PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 14, Fasc. 2 (1993), pp. 223–229

A NOTE ON THE ESTIMATION OF DEGREE OF DIFFERENCING IN LONG MEMORY TIME SERIES ANALYSIS

BY

M. S. PEIRIS AND J. R. COURT (SYDNEY)

Abstract. In this paper we investigate the properties of the estimator of degree of differencing the fractional d in long memory time series analysis via consistent spectral density estimation. It is shown that the proposed estimator is more efficient than some of the others in practice.

1. Introduction. The interest in fractional differencing and its applications to long memory time series modelling have become important to theoretical and applied time series analysis after the pioneering work by [5] and [6]. Since then a large number of authors have reported some interesting and mathematically elegant results on the modelling of long memory processes. This class of models is motivated by

(i) unbounded spectral densities at the frequency $\omega = 0$,

(ii) hyperbolic decay of the autocorrelation function (acf), and

(iii) the Hurst effect.

When modelling long memory time series, it is convenient to use the following two filters in tandem:

(a) transform the original long memory $\{X_t\}$ to a short memory $\{Y_t\}$ via

$$Y_t = (1-B)^d X_t, \quad d \in (-\tfrac{1}{2}, \tfrac{1}{2}), \tag{1.1}$$

where B is the backshift operator such that $B^j X_t = X_{t-j}, j \geqslant 0$;

(b) transform the short memory $\{Y_t\}$ to white noise $\{Z_t\}$ via

$$\Phi(B) Y_t = \Theta(B) Z_t, \tag{1.2}$$

where $\Phi(B) = 1 - \Phi_1 B - \ldots - \Phi_p B^p$ and $\Theta(B) = 1 - \Theta_1 B - \ldots - \Theta_q B^q$ are stationary autogressive (AR) and invertible moving average (MA) operators, respectively.

The resulting process is generated by

$$\Phi(B) \nabla^d X_t = \Theta(B) Z_t, \quad d \in (-\tfrac{1}{2}, \tfrac{1}{2}), \tag{1.3}$$

where $\nabla = 1-B$, and is said to constitute a family of fractionally integrated ARMA(p, q) or ARIMA(p, d, q) processes, $d \in (-\frac{1}{2}, \frac{1}{2})$. This family has a wide variety of applications in almost every area in scientific endeavour (e.g., stream flow, rainfall, temperature, economic). The problem of estimation of parameters has received a considerable attention in the last decade, and among others [4] suggests to estimate d (first) independent of the other parameters using the periodogram of the data near frequency zero. Once the d is confidently estimated from the observed series, the unique whitening filter for a short memory $\{Y_t\}$ may be obtained following [1]. Several other methods have also been proposed to estimate the degree of differencing (fractional), d, in the literature. However, the aim of this note is to improve the efficiency of estimation of d over [4] in the mean square sense when the estimate is obtained through a lag-window estimator of the spectral density of $\{X_t\}$. With that view in mind, in the next section we review some of the available methods of estimation of d.

2. Some available methods of estimating *d*. In this section, we review some existing methods of estimating d.

a. Janacek [7] proposes an estimator $\hat{d}_m$ based on the log spectrum. That is

$$\hat{d}_m = \Big(S - \sum_{k=1}^{M} \hat{c}_k/k\Big) \Big/ \sum_{k=1}^{M} k^{-2}, \tag{2.1}$$

where

$$S = \pi^{-1} \int_0^{\pi} W(\theta) \log \hat{f}(\theta)\, d\theta, \qquad W(\theta) = \sum_{k=1}^{\infty} \cos(k\theta)/k,$$

$$\hat{c}_k = n^{-1} \sum_{p=1}^{n-1} \log I_N(p, x) \cos(k\omega_p) + (2n)^{-1} [\log I_N(0, x) - \delta_N \log I_N(\pi, x)],$$

$n = (N-1)/2$, $\delta_N = 1$ when N is even and zero otherwise. $I_N(p, x)$ is the periodogram and $\omega_p = p/n$, M is the truncation point chosen large enough for c_k to be negligible.

b. Parzen [8] constructed the estimator $\hat{d}_k$ with the help of the non-parametric kernel density estimator

$$f(\omega) = \sum_{v=-\infty}^{\infty} k(v/M)\, \varrho_T(v) \exp(-2\pi i \omega v), \qquad |\omega| < 0.5,$$

where

$$k(t) = \begin{cases} 1-6t^2+6|t|^3, & |t| < 0.5, \\ 2(1-|t|)^3, & 0.5 \leqslant t \leqslant 1, \\ 0, & \text{otherwise,} \end{cases}$$

is the Parzen window, and $\varrho_T(v)$ is an estimator of the autocorrelation function at lag v. Then the resulting estimator for d_k is

$$\hat{d}_k = 0.5\left[k^{-1}\sum_{j=1}^{k}\log f((j+m)/N) - \log((k+1+m)/N)\right]. \tag{2.2}$$

c. A more popular method was suggested by [4] based on the regression analysis using the periodogram. Assume that $\{Y_t\}$ has the spectrum $f_Y(\omega)$. Then $\{X_t\}$ does not strictly have a spectrum for all ω, but can be expressed as (by filtering considerations as in [5])

$$f_X(\omega) = |1-e^{-i\omega}|^{-2d} f_Y(\omega), \tag{2.3}$$

where

$$f(\omega) = \frac{\sigma^2}{2\pi}\frac{|\Theta(e^{-i\omega})|^2}{|\Phi(e^{-i\omega})|^2} \quad \text{and} \quad \mathrm{Var}(Z_t) = \sigma^2.$$

Note. 1. $f_X(\omega)$ exists for all ω when $d < 0$.

2. $f_X(\omega)$ does not exist for all ω when $d > 0$ (specifically, $f_X(\omega) \to \infty$ as $\omega \to 0$ for $d > 0$).

Arguing as in [4] one can simplify the equality (2.3) as the simple linear regression equation

$$y_j = a + bx_j + e_j, \quad j = 1, 2, \ldots, m, \tag{2.4}$$

where

$$y_j = \ln I_n(\omega_j), \qquad x_j = \ln|1-\exp(-i\omega_j)|^2,$$

$$I_n(\omega_j) = \frac{1}{2\pi n}\Big|\sum_{t=1}^{n} X_t \exp(-i\omega_j t)\Big|^2, \qquad e_j = \ln\big(I_n(\omega_j)/f(\omega_j)\big),$$

$a = \ln f(0)$, $b = -d$, and m is a function of n such that $m/n \to 0$ as $n \to \infty$. This gives the estimator

$$\hat{d} = -\sum_{i=1}^{m}(x_i-\bar{x})(y_i-\bar{y})\Big/\sum_{i=1}^{m}(x_i-\bar{x})^2 \tag{2.5}$$

with

$$\hat{d} \sim \mathrm{AN}\Big(d, \pi^2/6\sum_{i=1}^{m}(x_i-\bar{x})^2\Big) \quad \text{as } n \to \infty.$$

The main reason for the popularity of this regression method is that it permits the estimation of d without any prior knowledge of p and q. However, the choice of m completely determines the accuracy of the estimate $\hat{d}$. It is obvious that the accuracy of $\hat{d}$ may be further improved by using a consistent spectral density estimator $\hat{h}(\omega_j)$ in place of the periodogram $I_n(\omega_j)$. The next section discusses the various properties of this proposed estimator for d.

3. Estimation of d using a lag-window spectral density estimate. It is well known that, for $d < 0$, $I_n(\omega_j)$'s are (asymptotically) independently distributed $\frac{1}{2} f_X(\omega_j)\chi_2^2$ variates when $\omega_j \neq 0, \pi$. This tells us that $I_n(\omega_j)$ is an unbiased but inconsistent estimator for $f_X(\omega_j)$ for $\omega_j \neq 0, \pi$. Hence we use the lag-window spectral density estimator in place of $I_n(\omega_j)$ in (2.4) in order to improve the quality of $\hat{d}$. Furthermore, we show that the resulting new estimator is more efficient (in the mean square sense) than that of (2.5).

Now we write (2.3) at $\omega_j = 2\pi j/n \in (0, \pi)$ in the form of

$$\ln f_X(\omega_j) = -d \ln |1-\exp(-i\omega_j)|^2 + \ln f_Y(\omega_j). \tag{3.1}$$

Let

$$\hat{h}(\omega) = (2\pi)^{-1} \sum_{|s|<n} \lambda(s)\hat{R}(s)e^{-is\omega}$$

be the lag-window spectral density estimator of $\{X_t\}$, where $\lambda(s)$ is the lag window, and $\hat{R}(s)$ is the estimated autocovariance at lag s.

Adding $\ln \hat{h}(\omega_j)$ to both sides of (3.1), we obtain

$$\ln \hat{h}(\omega_j) = -d \ln |1-\exp(-i\omega_j)|^2 + \ln\left[\frac{\hat{h}(\omega_j)}{f_X(\omega_j)} f(\omega_j)\right]. \tag{3.2}$$

If ω_j is near zero, (3.2) may be well approximated by

$$\ln \hat{h}(\omega_j) = \ln f(0) - d \ln |1-\exp(-i\omega_j)|^2 + \ln\left[\frac{\hat{h}(\omega_j)}{f_X(\omega_j)}\right]. \tag{3.3}$$

Let

$$y_j = \ln \hat{h}(\omega_j), \qquad x_j = \ln |1-\exp(-i\omega_j)|^2,$$

$$\alpha = \ln f(0), \qquad \beta = -d, \qquad e_j = \ln\left[\frac{\hat{h}(\omega_j)}{f_X(\omega_j)}\right].$$

Then we have

$$y_j = \alpha + \beta x_j + e_j, \qquad j = 1, 2, \ldots, [n^\theta], \qquad 0 < \theta < 1, \tag{3.4}$$

where $[n^\theta]$ denotes the integer part of n^θ.

It is known that $\{\hat{h}(\omega_j)/f_X(\omega_j)\}$ is not a sequence of independent random variables, but for most of the standard windows $\{\hat{h}(\omega_j)/f_X(\omega_j)\}$ is approximately uncorrelated under certain regularity conditions and for large n. Now we state the following lemma for later reference.

LEMMA 1. *For the regression model defined in* (3.4) *we have*
(i) $\mu_e = \mathrm{E}(e_j) \approx 0$;
(ii) $\sigma_e^2 = \mathrm{Var}(e_j) \approx n^{-1}\sum_{|s|<n} \lambda^2(s)$ *for all* $j \neq 0, n/2$ *(n even)*.

The proof of this lemma follows from the asymptotic distribution

$$\frac{\hat{h}(\omega)}{f_X(\omega)} \mathrel{\dot\sim} a\chi_\nu^2 \qquad \left(\nu = \frac{2n}{\sum_{|s|<n} \lambda^2(s)} \text{ and } a = \frac{1}{\nu}\right).$$

However, the asymptotic distribution of e_j is normal with mean 0 and variance σ_e^2 ([2], p. 364). Now it can be seen that the least squares estimator of d in (3.4), i.e.

$$\hat{d}_c = -\sum_{j=1}^{m} (x_j-\bar{x})(y_j-\bar{y}) / \sum_{j=1}^{m} (x_j-\bar{x})^2, \quad m = [n^\theta], \tag{3.5}$$

is consistent for d, and more efficient than that of $\hat{d}$ in (2.5).

4. Efficiency of $\hat{d}_c$. In this section we state and prove the following main theorem for the properties of the proposed estimator $\hat{d}_c$.

THEOREM. *The estimator $\hat{d}_c$ defined in* (3.5) *is consistent and more efficient than that of $\hat{d}$ given in* (2.5).

Proof. From (3.4) and (3.5) we have (approximately)

$$\operatorname{Var}(\hat{d}_c) \approx \frac{\sigma_e^2}{\sum_{j=1}^{m} (x_j-\bar{x})^2} = \frac{\sum_{|s|<n} \lambda^2(s)}{n \sum_{j=1}^{m} (x_j-\bar{x})^2}. \tag{4.1}$$

For many lag windows used in practice we have

$$\frac{1}{n} \sum_{|s|<n} \lambda^2(s) = k\frac{m}{n},$$

where k is a constant depending on the lag window. For example:

Lag window	k
Bartlett	2/3
Daniell	1
Parzen	0.5393
Tukey	$2(1-4a+6a^2)$
Tukey–Hanning	3/4
Tukey–Hamming	0.7948

Since the choice of m satisfies $m/n \to 0$ as $n \to \infty$, it follows that $\hat{d}_c$ is consistent for d. Now

$$\hat{d}_c \sim \mathrm{AN}\left(d, \frac{km}{n \sum_{j=1}^{m} (x_j-\bar{x})^2}\right)$$

and $km/n < \pi^2/6$ for any $m = [n^\theta]$, $0 < \theta < 1$. The increase in efficiency of $\hat{d}_c$ over $\hat{d}$ in (2.5) immediately follows.

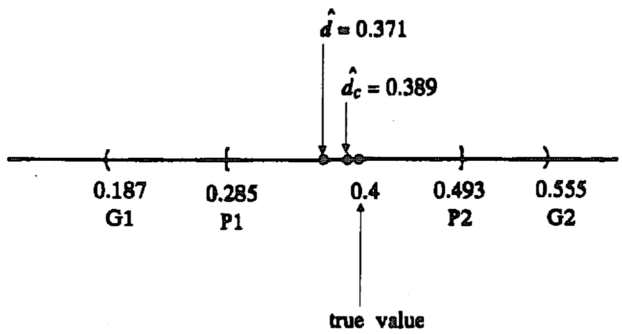
EXAMPLE. Consider the data $\{X_t;\, t = 1, 2, \ldots, 200\}$, which are given by [2], p. 530 (Example 13.2.1). The series $\{X_t\}$ was generated by $\nabla^{0.4} X_t = Z_t + 0.8\, Z_{t-1}$, $\{Z_t\} \sim \mathrm{WN}(0, 0.483)$.

Brockwell and Davies [2] used the method C of [4] with $m = 14$ and obtained the value $\hat{d} = 0.371$ with $\mathrm{Var}(\hat{\alpha}) = 0.00884$.

We use the Tukey–Hanning window,

$$\lambda(s) = \begin{cases} \frac{1}{2}\{1+\cos(\pi s/14)\}, & |s| \leqslant 14, \\ 0, & |s| > 14, \end{cases}$$

and obtain the corresponding $\hat{d}_c = 0.389$ with $\mathrm{Var}(\hat{d}_c) = 0.0028$. An approximate 95% confidence interval for d is (0.285, 0.493).



G1–G2: a 95% confidence interval for d by [3].

P1–P2: a 95% confidence interval for d by the proposed method.

Remark 1. In general, the spectral estimates corresponding to a given lag-window for two neighbouring Fourier frequencies are correlated. However, two estimates may be uncorrelated if the separation between their frequencies is appreciably greater than the "bandwidth" of the spectral window ([11], p. 456).

The above example shows that for large (but finite) n the assumption we made in the paper may not be "very" unrealistic.

Remark 2. In searching for a good model it is not reasonable to estimate d independently of the other parameters, since for any moderate size data set the behaviour of the spectral density near frequency zero may be affected not only by d, but also by the values of the autoregressive and moving average parameters. The maximum likelihood procedure appears to be more promising to this problem, and will be discussed in detail in a future paper.

Remark 3. In some cases the proposed estimate $\hat{d}_c$ may be more biased than that of $\hat{d}$. However, a large scale simulation by [3] recommended the use of $\hat{d}_c$ in practice.

Acknowledgements. We are grateful to the editor and the referee for valuable comments and useful suggestions on the earlier version of the paper. The first-named author is also thankful to Bovas Abraham for many helpful discussions.

REFERENCES

[1] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, 2nd edition, Holden-Day, San Francisco 1976.

[2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York 1990.

[3] G. Chen, B. Abraham and M. S. Peiris, *Lag window estimation of the degree of differencing in fractionally integrated time series models* (1993).

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

[5] C. W. J. Granger and R. Joyeux, *An introduction to long memory time series models and fractional differencing*, ibidem 1 (1980), pp. 15–30.

[6] J. R. M. Hosking, *Fractional differencing*, Biometrika 68 (1981), pp. 165–176.

[7] G. J. Janacek, *Determining the degree of differencing for time series via the long spectrum*, J. Time Ser. Anal. 3 (1982), pp. 177–183.

[8] E. Parzen, *Time series model identification by estimating information, memory and quartiles*, Technical Report, Texas A & M University, 1983.

[9] M. S. Peiris, *Some problems in time series analysis and forecasting*, unpublished Ph. D. Thesis, Monash University, Australia, 1987.

[10] — and B. J. C. Perera, *On the prediction with fractionally differenced ARMA models*, J. Time Ser. Anal. 9 (3) (1988), pp. 215–220.

[11] M. B. Priestley, *Spectral Analysis and Time Series*, Vols. 1 and 2, Academic Press, New York 1981.

School of Mathematics and Statistics
The University of Sydney, N.S.W. 2006, Australia

Received on 1.10.1992;
revised version on 11.10.1993