

BEHAVIOR OF SAMPLE MEANS AND
 PARAMETRIC TIME SERIES ESTIMATION

H. Joseph Newton
 Department of Statistics
 Texas A&M University
 College Station, TX 77843, U.S.A.

ABSTRACT

The standard error of the sample mean for autocorrelated data is directly proportional to the value of the spectral density evaluated at zero frequency of the process being sampled. Thus confidence intervals for the true mean using traditional formulas can be greatly in error.

This paper describes both parametric and nonparametric methods of spectral density estimation and illustrates numerically the basic results using a personal computer program called TIMESLAB which acts as a laboratory for studying such problems.

The results of the paper indicate that in many situations, adjusting for autocorrelation is easily performed.

1. INTRODUCTION AND BASIC THEORETICAL FRAMEWORK

Let Z be the set of integers and let X denote a discrete-time, continuous-space time series $\{X(t), t \in Z\}$, i.e. at each integer time point t , the random variable $X(t)$ is considered to be absolutely continuous.

We say that X is covariance stationary if the mean of $X(t)$ is a constant μ independent of t and the covariance of any two X 's separated by an integer time lag v is also independent of time. Thus

$$E(X(t)) = \mu,$$

$$\text{Cov}(X(t), X(t+v)) = R(v),$$

for $t, v \in Z$,

where E and Cov denote the expectation and covariance operator.

In order to obtain useful results about μ we need to assume that the autocovariance function $R(v)$ decays to zero at a reasonable

rate as $v \rightarrow \infty$. For example, if X 's fifty time units apart are correlated and we only have a sample of length thirty, then there is important information that cannot be obtained from the data without some assumption about the decay of $R(v)$.

If all joint distributions of a covariance stationary time series are multivariate normal, then knowledge of μ and R completely characterize the probabilistic behavior of X . Further, many of the basic results of time series analysis are valid as the sample length goes to infinity even if X is not a normal time series.

If $R(v)$ tends to zero as v tends to infinity quickly enough that

$$\sum_{v=-\infty}^{\infty} |R(v)| < \infty,$$

then we can define the Fourier Transform of R

$$f(\omega) = \sum_{v=-\infty}^{\infty} R(v)e^{-2\pi i v \omega}, \quad \omega \in [0, 1] \quad (1)$$

and can invert the Fourier transform to obtain

$$R(v) = \int_0^1 f(\omega)e^{2\pi i v \omega} d\omega, \quad v \in Z. \quad (2)$$

The function f is called the spectral density function of X and is mathematically equivalent to R as f and R are Fourier pairs as defined by (1) and (2).

If we normalize $R(v)$ by the variance $R(0)$ of X , we obtain the autocorrelation function $\rho(v)$, defined by

$$\rho(v) = \text{Corr}(X(t), X(t+v)) = \frac{R(v)}{R(0)}, \quad v \in Z.$$

Thus we can write

$$f(0) = \sum_{v=-\infty}^{\infty} R(v) \\ = R(0) \sum_{v=-\infty}^{\infty} \rho(v) .$$

The basic statistical properties for \bar{X} are given in the following Theorem, details of which can be found in standard Time Series Analysis texts (see Priestley (1981), section 5.2 for example).

Theorem 1.1

Let X be a covariance stationary time series having autocovariance function R , autocorrelation function ρ and spectral density function f . Let $X(1), \dots, X(n)$ be a sample from X and let $\bar{X} = \frac{1}{n} \sum_{t=1}^n X(t)$. Then

- a) $E(\bar{X}) = \mu$
- b) $n \text{Var}(\bar{X}) = \sum_{v=-(n-1)}^{n-1} (1 - \frac{|v|}{n}) R(v) \\ \rightarrow f(0) \text{ as } n \rightarrow \infty .$

Thus \bar{X} is unbiased and consistent and is also called ergodic since the time average \bar{X} for one possible sample converges in probability to the ensemble average $\mu = E(X(t))$. Under certain general assumptions we also have that \bar{X} is asymptotically normal, and thus we will write $\bar{X} \sim N(\mu, \frac{f(0)}{n})$.

Recall that the usual random sampling result is that $\bar{X} \sim N(\mu, \frac{\sigma_X^2}{n})$ where σ_X^2 is the variance of the population being sampled. Thus in the presence of autocorrelation, $f(0)$ plays the role of σ_X^2 . A simple example of this is when $\rho(v) = \rho^{|v|}$, i.e. the autocorrelation between $X(t)$ and $X(t+v)$ decays exponentially. This autocorrelation function arises from an autoregressive process of order one, i.e. when

$$X(t) = \rho X(t-1) + \epsilon(t) ,$$

where ϵ is a white noise series of zero mean uncorrelated random variables having constant variance σ_ϵ^2 . For this model we have

$$R(v) = \frac{\sigma_\epsilon^2}{1-\rho^2} \rho^{|v|} , \tag{3}$$

and thus $\rho(v) = \rho^{|v|}$, and some algebra gives

$$f(w) = \frac{\sigma_\epsilon^2}{|1-\rho e^{2\pi i v w}|^2} \tag{4}$$

which, since $\sigma_\epsilon^2 = (1-\rho^2) R(0)$ by (3), means

$$f(0) = \frac{\sigma_\epsilon^2}{(1-\rho)^2} = \frac{R(0)(1-\rho^2)}{(1-\rho)^2} = \frac{R(0)(1+\rho)}{1-\rho}$$

To finish the example, consider two models for data $X(1), \dots, X(n)$; 1) a random sample from a population having variance σ_X^2 , and 2) an autoregressive process of order one also having variance σ_X^2 . Then the ratio of the variance of \bar{X} under model one to that under model two is

$$e_n = \frac{\sigma_X^2/n}{\sigma_X^2(1+\rho)/n(1-\rho)} = \frac{1-\rho}{1+\rho} \tag{5}$$

which can range anywhere from zero to infinity. Thus autocorrelation must be considered very carefully when estimating μ .

2. NONPARAMETRIC AND PARAMETRIC SPECTRAL DENSITY ESTIMATION

From Theorem 1.1 we see that the problem of finding the properties of \bar{X} is actually a problem of estimating $f(0)$.

The parametric approach to estimating f consists of two steps: 1) find a model for the time series X that only has a small number of parameters and seems to match the properties of the observed data. 2) Estimate the parameters of this model and substitute the estimates into the formula for the spectral density of the model. Thus in the example of section 1, if the autoregressive model were determined suitable, we need only estimate σ_ϵ^2 and ρ and substitute these estimates into (4) to get an estimate of f .

2.1 The Nonparametric Approach

The nonparametric method ignores the possibility of a parametric model and operates directly on the general formula for f

$$f(\omega) = \sum_{v=-\infty}^{\infty} R(v)e^{-2\pi i v \omega} \quad (6)$$

In this approach, if we have data $X(1), \dots, X(n)$ we have only information about correlations between X 's at most $n-1$ lags apart. The basic approach is to estimate the true autocovariances $R(0), \dots, R(n-1)$ by the sample autocovariances

$$\begin{aligned} \hat{R}(v) &= \frac{1}{n} \sum_{t=1}^{n-v} (X(t) - \bar{X})(X(t+v) - \bar{X}) \\ &= \hat{R}(-v), |v| < n \end{aligned}$$

and then truncate the series (6) somehow and use \hat{R} instead of R . Note that in the parametric approach an assumption is made about the form of $R(v)$, see equation (3) for example, and thus in this sense we have information about $R(v)$ for all lags v .

We note again that the formula (6) is merely a Fourier series. There is a vast literature on speeding up convergence of Fourier series by using weighting functions (see Brillinger (1975), section 3.3 for example). Thus the most widely used nonparametric estimates of f are weighted truncated Fourier series of the form

$$\tilde{f}(\omega) = \sum_{v=-M}^M k\left(\frac{v}{M}\right) \hat{R}(v) e^{-2\pi i v \omega}$$

for some lag window k which is chosen to be symmetric about $v=0$. The most popular windows are the Parzen window k_p and Tukey window k_T given by

$$\begin{aligned} k_p(u) &= \begin{cases} 1 - 6u^2 + 6u^3, & 0 \leq u \leq \frac{1}{2} \\ 2(1-u)^3, & \frac{1}{2} \leq u \leq 1 \end{cases} \\ k_T(u) &= \frac{1}{2} (1 + \cos 2\pi u), \quad 0 \leq u \leq \frac{1}{2} \end{aligned}$$

See Priestley (1981), section 6.2 for a detailed discussion of windows.

The difficult part of nonparametric spectral density estimation is choosing the truncation point M . Typically one is advised to try various values and any features in f that are common to all values are considered real, while features that are not common to all require further investigation.

2.2 The Parametric Approach

The most popular class of parametric time series models is the ARMA class, *i.e.* the autoregressive-moving average process of order (p, q) which models X at time t as a linear combination of values of X at the previous p values of t plus a linear combination of random "shocks" that occur at time t and the previous q times:

$$\begin{aligned} X(t) + \alpha_1 X(t-1) + \dots + \alpha_p X(t-p) \\ = \varepsilon(t) + \beta_1 \varepsilon(t-1) + \dots + \beta_q \varepsilon(t-q) \end{aligned}$$

In order for this model to make sense for all integers t , the complex valued polynomial $g(z) = 1 + \sum_{j=1}^p \alpha_j z^j$ must have all of its zeros greater than one in modulus. The series of random shocks ε is taken to be white noise with variance σ^2 . The example in section 1 is just the case where $p=1$ and $q=0$.

The ARMA(p, q) model was perhaps popularized most by Box and Jenkins (1970) who use it for prediction purposes rather than spectral estimation. The model also has been widely used for spectral estimation particularly due to the work of Akaike (1969) and Parzen (1977). The parameters of the model are $p, q, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ and σ^2 as made clear by the formula for the spectral density (which is just the ratio of two finite degree trigonometric polynomials)

$$f(\omega) = \sigma^2 \frac{|1 + \sum_{k=1}^q \beta_k e^{2\pi i k \omega}|^2}{|1 + \sum_{j=1}^p \alpha_j e^{2\pi i j \omega}|^2}$$

This model also says that $\varrho(v)$ satisfies a p th degree difference equation (with coefficients $\alpha_1, \dots, \alpha_p$) as soon as v is greater than q . Since the zeros of g are greater than one in modulus this dictates an exponential rate of decay of ϱ .

There are two major problems in parametric estimation of f : 1) Estimating Orders, and 2) estimating the α 's, β 's, and σ^2 . If we assume that the ε series is normally distributed and we let $\hat{\alpha}_{j,1}, \dots, \hat{\alpha}_{j,j}, \hat{\beta}_{k,1}, \dots, \hat{\beta}_{k,k}$, and $\hat{\sigma}_{j,k}^2$ be the maximum

likelihood estimators of the parameters for order (j,k) , then the most widely used method for estimating the order of the ARMA process is to choose (\hat{p}, \hat{q}) to be the values of (j,k) minimizing the AIC criterion

$$\text{AIC}(j,k) = \log \hat{\sigma}_{j,k}^2 + \frac{2(j+k)}{n}$$

The calculation of the $\hat{\alpha}$'s and $\hat{\beta}$'s is a very difficult computational problem. Melard (1984) discusses an algorithm for doing the maximization problem that appears satisfactory even on personal computers.

Much of the emphasis in the parametric estimating of a spectral density has been on the special case of an ARMA(p,0) model, i.e. the autoregressive process of order p, denoted AR(p):

$$X(t) + \alpha_1 X(t-1) + \dots + \alpha_p X(t-p) = \epsilon(t),$$

teZ

This model has several useful features, including 1) almost any covariance stationary time series can be adequately modeled as an AR(p) for some (possibly large) order p, 2) the autoregressive spectral estimator can be interpreted as a maximum entropy spectral estimator (see Ulyrich and Bishop (1975), 3) the calculations involved in the estimation process are much simpler than those in the general ARMA case, and 4) there are several theoretical results available for the AR case that are not available for the ARMA case.

Newton and Pagano (1984, 1983) discuss the determination of simultaneous confidence bands for an AR spectral density and a confidence interval for the value ω where $f(\omega)$ is a relative maximum.

3. THE TIMESLAB TIME SERIES ANALYSIS LABORATORY

For several years we have been writing time series analysis software (see Newton (1983a,b), for example). The result of this effort is a command-driven, interactive-graphics program for IBM type personal computers called TIMESLAB which is to be published by Wadsworth and Brooks/Cole in early 1987. TIMESLAB is too extensive to

describe in detail here. In this section we give a brief description.

TIMESLAB consists of approximately 150 commands. The user can either issue these commands one at a time or else create a file containing several commands and then issue a command to successively issue each of the commands in the file. Such a file is called a MACRO file and the command used to invoke the MACRO is called MACRO.

One important TIMESLAB command is called DOS, which when invoked puts the user into what we call the "DOS mode", i.e. any command that the user could issue from the DOS prompt can be issued at this point while still in TIMESLAB. This allows users to use any utility that they want while still inside of TIMESLAB.

To illustrate what commands look like consider the following commands that could be invoked as a MACRO if they were first entered into a file (either by using the TIMESLAB built-in full screen editor or by using some external editor from the DOS mode). Note that the ? is the TIMESLAB "prompt" and is not typed by the user and that we have inserted line numbers for reference:

1. ?N=10
2. ?SEED=12345.
3. ?X=WN(SEED,N)
4. ?FNAME='WN1.SCN'
5. ?SAVESC(FNAME)
6. ?HIST(X,N,10,-4,4,.5)
7. ?RHO=CORR(X,N,36,256,1,RO,PER)
8. ?SAVESC(WN2.SCN)
9. ?PLOT(RHO,36,-1,1)
10. ?SAVESC(WN3.SCN)
11. ?PLOTSP(PER,256,RO)
12. ?RESCREEN(WN1.SCN,WN2.SCN,WN3.SCN,1)

Line 3 generates a Gaussian white noise series (henceforth referred to as X) of length 100 points, line 6 generates a high resolution (640x200 pixels) version of the histogram of X using 10 intervals so that the scale on the horizontal axis is -4 to 4 and the vertical axis is 0 to .5. Line 7 says to calculate the sample variance RO of X as well as its first 36 autocorrelations (called RHO)

Parametric Time Series Estimation

and its periodogram (PER) evaluated at the 256 equally spaced frequencies between 0 and 1. Then line 9 says to plot $RHO(i)$ vs i for $i=0, \dots, 36$ on a scale of -1 to 1, while line 11 says to produce a plot of the log PER/RO vs frequencies for the frequencies between 0. and .5 (there are actually 127 such frequencies but to decrease confusion almost all frequency specifications refer to the number between 0 and 1).

Lines 5, 8, and 10 tell TIMESLAB to save compressed images of the histogram, correlogram, and periodogram onto files called WN1.SCN, WN2.SCN, and WN3.SCN respectively, while line 12 says to redisplay these images one after another and to erase each one before doing the next. Using this feature essentially allows the user to produce a set of "slides" of an analysis for later display. Each such file requires between 3000 and 6000 bytes depending upon how dense the image is.

All of the plots that TIMESLAB produces on the screen are high resolution pixel graphs. Whenever one is produced, TIMESLAB pauses and gives the user the opportunity to either print the screen using a specially designed "screen dump", to save the screen onto a file for later display, to use an interactive locator/labeling utility called FIND, or to return to the TIMESLAB prompt for the next command. This pause can also be overridden by the BATCHON command which can be used for "production running" of MACROS.

To illustrate how TIMESLAB can be used for studying the sampling properties of the sample mean, consider the following MACRO which illustrates the AR(1) example in section 1 of this paper. The lines starting with a semicolon are comments and document what the MACRO does.

```

; MACRO to generate 100 AR(1) samples
; of length 100 and calculate confidence
; intervals with and without allowance
; for autocorrelation.
;
; Initialize:
;
ALPHA=<-.75>
LL1=LINE(100,0,0)
UL1=LINE(100,0,0)

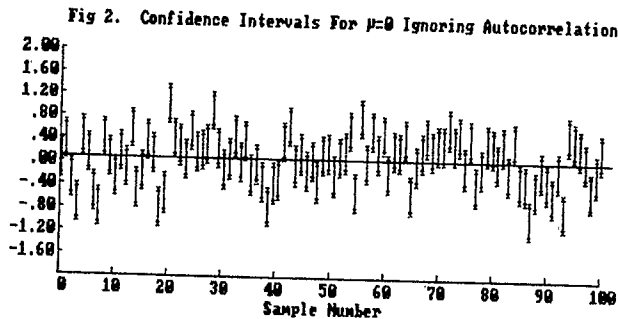
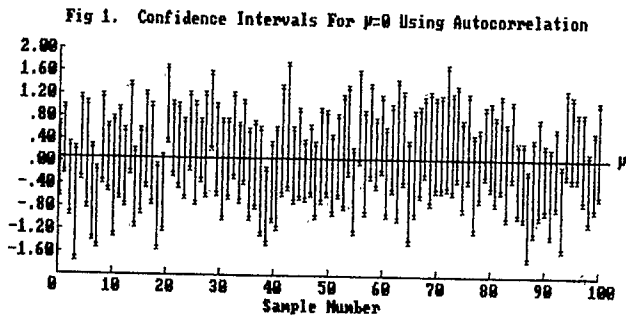
```

```

LL2=LINE(100,0,0)
UL2=LINE(100,0,0)
X=WN(SEED,10)
CNT=1
;
; START
;
X=ARDT(ALPHA,1,1,0,100,IER,RO)
X=SUBMNS(X,100,1,XBAR)
RHO=CORR(X,100,1,0,1,PER,RO)
RHO1=EXTRACT(RHO,1,1)
;
; Ignore Autocorrelation:
;
SE1=RO/100.
SE1=SE1^.5
SE1=2.*SE1
LL1[CNT]=XBAR-SE1
UL1[CNT]=XBAR+SE1
;
; Use Autocorrelation:
;
R1=1.+RHO1
R2=1.-RHO1
SE2=RO/100.
SE2=SE2*R1
SE2=SE2/R2
SE2=SE2^.5
SE2=2.*SE2
LL2[CNT]=XBAR-SE2
UL2[CNT]=XBAR+SE2
;
;
IF(CNT.EQ.100,END)
CNT=CNT+1
GOTO(START)
;
; END

```

The MACRO simulates 100 series of length 100 from a normal AR(1) process with $\alpha_1 = -.75$ and true mean 0. This corresponds to $\rho(1) = .75$. For each sample, a 95% confidence interval based on \bar{X} is calculated in two ways; the first using the traditional standard error formula, while the second uses the formula in section 12 which incorporates information provided by $\hat{f}(0)$. Figures 1 and 2 contain plots generated by TIMESLAB of the resulting sets of 100 confidence intervals and illustrate how incorrect the traditional confidence intervals can be if autocorrelation is ignored. In figure one, 93 of the confidence intervals contain the true value of μ while in figure two, only 50 of the intervals contain μ .



4. Numerical Examples

We used TIMESLAB to perform an experiment on the use of $\hat{f}(0)$ in finding 95% confidence intervals of μ for several different models and sample sizes. For each model/sample size combination we generated 100 realizations and kept track of the number of times the calculated confidence interval included μ . In each case the true mean was $\mu=0$ and the data generated was normally distributed.

The simulated data come from four AR models:

Model I: AR(1)

$$X(t) + \alpha_1 X(t-1) = \epsilon(t)$$

for $\alpha_1 = -.9, -.8, \dots, .8, .9$.

Model II: AR(2)

$$X(t) - .4X(t-1) - .45X(t-2) = \epsilon(t)$$

Model III: AR(5)

$$X(t) + 1.7X(t-1) + 2.4X(t-2) + 1.634X(t-3) + .872X(t-4) + .168X(t-5) = \epsilon(t)$$

Model IV: AR(4)

$$X(t) - 2.7607X(t-4) + 3.8106X(t-2) - 2.6535X(t-3) + .9238X(t-4) = \epsilon(t)$$

Model I was chosen because AR(1) autocorrelation seems to exist in many real applications while Models II, III, and IV provide examples of models that can be classified as easy, moderately difficult, and extremely difficult to estimate as reported in Newton and Pagano (1984). The true values of $f(0)$ for Models II, III, and IV are 44.44, .016547, 9.753431.

The Burg algorithm (Ulrych and Bishop (1975)) was used to estimate the parameters of the process. The experiment was performed both using the true order and an order estimated by the AIC criterion. Since the results were practically identical under both conditions, we only report the known order results.

Samples of sizes $n=50, 100, 200,$ and 400 were used.

Table 1 presents the results for the AR(1) case as a function of $\rho = -\alpha_1$ and n while Table 2 shows the results for Models II, III, and IV. Note that in general the observed coverages are certainly consistent with the nominal confidence level. Exceptions are Table 1 for $\rho=.8$ and $.9$ for $n=50$ and $n=100$ and Table 2 for Model II and $n=50, 100$. This is not surprising in view of equation (5) which shows that the "equivalent random sample sizes" are actually much smaller. For example, if $\rho=.9$ and $n=50$, then the equivalent sample size is $n(1-\rho)/(1+\rho) \doteq 2.5$.

TABLE 1. Confidence Interval Coverage for an AR(1) for Various n and ρ

	n				
	50	100	200	400	
ρ	-.9	94.	94.	96.	94.
	-.8	93.	94.	96.	96.
	-.7	92.	98.	96.	92.
	-.6	97.	97.	95.	96.
	-.5	98.	98.	96.	99.
	-.4	93.	96.	94.	90.
	-.3	98.	96.	98.	95.
	-.2	95.	93.	94.	93.
	-.1	94.	93.	96.	95.
	0	90.	97.	91.	94.
ρ	.1	95.	98.	98.	97.
	.2	91.	91.	94.	97.
	.3	90.	95.	91.	94.
	.4	94.	95.	97.	99.
	.5	92.	93.	95.	96.
	.6	89.	93.	96.	93.
	.7	91.	90.	95.	93.
	.8	79.	91.	95.	95.
	.9	76.	82.	90.	91.

TABLE 2. Confidence Interval Coverages for Models II, III, IV

		n			
		50	100	200	400
Model	II	74	83	92	97
	III	99	96	99	99
	IV	100	100	99	100

To get an idea of the sampling properties of $\hat{f}(0)$, we give in Figures 3 and 4 histograms of $\hat{f}(0)$ for Model II for $n=100$ and $n=400$. As expected, the variability for $n=400$ is less than that for $n=100$, and there is less skew as well. It would appear that the distribution of $\hat{f}(0)$ is chi-square that could be converging to normal as $n \rightarrow \infty$.

It should be pointed out that while the behavior of $\hat{f}(0)$ is interesting (and insightful results remain to be found), it is the studentized variable $\sqrt{n} \bar{X} / \sqrt{\hat{f}(0)}$ that is really of interest and the results of this section show that it performs well.

Fig 3. Histogram of $\hat{f}(0)$ for Model II, $n=100$
(Vertical axis is proportion/20). True $f(0)=44.44$

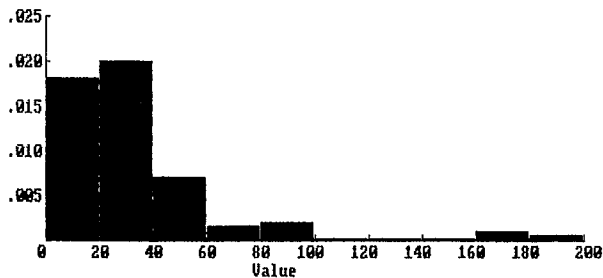
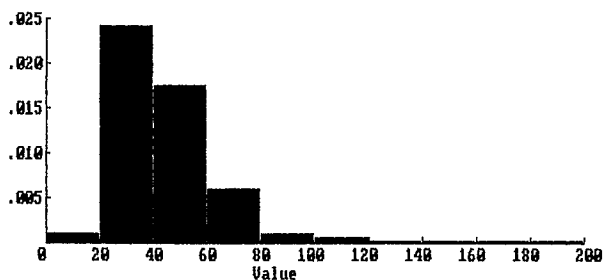


Fig 4. Histogram of $\hat{f}(0)$ for Model II, $n=400$
(Vertical axis is proportion/20). True $f(0)=44.44$



5. SUMMARY

In this paper we have shown how the estimation of the spectral density of a time series plays an important role in determining

the sampling properties of \bar{X} and that the whole range of modern time series analysis techniques can be brought to bear on the problem via a program such as TIMESLAB.

ACKNOWLEDGEMENT

Research supported by the Office of Naval Research.

REFERENCES

Akaike, H. (1969). Fitting autoregressive models for prediction. Annals of Institute of Statistical Mathematics, 21, 243-247.

Box, G. E. P. and Jenkins, G. M. (1970). Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco.

Brillinger, D. R. (1975). Time Series: Data Analysis and Theory. Holt, Rinehart, and Winston, New York.

Milard, G. (1984). A fast algorithm for the exact likelihood of autoregressive-moving average models. Applied Statistics, 35, 104-114.

Newton, H. J. (1983a). An Introduction to the Methods of Time Series Analysis in the Time and Frequency Domains. Statistics Department, Texas A&M.

Newton, H. J. (1983b). What should your time series analysis problem do? In Proceedings of the 15th Symposium on the Interface of Computer Science and Statistics (Ed. J. Gentle), 125-130, North-Holland, Amsterdam.

Newton, H. J. and Pagano, M. (1983). A Method for determining periods in time series. Journal of American Statistical Association, 78, 152-157.

Newton, H. J. and Pagano, M. (1984). Simultaneous confidence bands for autoregressive spectra. Biometrika, 71, 197-202.

Parzen, E. (1977). Multiple time series modeling: determining the order of approximating autoregressive schemes. In Multivariate Analysis, IV (Ed. P.

H. J. Newton

Krishnaiah), p.283-295. North-Holland, Amsterdam.

Priestley, M. (1981). Spectral Analysis and Time Series. Academic Press, New York.

Ulrych, T. L. and Bishop, T. N. (1975). Maximum entropy spectral analysis and autoregressive decomposition. Reviews of Geophysics and Space Physics, 13, 183-200.

AUTHOR'S BIOGRAPHY

H. JOSEPH NEWTON is an associate professor in the Department of Statistics at Texas A&M University. He received a B.S. in mathematics from Niagara University in 1971, and M.S. and Ph.D. degrees in Statistical Science from SUNY/Buffalo in 1973 and 1975 respectively. His interests range from the theory and practice of time series analysis to the development of graphics and user interfaces on small computers. His current research interests include spatial time series analysis, analysis of variance for correlated data, and the development of algorithms for time series computations. He is a member of ASA and IMS.

H. Joseph Newton
Department of Statistics
Texas A&M University
College Station, Texas 77843, U.S.A