BANDWIDTH SELECTION IN CLASSICAL AND ROBUST ESTIMATION OF LONG MEMORY

Lopes, S.R.C.[a] **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 semiparametric 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. Whenever there is no information about the data generating process, the simulation experiments suggest that we should select the $Bartlett$ or the $GPHT$ estimator, either based on all frequencies and robust LTS-estimation, or based on $\alpha = 0.50$ and robust MM-estimation.

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 Moving Average (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.

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 (FIGARCH, FIEGARCH, TGARCH, SW-ARCH, LM-ARCH, among many others). Many empirical papers have detected the presence of long memory in the volatility of risky assets, market indexes and 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 5 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). A total of 15 estimates are implemented in a Monte Carlo study. Our initial motivation was the possibility that the robust estimators would naturally downweight undesirable frequencies and would not need the trimming constant α.

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 optimal (that is, smaller mean squared error) 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.

We carried out several simulation experiments to identify the optimal bandwidth value, and considered two criteria for choosing the best estimator. The mean squared error criterion, denoted by $C1$, and the proportion of times within the total number of simulations, the estimates 90% confidence interval did not capture the true parameter value, denoted by $C2$.

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.

For models possessing long memory in volatility and no other form of short memory, both criteria selected the $GPHT.LS$ based on small α-values as the best estimator. It seems that short memory may act as contaminations for these models, since when we include short memory in the mean of the volatility models, we get as winners either the robust $BA.LTS$ or $BA.MM$. When it comes to ARFIMA models, and classical estimation, one should use just few frequencies, setting α between 0.50 and 0.60. Then either the $GPHT$ or the BA estimator may be used. Under criterion $C2$ we are able to select an overall winner for the ARFIMA models, the $GPHT.MM$ based on $\alpha = 0.50$. The vast majority of the winners under $C2$ (83%) are robust estimators. In summary, whenever no other information about the data generating process is available, we would select the BA or the $GPHT$ estimator, either based on all frequencies and LTS-estimation, or based on $\alpha = 0.50$ and MM-estimation.

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 35 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 $X_t|\mathcal{F}_{t-1}$ are independent random variables with zero mean and conditional variance 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

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

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

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

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

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

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*

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

where

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

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 $\omega > 0$, the parameters α_1 and β_1 must satisfy

- $\beta_1 - d \leq \phi_1 \leq \frac{2-d}{3}$
- $d(\phi_1 - \frac{1-d}{2}) \leq \beta_1(d + \alpha_1)$, 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 positive.

2.3 FISV Models

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

$$Y_t = g(X_t)\sigma_\varepsilon \varepsilon_t, \tag{2.12}$$

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 (2.12) 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 an ARFIMA(p,d,q) process with $d\in(-0.5,0.5)$, given by (2.1). Consider the set of harmonic frequencies $w_i=\frac{2\pi i}{n}$, $i=0,1,\cdots,[n/2]$, 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 (2.2), and adding $\ln(f_U(0))$, and $\ln(I(w_i))$ to both sides of this expression we obtain

$$\ln(I(w_i)) = \ln(f_U(0)) - d\ln\left(\left(2\,\sin\left(\frac{w_i}{2}\right)\right)^2\right) + \ln\left(\frac{f_U(w_i)}{f_U(0)}\right) + \ln\left(\frac{I(w_i)}{f_X(w_i)}\right), \quad (3.1)$$

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

$$I(w) = \frac{1}{2\pi}\left(\hat{\gamma}_X(0) + 2\sum_{l=1}^{n-1}\hat{\gamma}_X(l)\cos(l\,w)\right), \quad (3.2)$$

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 $\ln\left(\frac{f_U(w_i)}{f_U(0)}\right)$ may be discarded. Then, we may rewrite (3.1) in the context of a simple linear regression model:

$$y_i = a - d\,z_i + e_i\ , \quad i=1,\cdots,m \quad (3.3)$$

where $m=[n/2]$, $(a,-d)$ are the regression coefficients, $a=\ln(f_U(0))$, $y_i=\ln(I(w_i))$, $z_i=\ln((2\sin(w_i/2))^2)$, 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 MM-estimation method.

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

$$L_1(m) = \sum_{i=1}^{m}(r_i)^2, \quad (3.4)$$

where $r_i = y_i - a + d\, z_i$ is the residual related to the regression (3.3).

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

$$L_2(m) = \sum_{i=1}^{m^*} (r^2)_{i:m} \ , \tag{3.5}$$

where $(r^2)_{i:m}$ are the squared and then ordered residuals, that is, $(r^2)_{1:m} \leq \ldots \leq (r^2)_{m:m}$, 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

$$L_3(m) = \sum_{i=1}^{m} \rho_2 \left(\frac{r_i}{s} \right)^2 \ , \tag{3.6}$$

subject to the constraint

$$\frac{1}{m} \sum_{i=1}^{m} \rho_1 (\frac{r_i}{s}) \leq b \ , \tag{3.7}$$

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.

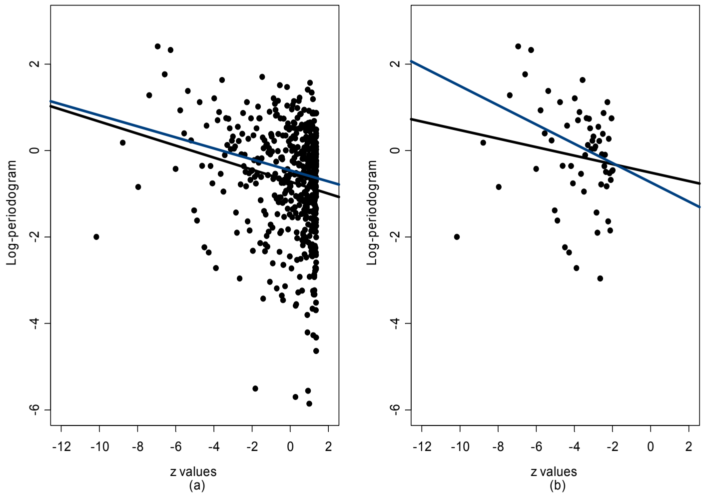

Figure 1: *OLS (black) and LTS (blue) estimates based on periodogram (3.2). The regression* (a) *at the left hand side uses all* $[n/2] = 500$ *frequencies and provided OLS and LTS d estimates equal to* 0.14 *and* 0.13*, respectively. The regression* (b) *at the right hand side uses* $m = (1000)^{0.59} = 60$ *frequencies and provided OLS and LTS d estimates equal to* 0.10 *and* 0.22*, respectively. Data simulated from a FISV model with true d value equal to* 0.30.

The regression data (y_i, z_i), $i = 1, ..., m$, used in Figure 1 are derived from 1000 observations simulated from a FISV process with $d = 0.30$. The regression (a) at the left hand side uses all $[n/2]$ frequencies, that is, $m = 500$. The data pattern depicted in graph (a) is typical: log-periodogram data contains a considerable amount of large z_i values. These are the values related to frequencies away from zero and, therefore, those less relevant in the estimation process. However, as graph (a) illustrates, they may have a large influence on the fits. The classical OLS (in black) and the robust LTS (in blue) slope estimates of (3.3), both based on (3.2), provided $\widehat{d}$ equal to 0.14 and 0.13, respectively.

According to the theory, the most influent points should be those associated to smaller z_i values. This suggests trimming the points associated to large frequencies, technique implemented at the right hand side of Figure 1. The regression (b) uses $m = (1000)^{0.59} = 60$ frequencies and provided OLS and LTS estimates $\widehat{d}$ equal to 0.10 and 0.22, respectively. The robust LTS procedure provides now an estimate closer to the true value. However, some points still tilt the classical OLS regression line, 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 (3.2), 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 (3.3) based on (3.2), which we denote by *GPH-LS*. The number of frequencies m used in (3.3) depends on a trimming constant $0 < \alpha < 1$. 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.

To obtain the robust versions of the *GPH* estimator we just apply the *LTS* and the *MM* methodologies to the regression model (3.3) with $m = n^{\alpha}$, based on (3.2). 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 of the spectral density function. Reisen (1994) proposed using a consistent estimator which is a smoothed version of the periodogram function (3.2), the *SPR* estimator.

More specifically, the regression estimator *SPR* is obtained by replacing the spectral density function in the expression (2.2), by the smoothed periodogram function, denoted by $I_s(\cdot)$, given by

$$I_s(w) = \frac{1}{2\pi} \sum_{j=-\nu}^{\nu} \kappa\left(\frac{j}{\nu}\right) \hat{\gamma}_X(j) \cos(jw), \tag{3.8}$$

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

$$\kappa(u) = \begin{cases} 1 - 6u^2 + 6|u|^3, & \text{if } |u| \leq \frac{1}{2}, \\ 2(1 - |u|)^3, & \text{if } \frac{1}{2} < |u| \leq 1, \\ 0, & \text{otherwise.} \end{cases} \tag{3.9}$$

The *SPR* estimator proposed by Reisen (1994) is obtained by applying the *OLS* procedure to the regression model (3.3) based on (3.8) and (3.9). 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). The robust versions are obtained by applying the *LTS* and the *MM* methodologies to (3.3) based on (3.8) and (3.9), 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 (3.8), where now the function $\kappa(\cdot)$ is defined as

$$\kappa(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{3.10}$$

The classical and robust versions are obtained by applying the *OLS*, the *LTS* and the *MM* methodologies to the regression model (3.3) based on (3.8) and (3.10), producing the *BA-LS*, the *BA-LTS*, and the *BA-MM* estimators. The value of m in (3.3) 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 (3.3) based on (3.2), but considering only the frequencies w_i, for $i \in \{l, l+1, \cdots, m\}$, where $l > 1$ is a trimming value that tends to infinity more slowly than m.

It would be interesting to compare the R and the *LTS* concepts. The R concept trims the extreme z_i 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 not be associated to small frequencies, but certainly are associated to leverage points. In other words, the *LTS* procedure can identify which data points associated with 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

$$I(w_i) = \frac{1}{\sum_{t=0}^{n-1} g(t)^2} \left| \sum_{t=0}^{n-1} g(t) X_t e^{-iw_i t} \right|^2, \tag{3.11}$$

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

$$g(t) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi(t+0.5)}{n}\right)\right]. \tag{3.12}$$

We obtain the classical *GPHT-LS* and the robust versions *GPHT-LTS* and *GPHT-MM* by applying the classical and the robust methodologies to model (3.3) based on (3.11) and (3.12), and setting $m = n^{\alpha}$.

4 Assessing the Estimators Performances

In this section we carry on a large simulation study to compare the five semi-parametric (classical) estimators and their robust versions[1] when estimating the fractional parameter d in 35 data generating processes (DGP).

The notations and detailed specifications of all DGP's considered are given in Table 1. Models $M1$ to $M22$ possess long memory in volatility and short memory in mean, being combinations of ARMA(p, q) and FIGARCH(r, d, s) processes. Models $M23$ to $M28$ are FISV$(p, d, q, \sigma_{\varepsilon})$ processes, and models $M29$ to $M35$ possess just long memory in the mean, being ARFIMA(p, d, q) processes. 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 model. α_1 and β_1 correspond to the autoregressive and moving average parameters in the FIGARCH process. The processes $\{\epsilon_t\}_{t\in\mathbb{Z}}$ in the ARFIMA and FISV models ($M23$ through $M35$) are Gaussian.

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 $\omega = 0.10$.
- 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.

For each of the three versions of the semiparametric estimators *GPH*, *SPR*, *BA*, *R* and *GPHT*, we considered 19 possibilities for the trimming constant α. Specifically, we set $\alpha \in \{0.50, 0.52, ..., 0.84, 0.86\}$. The version not tuned by α,

[1] In a previous version of the paper we had considered the parametric Whittle estimator (Whittle, 1953). The simulation results concerning this estimator were withdrawn from the present paper since we now focus on bandwidth selection of semiparametric estimators. The results, however, are avaliable upon request from the authors.

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.

Table 1: *Notations and specifications of all data generating processes considered in simulations.*

	ARFIMA			FIGARCH			FISV				Distr.	Distr.
Model	ϕ	d	θ	α_1	d	β_1	ϕ	d	θ	σ_ε	Z_t	ε_t
$M1$		——		0.00	0.50	0.00		——			$N(0,1)$	——
$M2$		——		0.00	0.50	0.00		——			$t_4(0,1)$	——
$M3$	0.50	0.00	0.00	0.00	0.50	0.00		——			$N(0,1)$	——
$M4$	0.50	0.00	0.00	0.00	0.50	0.00		——			$t_4(0,1)$	——
$M5$		——		-0.20	0.50	0.00		——			$N(0,1)$	——
$M6$		——		-0.20	0.50	0.00		——			$t_4(0,1)$	——
$M7$	0.00	0.00	0.50	-0.20	0.50	0.00		——			$N(0,1)$	——
$M8$	0.00	0.00	0.50	-0.20	0.50	0.00		——			$t_4(0,1)$	——
$M9$	0.50	0.00	0.00	0.00	0.25	0.20		——			$N(0,1)$	——
$M10$	0.50	0.00	0.00	0.00	0.25	0.20		——			$t_4(0,1)$	——
$M11$	0.00	0.00	0.50	0.00	0.25	0.20		——			$N(0,1)$	——
$M12$	0.00	0.00	0.50	0.00	0.25	0.20		——			$t_4(0,1)$	——
$M13$	0.20	0.00	0.20	-0.20	0.75	0.20		——			$N(0,1)$	——
$M14$	0.20	0.00	0.20	-0.20	0.75	0.20		——			$t_4(0,1)$	——
$M15$	0.20	0.00	0.20	-0.20	0.50	0.20		——			$N(0,1)$	——
$M16$	0.20	0.00	0.20	-0.20	0.50	0.20		——			$t_4(0,1)$	——
$M17$	0.20	0.00	0.20	-0.20	0.25	0.20		——			$N(0,1)$	——
$M18$	0.20	0.00	0.20	-0.20	0.25	0.20		——			$t_4(0,1)$	——
$M19$		——		0.15	0.00	0.80		——			$N(0,1)$	——
$M20$		——		0.15	0.00	0.80		——			$t_4(0,1)$	——
$M21$	0.50	0.00	-0.50	0.15	0.00	0.80		——			$N(0,1)$	——
$M22$	0.50	0.00	-0.50	0.15	0.00	0.80		——			$t_4(0,1)$	——
$M23$		——			——		0.60	0.30	0.00	0.30		$N(0,1)$
$M24$		——			——		0.60	0.30	0.00	0.30		$t_4(0,1)$
$M25$		——			——		0.00	0.30	0.70	0.30		$N(0,1)$
$M26$		——			——		0.00	0.30	0.70	0.30		$t_4(0,1)$
$M27$		——			——		0.60	0.30	0.70	0.30		$N(0,1)$
$M28$		——			——		0.60	0.30	0.70	0.30		$t_4(0,1)$
$M29$	0.60	0.45	0.00		——			——				
$M30$	0.60	0.30	0.00		——			——				
$M31$	0.00	0.45	0.90		——			——				
$M32$	0.00	0.30	0.90		——			——				
$M33$	0.60	0.45	0.70		——			——				
$M34$	0.60	0.30	0.70		——			——				
$M35$	0.60	0.00	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 estimators sensitivity 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 $\widehat{d^j}$, $j = 1, \cdots, 300$, 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 $\widehat{d_1^j}, \widehat{d_2^j}, ..., \widehat{d_S^j}$;
- 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.
- $C1^*$*:* 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 $C1^*$ are reported. In what follows we summarize the results for each model considered.

4.1 Simulations results

We provide detailed analysis of the results from models $M1$ and $M2$, and then summarize the results from the other models. In the tables that follow, whenever the value for α used is the maximum possible we write $[n/2]$.

Results from model $M1$*: ARFIMA*$(0,0,0)$ *combined with a FIGARCH*$(0,0.50,0)$ *process with Gaussian conditional distribution.* Figure 2 illustrates the trade off bias-variance, and how difficult is choosing an optimality criterion. This figure shows the simulated distributions of the three best estimators according to the following set up. In graph (a) we show the distributions of estimators possessing smaller $|B^j|$, and in graph (b) of those possessing smaller $|B_M^j|$. In the second row, we show the simulated distributions of the estimators possessing smaller standard deviation sd^j in (c), and in (d) of those presenting the smaller confidence interval length. Finally, the third row shows the winners from criteria *C1*, graph (e), and $C1^*$, graph (f). The horizontal dotted line in all graphs corresponds to the true d value, 0.50.

As we can see in the second row of Figure 2, the estimators possessing smaller variability are based on OLS estimation[2]. However they show unacceptable large biases. If the primary concern is just bias, we observe that the $GPHT$ estimator shows up 5 times. When we combine an accuracy measure and a variability measure, the $GPHT$ shows up 4 times. It seems that to correct the bias of the $GPHT$ estimator, for this model, one needs small α values. We decided to choose the classical $GPHT.LS$ with $\alpha = 0.54$, the third place under *C1*, as the overall winner. The results for model $M1$ illustrated in Figure 2 are given in detail in Table 2.

Table 2: *Model $M1$: Three best results under 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)	$GPHT.LS$(0.54)
abs(B^j)	0.0067	0.0129	0.0136	0.0303
	$GPHT.LS$(0.50)	$GPHT.LS$(0.52)	$GPHT.MM$(0.50)	
abs(B^j_M)	0.0089	0.0131	0.0140	0.0434
	$SPR.LS$([n/2])	$SPR.LS$(0.86)	$BA.LS$([n/2])	
sd^j	0.0518	0.0539	0.0541	0.1333
	$SPR.LS$([n/2])	$BA.LS$([n/2])	$SPR.LS$(0.86)	
$\|\|CI^j\|\|$	0.1650	0.1704	0.1746	0.4304
	$GPHT.LS$(0.58)	$GPHT.LS$(0.56)	$GPHT.LS$(0.54)	
$C1$	0.0179	0.0182	0.0187	0.0187
	$SPR.LS$([n/2])	$GPHT.LS$([n/2])	$BA.LS$([n/2])	
$C1^*$	0.3232	0.3297	0.3306	0.4738

Results from model $M2$: ARFIMA$(0,0,0)$ combined with a FIGARCH$(0,0.50,0)$ process with t_4 conditional distribution. The results are summarized in Table 3. Again, the $GPHT.LS$ estimator comes out as the overall winner, but now based on a smaller α-value, probably to improve bias-robustness of the estimates, which might be affected by the heavier tails of the conditional distribution. Even though, the bias of winner from model $M2$ is larger than that from model $M1$.

Table 3: *Model $M2$: 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)	$GPHT.LS$(0.52)	$GPHT.LS$(0.54)	$GPHT.LS$(0.50)
abs(B^j)	0.0759	0.0920	0.1040	0.0759
	$GPHT.LS$(0.50)	$GPHT.LS$(0.52)	$GPHT.LS$(0.56)	
abs(B^j_M)	0.0799	0.0954	0.1131	0.0799
	$BA.LS$(0.86)	$BA.LS$(0.84)	$BA.LS$([n/2])	
sd^j	0.0600	0.0601	0.0606	0.1560
	$BA.19.LS$	$BA.LS([n/2]$	$SPR.LS([n/2])$	
$\|\|CI^j\|\|$	0.1876	0.1888	0.1899	0.5015
	$GPHT.LS$(0.50)	$GPHT.LS$(0.52)	$GPHT.LS$(0.56)	
$C1$	0.0301	0.0312	0.0316	0.0301
	$SPR.LS$	$BA.LS$	$GPH.LTS$(0.86)	
$C1^*$	0.3811	0.3913	0.3945	0.5814

[2]This is somehow expected since we have a Gaussian process with no contaminations.

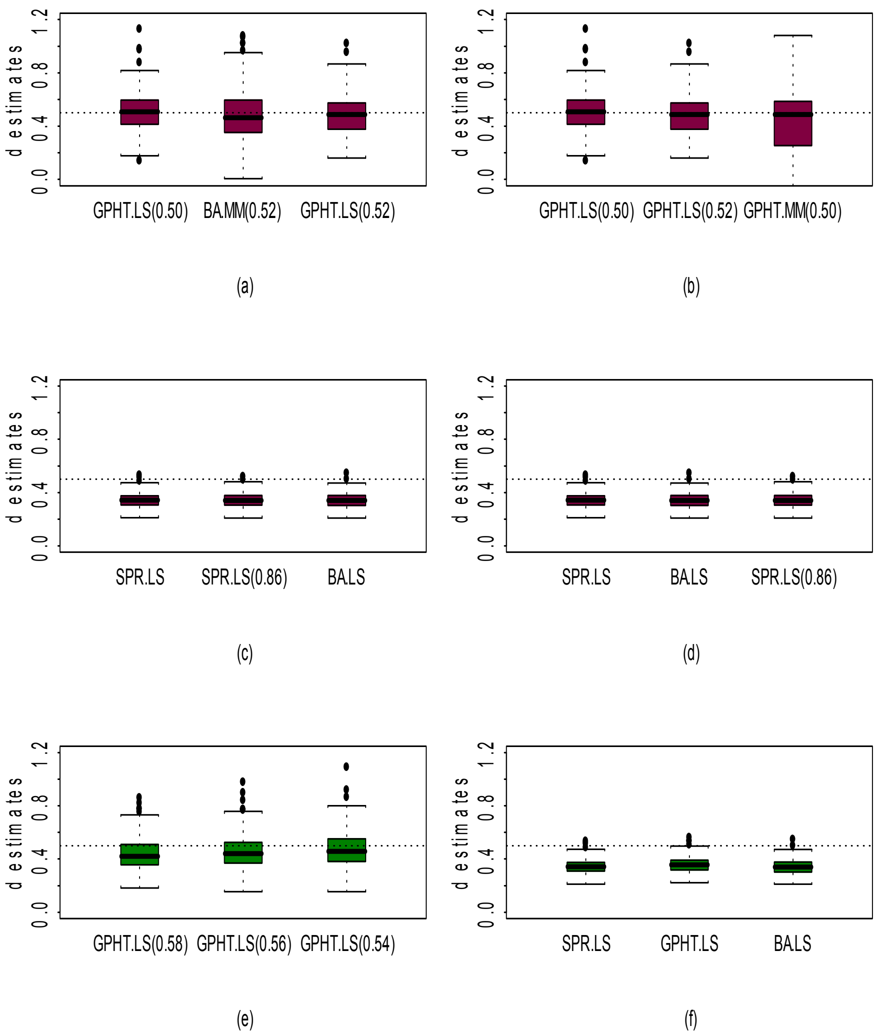

Figure 2: *The simulated distributions of best estimators from model* $M1$. *In the first row, we show the estimators possessing smaller* $|B^j|$ *in* (a), *and those possessing smaller* $|B^j_M|$ *in* (b). *In the second row, we show the estimators possessing smaller standard deviation* sd^j *in* (c), *and those presenting the smaller confidence interval length in* (d). *Finally, the third row shows the winners from criteria C1, graph* (e), *and* $C1^*$, *graph* (f). *Horizontal dotted line corresponds to the true d value,* 0.50.

In Table 2 as well as in Table 3 we observe that small α values are related to smaller biases, and large α values are related to smaller variability. This trade off bias-variance is illustrated in Figure 3. This figure shows the squared bias and variance of a classical and a robust estimator, as functions of their α values. At the left hand side we plot the simulation results for the $GPHT.LS$, winner in model $M1$, and at the right hand side we show the results for the $BA.LTS$, winner in model $M3$. The triangles represent the squared mean bias and the diamonds represent the variances from the 300 repetitions of each one. The circles point out the final choices, the $GPHT.LS$ with $\alpha = 0.54$, and the $BA.LTS$ using all $[n/2]$ frequencies. We observe that the robust procedure LTS shows better performance when all $[n/2]$ frequencies are used. This is actually expected since the LTS procedure already trims the frequencies related to atypical points.

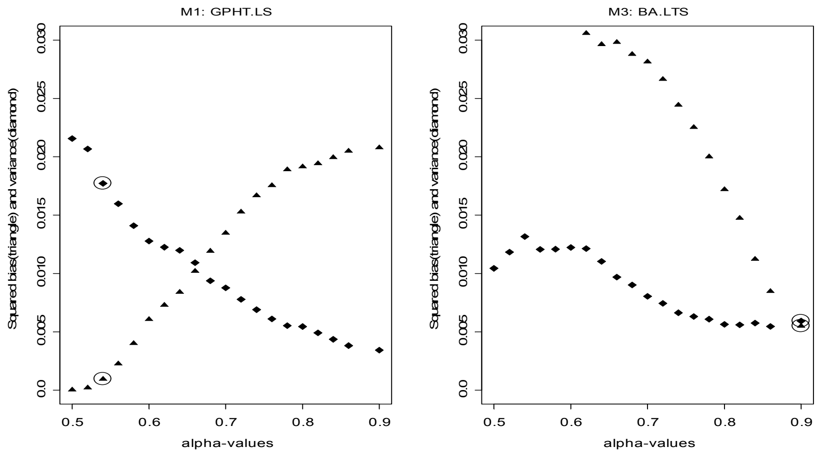

Figure 3: *The trade-off bias-variance. Figure shows for each α-value in the x-axis, the squared bias and variance in the y-axis of the estimators $GPHT.LS$ in model $M1$ in the left, and for the robust $BA.LTS$ in model $M3$ in the right hand side. Points circled correspond to the winners for each model.*

Summary of results from all models considered. For all models considered we observed an unacceptable large bias for the estimators winning under criterion $C1^*$. Based on that we decided to drop this criterion. Instead, following the suggestion of a referee, we implemented a third criterion for choosing the best estimator. This criterion, denoted by *C2*, considers the number of times out of the S simulations, the estimates 90% confidence interval did not capture the true parameter value. This would measure not only the bias of the estimators, but also the behavior of the estimator's variance and its asymptotic distribution.

Table 4 summarizes the simulation results for all DGP's considered. For each model, we report the winners under $C1$ (column 2) and $C2$ (column 7). When all frequencies are used we report $[n/2]$ instead of their α values. The third to sixth columns provide the measures of bias and variability for the winner under $C1$.

The application of criterion $C2$ resulted in interesting findings. We first note that the vast majority of the winners under $C2$ (83%) are robust estimators. For models $M1$ and $M2$, criterion $C2$ confirmed the $GPHT.LS$ based on small α-values as the best estimator. Criterion $C2$ also indicates that the $GPHT.LS$ could be the best option whenever just long memory (in volatility) is present and there is no other form of short memory, either in mean or in volatility. Models falling in this category are the above cited $M1$ and $M2$ and FISV models (see Table 4). This suggests that short memory may act as contaminations breaking down the estimates based on the least squares method.

Table 4: *Summary of results from all models.*

DGP	$C1$: Winner(α)	abs(B^j)	abs(B_M^j)	sd^j	$\|\|CI^j\|\|$	$C2$: Winner(α)
$M1$	$GPHT.LS(0.54)$	0.063	0.080	0.119	0.395	$GPHT.LS(0.52)$
$M2$	$GPHT.LS(0.50)$	0.076	0.080	0.156	0.502	$GPHT.LS(0.50)$
$M3$	$BA.LTS([n/2])$	0.074	0.081	0.080	0.248	$BA.MM([n/2])$
$M4$	$BA.LTS([n/2])$	0.019	0.035	0.118	0.384	$BA.MM(0.84)$
$M5$	$GPHT.LS(0.62)$	0.071	0.068	0.105	0.343	$GPHT.MM(0.50)$
$M6$	$GPHT.LS(0.54)$	0.065	0.079	0.137	0.449	$BA.MM(0.54)$
$M7$	$GPHT.LS(0.52)$	0.023	0.030	0.143	0.450	$BA.MM(0.54)$
$M8$	$BA.LTS([n/2])$	0.133	0.142	0.080	0.249	$BA.MM(0.54)$
$M9$	$BA.LS([n/2])$	0.013	0.011	0.044	0.146	$GPHT.LTS(0.84)$
$M10$	$BA.LS(0.82)$	0.004	0.011	0.055	0.170	$GPHT.MM(0.60)$
$M11$	$SPR.LS([n/2])$	0.021	0.024	0.037	0.121	$SPR.MM(0.54)$
$M12$	$SPR.LS(0.86)$	0.004	0.011	0.055	0.181	$SPR.LTS(0.86)$
$M13$	$BA.LTS([n/2])$	0.096	0.112	0.133	0.437	$GPHT.LTS(0.84)$
$M14$	$BA.LTS([n/2])$	0.030	0.019	0.166	0.518	$GPHT.MM(0.60)$
$M15$	$SPR.LS([n/2])$	0.040	0.044	0.054	0.182	$GPHT.MM([n/2])$
$M16$	$SPR.LS([n/2])$	0.004	0.003	0.066	0.201	$GPHT.MM(0.86)$
$M17$	$BA.LS(0.84)$	0.001	0.002	0.035	0.117	$BA.MM(0.76)$
$M18$	$BA.LS(0.80)$	0.005	0.010	0.056	0.182	$BA.MM(0.76)$
$M19$	$BA.LTS([n/2])$	0.099	0.098	0.066	0.216	$BA.MM([n/2])$
$M20$	$BA.LTS([n/2])$	0.103	0.102	0.087	0.267	$BA.MM([n/2])$
$M21$	$BA.LS(0.50)$	0.119	0.116	0.052	0.176	$SPR.MM(0.50)$
$M22$	$BA.LS(0.50)$	0.126	0.121	0.061	0.208	$SPR.MM(0.50)$
$M23$	$SPR.LS(0.86)$	0.000	0.001	0.032	0.106	$SPR.MM(0.60)$
$M24$	$SPR.LS(0.84)$	0.004	0.003	0.036	0.117	$SPR.MM(0.50)$
$M25$	$GPHT.LS(0.58)$	0.057	0.059	0.111	0.337	$GPHT.LS(0.54)$
$M26$	$GPHT.LS(0.58)$	0.038	0.023	0.114	0.376	$GPHT.LS(0.56)$
$M27$	$GPHT.LS(0.60)$	0.033	0.035	0.110	0.358	$GPHT.LS(0.60)$
$M28$	$GPHT.LS(0.60)$	0.025	0.030	0.102	0.336	$GPHT.LS(0.60)$
$M29$	$BA.LS(0.64)$	0.008	0.009	0.059	0.189	$GPHT.MM(0.50)$
$M30$	$BA.LS(0.62)$	0.003	0.003	0.055	0.180	$GPHT.MM(0.52)$
$M31$	$GPHT.LS(0.50)$	0.193	0.191	0.171	0.555	$GPHT.MM(0.50)$
$M32$	$GPHT.LS(0.50)$	0.218	0.214	0.166	0.534	$GPHT.MM(0.50)$
$M33$	$BA.LTS(0.68)$	0.040	0.039	0.075	0.252	$GPHT.MM(0.52)$
$M34$	$BA.LTS(0.64)$	0.067	0.069	0.075	0.257	$SPR.MM(0.50)$
$M35$	$BA.LS(0.50)$	0.020	0.021	0.038	0.122	$GPHT.MM(0.50)$

Actually, by including short memory in the mean of models $M1$ and $M2$, what would give rise to models $M3$ and $M4$, results as the winner a robust version of the BA estimator, either the $BA.LTS$ or the $BA.MM$, with large α. The results are no so clear when we add short memory in volatility to models $M1$ and $M2$ (models $M5$ and $M6$), when one may apply either a classical or a robust estimation procedure on the BA or on the $GPHT$ estimator, and choose a small α value.

Models $M7$ through $M18$ are all long memory processes in the volatility, combined with short memory in the mean and in the volatility. For these models, and according to criterion $C2$, we would select either the $GPHT.LTS$ based on all frequencies, or a robust MM-estimation based semiparametric estimator with moderate α, either the $GPHT.MM$ or the $BA.MM$. We note that in models $M13$ and $M14$ the true d value is very large, 0.75. For these models the criteria $C1$ and $C2$ agree that a robust estimator is needed.

In the case of no long memory in the volatility, simple GARCH model ($M19$ and $M20$), the winner is a robust version of the BA estimator, either LTS or MM based on the $[n/2]$ frequencies. However, when the short range effects are included to these GARCH(1,1) processes, giving rise to models $M21$ and $M22$, the winner under $C1$ becomes the classical $BA.LS$ tuned with the smaller α value.

When it comes to ARFIMA models ($M29$ to $M35$), according to $C1$, one should use either the classical $GPHT$ with $\alpha = 0.50$, or the BA estimator. The range for α for the $BA.LS$ is not so clear, and could vary from 0.50 to 0.68. However, for the ARFIMA models, the $C2$ criterion was able to select an overall winner, the $GPHT.MM$ based on $\alpha = 0.50$.

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).

Financial returns typically exhibit short memory in mean and volatility just for the first few lags, and weak long memory in volatility. These characteristics would correspond to specification of models $M17$ and $M18$. Thus, to estimate d without having fitted yet a fully parametric model, we decided to apply the $BA.LS(0.82)$ estimator, the winner of models $M17$ and $M18$, which yielded a d estimate of 0.1706. We also computed the winner for these models under criterion $C2$, the $BA.MM$ tuned with $\alpha = 0.76$, which yielded the value 0.1630.

As an exploratory analysis, and for the sake of completeness, we examined the plot suggested in Taqqu and Teverovsky (1996). According to this technique, given a semiparametric estimator, to choose its best α value, one could examine the d estimates as functions of their corresponding α values. This figure (not shown here) provided some indication of flatness for α in $[0.64, 0.74]$. The Taqqu and Teverovsky (1996) estimates would be $BA.LS(0.70) = 0.2045$. We note that any graphical procedure, though very interesting, is clearly subjective, and are difficult to be applied within a more complex decision based procedure.

To complete this analysis, we then fitted a fully parameterised model to the Taiwan daily returns. To model the serial dependence in the mean and variance of the daily returns we considered 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 were estimated by maximum likelihood using the SPlus module FinMetrics, based on Gaussian conditional distribution (the only one available for estimation of FIGARCH models), and the AIC criterion was used to select the best model. The best fit turned out to be an ARMA$(1, 1)$ combined with a FIGARCH$(1, d, 0)$ with all parameters estimates highly significant, see Table 5, estimating d as 0.2566.

Table 5: *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
ω	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

Actually, none of the models used in the simulations possesses the specification ARMA$(1, 1)$ combined with FIGARCH$(1, d, 0)$ found for Taiwan. Thus we carried out another simulation experiment considering this model found by the fully parametric approach, setting as true parameters values those given in Table 5, i.e., $d = 0.26$, $\alpha_1 = -0.20$, $\omega = 0.47$, $\theta = -0.54$, and $\phi = 0.58$, and Gaussian innovations. The same estimators were computed, 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).

6 Conclusions

Semiparametric methods seem to be very suitable for empirical analysis of long memory in volatility processes, 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 bandwidth value. Additional complications arise from the lack of robustness of the least squares estimation methodology. In this paper we adressed the issue of tuning a selection of semiparametric estimators 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.

We carried out several simulation experiments to identify the optimal bandwidth value, and considered two criteria for choosing the best estimator. The mean squared error criterion, denoted by $C1$, and the proportion of times within the total number of simulations, the estimates 90% confidence interval did not capture the true parameter value, denoted by $C2$.

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.

For models possessing long memory in volatility and no other form of short memory, both criteria selected the $GPHT.LS$ based on small α-values as the best estimator. Models falling in this category may be FIGARCH and FISV models. It seems that short memory may act as contaminations for these models, since when we include short memory in the mean for these models, we get as winners either the robust $BA.LTS$ or $BA.MM$.

When it comes to ARFIMA models, and classical estimation, one should use just few frequencies, setting α between 0.50 and 0.60. Then either the $GPHT$ or the BA estimator may be used. Under criterion $C2$ we are able to select an overall winner for the ARFIMA models, the $GPHT.MM$ based on $\alpha = 0.50$.

The vast majority of the winners under $C2$ (83%) are robust estimators. The less convincing results from the robust estimators under $C1$ may be related to their specification, based on a 0.50 breakdown point. More efficient estimates mat be obtained if smaller breakdown point versions are specified. We expect that under contaminated models the robust estimators will present an even better performance. We plan to investigate this issue in a future research.

In summary, whenever no other information about the data generating process is available, we would select the BA or the $GPHT$ estimator, based on all frequencies and LTS-estimation, or based on $\alpha = 0.50$ and MM-estimation.

Acknowledgements

The authors wish to thank the Editor and the two referees for their valuable comments and suggestions that have greatly improved the paper. S.R.C. Lopes was partially supported by CNPq-Brazil, by Pronex *Probabilidade e Processos Estocásticos* (Convênio MCT/CNPq/ FAPERJ - Edital 2003), by Edital Universal *Modelos com Dependência de Longo Alcance: Análise Probabilística e Inferência* (CNPq-No. 476781/2004-3) 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, Edital CNPq 019/2004 Universal, 470233/ 2004-4-APQ *Modelagem de Extremos em Atuária, Finanças e Meio Ambiente*, and Convênio MCT/CNPq/FAPERJ - Edital 2004, *Cópulas: Permutabilidade, Assimetrias, Inferência Estatística*.

References

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

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.

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*, Vol. **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.

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*, Vol. **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.