procedures for the fractional parameter d. In this section we recall some well known regression estimation methods based on the periodogram function and propose new ones.

Let $\{X_t\}_{t\in\mathbb{Z}}$ be a ARFIMA(p, d, q) process with $d \in (-0.5, 0.5)$, given by (2.1). Its spectral density function is given by

$$f_X(\omega) = \left[2\,\sin\left(\frac{\omega}{2}\right)\right]^{-2d} f_U(\omega), \text{ for } 0 < \omega \leq \pi, \tag{3.1}$$

where $f_U(\cdot)$ is the spectral density function of the ARMA process.

Consider the set of harmonic frequencies $\omega_j = \frac{2\pi j}{n}$, $j = 0, 1, \cdots, [n/2]$, where n is the sample size, and $[x]$ means the integer part of x. By taking the logarithm of the spectral density function $f_X(\cdot)$ given by (3.1), and adding $\ln f_U(0)$, and $\ln I(\omega_j)$ to both sides of this expression we obtain

$$\ln I(\omega_j) = \ln f_U(0) - d\ln\left[2\,\sin\left(\frac{\omega_j}{2}\right)\right]^2 + \ln\left\{\frac{f_U(\omega_j)}{f_U(0)}\right\} + \ln\left\{\frac{I(\omega_j)}{f_X(\omega_j)}\right\}, \tag{3.2}$$

where $I(\cdot)$ is the periodogram function given by

$$I(\omega) = \frac{1}{2\pi}\left(\hat{\gamma}_X(0) + 2\sum_{l=1}^{n-1}\hat{\gamma}_X(l)\cos(l\,\omega)\right), \tag{3.3}$$

where $\hat{\gamma}_X(k) = \frac{1}{n}\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})$, for $k \in \{0, 1, \cdots, n-1\}$, is the sample autocovariance function of the process X_t in (2.1).

When considering only the frequencies close to zero, the term $\ln\{\frac{f_U(w_j)}{f_U(0)}\}$ may be discarded. Then, we may rewrite (3.2) in the context of a simple linear regression model:

$$y_i = a - d\,z_i + e_i\ , \quad i = 1, \cdots, m \tag{3.4}$$

where $m = [n/2]$, $(a, -d)$ are the regression coefficients, $a = \ln f_U(0)$, $y_i = \ln I(\omega_i)$, $z_i = \ln\{2\sin(\omega_i/2)\}^2$, and the errors $e_i = \ln\{\frac{I(w_i)}{f_X(w_i)}\}$ are noncorrelated random variables centered at zero with constant variance.

We recall that when Y_t follows a FISV process with $d \in (-0.5, 0.5)$, $\ln(Y_t^2)$ is the sum of a zero mean Gaussian ARFIMA process and independent non-Gaussian innovation process. Also, the FIGARCH(r, d, s) process, $d \in (0, 1)$, has been defined in expression (8) of Baillie et al. (1996) as an ARFIMA process on the squared data with a more complicated error structure. Thus, the regression based method also applies to these processes.

A semiparametric regression estimator may be obtained by minimizing some loss function of the residuals $r_i = y_i - a + d\,z_i$. We will consider three different loss functions. They give rise to the classical Ordinary Least Squares method (OLS), and two high breakdown point robust methods, the Least Trimmed Squares method (LTS), and the the MM-estimation method.

The OLS estimators are the values $(\hat{a}, -\hat{d})$ which minimize the loss function

$$L_1(m) = \sum_{i=1}^{m}(r_i)^2, \tag{3.5}$$

where $r_i = y_i - a + d\,z_i$ is the residual related to the regression (3.4).

Whenever the errors e_i follow a normal distribution, the OLS estimates have the minimum variance among all unbiased estimates (see Rao, 1973). If the errors follow another distribution (as in the cases considered here), non-linear estimates may possess better statistical properties. In fact, it is well known (see Huber, 1981) that regression outliers, leverage points, and gross errors are responsible for considerable bias and inefficiency (even in the Gaussian environment) in the OLS estimates.

How biased an estimate can become at the presence of outliers and leverage points can be measured by the value of its breakdown point. Loosely speaking, the breakdown point of an estimator represents the smallest proportion of atypical points in the sample that makes the estimates meaningless, that is, estimates providing distorted information about the parameters being estimated. The OLS estimator has zero breakdown point, meaning that just one spurious observation is able to completely distort the OLS estimator.

Robust alternatives to OLS may be obtained by minimizing a robust version of the dispersion of the residuals. The Least Trimmed Squares (LTS) estimates of Rousseeuw (1984) minimize the loss function

$$L_2(m) = \sum_{i=1}^{m^*} (r^2)_{i:m} \ , \tag{3.6}$$

where $(r^2)_{i:m}$ are the squared and then ordered residuals, that is, $(r^2)_{1:m} \le ... \le (r^2)_{m:m}$, and m^* is the number of points used in the optimization procedure. The constant m^* is responsible both for the breakdown point value and efficiency. When m^* is approximately $m/2$ the breakdown point is approximately 50%. The LTS estimates have been previously used by Taqqu, Teverovsky, and Willinger (1995) for the estimation of the long range parameter in ARFIMA models.

The MM-estimates (see Yohai, 1987) may possess simultaneously high breakdown point and high efficiency. They are defined as the solution $(\hat{a}, -\hat{d})$ which minimizes the loss function

$$L_3(m) = \sum_{i=1}^{m} \rho_2 \left(\frac{r_i}{s}\right)^2 \ , \tag{3.7}$$

subject to the constraint

$$\frac{1}{m} \sum_{i=1}^{m} \rho_1 (\frac{r_i}{s}) \le b \ , \tag{3.8}$$

where ρ_2 and ρ_1 are symmetric, bounded, nondecreasing on $[0, \infty)$ with $\rho_i(0) = 0$ and $\lim_{u \to \infty} \rho_i(u) = 1$, $i = 1, 2$, s is a scale parameter, and b is a tuning constant. The breakdown point of the MM-estimator only depends on ρ_1 and it is given by $\min(b, 1 - b)$.

The two robust methods chosen possess appealing definitions, well established asymptotic properties, and can be rapidly computed using the SPlus software. The only references we are aware of on robust estimation of the long memory parameter are Beran (1994), Agostinelli and Bisaglia (2004), and the already cited Taqqu, Teverovsky, and Willinger (1995). All of them considered just ARFIMA processes. Agostinelli and Bisaglia (2004) approach differs from ours since they propose a robustification of the maximum likelihood functions. Figure 1 illustrates the role of a robust estimate and the data type we are dealing with.

The data used in Figure 1 is from a simulated FISV process with $d = 0.30$. Log-periodogram data typically shows a considerable amount of large z_i values, $i = 1, ..., [n/2]$. These are the values related to frequencies away from zero and, therefore, those less relevant in the estimation process. However, they may have a large influence on the fits. The left hand side of Figure 1 illustrates this fact, and show the effect of the large amount of large z_i values on the classical OLS estimate (black) and the robust LTS (red) slope estimates of (3.4), both based on (3.3).

According to the theory, the more influent points should be those associated to the smaller z_i values. This suggests trimming the points associated to large frequencies, technique implemented at the right hand side of Figure 1, where we

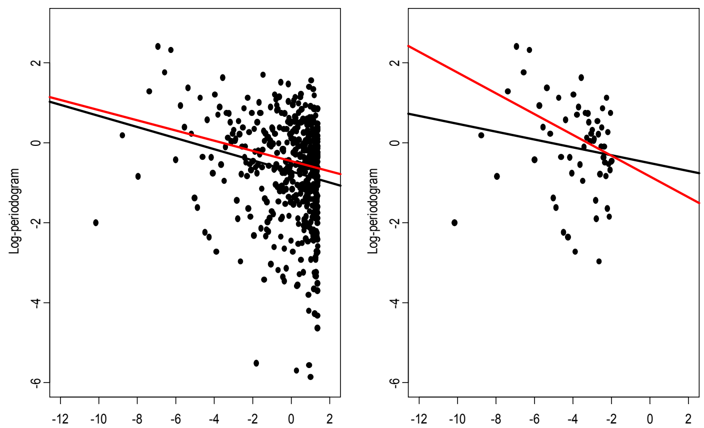

Figure 1: *The OLS (black) and the LTS (red) estimates based on periodogram (3.3). The left hand side uses all* $[n/2]$ *data points, and the right hand side uses half of the data points.*

use just half of the data. Now, we can see that the robust LTS procedure (in red) provides an estimate close to the true value. However, some points still tilt the classical OLS regression line (in black), distorting the slope estimate, resulting in an under-estimation of d.

Thus, a critical issue is how many (m) frequencies should be used by the regression type estimators. The choice of m affects the estimators properties, such as unbiasedness and efficiency. We address this issue in Section 4. By considering variations of (3.3), periodogram based methodologies have been proposed. In the following subsections we summarize the most important ones.

3.1 Classical and robust GPH estimators

The first estimation method based on the periodogram function was proposed by Geweke and Porter-Hudak (1983). To obtain an estimate for d, these authors proposed applying the Ordinary Least Squares method in (3.4) based on (3.3). The classical $GPH\text{-}LS$ estimator of d is then given by

$$GPH\text{-}LS = -\frac{\displaystyle\sum_{j=1}^{g(n)}(z_j - \bar{z})(y_j - \bar{y})}{\displaystyle\sum_{j=1}^{g(n)}(z_j - \bar{z})^2}, \tag{3.9}$$

where the trimming value $g(n)$ is usually $g(n) = n^{\alpha}$, $0 < \alpha < 1$, y_j is based on (3.3), and z_j is as previously defined. Lopes et al. (2004) considered α in the interval $[0.55, 0.65]$, and Porter-Hudak (1990) considered $\alpha \in \{0.62, 0.75\}$ for the case of seasonal fractionally integrated time series data.

Robinson (1995) established consistency properties of semiparametric estimators of the long memory parameter, including the GPH, within the context of ARFIMA models. He also provided an asymptotic distribution theory for any value of d under mild conditions. Based on results in Andersen and Bollerslev (1997) and Robinson (1999), Bollerslev and Wright (2000) argue that log-periodogram estimates calculated from the log-squared, squared, and absolute data may be considered to be consistent.

To obtain the robust versions of the GPH estimator we just apply the LTS and the MM methodologies to the regression model (3.4) with $m = n^\alpha$, based on (3.3). This gives rise to the GPH-LTS and the GPH-MM estimators. For the GPH and all other regression based estimators that follow, we will investigate the effect of $\alpha \in [0.50, 0.86]$ on the estimates bias and variance.

3.2 Classical and robust SPR estimators

As shown in Brockwell and Davis (1991), the periodogram function is not a consistent estimator for the spectral density function. Reisen (1994) proposed using a consistent estimator for the spectral density function, which is a smoothed version of the periodogram function (3.3), the SPR estimator.

More specifically, the regression estimator SPR is obtained by replacing the spectral density function in the expression (3.1), by the smoothed periodogram function, denoted by $I_s(\cdot)$, given by

$$I_s(\omega) = \frac{1}{2\pi} \sum_{j=-\nu}^{\nu} \kappa\left(\frac{j}{\nu}\right) \hat{\gamma}_X(j) \cos(j\omega), \tag{3.10}$$

where $\kappa(\cdot)$ is the Parzen lag window given by

$$\kappa(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & \text{if } |u| \leq \frac{1}{2}, \\ 2(1-|u|)^3, & \text{if } \frac{1}{2} < |u| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.11}$$

The SPR estimator proposed by Reisen (1994) is obtained by applying the OLS procedure to the regression model (3.4) based on (3.10) and (3.11). We call these estimates the SPR-LS. The truncation point in the Parzen lag window is defined by $\nu = n^\beta$, $0 < \beta < 1$. Here, we consider $\beta = 0.9$ (see Reisen, 1994 for a discussion on the value of β). Again, the robust versions may be obtained by applying the LTS and the MM methodologies to (3.4) based on (3.10) and (3.11), producing the SPR-LTS and the SPR-MM.

3.3 Classical and robust BA estimators

By considering the Bartlett lag window, another consistent estimator for the spectral density function may be obtained. This spectral window will provide a smoothed version of the periodogram function (3.10), where now the function $\kappa(\cdot)$ is defined as

$$\kappa(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.12}$$

The classical and robust versions are obtained by applying the *OLS*, the *LTS* and the *MM* methodologies to the regression model (3.4) based on (3.10) and (3.12), producing the *BA-LS*, the *BA-LTS*, and the *BA-MM* estimators. The value of m in (3.4) is again given by n^{α}, and the truncation point ν is set equal to 30 (see Bollerslev and Wright, 2000).

3.4 Classical and robust R estimators

The regression estimator R, proposed by Robinson (1995) is obtained by applying the Ordinary Least Squares method in (3.4) based on (3.3), but considering only the frequencies $i \in \{l, l+1, \cdots, g(n)\}$, where $l > 1$ is a trimming value that tends to infinity more slowly than $g(n)$.

It is interesting to compare the R and the *LTS* concepts. The R concept trims the extreme z_j values associated with the frequencies close to zero, which we know are the important ones. On the other hand, the *LTS* concept trims the extreme ordered residuals which may or may be not be associated to small frequencies, but certainly are associated to leverage points. In other words, the *LTS* procedure identifies which data points associated with a small frequencies are outliers and, if they exist, excludes them from the calculations. The *R-LTS* and *R-MM* versions are obtained by applying the robust methodologies, as previously.

3.5 Classical and robust *GPHT* estimators

The *GPHT* method (see Hurvich and Ray (1995) and Velasco (1999b)) uses a modified periodogram function given by

$$I(\omega_j) = \frac{1}{\sum_{t=0}^{n-1} g(t)^2} \left| \sum_{t=0}^{n-1} g(t) X_t e^{-i\omega_j t} \right|^2, \tag{3.13}$$

where the tapered data is obtained from the cosine-bell function

$$g(t) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(t+0.5)}{n}\right)\right]. \tag{3.14}$$

We obtain the classical *GPHT-LS* and the robust versions *GPHT-LTS* and *GPHT-MM* by applying the classical and the robust methodologies on model (3.4) based on (3.13) and (3.14), and setting $m = n^{\alpha}$.

3.6 Classical W estimator

The W estimator was proposed by Whittle (see Whittle, 1953). He considered the function

$$Q(\eta) = \int_{-\pi}^{\pi} \frac{I(\omega)}{f_X(\omega;\eta)} dw,$$

where η denotes the vector of unknown parameters, and $f_X(\cdot\,;\eta)$ is the spectral density function of $\{X_t\}_{t\in\mathbb{Z}}$, given by (3.1).

The W estimator is the value of η which minimizes the function $Q(\cdot)$. Here $\eta = d$. The estimation procedure is carried out by essentially minimizing

$$\mathcal{L}_n(\hat{\eta}) = \sum_{j=1}^{[\frac{n-1}{2}]} \frac{I(\omega_j)}{f_X(\omega_j;\hat{\eta})}. \tag{3.15}$$

Robinson and Zaffaroni (1998) proposed a version of the Whittle estimator for squared FIGARCH data, deriving a formula for the spectral density and autocovariance function. The estimator is not asymptotically efficient, but Zaffaroni (1999) has developed a central limit theorem distributional result. More details of this estimator can be found in Fox and Taqqu (1986). Differently from all other estimators considered, the W estimator is in the parametric class.

4 Assessing the Estimators Performances

In this section we assess the performance of 301 estimators when estimating the fractional parameter d in 31 specifications of ARFIMA(p,d,q)-FIGARCH(r,d,s) processes. The same notation d for the fractional parameter in the mean and volatility specification will not cause any confusion, because the models to be considered possess either long memory in mean or in volatility.

To simulate the data and to compute the estimates we used the S language and SPlus programs. We hold fixed the following specifications:

- For each model considered the number of replications S is 300. All series have length $n = 1000$.
- In all FIGARCH models considered $w = 0.10$.
- The trimming constant l in the R estimator is fixed equal to 3.
- The constant $\nu = n^{\beta}$ for the SPR estimators is found by putting $\beta = 0.90$. Since $n = 1000$, $\nu = 501.19$.
- Both loss functions ρ_i, $i = 1, 2$, for the MM-estimator are chosen as the Tukey Biweighted function (see Yohai, 1987). They are tuned such that the resulting estimates possess 0.50 breakdown point and an efficiency of 85% at the normal model.

We compute a total of 300 regression type (semiparametric) estimators. They are obtained by varying the estimation method and the value of α for the 5 regression type estimators GPH, SPR, BA, R and $GPHT$. We consider 19 possibilities for the trimming constant α used to define $m = g(n) = n^{\alpha}$. Specifically, we set $\alpha \in \{0.50, 0.52, ... \ 0.84, 0.86\}$. The version not tunned by α, that is, based on the $[n/2]$ data points is also computed, and it is equivalent to set $\alpha = 0.8997$. Thus m varies through a fairly wide range, between 31.6 and 500. For each resulting estimator we consider the 3 estimation methods: classical LS, and robust LTS and MM. We also compute the parametric W estimator.

We consider a total of 31 different data generating processes (DGP). They are 18 combinations of ARMA and FIGARCH processes ($M1$ to $M18$); 6 FISV models ($M19$ to $M24$); and 6 ARFIMA processes ($M25$ to $M31$). We give next their notations and detailed specifications. The notation t_4 means a t-student distribution with 4 degrees of freedom. ϕ and θ are the autoregressive and moving average parameters in the ARFIMA part. α_1 and β_1 correspond to the autoregressive and moving average parameters in the FIGARCH part.

$M1$: ARFIMA$(0,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.N(0,1)$.

$M2$: ARFIMA$(0,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.t_4(0,1)$.

$M3$: ARFIMA$(1,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\phi = 0.50$

$M4$: ARFIMA$(1,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.t_4(0,1)$, $\phi = 0.50$.

$M5$: ARFIMA$(0,0,0)$-FIGARCH$(1,0.50,0)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = -0.20$.

$M6$: ARFIMA$(0,0,0)$-FIGARCH$(1,0.50,0)$ process, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = -0.20$.

$M7$: ARFIMA$(0,0,1)$-FIGARCH$(1,0.50,0)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = -0.20$, $\theta = 0.50$.

$M8$: ARFIMA$(0,0,1)$-FIGARCH$(1,0.50,0)$ process, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = -0.20$, $\theta = 0.50$.

$M9$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.75,1)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M10$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.75,1)$ process with, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M11$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.50,1)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M12$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.50,1)$ process with, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M13$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.25,1)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M14$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.25,1)$ process with, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = -0.20$, $\beta_1 = 0.20$, $\phi = 0.20$, $\theta = 0.20$.

$M15$: ARFIMA$(0,0,0)$-FIGARCH$(1,0.00,1)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = 0.15$, $\beta_1 = 0.80$.

$M16$: ARFIMA$(0,0,0)$-FIGARCH$(1,0.00,1)$ process, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = 0.15$, $\beta_1 = 0.80$.

$M17$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.00,1)$ process, with $Z_t \sim i.i.d.N(0,1)$, $\alpha_1 = 0.15$, $\beta_1 = 0.80$, $\phi = 0.50$, $\theta = 0.50$.

$M18$: ARFIMA$(1,0,1)$-FIGARCH$(1,0.00,1)$ process, with $Z_t \sim i.i.d.t_4(0,1)$, $\alpha_1 = 0.15$, $\beta_1 = 0.80$, $\phi = 0.50$, $\theta = 0.50$.

$M19$: FISV$(1, 0.30, 0, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$, $\sigma_\varepsilon = 0.3$.

$M20$: FISV$(1, 0.30, 0, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.t_4(0,1)$, $\phi = 0.60$, $\sigma_\varepsilon = 0.3$.

$M21$: FISV$(0, 0.30, 1, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.N(0,1)$, $\theta = 0.70$, $\sigma_\varepsilon = 0.3$.

$M22$: FISV$(0, 0.30, 1, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.t_4(0,1)$, $\theta = 0.70$, $\sigma_\varepsilon = 0.3$.

$M23$: FISV$(1, 0.30, 1, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$, $\theta = 0.70$, $\sigma_\varepsilon = 0.3$.

$M24$: FISV$(1, 0.30, 1, \sigma_\varepsilon)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\varepsilon_t \sim i.i.d.t_4(0,1)$, $\phi = 0.60$, $\theta = 0.70$, $\sigma_\varepsilon = 0.3$.

$M25$: ARFIMA$(1, 0.45, 0)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$.

$M26$: ARFIMA$(1, 0.30, 0)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$.

$M27$: ARFIMA$(0, 0.45, 1)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\theta = 0.90$.

$M28$: ARFIMA$(0, 0.30, 1)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\theta = 0.90$.

$M29$: ARFIMA$(1, 0.45, 1)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$, $\theta = 0.70$.

$M30$: ARFIMA$(1, 0.30, 1)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$, $\theta = 0.70$.

$M31$: ARFIMA$(1, 0.00, 1)$-FIGARCH$(0, 0, 0)$ process, with $\epsilon_t \sim i.i.d.N(0,1)$, $\phi = 0.60$, $\theta = 0.70$.

When estimating d in volatility models, some authors had used the absolute, the log-squared, or squared data (see Bollerslev and Wright, 2000) as volatility measures. Ding and Granger (1996) define the long memory property of ARCH models as the limiting case of a model with N volatility components, a GARCH(N, N) model, as $N \to \infty$. This model displays the long range memory in powers of the absolute data. Based on these considerations, we use here the absolute data to estimate d in the FIGARCH processes. To estimate d in the FISV processes we used the log squared data, as in Breidt et al. (1998) and Bollerslev and Wright (2000). An issue not touched in the present paper is the sensitivity of estimators to series lengths or to the choice of the volatility measure.

Let d_0 represent the parameter d true value in each model. For each estimator $\widehat{d^j}$, $j = 1, \cdots, 301$, the following statistics were computed to summarize its simulated probability distribution:

- The mean bias: for each $\widehat{d^j}$ we compute $B^j = \frac{1}{S}\sum_{i=1}^{S}(\widehat{d_i^j} - d_0)$;
- The median bias: for each $\widehat{d^j}$ we compute $B_M^j = \text{Median}_i(\widehat{d_i^j} - d_0)$;
- The sample standard deviation sd^j: for each $\widehat{d^j}$ we compute the square root of $V^j = \frac{1}{S-1}\sum_{i=1}^{S}(\widehat{d_i^j} - \bar{\widehat{d^j}})^2$, where $\bar{\widehat{d^j}}$ is the arithmetic mean of the S $\widehat{d_i^j}$;
- The 0.90% percentile confidence interval: for each $\widehat{d^j}$ we compute the $CI^j = [q_{0.05}^j, q_{0.95}^j]$, where q_p^j is the empirical p-quantile of estimator $\widehat{d^j}$.

For each model the following criteria were used to find out the best estimator:

- *C1:* Find the $\widehat{d^j}$ for which the value of ${B^j}^2 + V^j$ is minimum.
- *C2:* Find the $\widehat{d^j}$ for which the value of $|B_M^j| + ||CI^j||$ is minimum. Here, the notation $|B_M^j|$ means the absolute value of B_M^j, and $||CI^j||$ means the lenght of CI^j, that is, $q_{0.95}^j - q_{0.05}^j$.

For a given model and each criterion, the estimators are ranked and the 3 best ones are recorded. By noting that there is little difference among the criteria values obtained for the three highest ranked competitors, we decided to choose as the overall winner the one (or the ones) selected by both criteria, despite its position. In the case of ties, both (or the three) estimators are reported. In addition, in the case that all six positions are occupied by different estimators, the winners under $C1$ and $C2$ are reported. In what follows we summarize the results for each model considered.

4.1 Simulations results

We will provide a detailed analysis of the results from models $M1$ and $M2$, and then summarize the results from the other models. In the tables and figures that follow, whenever the value for α used is the maximum possible, we report $[n/2]$ (in the tables) or nothing (in the figures) instead.

Results from model $M1$: ARFIMA$(0,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.N(0,1)$. Figure 2 illustrates how difficult is choosing an optimality criterion, due to the trade off bias-variance. This figure shows the simulated distribution of the three best estimators according to the following set up. In the first row we show the estimators possessing smaller $|B^j|$ (left), and the estimators possessing smaller $|B^j_M|$ (right). In the second row, we show the estimators possessing smaller standard deviation sd^j on the left, and those presenting the smaller confidence interval length on the right hand side. Finally, the third row shows the winners from criteria *C1* (left) and *C2* (right).

Table 1: $M1$*: Three best results under the all criteria used, and overall winner.*

Criterion	1st. Estimator(α)	2nd. Estimator(α)	3rd. Estimator(α)	Winner(α)
	$GPHT.LS(0.50)$	$BA.MM(0.52)$	$GPHT.LS(0.52)$	$GPHT.LS(0.58)$
abs(B^j)	0.0067	0.0129	0.0136	0.0627
	$GPHT.LS(0.50)$	$GPHT.LS(0.52)$	$GPHT.MM(0.50)$	
abs(B^j_M)	0.0089	0.0131	0.0140	0.0796
	$SPR.LS$([n/2])	$SPR.LS(0.86)$	$BA.LS$([n/2])	
sd^j	0.0518	0.0539	0.0541	0.1186
	$SPR.LS$([n/2])	$BA.LS$([n/2])	$SPR.LS(0.86)$	
$\|\|CI^j\|\|$	0.1650	0.1704	0.1746	0.3945
	$GPHT.LS(0.58)$	$GPHT.LS(0.56)$	$GPHT.LS(0.54)$	
$C1$	0.0179	0.0182	0.0187	0.0179
	$SPR.LS$([n/2])	$GPHT.LS$([n/2])	$BA.LS$([n/2])	
$C2$	0.3232	0.3297	0.3306	0.4741

The results shown in Figure 2 are given in detail in Table 1. As we can see, for model $M1$ there is no overall winner. However the GPHT estimator shows up more frequently. It seems that if the primary concern is just bias, one should use either the classical or the robust (MM) $GPHT$ with small α values, close to 0.50. If a small variability is more important, then one should move to the classical SPR or BA. However, as Figure 2 illustrates, these low standard deviation estimates

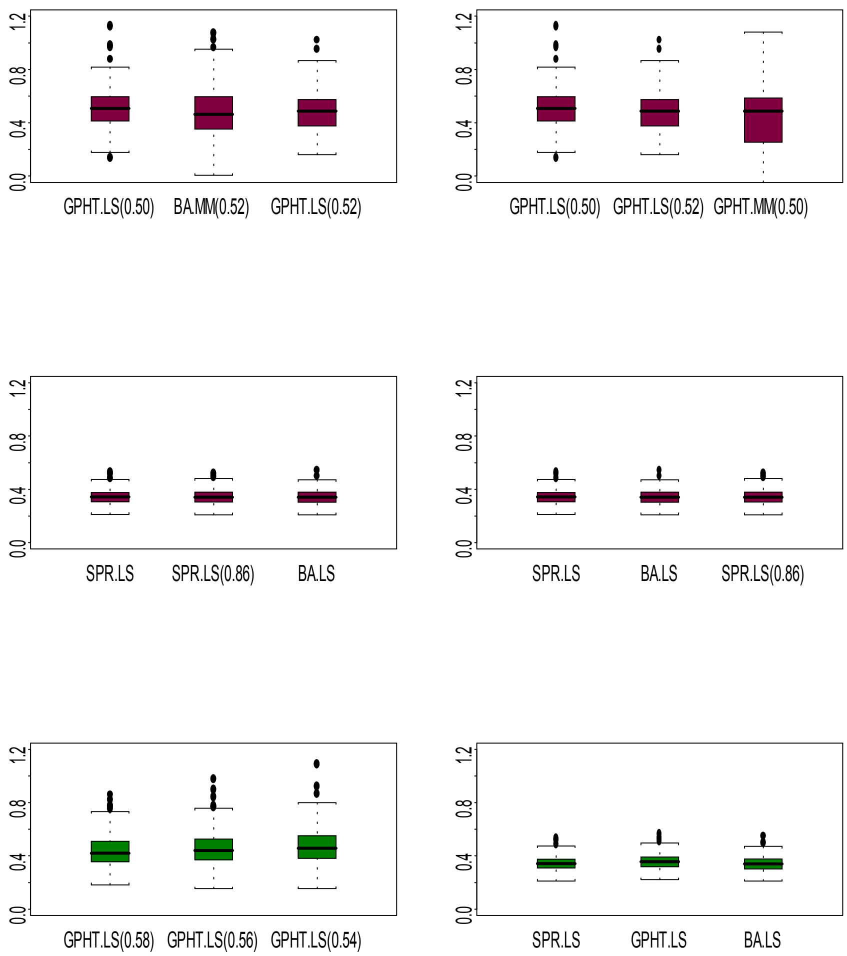

Figure 2: *The simulated distributions of the three best estimators from model $M1$. In the first row, we show the estimators possessing smaller $|B^j|$ at the left hand side, and those possessing smaller $|B^j_M|$ at the right hand side. In the second row, we show the estimators possessing smaller standard deviation sd^j on the left, and those presenting the smaller confidence interval length on the right. Finally, the third row shows the winners from criteria C1 (left) and C2 (right).*

may possess an unacceptable large bias. When we combine an accuracy measure and a variability measure, the $GPHT$ shows up 4 times. We decided to choose the classical $GPHT.LS$ with $\alpha = 0.58$ as the overall winner.

Table 2: $M2$: *Three best results under the all criteria used, and overall winner.*

Criterion	1st. Estimator(α)	2nd. Estimator(α)	3rd. Estimator(α)	Winner(α)
	$BA.MM(0.54)$	$BA.MM(0.56)$	$BA.MM(0.52)$	$BA.LTS$([n/2])
abs(B^j)	0.1215	0.1240	0.1307	0.1606
	$BA.MM(0.54)$	$BA.MM(0.56)$	$BA.MM(0.58)$	
abs(B^j_M)	0.1384	0.1443	0.1531	0.1727
	$BA.LS(0.86)$	$BA.LS(0.84)$	$BA.LS$([n/2])	
sd^j	0.0600	0.0601	0.0606	0.0867
	$BA.LS(0.86)$	$BA.LS$([n/2])	$SPR.LS$([n/2])	
$\|\|CI^j\|\|$	0.1876	0.1888	0.1899	0.2573
	$BA.LTS$([n/2])	$BA.LTS(0.82)$	$BA.LTS(0.84)$	
$C1$	0.0333	0.0340	0.0340	0.0333
	$SPR.LS$([n/2])	$BA.LS$([n/2])	$R.LTS$([n/2])	
$C2$	0.3811	0.3913	0.3943	0.4300

Results from model $M2$: ARFIMA$(0,0,0)$-FIGARCH$(0,0.50,0)$ process, with $Z_t \sim i.i.d.t_4(0,1)$. It is impressive the excellent performance of the BA-estimator in the second experiment. See Table 2 and Figure 3. When bias is the concern, the robust $BA.MM$ estimator tuned with small α values turns out the best option. When variability is taken into account, the classical $BA.LS$ estimator with large α values, or no-α are the winners. The trade off between bias and variance result in the choice of the robust $BA.LTS$ estimator tuned with a large α or no-α (three winners under $C1$). We take as the overall winner the robust $BA.LTS$, the best under *C1*, which does not need any tuning constant, a very interesting result. We must note though, that for this model all estimators presented a (probably unacceptable) large bias.

The impressive trade off bias-variance observed for models $M1$ and $M2$ is illustrated in Figure 4. This figure shows the winners under models $M1$ (left, the $GPHT.LS$ with $\alpha = 0.58$) and $M2$ (right, the $BA.LTS$ using all $[n/2]$ frequencies) plotted as functions of the tuning constant α. The triangles represent the biases, and the diamonds represent the variances of the 20 estimators. The circles point out the final choices.

Summary of results from other models considered. Table 3 summarizes the results from all data generating processes considered and specified in the first column (DGP). For each model, we report the winners under $C1$ and $C2$, unless there is an overall winner. The second column names the winner(s), and the third one provides its (their) α value. When all $[n/2]$ frequencies are used we report $[n/2]$ instead of α. The fourth to ninth columns provide the criteria values attained by the winner(s).

We first note in Table 3 the unacceptable large bias of all estimators winning under criterion $C2$. We thus continue our analysis considering just the winner chosen by $C1$.

For estimating d in the processes $M1$ to $M4$ (those possessing $\alpha_1 = \beta_1 = 0$), we would select the robust $BA.LTS$ using all $[n/2]$ frequencies, even though it was not an option for model $M1$. When a autoregressive part is included in the volatility dynamics (models $M5$ to $M8$), the $GPHT.LS$ tuned with small α values may be considered the best choice.

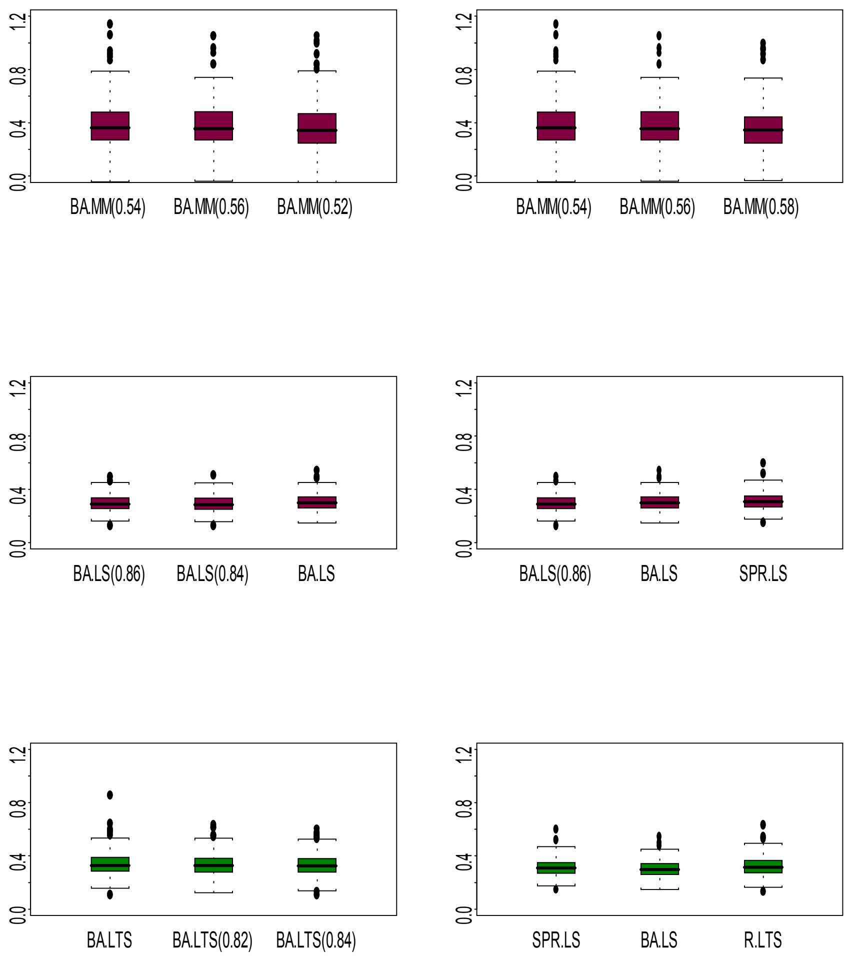

Figure 3: *The simulated distributions of the three best estimators from model $M2$. In the first row, we show the estimators possessing smaller $|B^j|$ at the left hand side, and those possessing smaller $|B^j_M|$ at the right hand side. In the second row, we show the estimators possessing smaller standard deviation sd^j on the left, and those presenting the smaller confidence interval length on the right. Finally, the third row shows the winners from criteria C1 (left) and C2 (right).*

The robust $BA.LTS$ using all $[n/2]$ frequencies wins again for models $M9$ and $M10$, which possess the autoregressive and moving average components for the mean and volatility, as well as strong long memory, $d = 0.75$. The robust estimator also wins for models $M15$ and $M16$, another extreme situation (no long memory), actually pure GARCH(1,1) models. Even though the $BA.LTS$ shows large bias, the t-test does not reject the null hypothesis $d = 0$.

However, when the short range effects are included to these GARCH(1,1) pro-

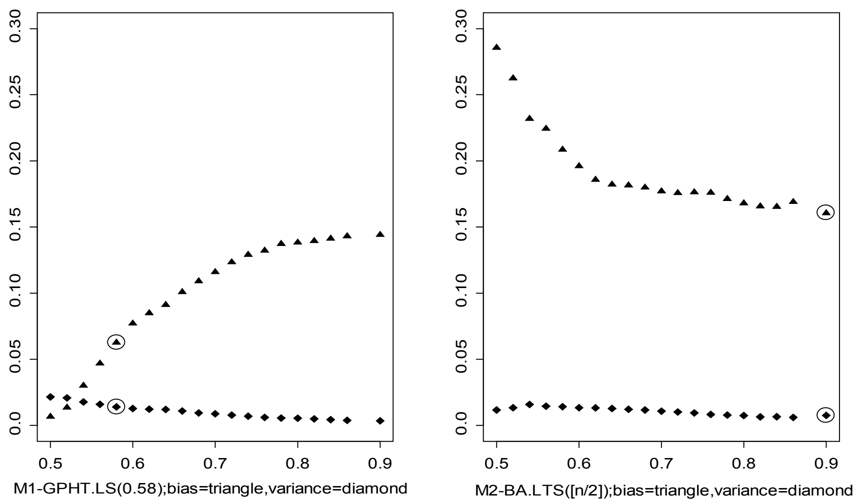

Figure 4: *The trade-off bias-variance for models $M1$ and $M2$. Robust estimator does not need tuning.*

cesses, giving rise to models $M17$ and $M18$, the winner becomes the classical $BA.LS$ tuned with the smaller α value.

In summary, for FIGARCH processes, simulations indicate that we should use the $GPHT.LS$ tuned with small α values (around 0.60) for series presenting no short range dependence in the mean, or the $BA.LS$ tuned with moderate α values (around 0.82), for series possessing short range dependence in the mean, as in models $M13$ and $M14$. Alternatively, if one suspects of strong long range dependence, one may let the robust procedure LTS automatically select the frequencies for trimming.

The FISV processes are models $M19$ to $M24$. For those processes generated according to a simple autoregressive process, the winner is the classical $SPR.LS$ estimator based on moderate α values $(0.82-0.86)$. When both short range effects are included, the best option is the classical $GPHT$ estimator tuned with small α values $(0.58-0.60)$.

When it comes to ARFIMA models ($M25$ to $M31$) and classical estimation, it seems that it is very important to use just few frequencies, setting $\alpha = 0.50$. Then either the $GPHT$ or the BA estimator may be used. If the robust estimation procedure LTS is applied, then one may use the BA estimator tuned with larger α values, say $\alpha = 0.66$.

None of the experiments resulted in a winner type R-estimator. This is in line with Deo and Hurvich (2003) remark that when computing the GPH estimator it is crucial for the finite sample performance of this estimator (which may also be true for all regression type estimators) that the lowest frequencies not be dropped.

5 Real Data

In this section we provide an illustration using an emerging market returns series. The data consist of 2608 observations of the Taiwan daily index returns from January, 3, 1994 to December, 31, 2003. This period includes examples of extreme market events such as the Asian series of devaluation during 1997. Crises in East Asian economies usually result in considerable depreciations of national currencies and have important global repercussions. Taiwan is the largest emerging market, with a total market capitalization of US\$ 379 billion, followed by Korea (US\$ 298 billion) and India (US\$ 252 billion).

According to our simulations results (winners from models $M13$ and $M14$) we estimate d computing the classical $BA.LS(0.82)$, which yields the value 0.1706.

According to Taqqu and Teverovsky (1996), we should examine the plot in Figure 5 to choose the best α value. We do that for the $BA.LS$ (triangles) and the $BA.LTS$ (diamonds). There is some indication of flatness from 0.64 to 0.74 for the classical, and from 0.60 to 0.68 for the robust. The Taqqu and Teverovsky (1996) estimates could then be, respectively, $BA.LS(0.70) = 0.2045$ and $BA.LTS(0.64) = 0.2966$. The classical value is slightly larger than our result. We should note, however, that this graphical procedure, though very interesting, is clearly subjective, and could not be used within a more complex decision based procedure.

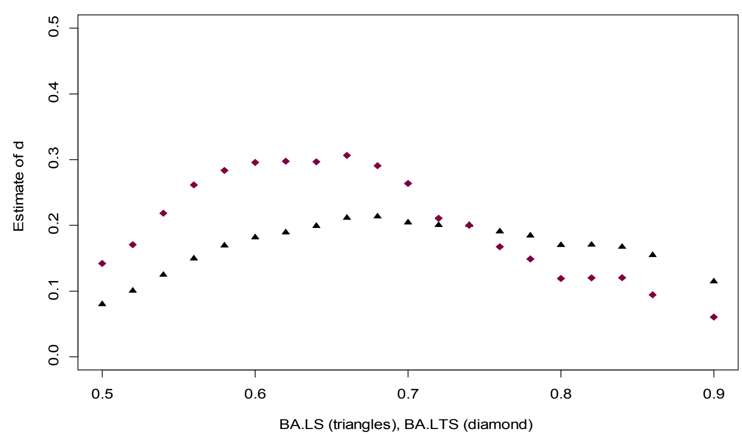

Figure 5: *BA.LS and BA.LTS estimates of d plotted as functions of alpha.*

To complete this analysis, we fit a fully parametrized model to the Taiwan daily returns. To model the serial dependence in the mean and variance of the daily returns we consider all combinations of ARMA(p, q) and FIGARCH (r, d, s) processes derived from setting $p = 0, 1, 2$, $q = 0, 1, 2$, $r = 0, 1, 2$, and $s = 0, 1, 2$. Models are estimated by maximum likelihood using the FinMetrics module of SPlus, and the AIC criterion is used to select the best model. The best fit turned out to be an ARMA$(1, 1)$-FIGARCH$(1, d, 0)$ with all parameters estimates highly significant, see Table 5. Note $d = 0.2566$ which is half way the classical and robust estimates indicated by the graphical analysis.

Actually, none of the models used in the simulations possesses the specification ARMA$(1,1)$-FIGARCH$(1,d,0)$ found for Taiwan. Thus we carried out another simulation assuming this model found by the fully parametric approach, setting as true values those given in Table 5, i.e., $d = 0.26$, $\alpha_1 = -0.20$, $w = 0.47$, $\theta = -0.54$, and $\phi = 0.58$. The same 301 estimators were used and the winner according to criterion $C1$ was the classical $BA.LS(0.82)$ (absolute bias = 0.0107, and standard error equal to 0.0407).

We note that using the standard error 0.0407 from the simulations and the point estimate 0.1706 we compute a confidence interval of $[0.0892, 0.2520]$ for the semiparametric estimator, which we compare with the fully parametrized maximum liklihood estimator confidence interval of $[0.2154, 0.2978]$.

Table 4: ARMA$(1,1)$-FIGARCH$(1,d,0)$ fit to daily returns from Taiwan.

	Estimate	Std.Error	t value	$Pr(>\|t\|)$
ϕ	0.5835	0.26251	2.223	1.316e-002
θ	-0.5446	0.27165	-2.005	2.254e-002
w	0.4721	0.04669	10.111	0.000e+000
α_1	-0.1983	0.02487	-7.974	1.110e-015
d	0.2566	0.02062	12.446	0.000e+000

6 Conclusions

Semiparametric methods seem to be very suitable for empirical anaysis of long memory in volatility, specially because the high complexity of fully parametric approach based on the joint modeling of volatility and mean. However, care is needed when using semiparametric regression type estimators, as their statistical properties also depend on a bandwith value. Additional complications arise from the lack of robustness of the least squares estimation methodology. In this paper we adreessed the issue of tunig semiparametric estimates in order to balance their bias and variance. We considered models with long memory in mean (ARFIMA) and in the volatility (FIGARCH and FISV processes), with innovations following either a Gaussian or a t-student distribution.

A result from the simulations is that the best number m of frequencies to be used (or best α value) is completely dependent on the data generating process. For the same FIGARCH specification, different models for the conditional mean will lead to a different tuning choice. Another conclusion is that the range $[0.50, 0.86]$ for specifying α seems to be adequate.

In summary, for FIGARCH processes, simulations indicate that we should use the $GPHT.LS$ tuned with small α values (around 0.60) for series presenting no short range dependence in the mean, and the $BA.LS$ tuned with moderate α values (around 0.82), for series possessing short range dependence in the mean. Alternatively, if one suspects of strong (or no) long range dependence, one may let the robust procedure LTS automatically select the frequencies for trimming.

When the FISV processes are generated based on a simple autoregressive process, the winner is the SPR-estimator, and the corresponding classical regression procedure does need too much trimming. When both short range effects are included, the best option is the classical $GPHT$ estimator tuned with small α values.

Best results for data following ARFIMA processes seem to be those based on classical estimation using just few frequencies, setting $\alpha = 0.50$. Then either the $GPHT$ or the BA estimator may be used. If the robust estimation procedure LTS is applied, then one may use the BA estimator tuned with larger α values, say $\alpha = 0.66$.

The slightly less convining results from the robust estimators do not eliminate their usefulness in this environment. It is possible that better results are obtained if smaller breakdown point versions are used, which would warrant a more efficient procedure.

Table 3: *Summary of results from all models.*

DGP	Winner(s)	α	abs(B^j)	abs(B^j_M)	sd^j	$\|\|CI^j\|\|$	$C1$	$C2$
$M1$	*GPHT.LS*	0.58	0.0627	0.0796	0.1186	0.3945	0.0179	0.4741
$M1$	*SPR.LS*	$[n/2]$	0.1557	0.1582	0.0518	0.1650	0.0269	0.3232
$M2$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.1606	0.1727	0.0867	0.2573	0.0333	0.4300
$M2$	*SPR.LS*	$[n/2]$	0.1886	0.1913	0.0623	0.1899	0.0395	0.3811
$M3$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.0680	0.0859	0.0799	0.2647	0.0110	0.3506
$M3$	*SPR.LS*	$[n/2]$	0.1219	0.1256	0.0523	0.1691	0.0176	0.2947
$M4$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.0314	0.0492	0.1066	0.3370	0.0123	0.3862
$M4$	*SPR.LS*	$[n/2]$	0.1111	0.1143	0.0615	0.2000	0.0161	0.3142
$M5$	*GPHT.LS*	0.64	0.0682	0.0712	0.1083	0.3519	0.0164	0.4231
$M5$	*SPR.LS*	0.82	0.1657	0.1706	0.0589	0.1879	0.0309	0.3585
$M6$	*GPHT.LS*	0.60	0.0953	0.1089	0.1239	0.3901	0.0244	0.4989
$M6$	*GPHT.LS*	0.78	0.1510	0.1591	0.0883	0.2640	0.0306	0.4231
$M7$	*GPHT.LS*	0.58	0.0570	0.0567	0.1233	0.4191	0.0184	0.4758
$M7$	*W*	—	0.1983	0.2046	0.0450	0.1391	0.0413	0.3436
$M8$	*GPHT.LS*	0.54	0.0708	0.0865	0.1429	0.4585	0.0254	0.5450
$M8$	*W*	—	0.1764	0.1811	0.0568	0.1833	0.0343	0.3644
$M9$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.0962	0.1121	0.1329	0.4367	0.0269	0.5488
$M9$	*W*	—	0.2694	0.2501	0.0321	0.0975	0.0736	0.3475
$M10$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.0299	0.0193	0.1661	0.5185	0.0285	0.5378
$M10$	*W*	—	0.2546	0.2501	0.0137	0.0365	0.0650	0.2866
$M11$	*W*	—	0.0485	0.0439	0.0435	0.1309	0.0042	0.1748
$M12$	*W*	—	0.0122	0.0001	0.0268	0.0757	0.0009	0.0758
$M13$	*BA.LS*	0.84	0.0012	0.0015	0.0350	0.1166	0.0012	0.1180
$M14$	*BA.LS*	0.80	0.0051	0.0095	0.0557	0.1824	0.0031	0.1919
$M15$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.0998	0.0979	0.0659	0.2163	0.0143	0.3141
$M16$	***BA.LTS***	$[\boldsymbol{n/2}]$	0.1033	0.1016	0.0874	0.2670	0.0183	0.3686
$M16$	*W*	—	0.1835	0.1756	0.0471	0.1454	0.0359	0.3209
$M17$	*BA.LS*	0.50	0.1190	0.1160	0.0519	0.1757	0.0168	0.2918
$M18$	*BA.LS*	0.50	0.1262	0.1211	0.0614	0.2078	0.0197	0.3288
$M19$	*SPR.LS*	0.86	0.0002	0.0005	0.0319	0.1055	0.0010	0.1060
$M20$	*SPR.LS*	0.82	0.0042	0.0030	0.0355	0.1166	0.0013	0.1196
$M21$	*GPHT.LS*	0.58	0.0569	0.0591	0.1110	0.3373	0.0156	0.3964
$M22$	*GPHT.LS*	0.58	0.0381	0.0232	0.1139	0.3764	0.0144	0.3996
$M23$	*GPHT.LS*	0.60	0.0331	0.0349	0.1104	0.3577	0.0133	0.3926
$M23$	*SPR.LS*	$[n/2]$	0.2530	0.2534	0.0280	0.0919	0.0648	0.3454
$M24$	*GPHT.LS*	0.60	0.0245	0.0295	0.1015	0.3361	0.0109	0.3656
$M24$	*BA.LS*	$[n/2]$	0.2639	0.2630	0.0258	0.0854	0.0703	0.3484
$M25$	*W*	—	0.0499	0.0499	0.0000	0.0000	0.0025	0.0500
$M26$	*BA.LS*	0.62	0.0026	0.0031	0.0554	0.1803	0.0031	0.1834
$M27$	*GPHT.LS*	0.50	0.1932	0.1914	0.1711	0.5554	0.0666	0.7468
$M27$	*BA.LS*	0.50	0.4059	0.4035	0.0501	0.1668	0.1673	0.5703
$M28$	*GPHT.LS*	0.50	0.2182	0.2139	0.1660	0.5341	0.0752	0.7480
$M28$	*BA.LS*	0.50	0.3236	0.3227	0.0317	0.1012	0.1057	0.4239
$M29$	***BA.LTS***	**0.68**	0.0399	0.0389	0.0754	0.2522	0.0073	0.2911
$M29$	*SPR.LS*	$[n/2]$	0.1001	0.1009	0.0270	0.0893	0.0107	0.1902
$M30$	***BA.LTS***	**0.64**	0.0674	0.0692	0.0751	0.2572	0.0102	0.3264
$M30$	*SPR.LS*	$[n/2]$	0.1058	0.1064	0.0261	0.0858	0.0119	0.1922
$M31$	*BA.LS*	0.50	0.0203	0.0213	0.0378	0.1222	0.0018	0.1435

Acknowledgements

S.R.C. Lopes was partially supported by CNPq-Brazil, by Pronex *Probabilidade e Processos Estocásticos* (Convênio MCT/CNPq/FAPERJ - Edital 2003) and also by *Fundação de Amparo à Pesquisa no Estado do Rio Grande do Sul* (FAPERGS Foundation). B.V.M. Mendes was partially supported by CNPq-Brazil and by Pronex *Cópulas: Permutabilidade, Assimetrias, Inferência Estatística* (Convênio MCT/CNPq/FAPERJ - Edital 2004).

References

Agostinelli, C. and Bisaglia, L. (2003). "Robust Estimation of ARFIMA Processes´´. Working paper of *Dipartimento di Statistica, Università di Venezia.*

Andersen, T.G., and Bollerslev, T. (1997). "Heterogeneius information arrivals and return volatility dynamics: uncovering the long run in high frequency returns´´. *Journal of Finance*, **52**, 975-1005.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). "Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, Vol. **74**, 3-30.

Beran, J. (1994). *Statistics for Long-Memory Process.* New York: Chapman and Hall.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity". *Journal of Econometrics*, Vol. **31**, 307-327.

Bollerslev, T. and Mikkelsen, H.O. (1996). "Modeling and pricing long memory in stock market volatility". *Journal of Econometrics*, Vol. **73**, 151-184.

Bollerslev, T. and Wright, J.H. (2000). "Semiparametric estimation of long-memory volatility dependencies: The role of high-frequency data". *Journal of Econometrics*, Vol. **98**, 81-106.

Breidt, F.J., Crato, N. and de Lima, P. (1998). "The detection and estimation of long memory in stochastic volatility". *Journal of Econometrics*, Vol. **83**, 325-348.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods.* New York: Springer-Verlag.

Deo, R.S., and Hurvich, C.M. (2003). "Estimation of Long Memory in Volatility´´. In *Long-Range Dependence* (eds.) P. Doukhan, G. Oppenheim and M.S. Taqqu. Boston: Birkhäuser, 313-324.

Ding, Z. and Granger, C.W.J. (1996). "Modeling volatility persistence of speculative returns: A new approach´´. *Journal of Econometrics*, Vol. **73**, 185-215.

Engle, R.F. (1982). "Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation". *Econometrica*, Vol. **50**, 987-1008.

Fox, R. and Taqqu, M.S. (1986). "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series". *The Annals of Statistics*, Vol. **14**, 517-532.

Geweke, J. and Porter-Hudak, S. (1983). "The Estimation and Application of Long Memory Time Series Model". *Journal of Time Series Analysis*, Vol. **4**, 221-238.

Granger, C.W.J. and Joyeux, R. (1980) "An Introduction to Long Memory Time Series Models and Fractional Differencing´´. *Journal of Time Series Analysis*, Vol. **1**, 15-29.

Henry, M. (2001). "Robust Automatic Bandwidth for Long Memory´´. *Journal of Time Series Analysis*, **22**, 3, 293-316.

Hosking, J. (1981). "Fractional Differencing". *Biometrika*, Vol. **68**, 165-167.

Huber, P.J. (1981). *Robust Statistics.* New York: Wiley.

Hurvich, C.M. and Ray, B.K. (1995). "Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes". *Journal of Time Series Analysis*, Vol. **16**, 017-042.

Hurst, H.R. (1951). "Long-term storage in reservoirs". *Trans. Am. Soc. Civil Eng.*, Vol. **116**, 770-799.

Lopes, S.R.C., Olbermann, B.P. and Reisen, V.A. (2004). "A Comparison of Estimation Methods in Non-Stationary ARFIMA Processes". *Journal of Statistical Computation and Simulation*, Vol. **74**, 339-347.

Lopes, S.R.C and Mendes, B.V.M. (2005). "FIGARCH Modeling in Finance: a Review of the Theory and Empirical Evidence". Submitted.

Porter-Hudak, S. (1990). "An Application of the Seasonal Fractionally Differenced Model to the Monetary Aggregates". *Journal of American Statistical Association*, Vol. **85**, 338-344.

Rao, C. R. (1973). *Linear Statistical Inference and its Applications.* 2nd. Edition. New York: Wiley.

Reisen, V.A. (1994). "Estimation of the Fractional Difference Parameter in the ARIMA(p,d,q) model using the Smoothed Periodogram". *Journal of Time Series Analysis*, Vol. **15**, 335-350.

Reisen, V.A., Abraham, B. and Lopes, S.R.C. (2001). "Estimation of Parameters in ARFIMA Processes: A Simulation Study". *Communications in Statistics: Simulation and Computation*, Vol. **30**, 787-803.

Robinson, P.M. (1995). "Log-periodogram regression of time series with long range dependence". *The Annals of Statistics*, Vol. **23**, 1048-1072.

Robinson, P.M. (1999). "The memory of stochastic volatility models". Unpublished manuscript, London School of Economics.

Robinson, P.M. and Zaffaroni, P. (1998). "Nonlinear Time Series with Long Memory: a Model for Stochastic Volatility". *Journal of Statistical Planning and Inference*, Vol. **68**, 359-371.

Rousseeuw, P. J. (1984). "Least Median of Squares Regression". *Journal of the American Statistical Association*, Vol. **79**, 871-880.

Taqqu, M. S., and Teverovsky, V. (1996). "Semi-parametric graphical estimation techniques for long-memory data´´. In *Athens Conference on Applied Probability and Time Series Analysis* (eds.) P. M. Robinson and M. Rosenblatt.

Taqqu, M. S., Teverovsky, V., Willinger, W. (1995). "Estimators for long range dependence: an empirical study´´. *Fractals*, 3(4):785-798.

Velasco, C. (1999a). "Gaussian Semiparametric Estimation of Non-stationary Time Series". *Journal of Time Series Analysis*, Vol. **20**, 87-127.

Velasco, C. (1999b). "Non-stationary log-periodogram regression". *Journal of Econometrics*, Vol. **91**, 325-371.

Whittle, P. (1953). "Estimation and information in stationary time series". *Arkiv för Mathematik*, Vol. **2**, 423-434.

Yohai, V. J. (1987). "High Breakdown point and high efficiency robust estimates for regression. *Annals of Statistics*, Vol. **15**, 642-656.

Zaffaroni, (1999). "Gaussian estimation of long-range dependent volatility in asset prices". Preprint.