New Methods for the Analysis of Long-memory Time-series: Application to Spanish Inflation

MIGUEL A. DELGADO
Universidad Carlos III, Madrid, Spain

and

PETER M. ROBINSON
London School of Economics, U.K.

ABSTRACT

Models for long-memory time series are considered in which the autocovariance sequence is parameterized only at very long lags or the spectral density is parameterized only at very low frequencies. Various recently proposed methods for estimating the differencing parameters are reviewed and are applied to an economic time series of prices in Spain.

KEY WORDS Long-memory Differencing parameters Semi-parametric estimation Autocovariance Averaged periodogram Log-periodogram regression Inflation rate

INTRODUCTION

This paper describes and applies to real economic data some very recent developments in the analysis of time series. Consider a real-valued time series x_t, $t = 1, 2, \ldots$ which is observed at $t = 1, 2, \ldots, n$. We assume that x_t is at least covariance stationary so that the mean $\mu = E(x_t)$ and the autocovariances

$$\gamma_j = E\{(x_t - \mu)(x_{t-j} - \mu)\}, \; j = 1, 2, \ldots$$

do not depend on t. We also assume that there exists a spectral density given by

$$f(\lambda) = (2\pi)^{-1} \sum_{j=-\infty}^{\infty} \gamma_j \cos(j\lambda), \; -\pi \leqslant \lambda \leqslant \pi$$

Thus it is assumed that any stochastic or non-stochastic trends have been removed from the raw observed time series. The models most frequently used in the analysis and forecasting of time series impose strong conditions on the rate of decay of the γ_j's as j tends to infinity, or equivalently, boundedness and strong smoothness conditions on $f(\lambda)$. These models involve stronger conditions than the summability condition

$$\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty \tag{1}$$

or the boundedness restriction

$$f(0) < \infty \tag{2}$$

In particular, stationary autoregressive moving average (ARMA) models imply autocovariances that decay exponentially as $j \to \infty$, and a spectral density which is analytic at all frequencies. The ARMA models are the ones which have been most extensively studied and applied, but in fact there exist many other time-series models which satisfy equations (1) and (2).

Empirical observation, nevertheless, is sometimes consistent with models which do not satisfy these conditions. Figures 1 and 2 plot the sample autocovariances and periodogram for the differenced logs of the Spanish monthly general price index, recorded from July 1939 to October 1991 (see de Ojeda Eiseley, 1988, for a description of the series).[1] Thus, we have $n + 1 = 628$ observations. The original series, P_t, $t = 0, 1, \ldots$, appears non-stationary, but the series

$$x_t = \log(P_t) - \log(P_{t-1}),\ t = 1, \ldots, n \tag{3}$$

appears more stationary. Figure 1 plots the correlogram $\hat{\rho}_j = \hat{\gamma}_j / \hat{\gamma}_0$, where $\hat{\gamma}_j$ is the sample autocovariance

$$\hat{\gamma}_j = n^{-1} \sum_{t=1}^{n-j} (x_t - \bar{x})(x_{t-j} - \bar{x}),\ j = 0, 1, \ldots, n-1 \tag{4}$$

where $\bar{x} = n^{-1} \Sigma_{t=1}^{n} x_t$. While the $\hat{\rho}_j$ do appear to decay as j increases, they do so slowly, in

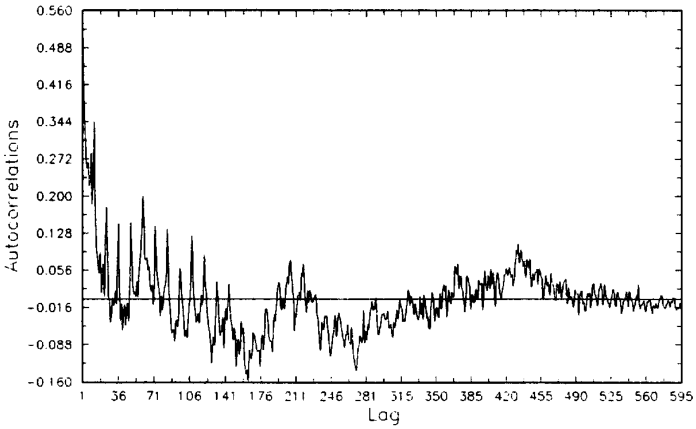

Figure 1. Correlogram (inflation rate x_t)

---

[1] The series has been created linking the five price indices for different periods and with a different basis. These series are not homogeneous because the weights of the goods and the goods entering differ from one index to other. In order to link the series, certain linking coefficients were calculated based on the periods where two indices overlap. The components of the general (aggregate) price index are the groups Food, No-food, Clothing, Housing, Domestic Goods and Other Goods.

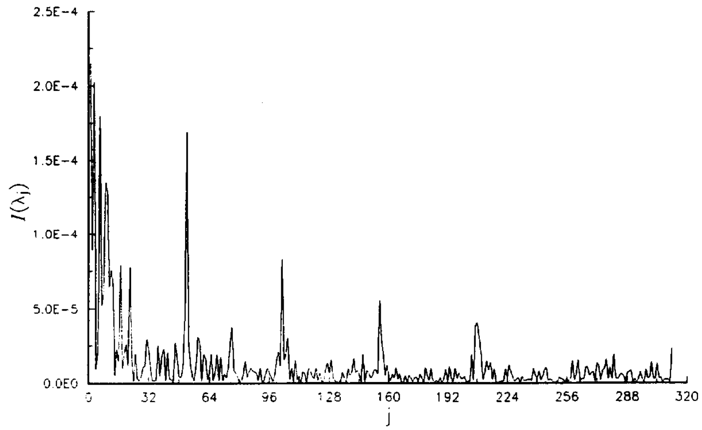

Figure 2. Periodogram (inflation rate x_t)

a manner that could be consistent with the failure of equation (1). Figure 2 plots the periodogram

$$I(\lambda) = (2\pi n)^{-1}\left|\sum_{t=1}^{n} x_t e^{it\lambda_j}\right|^2 \tag{5}$$

at frequencies $\lambda = \lambda_j = 2\pi j/n$, for $j = 1, \ldots, (n-1)/2$. Although $I(\lambda)$ does not provide a consistent estimate of $f(\lambda)$, Figure 2 suggests that condition (2) may fail (note that equation (5) is unaffected by a non-zero $\bar{x}$ when evaluated at frequency λ_j for integer j).

Parametric models for stationary series which violate equations (1) and (2) have long been available. One of these is the 'fractional noise' process with autocovariances

$$\gamma_j = \tfrac{1}{2}\gamma_0\{|j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d-1}\},\ j = 1, 2, \ldots \tag{6}$$

where

$$0 < d < \tfrac{1}{2} \tag{7}$$

It may be shown that

$$\gamma_j \approx cj^{2d-1}, \qquad \text{as } j \to \infty \tag{8}$$

for some c satisfying $0 < c < \infty$. In view of equations (7) and (8) it is clear that condition (1) does not hold. It may also be shown that

$$f(\lambda) \approx C\lambda^{-2d}, \qquad \text{as } \lambda \to 0^+ \tag{9}$$

for some C satisfying $0 < C < \infty$. In view of equations (7) and (9) it is clear that condition (2) also does not hold. We call series which violate equations (1) and (2) 'long-memory'. Early work on long-memory time series by Mandelbrot and his co-authors (see e.g. Mandelbrot and van Ness, 1968) stressed model (6). However, (6) is a very parsimonious model which implies that γ_j and $f(\lambda)$ decay monotonically as j and λ increase, and these properties do not seem relevant in Figures 1 and 2. A much weaker class of parametric models are the autoregressive fractionally integrated moving averages, given by

$$(1 - L)^d a(L)(x_t - \mu) = b(L)e_t,\ t = 1, 2, \ldots \tag{10}$$

where L is the lag operator, $a(.)$ and $b(.)$ are polynomials of degree p and q, respectively, having no roots in common or on the unit circle, and $\{e_t, t \geqslant 1\}$ is a sequence of uncorrelated random variables with zero mean and unknown, positive, finite variance σ^2. Again, it may be shown that conditions (8) and (9) hold and thus that (1) and (2) do not. The model contains the simple model $(1 - L)^d(x_t - \mu) = e_t$, considered by Adenstedt (1974), and has been applied in practice by a number of researchers. For suitable p and q it can describe a variety of non-monotonic behaviour in γ; and $f(\lambda)$, and thus has the potential to model the phenomena exhibited in Figures 1 and 2. However, correct choice of the autoregressive and moving average orders p and q is important; if either is misspecified, then estimates of d in equation (10) are liable to be inconsistent. The autoregressive and moving average components $a(.)$ and $b(.)$ are designed to model the short- and medium-run components of x_t, and it is unfortunate that their orders are important to the estimation of the long-run parameter d.

The preceding discussion suggests that there are advantages in estimating d on the basis of the limiting relationships (8) and (9). These can be called 'semi-parametric' models because they parameterize only the long-run characteristics of x_t, while allowing the short- and medium-run characteristics to be non-parametric. There is a price to be paid in terms of efficiency in not using a correct parametric model, but when n is large the greater robustness of semi-parametric model-based procedures is relevant.

Several methods of estimating the semi-parametric models (8) and (9) have been introduced or developed by Robinson (1990, 1991, 1992). These are described in the following section. In the third section applications of the methods to the Spanish price index series are reported.

PARAMETER ESTIMATES

In this section four alternative estimates of the differencing parameter d are described, based on the relations (8) or (9).

Log autocovariance estimate

Because relation (1.8) implies that the γ_j are eventually all positive, we can take logs for large enough j,

$$\log \gamma_j \approx \log c + (2d - 1)\log j, \text{ as } j \to \infty$$

This relation has the advantage of being linear in d. Robinson (1990) proposed substituting $\hat{\gamma}_j$ for γ_j and then carrying out an ordinary least squares regression of $\log \hat{\gamma}_j$ on $\log j$ for large j, with $\hat{\gamma}_j$ given in equation (4). This leads to the estimate

$$\hat{d}_1 = \frac{1}{2}\left\{1 + \frac{\sum_{j=n-r}^{n-1} \log \hat{\gamma}_j(\log j - \overline{\log j})}{\sum_{j=n-r}^{n-1} (\log j - \overline{\log j})^2}\right\} \tag{11}$$

where $\overline{\log j} = \Sigma_{j=n-r}^{n-1} \log j$, and r is a large integer less than n. No asymptotic distributional properties of $\hat{d}_1$ seem yet to have been obtained. However, it is anticipated that under condition (8) and additional regularity conditions there exist sequences r increasing more slowly than n such that $\hat{d}_1$ is consistent for d.

Minimum distance autocovariance estimate

Despite its computational advantages, $\hat{d}_1$ has the disadvantage that even if the γ_j are all positive for large j, some $\hat{\gamma}_j$ can be negative, especially when γ_j is close to zero. An alternative procedure due to Robinson (1990) is to minimize the squared distance between $\hat{\gamma}_j$ and cj^{2d-1} for large j, so that d and c are estimated by

$$(\hat{d}_2, \hat{c}_2) = \underset{c,d}{\text{argmin}} \sum_{j=n-r}^{n-1} (\hat{\gamma}_j - cj^{2d-1})^2 \tag{12}$$

Concentrating out c, we have

$$\hat{d}_2 = \underset{d}{\text{argmax}} \left\{ \sum_{j=n-r}^{n-1} \hat{\gamma}_j j^{2d-1} \right\}^2 \Big/ \sum_{j=n-r}^{n-1} j^{2(2d-1)} \tag{13}$$

The sets over which maximization/minimization is carried out in equations (12) and (13) will be typically compact with respect to d, such as the interval $[\varepsilon, \frac{1}{2} - \varepsilon]$ for some small ε. Again, no asymptotic properties of $\hat{d}_2$ seem yet to have been obtained, but once more it is likely that $\hat{d}_2$ is consistent for d, under regularity conditions and for a suitable sequence r.

Averaged periodogram estimate

An estimate of d which has been shown to be consistent for d, and under mild conditions, is due to Robinson (1991). This estimate is based on the limiting relation (9) for $f(\lambda)$ rather than on (8). Incidentally, while both conditions (8) and (9) hold simultaneously in case of the parametric models (6) and (10), these properties are not precisely equivalent, and, in particular, condition (9), unlike (8), does not imply that the autocovariances γ_j are all eventually positive. The estimate of d of Robinson (1991) employs an average of the periodogram (5) near zero frequency,

$$\hat{F}(\lambda_m) = 2\pi n^{-1} \sum_{j=1}^{m} I(\lambda_j)$$

where m is a positive integer less than n. Robinson (1991) showed under regularity conditions and the condition that $m \to \infty$ and $m/n \to 0$ as $n \to \infty$, that

$$(1 - 2d)\hat{F}(\lambda_m)/C\lambda_m^{1-2d} \to_p 1, \text{ as } \lambda \to 0^+ \tag{14}$$

indicating convergence in probability to the right-hand side. Now for a constant $q \in (0, 1)$ we likewise have

$$(1 - 2d)\hat{F}(q\lambda_m)/C(q\lambda_m)^{1-2d} \to_p 1, \text{ as } \lambda \to 0^+ \tag{15}$$

Differencing the logs of the left hand sides of (14) and (15) eliminates the scale factor C and suggests the estimate

$$\hat{d}_3 = \tfrac{1}{2} - \log(\hat{F}(q\lambda_m)/\hat{F}(\lambda_m))/(2 \log q) \tag{16}$$

Then Robinson (1991) showed that $\hat{d}_3$ is consistent for d under the same conditions as those underlying equation (14). Incidentally, these conditions seem mild, including only a moment condition on x of degree 2.

Log-periodogram regression estimate

Unfortunately no limiting distribution theory is yet available for the averaged periodogram estimate $\hat{d}_3$, and Robinson (1991) conjectured that although it may be asymptotically normal for $0 < d < \frac{1}{4}$, it may be asymptotically non-normal for $\frac{1}{4} \leqslant d < \frac{1}{2}$, and thus rather difficult to use as a basis for statistical inference. An alternative semi-parametric estimate of d, using the periodogram, was proposed by Geweke and Porter-Hudak (1983). They suggested regressing $\log I(\lambda_j)$ on $-\log(4\sin^2\lambda_j/2)$ over frequencies λ_j, $j = 1, \ldots, m$. The resulting d estimate is

$$\hat{d}_4 = -\frac{\sum_{j=1}^{m}\left(\log(4\sin^2\lambda_j/2) - m^{-1}\sum_{j=1}^{m}\log(4\sin^2\lambda_j/2)\right)\log I(\lambda_j)}{\sum_{j=1}^{m}\left(\log(4\sin^2\lambda_j/2) - m^{-1}\sum_{j=1}^{m}\log(4\sin^2\lambda_j/2)\right)^2} \tag{17}$$

Geweke and Porter-Hudak (1983) attempted a proof of asymptotic statistical properties of $\hat{d}_4$ only when $-0.5 < d < 0$ (in which $f(\lambda)$ is zero, not infinity, at zero frequency), but even in this case their proof was incorrect, as shown by Robinson (1992). Following a suggestion of Künsch (1986) that the very lowest frequencies be omitted from the regression, Robinson (1992) established the consistency and asymptotic normality of the estimate

$$\hat{d}_4 = -\frac{1}{2}\frac{\sum_{j=l+1}^{m}\left(\log\lambda_j - m^{-1}\sum_{j=l+1}^{m}\log\lambda_j\right)\log I(\lambda_j)}{\sum_{j=l+1}^{m}\left(\log\lambda_j - m^{-1}\sum_{j=l+1}^{m}\log\lambda_j\right)^2} \tag{18}$$

where l is a 'trimming number' which tends to infinity with m, but more slowly, where again m tends to infinity slower than n. (Note that $4\sin^2\lambda/2 \approx \lambda^2$ as $\lambda \to 0^+$, so there is no great significance in the use, in equation (18), of $-2\log\lambda_j$ as a regressor in place of $-\log(4\sin^2\lambda_j/2)$ in (17).) Specifically, Robinson (1992) showed that

$$2m^{1/2}(\hat{d}_4 - d) \xrightarrow{d} N(0, \pi^2/6) \tag{19}$$

The major drawback in the statistical theory provided by Robinson (1992) is that Gaussianity of x_t was assumed, unlike in the consistency proof of $\hat{d}_3$.

APPLICATION TO THE SPANISH INFLATION RATE

In this section we report applications of the several semi-parametric methods of estimating d to the difference log price series whose autocorrelations and periodogram were displayed in Figures 1 and 2. The peaks in the periodogram in Figure 2 at higher frequencies suggest some seasonal effects. Results for the seasonally differenced series $(1 - L^{12})\log(P_t)$ and $(1 - L)(1 - L^{12})\log(P_t)$ were also obtained but are not reported. The peaks are not very large and are not inconsistent with assumption (9), and possibly too small to warrant seasonal differencing.

There is an obvious problem with the application of $\hat{d}_1$, that even for large j many of the $\hat{\gamma}_j$ in Figure 1 are negative, while the model calls for all positive $\hat{\gamma}_j$ for large enough j up to $n - 1$. Thus $\hat{d}_1$ is not operational here, indeed Figure 1 may suggest that the eventually positive γ_j implication of relation (8) is unsatisfied. At the same time, it may be the case that the negative $\hat{\gamma}_j$ for j larger than about 350 are very close to zero and thus possibly not significantly negative. However, for an interval of 'large' j values between $j = 416$ and $j = 456$ all $\hat{\gamma}_j$ are positive, so one could 'trim out' the $\hat{\gamma}_j$ for $j > 456$ from $\hat{d}_1$. Figure 3 displays $\hat{d}_1$ for $r = 456 - 440 = 16$ to $r = 456 - 416 = 40$ with $n - 1$ the upper limit of summation in $\hat{d}_1$, replaced by 456. The estimates presented in this figure are very different from those obtained with the other three methods.

Next, $\hat{d}_2$ was implemented. Now the negative autocorrelations cause no problem and Figure 4 presents the results for $r = 540$ to $r = 572$. The d estimates are similar to those using

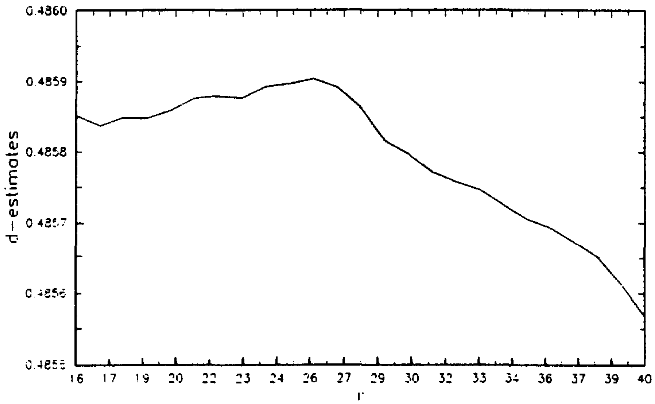

Figure 3. Log-autocovariance estimates (inflation rate x_t)

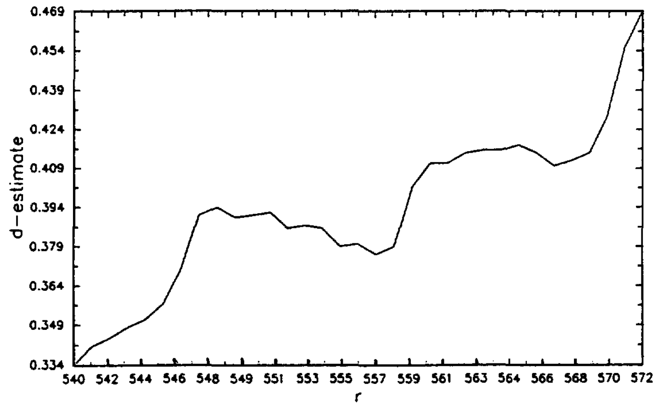

Figure 4. Minimum distance autocovariance estimates (inflation rate x_t)

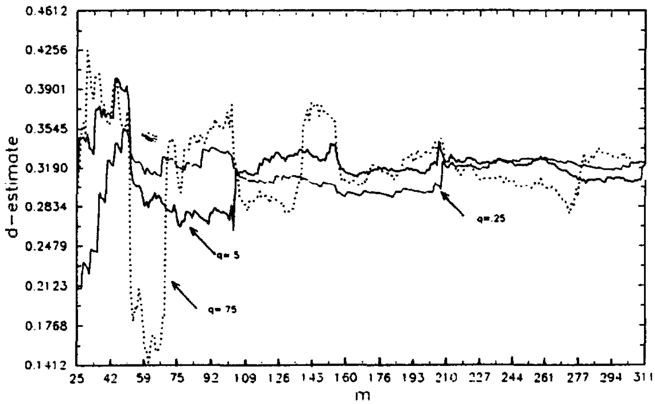

Figure 5. Average periodogram estimate (inflation rate x_t)

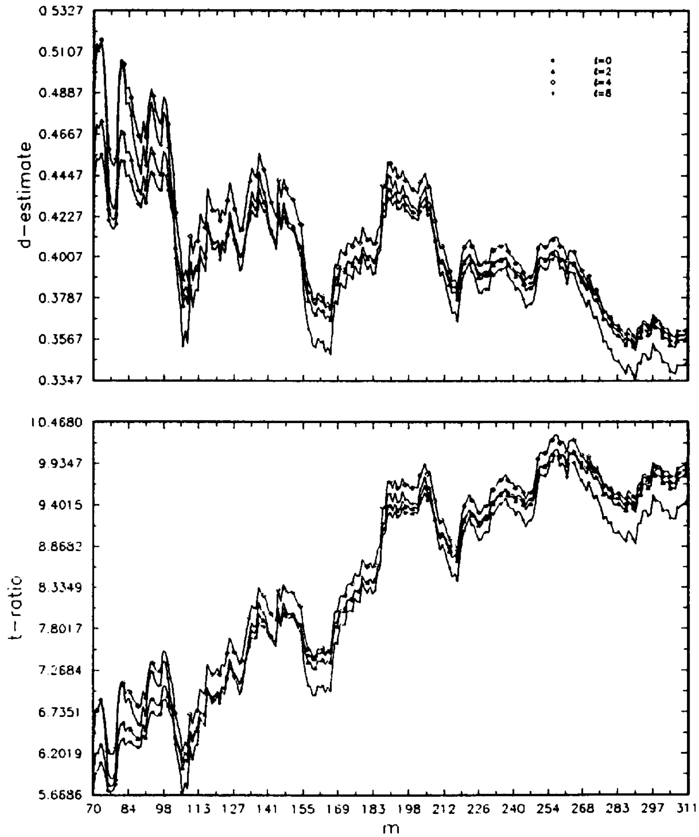

Figure 6. Log-periodogram regression estimate (inflation rate x_t)

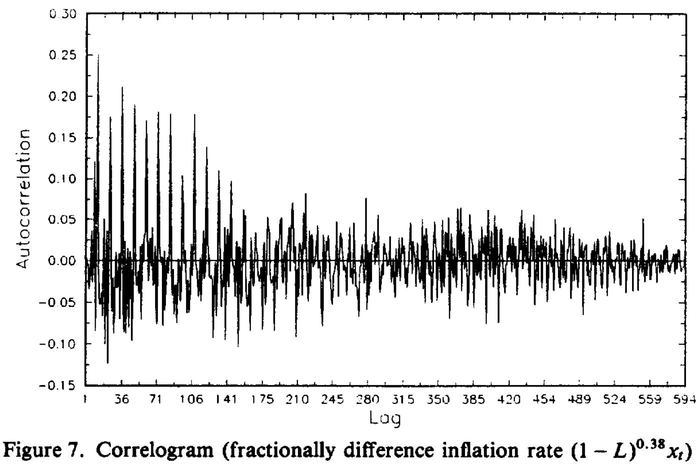

Figure 7. Correlogram (fractionally difference inflation rate $(1-L)^{0.38}x_t$)

$\hat{d}_4$, when r is in the interval (547, 569). The d estimates in this interval vary between 0.37 and 0.41. Outside this interval, the d estimates seem unreliable, varying monotonically with r.

The averaged periodogram estimate was computed for three different values of q ($q = 0.25$, 0.5 and 0.75) and for m between 17 and 300. The results are shown in Figure 5. Positive estimates of d were obtained throughout, usually ones of about 0.3, suggesting a substantial degree of long-memory. However, there is also a substantial degree of volatility for m less than 70 when $q = 0.25$ and in the case of the other values of q for somewhat smaller values of m. Even for the larger values of m, there is a fairly distinct sensitivity to q, although the estimates do seem to stabilize to values between about 0.3 and 0.35.

Results for the log-periodogram regression estimates $\hat{d}_4$ are displayed in the upper part of Figure 6 for values $l = 0$, 2, 4,, 8 of the trimming number and m between 70 and 311. The

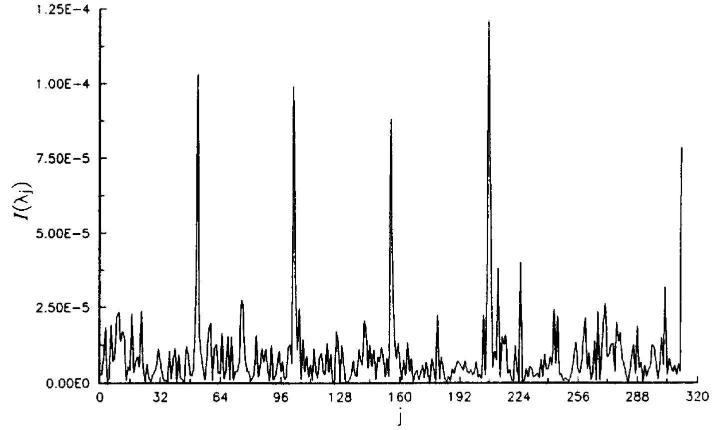

Figure 8. Periodogram (fractionally difference inflation rate $(1 - L)^{0.38} x_t$)

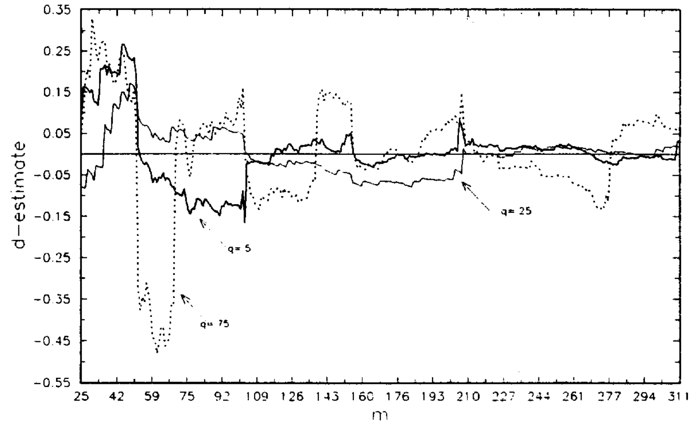

Figure 9. Average periodogram estimate (fractionally difference inflation rate $(1 - L)^{0.38} x_t$)

estimates are not very sensitive to l. The estimate is very unstable for m's in the interval (70, 113). However, there is a reasonable degree of stability over m when $m > 113$. All the estimates were above 0.34 and in the lower part of Figure 6, t-ratios based on the central limit result (19) are displayed, suggesting that d is significantly larger than zero.

The results described above consistently indicate that this inflation series suffers from long-memory. We fractionally differenced the inflation rate series with a $d = 0.38$. The correlogram and periodogram for the resulting filtered series $e_t = (1 - L)^{0.38} x_t$, $t > 0$, are plotted in Figures 7 and 8. The autocorrelation estimates are very close to zero. However, the seasonal peaks are still present.

Figures 9 and 10 show the averaged periodogram and log-periodogram d estimates for the e_t series. The values of m used are the same as in the application to the original series. The d estimates employing both methods are very close to zero. The t-ratios shown in Figure 10, based on the log-periodogram estimate, are all below the asymptotic normal critical values at 1% of significance. Thus it seems that the long-memory has been removed by fractional differencing.

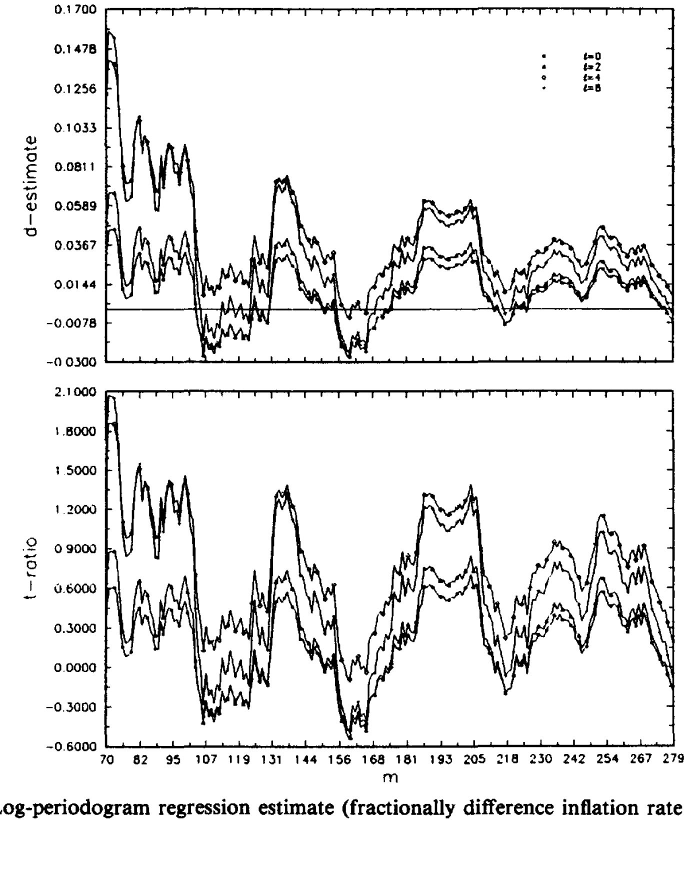

Figure 10. Log-periodogram regression estimate (fractionally difference inflation rate $(1 - L)^{0.38} x_t$)

ACKNOWLEDGEMENTS

This article is based on research funded by the Economic and Social Research Council (ESRC) reference number R000233609 and by the Spanish Dirección General de Investigación Científica y Técnica (DGYCIT) reference number PB92-0247. We thank Miguel A. Benito for excellent research assistance.

REFERENCES

Adenstedt, R. 'On large-sample estimation for the mean of a stationary random sequence', *Annals of Statistics*, **2** (1984), 1095–1107.

De Ojeda Eiseley, A., 'Indices de precios en España en el periodo 1913–1987', *Estudios de Historia Económica*, No. 17 (1988), Banco de España. Servicio de Estudios.

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

Künsch, H. R., 'Discrimination between monotonic trends and long-range dependence', *Journal of Applied Probability*, **23** (1986), 1025–1030.

Mandelbrot, B. and Van Ness, J. V., 'Fractional Brownian motion, fractional noise and applications', *SIAM Review*, **10** (1968), 422–437.

Robinson, P. M., 'Time series with strong dependence', Invited Paper, 1990 World Congress of the Econometric Society, Forthcoming in *Advances in Econometrics: Six World Congress*, Cambridge: Cambridge University Press.

Robinson, P. M., 'Semiparametric analysis of long-memory time series', Preprint (1991).

Robinson, P. M., 'Log-periodogram regression for time series with long range dependence', Preprint (1992).

Authors' biographies:
Miguel A. Delgado is Professor of Econometrics at Universidad Carlos III de Madrid. He received his PhD from the London School of Economics. His primary research interests are econometric theory and non-parametric and semi-parametric inference.
Peter M. Robinson is Professor of Econometrics at the London School of Economics. He received his PhD from The Australian National University. His primary research interests are econometric theory, time-series analysis and non-parametric and semi-parametric estimation.

Authors' addresses:
Miguel A. Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Calle Madrid 126-128, Getafe, Madrid 28903, Spain.
Peter M. Robinson, Department of Economics, The London School of Economics, Houghton Street, London WC2A 2AE, UK.