Working Paper 93-02
Statistics and Econometrics Series 02
February 1993

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

NEW METHODS FOR THE ANALYSIS OF LONG MEMORY TIME SERIES: APPLICATION TO SPANISH INFLATION

Miguel A. Delgado and Peter M. Robinson*

Abstract

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

Key Words

Long Memory; Differencing Parameters; Semiparametric Estimation; Autocovariance; Averaged Periodogram Regression; Inflation Rate.

*Delgado, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, Robinson, London School of Economics. This article is based on research funded by the Economic and Social Research Council (ESRC) reference number: R00023609. We thank Miguel A. Benito for excellent research assistantship.

1. 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,... which is observed at t=1,2,...,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

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

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, on the one hand, the summability condition

$$\sum_{j=-\infty}^{\infty} |\gamma_j| < \infty, \tag{1.1}$$

or the boundedness restriction

$$f(0)< \infty. \tag{1.2}$$

In particular, stationary autoregressive moving average (ARMA) models imply autocovariances that decay exponentially as $j\longrightarrow\infty$, 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 (1.1) and (1.2).

Empirical observation, nevertheless, is sometimes consistent with models which do not satisfy (1.1) or (1.2). 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)[2]. Thus, we have n+1= 628 observations. The original series, P_t, t=0, 1, ..., appears nonstationary, but the series

$$x_t = \log(P_t) - \log(P_{t-1}), \quad t=1,\dots,n, \tag{1.3}$$

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

$$\hat{\gamma}_j = n^{-1} \sum_{t=1}^{n-j} (x_t - \bar{x})(x_{t-j} - \bar{x}), \quad j= 0,1,\dots,n-1, \tag{1.4}$$

where $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$. While the $\hat{\rho}_j$ do appear to decay as j increases, they do so slowly, in a manner that could be consistent with the failure of (1.1). Figure 2 plots the periodogram

$$I(\lambda) = (2\pi n)^{-1} \left|\sum_{t=1}^{n} x_t \, e^{it\lambda_j}\right|^2, \tag{1.5}$$

at frequencies $\lambda = \lambda_j = 2\pi j/n$, for j=1,...,(n-1)/2. Although $I(\lambda)$ does not provide a consistent estimate of $f(\lambda)$, smoothed versions of $I(\lambda)$ used to estimate $f(\lambda)$ and the spectral peak as λ approaches 0 suggests that (1.2) may fail (note that formula (1.5) is unaffected by a nonzero $\bar{x}$ when evaluated at frequency λ_j for integer j).

Figures 1 and 2 about here

Parametric models for stationary series which violate (1.1) and (1.2) have long been available. One of these is the "fractional noise" process with autocovariances

[2] The series has been created linking 5 price indeces for different periods and with 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.

$$\gamma_j = \frac{1}{2}\gamma_0 \left\{ |j+1|^{2d+1} - 2|j|^{2d+1} + |j-1|^{2d-1} \right\}, \quad j=1,2,\ldots, \tag{1.6}$$

where

$$0 < d < 1/2. \tag{1.7}$$

It may be shown that

$$\gamma_j \approx c\, j^{2d-1}, \text{ as } j \to \infty, \tag{1.8}$$

for some c satisfying $0 < c < \infty$. In view of (1.7) and (1.8), it is clear that (1.1) does not hold. It may also be shown that

$$f(\lambda) \approx C\, \lambda^{-2d}, \text{ as } \lambda \longrightarrow 0^+, \tag{1.9}$$

for some C satisfying $0 < C < \infty$. In view of (1.7) and (1.9), it is clear that (1.2) also does not hold. We call series which violate (1.1) and (1.2) "long-memory". Early work on long-memory time series of B. Mandelbrot and his coauthors (see e.g. Mandelbrot and van Ness, 1968) stressed model (1.6). However, (1.6) is a very parsimonious model which implies that γ_j and $f(\lambda)$ decay monotonically as j and λ increases, 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

$$(1-L)^d\, a(L)\, (x_t - \mu) = b(L)\, e_t, \quad t=1,2,\ldots, \tag{1.10}$$

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 $\{e_t,\ t \geq 1\}$ is a sequence of uncorrelated random variables with zero mean and unknown, positive, finite variance σ^2. Again, it may be shown that (1.8) and (1.9) hold, and thus that (1.1) and (1.2) 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 behavior in γ_0 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 (1.10) 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 (1.8) and (1.9). These can be called "semiparametric" models because they parameterize only the long run characteristics of x_t, while allowing the short- and medium-run characteristics to be nonparametric. 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 semiparametric model-based procedures is relevant.

Several methods of estimating the semiparametric models (1.8) and (1.9) have been introduced or developed by Robinson (1990, 1991, 1992). These are described in the following section. In Section 3, applications of the methods to the Spanish price index series are reported.

2. PARAMETER ESTIMATES

In the current section four alternative estimates of the differencing parameter d are described, based on the relations (1.8) or (1.9).

1. Log autocovariance estimate

Because (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 (1.4). This leads to the estimate

$$\hat{d}_1 = 1/2 \left\{ 1 + \frac{\sum_{j=n-r}^{n-1} \log \hat{\gamma}_j \, (\log j - \overline{\log j})}{\sum_{j=n-r}^{n-1} (\log j - \overline{\log j})^2} \right\}, \tag{2.1}$$

where $\overline{\log j} = \sum_{j=n-r}^{n-1} \log j$, 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 (1.8) and additional regularity conditions, there exist sequences r increasing more slowly than n such that $\hat{d}_1$ is consistent for d.

2. 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, specially 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 c j^{2d-1} for large j, so that d and c are estimated by

$$(\hat{d}_2, \hat{c}_2) = \underset{c,\, d}{\operatorname{argmin}} \sum_{j=n-r}^{n-1} (\hat{\gamma}_j - c\, j^{2d-1})^2. \tag{2.2}$$

Concentrating out c, we have

$$\hat{d}_2 = \underset{d}{\operatorname{argmax}} \left\{ \sum_{j=n-r}^{n-1} \hat{\gamma}_j \, j^{2d-1} \right\}^2 / \sum_{j=n-r}^{n-1} j^{2(2d-1)}. \tag{2.3}$$

The sets over which maximization/minimization is carried out in (2.2) and (2.3) will be typically compact with respect to d, such as the interval $[\varepsilon, 1/2-\varepsilon]$ for some small ε. Again, no asymptotic properties of $\hat{d}_2$ seem yet to have been obtained, but again it is likely that $\hat{d}_2$ is consistent for d, under regularity conditions and for a suitable sequence r.

3. 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 (1.9) for $f(\lambda)$, rather than on (1.8). Incidentally, while both (1.8) and (1.9) hold simultaneously in case of the parametric models (1.6) and

(1.10), these properties are not precisely equivalent, and in particular (1.9), unlike (1.8), 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 (1.5) 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 $m/n \longrightarrow 0$ as $n \longrightarrow \infty$, that

$$(1-2d)\, \hat{F}(\lambda_m)/\, C\, \lambda_m^{1-2d} \longrightarrow_p 1, \text{ as } \lambda \longrightarrow 0^+, \tag{2.4}$$

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

$$(1-2d)\, \hat{F}(q\, \lambda_m)/\, C\, (q\, \lambda_m)^{1-d} \longrightarrow_p 1, \text{ as } \lambda \longrightarrow 0^+. \tag{2.5}$$

Differencing the logs of (2.4) and (2.5) eliminates the scale factor C and suggests the estimate

$$\hat{d}_3 = 1/2 - \mathit{log}\, (\hat{F}(q\, \lambda_m)/\hat{F}(\lambda_m))/(2\mathit{log}\, q). \tag{2.6}$$

Then Robinson (1991) showed that $\hat{d}_3$ is consistent for d under the same conditions as those underlying (2.4). Incidentally, these conditions seem mild, including only a moment condition on ε of degree 2.

4. <u>Log-periodogram regression estimate</u>.

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

$$\hat{d}_4 = - \frac{\sum_{j=1}^{m}\left[log(4 \sin^2 j/2) - m^{-1} \sum_{j=1}^{m} log(4 \sin^2 \lambda_j/2)\right] log\ I(\lambda_j)}{\sum_{j=1}^{m}\left[log(4 \sin^2 j/2) - m^{-1} \sum_{j=1}^{m} log(4 \sin^2 \lambda_j/2)\right]^2} . \quad (2.7)$$

Geweke and Porter-Hudak (1983) attempted a proof of asymptotic statistical properties of $\hat{d}_4$ only in case $-0.5 < d < 0$ (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

$$\hat{d}_4 = - 1/2 \frac{\sum_{j=\ell+1}^{m}\left(log\ \lambda_j - m^{-1}\sum_{j=\ell+1}^{m} log\ \lambda_j\right) log\ I(\lambda_j)}{\sum_{j=\ell+1}^{m}\left(log\ \lambda_j - m^{-1}\sum_{j=\ell+1}^{m} log\ \lambda_j\right)^2}, \quad (2.8)$$

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 $\lambda \longrightarrow 0^+$, so there is not great significance in the use, in (2.8), of $-2log\ \lambda_j$ as regressor in place of $-\ log\ (4\ sin^2\ \lambda_j/2)$ in (2.9)). Specifically, Robinson (1992) showed that

$$2m^{1/2}\ (\hat{d}_4 - d) \longrightarrow_d N(0,\ \pi^2/6), \quad (2.9)$$

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 $\hat{d}_3$.

3. APPLICATION TO SPANISH INFLATION RATE

In this section we report applications of the several semiparametric methods of estimating d to the differenced log price series whose autocorrelations and periodogram were displayed in Figures 1 and 2.

The bumps in the periodogram in Figure 2 at higher frequencies are suggestive of some seasonal effects. Results for the seasonally differenced series $(1-L^{12})\ P_t$ and $(1-L)\ (1-L^{12})\ P_t$ were also obtained but are not reported. The bumps are not very large and not inconsistent with assumption (1.9), 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 (1.8) 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 though 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 through 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.

Figure 3 about here

Next, $\hat{d}_2$ was implemented. Now the negative autocorrelations cause no problem and Figure 4 presents the results for r=540 through r=572. The d estimates are similar to those using $\hat{d}_4$, 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.

Figure 4 about here

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

Figure 5 about here

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 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 (2.9) are displayed suggesting, that d is significantly larger than zero.

The results described above, consistently suggest 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.

Figure 7 and 8 about here

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 showed 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 the fractional differencing.

Figure 9 and 10 about here

REFERENCES

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

DE OJEDA EISELEY, A. (1988). Indices de Precios en España en el Periodo 1913-1987. *Estudios de Historia Económica, nº 17, Banco de España. Servicio de Estudios.*

GEWEKE, J. and S. PORTER-HUDAK (1983) The estimation and application of long-memory time series models. *Journal of Time Series Analysis* 4, 221-238.

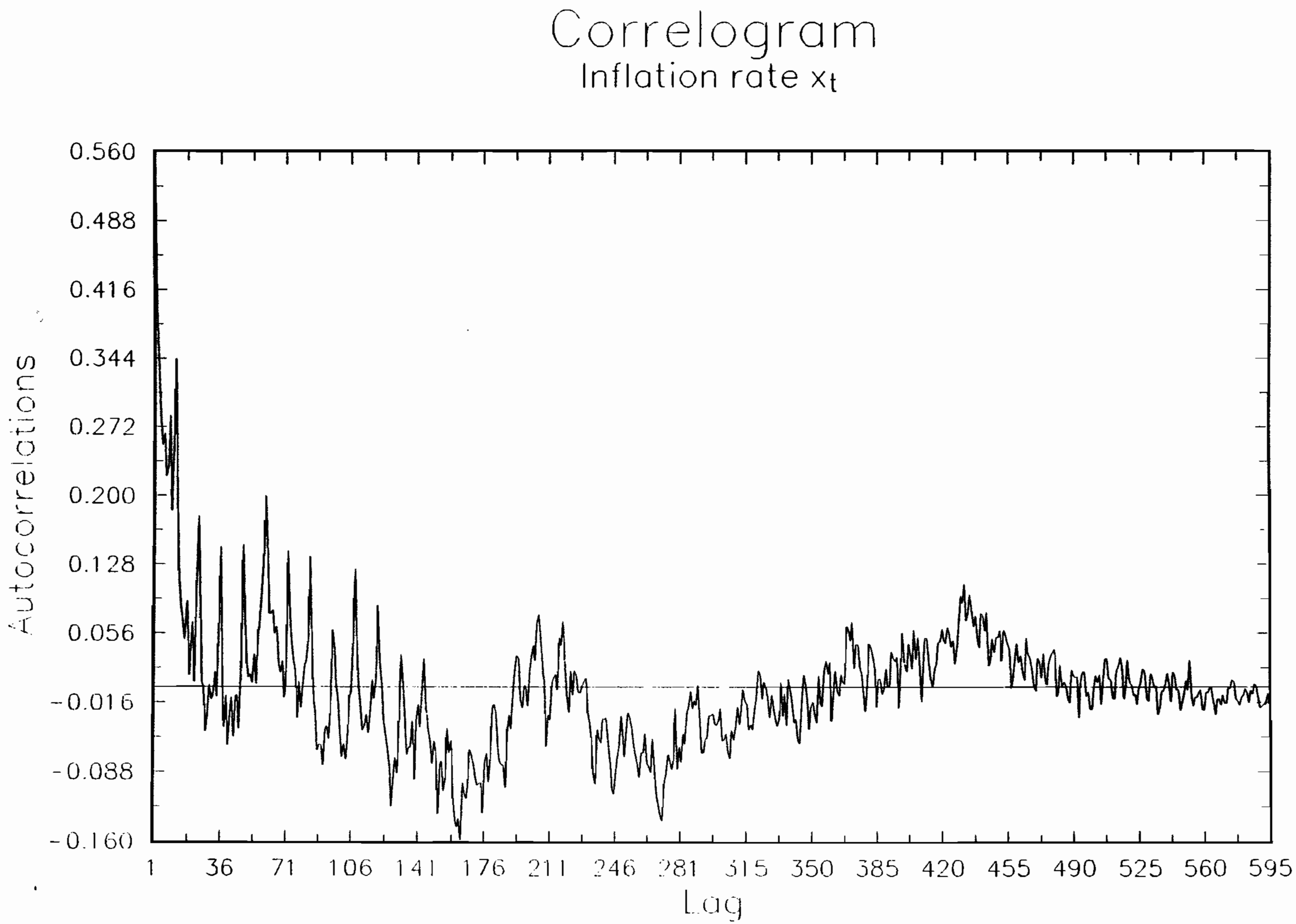
KÜNCH, H.R. (1986). Discrimination between monotonic trends and long-range dependence. *Journal of Apply Probability* 23, 1025-1030.

MANDELBROT, B., and J.V. VAN NESS (1968) Fractional Brownian motion, fractional noise and applications. *SIAM Review* 10, 422-437.

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

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

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



Correlogram
Inflation rate x_t
Autocorrelations
0.560
0.488
0.416
0.344
0.272
0.200
0.128
0.056
-0.016
-0.088
-0.160
1
36
71
106
141
176
211
246
281
315
350
385
420
455
490
525
560
595
Lag

Figure 1

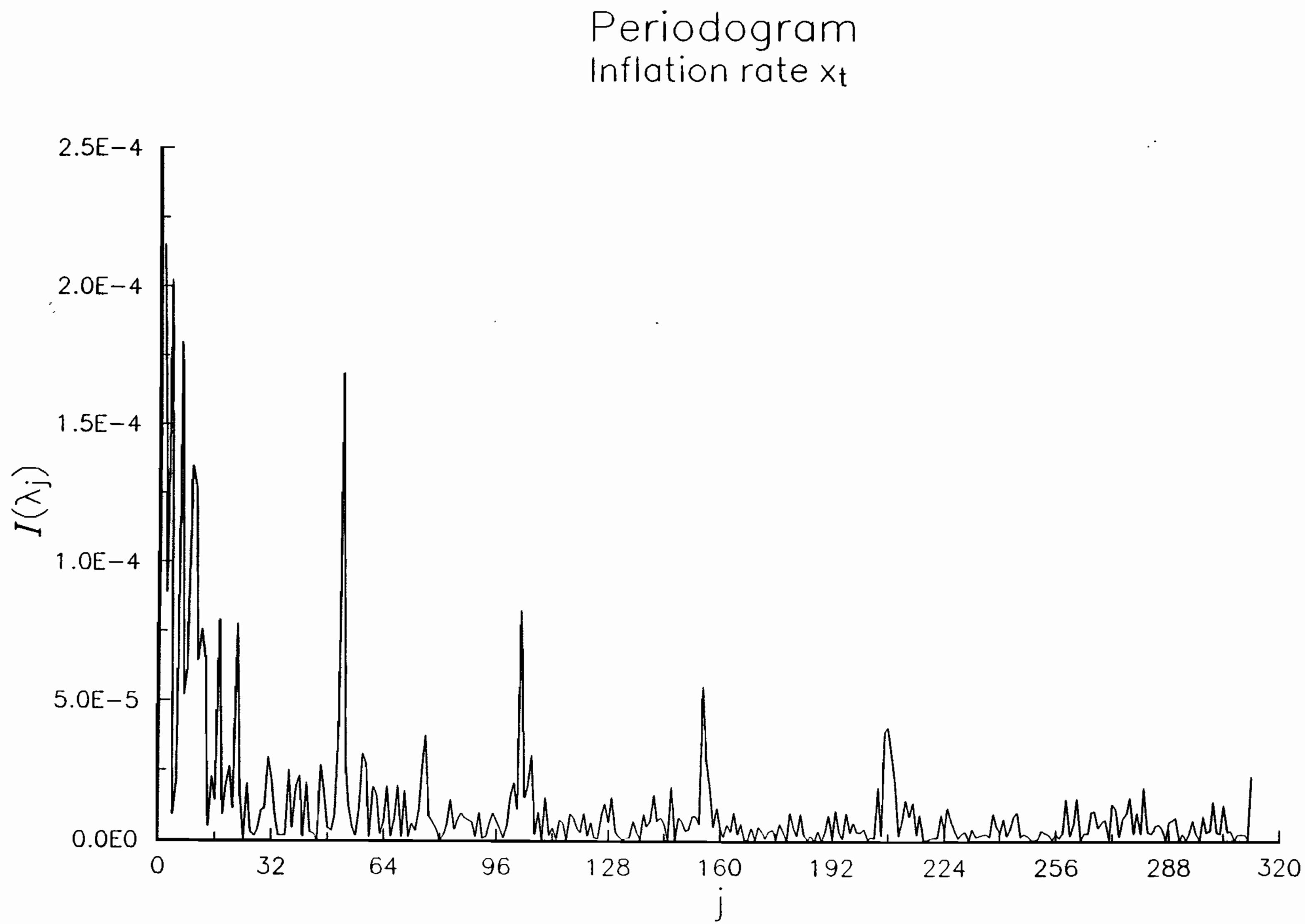

Periodogram
Inflation rate xt
2.5E−4
2.0E−4
1.5E−4
1.0E−4
5.0E−5
0.0E0
I(λj)
0
32
64
96
128
160
192
224
256
288
320
j

Figure 2

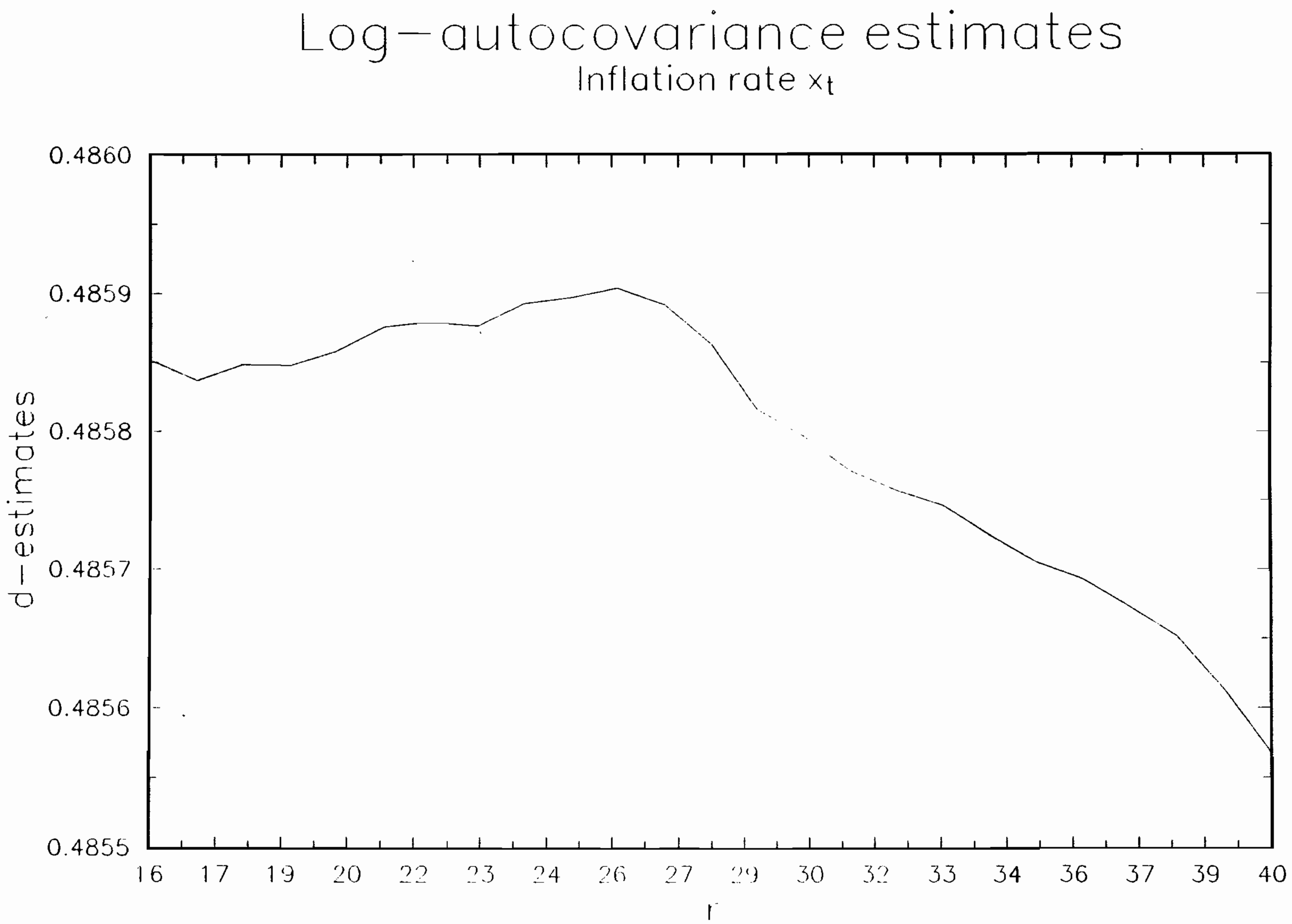

Log−autocovariance estimates
Inflation rate x_t
d−estimates
0.4860
0.4859
0.4858
0.4857
0.4856
0.4855
16 17 19 20 22 23 24 26 27 29 30 32 33 34 36 37 39 40
r

Figure 3

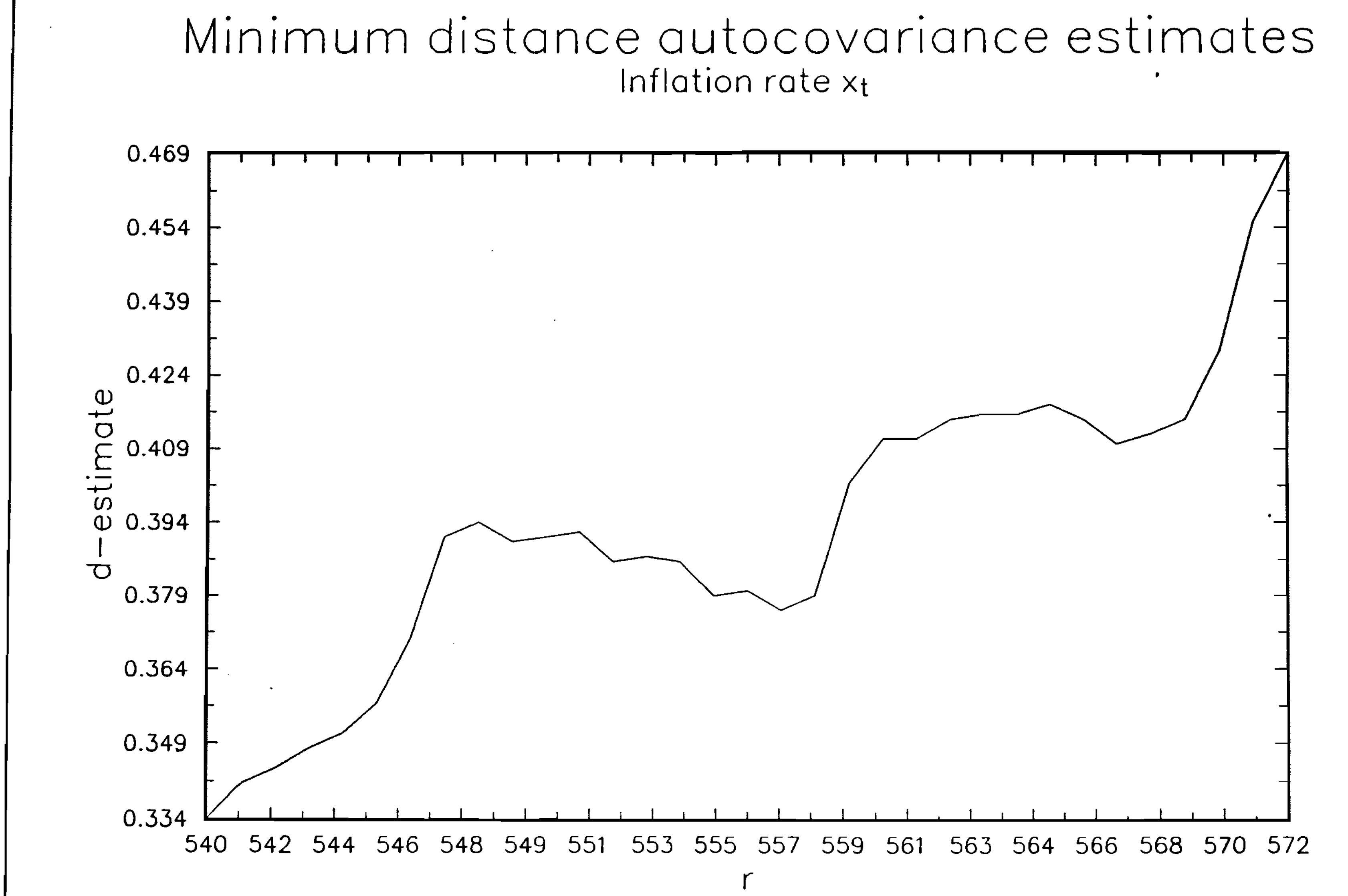

Minimum distance autocovariance estimates
Inflation rate x_t
d−estimate
0.469
0.454
0.439
0.424
0.409
0.394
0.379
0.364
0.349
0.334
540 542 544 546 548 549 551 553 555 557 559 561 563 564 566 568 570 572
r

Figure 4

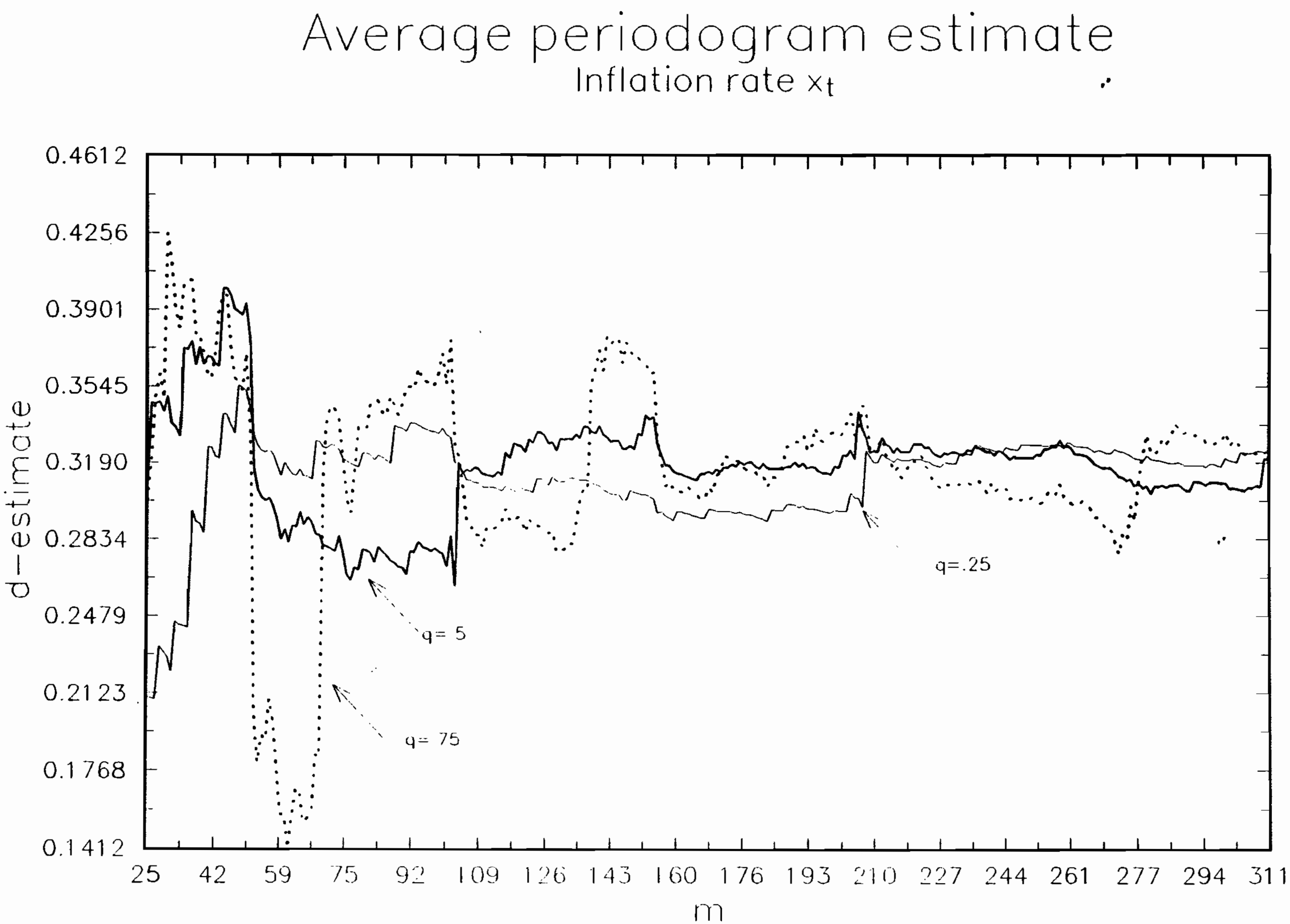
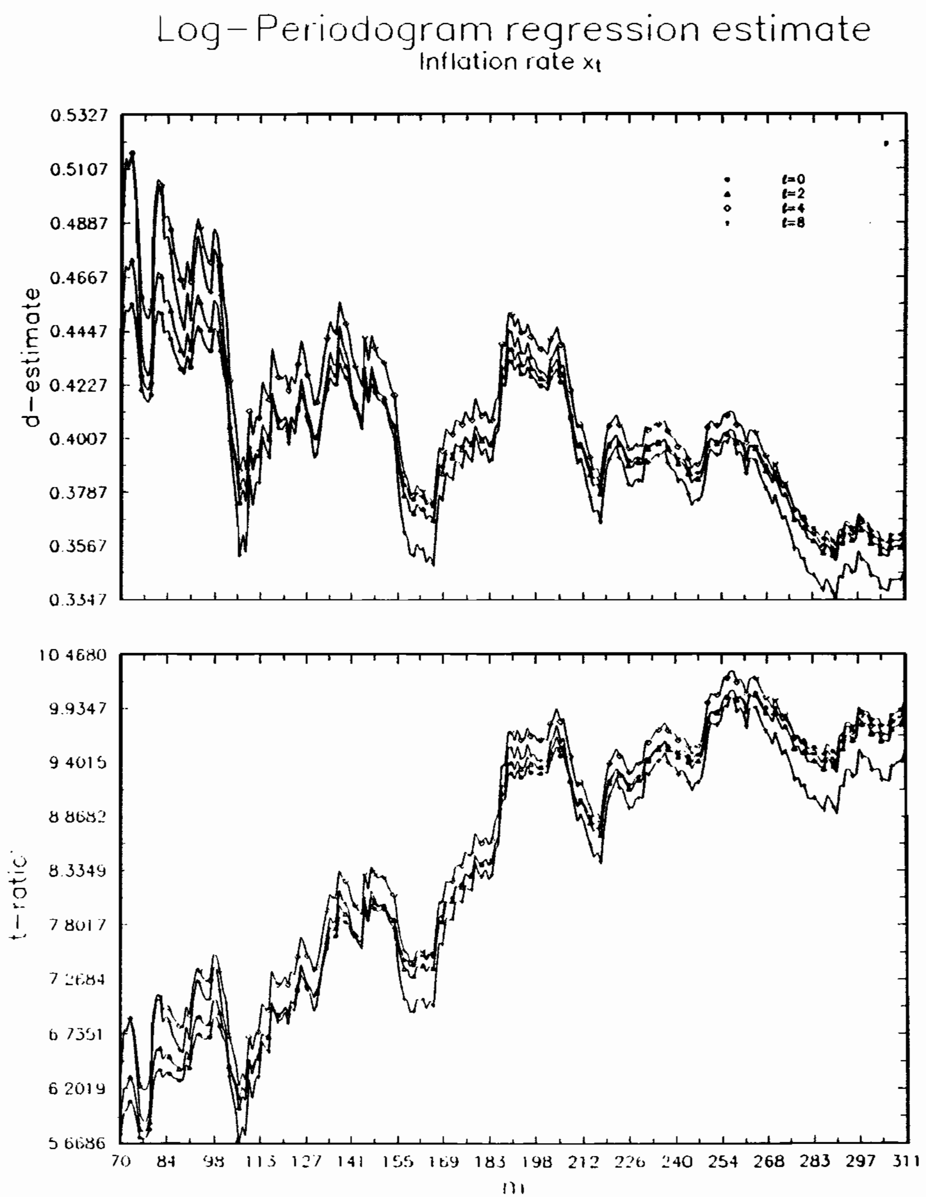
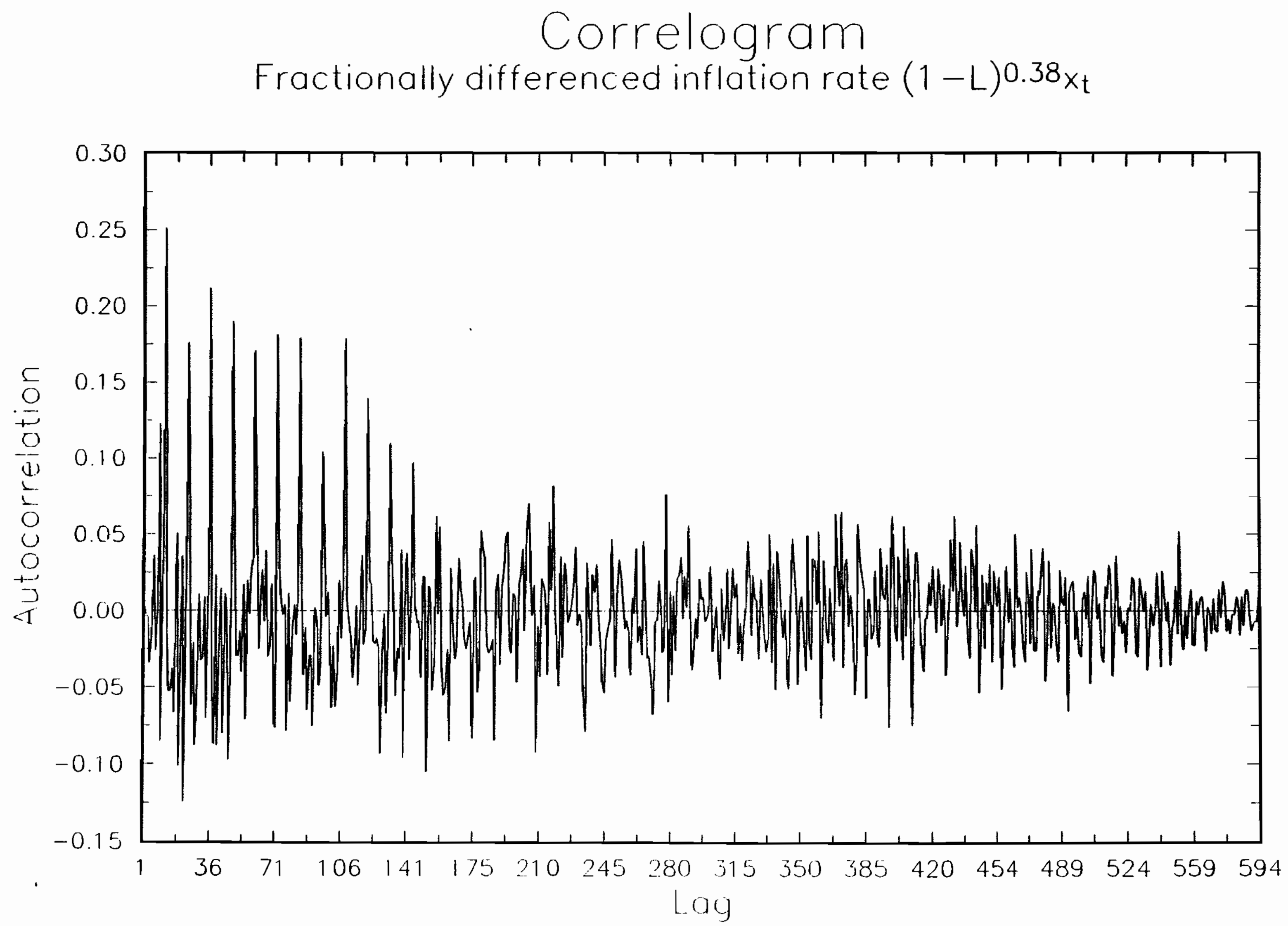

Average periodogram estimate
Inflation rate x_t
d-estimate
0.4612
0.4256
0.3901
0.3545
0.3190
0.2834
0.2479
0.2123
0.1768
0.1412
25 42 59 75 92 109 126 143 160 176 193 210 227 244 261 277 294 311
m
q= 5
q= 75
q=.25

Figure 5



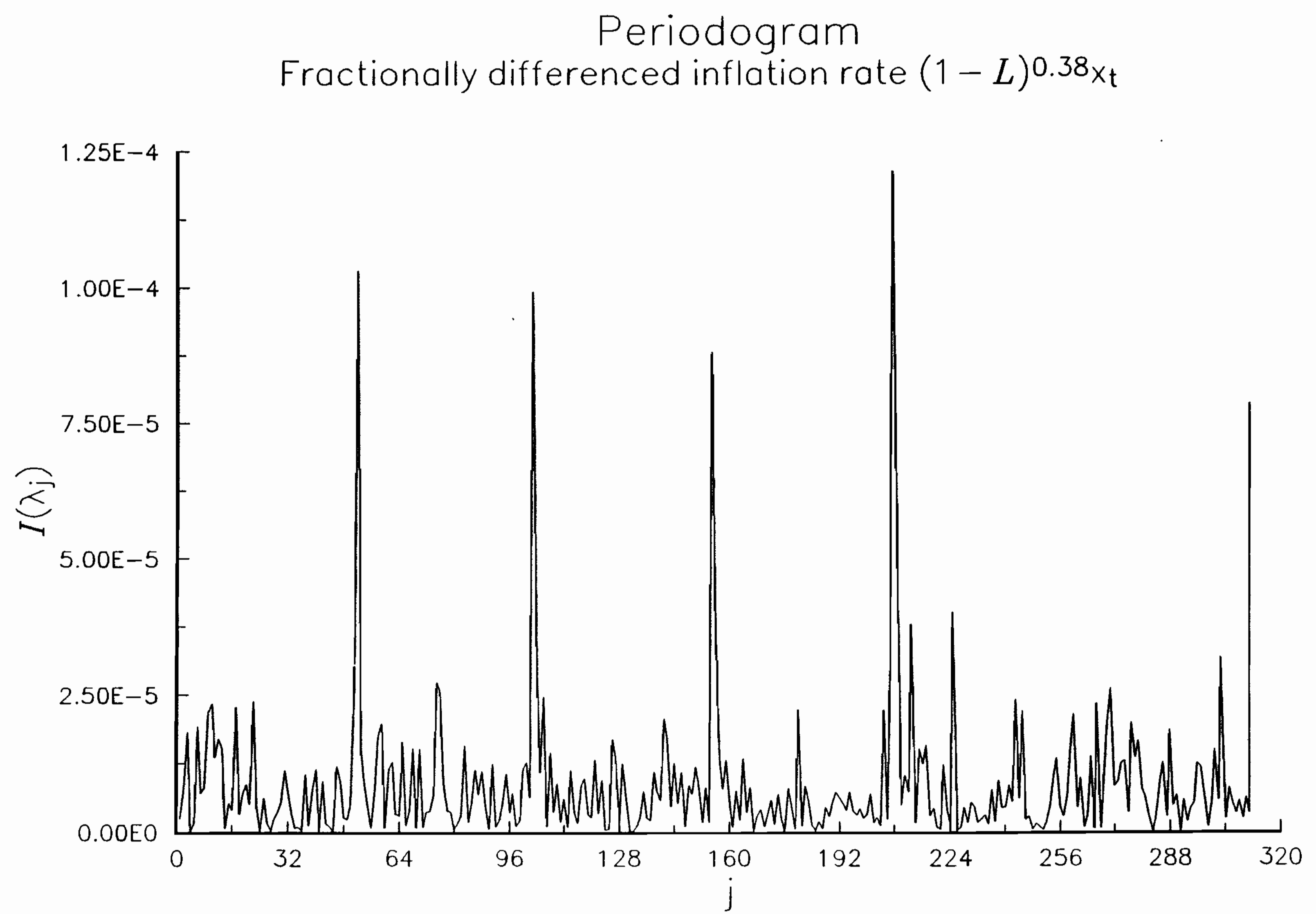
Log–Periodogram regression estimate
Inflation rate x_t
d–estimate
0.5327
0.5107
0.4887
0.4667
0.4447
0.4227
0.4007
0.3787
0.3567
0.3347
ℓ=0
ℓ=2
ℓ=4
ℓ=8
t–ratio
10.4680
9.9347
9.4015
8.8682
8.3349
7.8017
7.2684
6.7351
6.2019
5.6686
70 84 98 113 127 141 155 169 183 198 212 226 240 254 268 283 297 311
m

Figure 6



Correlogram
Fractionally differenced inflation rate (1 −L)0.38xt
Autocorrelation
0.30
0.25
0.20
0.15
0.10
0.05
0.00
−0.05
−0.10
−0.15
1
36
71
106
141
175
210
245
280
315
350
385
420
454
489
524
559
594
Lag

Figure 7



Periodogram
Fractionally differenced inflation rate $(1-L)^{0.38}x_t$
1.25E−4
1.00E−4
7.50E−5
5.00E−5
2.50E−5
0.00E0
$I(\lambda_j)$
0
32
64
96
128
160
192
224
256
288
320
j

Figure 8

Average periodogram estimate

Fractionally differenced inflation rate $(1-L)^{0.38}x_t$

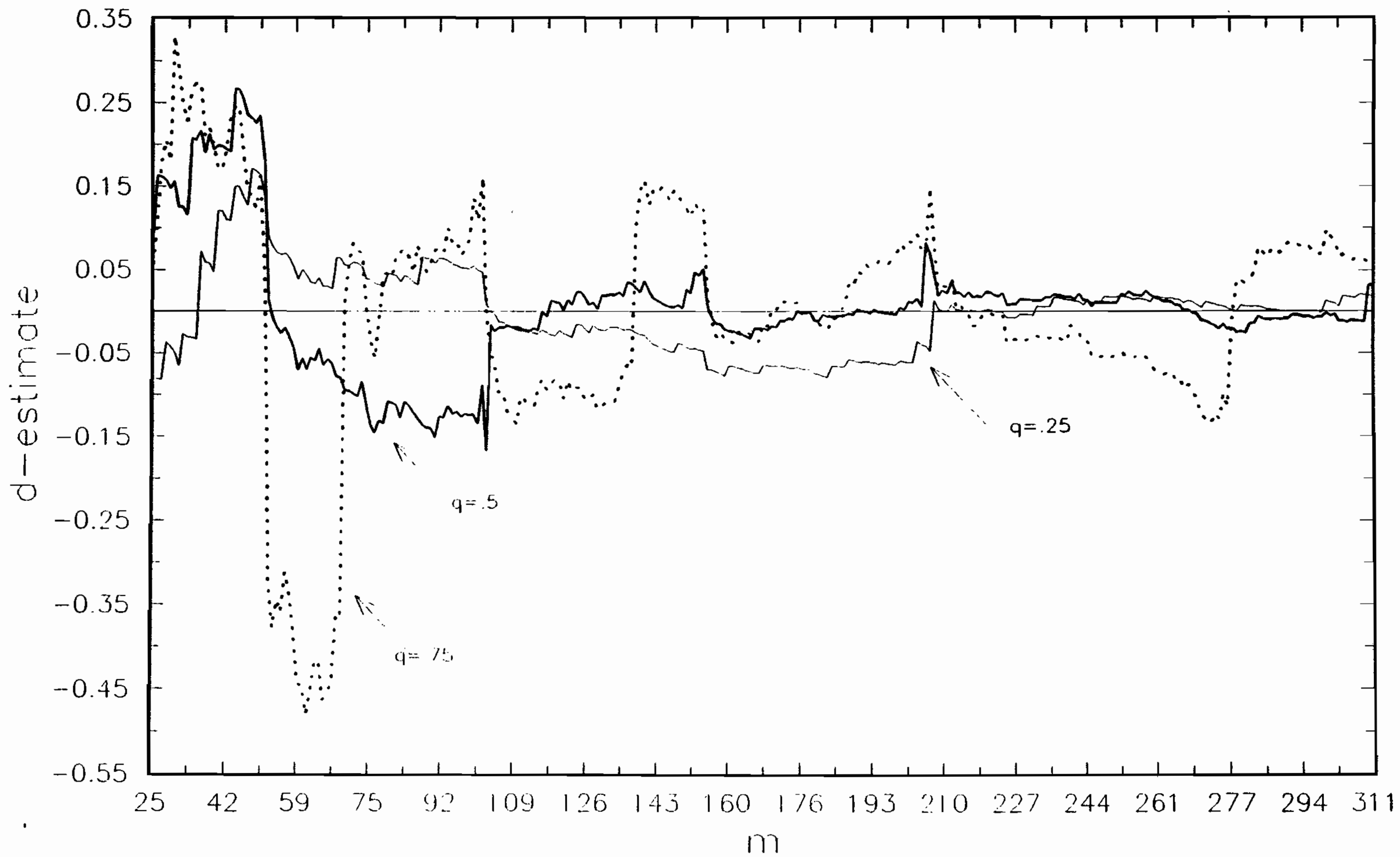
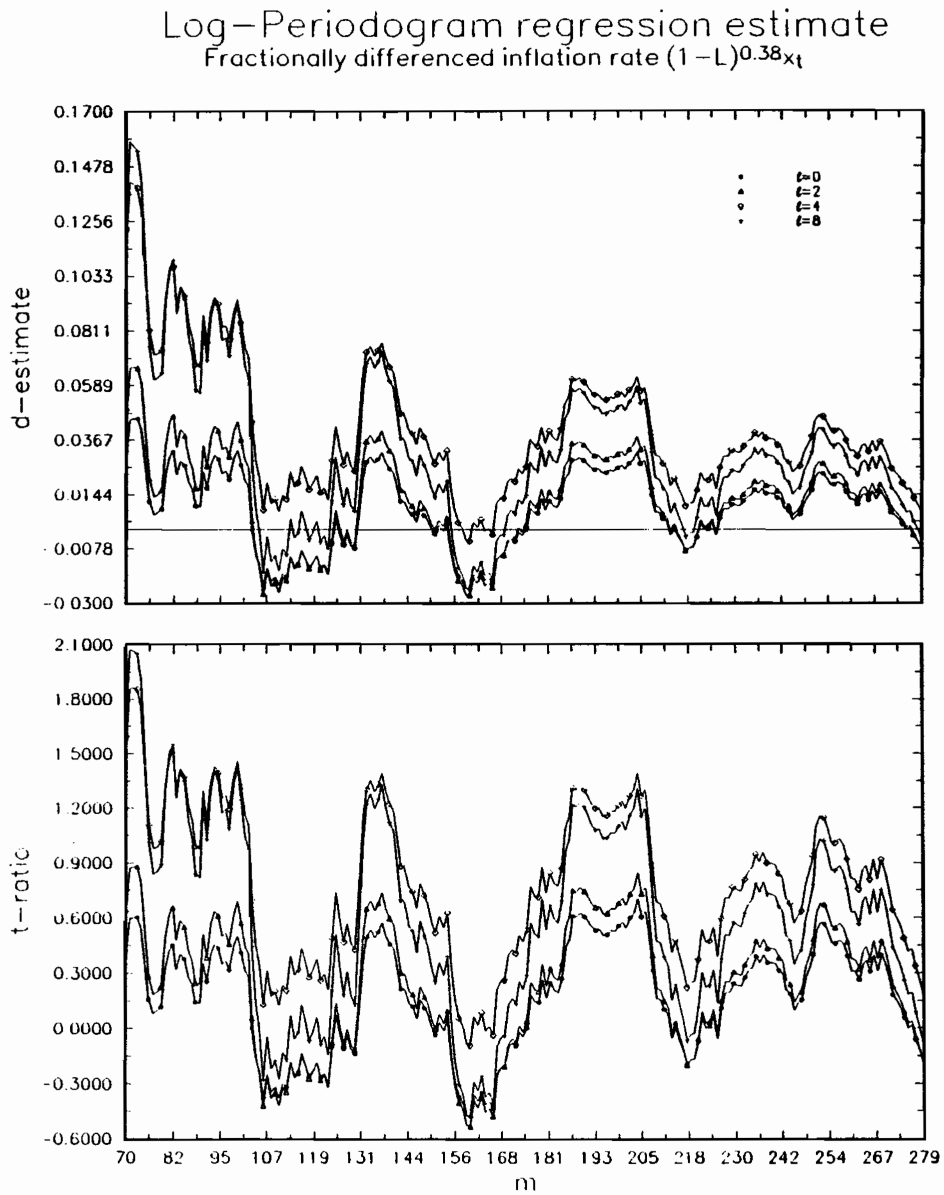


Figure 9



Log−Periodogram regression estimate
Fractionally differenced inflation rate (1 − L)0.38xt
d−estimate
t−ratio
ℓ=0
ℓ=2
ℓ=4
ℓ=8
m

Figure 10