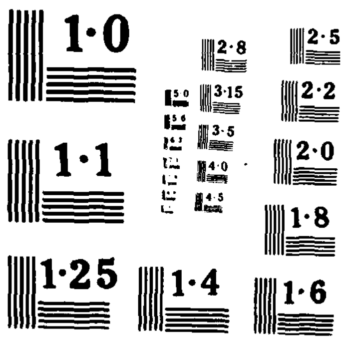
AD-A170 094 A FAST GRAPHICAL GOODNESS OF FIT TEST FOR TIME SERIES 1/1
MODELS(U) MARYLAND UNIV COLLEGE PARK DEPT OF
MATHEMATICS B KEDEM AUG 85 MD-85-38-BK
UNCLASSIFIED AFOSR-TR-86-0274 AFOSR-82-0187 F/G 12/1 NL


1·0
2·8
2·5
3·15
2·2
3·5
1·1
2·0
4·0
4·5
1·8
1·25
1·4
1·6

AFOSR-TR- 86-0274

AD-A170 094

A Fast Graphical Goodness of Fit Test For

Time Series Models

by

Benjamin Kedem

Department of Mathematics
University of Maryland
College Park, MD 20742

August 1985

MD-85-38-BK

TR85-32

* The support of Grant AFOSR 82-0187 is gratefully acknowledged.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release; distribution unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AFOSR-TR- 86-0274

6a. NAME OF PERFORMING ORGANIZATION University of Maryland

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AFOSR, Major Brian W. Woodruff
Program Manager/NM

6c. ADDRESS *(City, State and ZIP Code)*
College Park, MD 20742

7b. ADDRESS *(City, State and ZIP Code)*
Bolling AFB
Washington, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Office of Scientific Research

8b. OFFICE SYMBOL *(If applicable)* AFOSR

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
AFOSR 82-0187

8c. ADDRESS *(City, State and ZIP Code)*
Bolling AFB
Washington, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	~~2304~~/A5	

11. TITLE *(Include Security Classification)* A fast graphical goodness of fit test for time series models

12. PERSONAL AUTHOR(S) Benjamin Kedem

13a. TYPE OF REPORT Technical Report

13b. TIME COVERED FROM ________ TO ________

14. DATE OF REPORT *(Yr., Mo., Day)* August 1985

15. PAGE COUNT 18

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Higher Order Crossings

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

There has been a growing interest in graphical methods in time series analysis and especially so since the popularization of electronic devices with graphics capabilities. In following this trend, the present article discusses a certain zero-crossings based graphical technique useful in testing for goodness of fit of time series models. The idea is to use plots of higher order crossings which are akin to plots of the correlogram and spectral densities or the periodogram, but with the advantage of great simplicity. Under the Gaussian assumption, the sequence of expected higher order crossings is equivalent to the autocorrelation function and hence to the normalized spectral distribution function, but it summarizes the data differently. In this regard, the monotone property of higher crossings which plays an instrumental role for the initial rate of increase exhibited by higher crossings proves to be an effective summary feature. As the higher crossings continue to increase, their rate loses its discrimination potency and

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Maj Woodruff ~~Br Kedem~~

22b. TELEPHONE NUMBER *(Include Area Code)*
(301) 454-7055

22c. OFFICE SYMBOL
NM

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

and different processes seem to share similar rates. This is why, in general, very few higher crossings are used in testing goodness of fit.

The present paper gives an overview of our previous work particularly Kedem and Reed (1985) and Kedem (1985) to which the reader is referred for mathematical details and more examples.

Abstract

The oscillatory appearance of stationary time series is captured very economically by only a few higher order crossings which in addition contain a great deal of the spectral content of the process. A useful approximation to the variances of higher order crossings is discussed and is applied in the construction of probability limits for the hypothesized higher order crossings. From this, a graphical display of higher order crossings together with their probability limits provide a fast goodness of fit test. Examples illustrate the applicability of this device.

1. Introduction.

There has been a growing interest in graphical methods in time series analysis and especially so since the popularization of electronic devices with graphics capabilities. In following this trend, the present article discusses a cetain zero-crossings based graphical technique useful in testing for goodness of fit of time series models. The idea is to use plots of higher order crossings which are akin to plots of the correlogram and spectal densities or the periodogram, but with the advantage of great simplicity. Under the Gaussian assumption, the sequence of expected higher order corssings is equivalent to the autocorrelation function and hence to the normalized spectral distribution function, but it summarizes the data differently. In this regard, the monotone property of higher crossings plays an instrumental role for the initial rate of increase exhibited by higher crossings proves to be an effective summary feature. As the higher crossings continue to increase their rate loses its discrimination potency and different processes seem to share similar rates. This is why in general very few higher crossings are used in testing goodness of fit.

The present paper gives an overview of our previous work particularly Kedem and Reed (1985) and Kedem (1985) to which the reader is referred for mathematical details and more examples.

2. Plots of Higher Order Crossings.

Let $\{z_t\}$, $t = 0, \pm 1, \ldots,$ be a zero mean stationary Gaussian process with correlation function ρ_j and normalized spectral distribution function F, and let ∇ be the difference operator, $\nabla z_t = z_t - z_{t-1}$. It is convenient to introduce the clipped binary process

$$x_t^{(k)} = \begin{cases} 1, & \nabla^{k-1} z_t \geq 0 \\ 0, & \text{otherwise} \end{cases}, \quad k = 1,2,\ldots$$

which gives rise to the indicator at time t

$$d_t^{(k)} = \begin{cases} 1, & x_t^{(k)} \neq x_{t-1}^{(k)} \\ 0, & \text{otherwise.} \end{cases}$$

The higher order crossings of order k, $D_{k,n}$, is defined by

$$D_{k,n} = d_2^{(k)} + \ldots + d_n^{(k)}.$$

It is seen that $D_{k,n}$ counts the number of axis-crossings in the $(k-1)$'th differenced series $\nabla^{k-1} z_1, \ldots, \nabla^{k-1} z_n$. $D_{1,n}$ then is the usual number of zero- or axis-crossings by the original series $z_1, \ldots, z_n$.

From the point of view of the theory of stationary Gaussian processes, the sequence of higher order crossings is equivalent to the correlation and spectral structures. This is stated precisely in

<u>Theorem 1</u>. Let $\{z_t\}$ be a zero mean stationary Gaussian process with correlation function ρ_j. Then the sequence $\{\rho_j\}$ is completely ditermined from the sequence $\{E(D_{j,n})\}$. That is, ρ_k is determined by $E(D_{1,n}), \ldots, E(D_{k,n})$.

<u>Proof</u>: From Kedem and Slud (1981),

$$\cos\left(\frac{\pi E(D_{k+1,n})}{n-1}\right) = \frac{-\binom{2k}{k-1}+\rho_1\left[\binom{2k}{k}+\binom{2k}{k-2}\right]-\ldots+(-1)^k\rho_{k+1}}{\binom{2k}{k}-2\rho_1\binom{2k}{k-1}+\ldots+(-1)^k 2\rho_k} \tag{1}$$

and the ρ_j can be determined recursively from the $E(D_{k,n})$. □

Obviously it is also true, from (1), that knowledge of $\{\rho_j\}$ is equivalent to knowledge of the sequence $\{E(D_{j,n})\}$. It follows that F is completely determined by the sequence of expected higher order crossings. This is summarized by the symbolism

$$\{E(D_{j,n})\} \Leftrightarrow \{\rho_k\} \Leftrightarrow F.$$

Thus, exactly for the same reasons that plots of ρ_k and F are extensively used in time series analysis, it is useful to observe plots of higher order crossings too.

The main thing to observe in plots of higher order corssings is the rate at which they increase and the starting point $D_{1,n}$. The fact that higher order crossings tend to increase can be attributed to the general fact that

$$D_{j,n} \leq D_{j+1,n} + 1$$

with probability one. Hence the $D_{j,n}$ tend to increase with j for fixed but large n. See also Kedem and Slud (1981).

It is instructive to observe plots of higher order crossings and thus motivate the central idea of this paper. Figure 1 displays plots of ten higher order crossings $D_{1,1000},\ldots,D_{10,1000}$, obtained from first order autoregressive processes with different parameter values ϕ. It is seen that the initial rate of increase and starting point differ from process to process, but that as the order increases the rate is almost independent of the parameter. This same behavior has been observed in numerous cases which may be interpreted to mean that only the very first few higher crossings carry sufficient information which discriminates clearly between different processes.

Accordingly, it is suggested that plots with as few as six $D_{j,n}$ can be useful in goodness of fit testing. At the same time it should be noted that higher order crossings of high order carry information too but this information is less amenable and will not be used here.

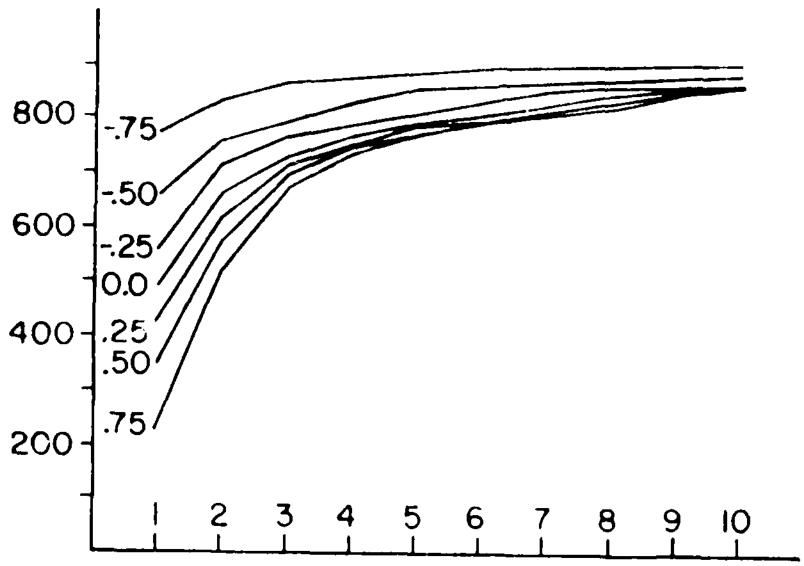


Figure 1. Plots of $D_{j,1000}$, $j=1,\ldots,10$, from $z_t = \phi z_{t-1} + u_t$, u_t are independent $N(0,1)$ random variables and $\phi = 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75$.

3. The variance of higher order crossings.

The probability distribution of the $D_{j,n}$ is quite intractable and we shall concentrate on the more modest problem of approximating the variance of higher order crossings needed for the proposed goodness of fit test.

In general, the variance of $D_{j,n}$ is a function of the fourth order cumulant function $\kappa_x^{(j)}(r,s,t)$ of $\{X_t^{(j)}\}$ which is summable under appropriate moment conditions. Thus for $j = 1$ the following asymptotic result was proved in Kedem (1980).

Theorem 2. If ρ_j is absolutely summable then

$$\sum_{k=-\infty}^{\infty} |\kappa_x^{(1)}(1,-k,1-k)| < \infty$$

and

$$\frac{D_{1,n} - E(D_{1,n})}{\sqrt{n}} \xrightarrow{L} N(0,\sigma_1^2),\ n \to \infty,$$

where

$$\sigma_1^2 = \frac{1}{\pi^2} \sum_{k=-\infty}^{\infty} [(\sin^{-1}\rho_k)^2 + \sin^{-1}\rho_{k-1}\sin^{-1}\rho_{k+1} + 4\pi^2\kappa_x^{(1)}(1,-k,1-k)].$$

The same result applies to every $D_{j,n}$ provided the correlation functions of $\{\nabla^{j-1}z_t\}$ decays to zero fast. However $\kappa_x^{(j)}$ is not known in general which makes the above result impractical.

Another approach is to hold n fixed and let j increase. In this case it is possible to obtain a useful asymptotic result under the assymption of m-dependence. Assume that π is a point of increase for F and let

$$\lambda_j^{(k)} = P_r(x_t^{(k)}=1 \mid x_{t-j}^{(k)}=1).$$

Then $\lambda_1^{(k)} \to 0$ as $k \to \infty$ and it was shown by Kedem and Reed (1985) that

$$\operatorname{cov}(d_t^{(k)}, d_s^{(k)}) = o(\lambda_1^{(k)}). \tag{2}$$

The proof of this fact depends on the differential properties of $\rho_j^{(k)}$, the correlation function of $\{\nabla^k z_t\}$. (2) readily yields.

<u>Theorem 3</u>. Let $\{z_t\}$ be an m-dependent stationary Gaussian process and assume that π is a point in the support of F. Then for fixed n

$$\lim_{k\to\infty} \frac{\operatorname{Var}(D_{k,n})}{(n-1)\lambda_1^{(k)}(1-\lambda_1^{(k)})} = 1.$$

This result was used in the construction of probability limits for the higher crossings under the hypothesis of white noise. However the assumption of m-dependence cannot always be verified and another approximation is called for.

A rather close approximation to the variance of $D_{j,n}$ can be provided if it is assumed that the binary sequence $\{d_t^{(k)}\}$ is a Markov chain. This first order approximation has been found very satisfactory by an extensive simulation.

Define the two parameters associated with the chain,

$$p^{(k)} \equiv 1 - \lambda_1^{(k)}, \quad q^{(k)} \equiv \lambda_1^{(k)}$$

$$\nu^{(k)} \equiv \frac{1-2\lambda_1^{(k)} + \lambda_2^{(k)}}{2(1-\lambda_1^{(k)})}$$

When the process is a stationary Gaussian autoregressive-moving average process with known (or hypothesized) parameters, $p^{(k)}$ and $\nu^{(k)}$ are known too explicityly. Then if $\{d_t^{(k)}\}$ is a Markov chain it can be shown (Kedem (1985))

That

$$\mathrm{Var}(D_{k,n}) = (n-1)p^{(k)}q^{(k)} + \frac{2p^{(k)}q^{(k)}(\nu^{(k)}-p^{(k)})}{1-\nu^{(k)}} \cdot [(n-1)-V_{k,n}] \tag{3}$$

where

$$V_{k,n} = q^{(k)}[1-(\frac{\nu^{(k)}-p^{(k)}}{q^{(k)}})^{n-1}]/(1-\nu^{(k)}).$$

This approximation has been compared (Kedem (1985)) with actual estimates obtained from 100 independent realizations each of length $n = 1000$. The results are given in Table I. Although $E(D_{j,1000})$ are known explicitly when the parameters are known, these expectations are estimated too as a check of the whole simulation. It is seen that (3) agrees well with the simulation results. An algorithm for obtaining $p^{(k)}$, $\nu^{(k)}$ is given in Kedem (1985).

Series	j	$E(D_{j,1000})$	$\hat{E}(D_{j,1000})$ From 100 Realizations	$\{Var(D_{j,1000})\}^{\frac{1}{2}}$ From (3)	$\{\hat{Var}(D_{j,1000})\}^{\frac{1}{2}}$ From 100 Realizations
White Noise	1	500	497	15.81	15.96
	2	666	666	13.15	13.63
	3	732	732	12.16	12.53
	4	769	770	11.57	11.49
	5	794	795	11.18	11.05
	6	813	814	10.82	10.00
AR(2) $\phi_1 = 0.4$ $\phi_2 = -0.7$	1	424	425	9.64	9.67
	2	484	485	9.38	9.13
	3	536	537	10.29	10.81
	4	594	594	11.27	12.72
	5	651	652	11.87	12.02
	6	702	701	12.04	11.34
ARMA(1,1) $\phi = 0.5$ $\theta = 0.7$	1	552	552	14.62	14.74
	2	679	679	12.96	12.87
	3	737	737	12.09	12.05
	4	773	772	11.27	11.52
	5	797	797	10.70	11.12
	6	814	814	10.15	10.80
ARMA(2,2) $\phi_1 = -1.4$ $\phi_2 = -0.5$ $\theta_1 = 0.2$ $\theta_2 = 0.1$	1	884	883	10.04	10.51
	2	897	897	9.20	9.53
	3	903	903	8.84	9.01
	4	908	908	8.60	8.50
	5	911	911	8.43	8.47
	6	914	914	8.29	8.38

Table 1. Comparison of (3) with the variance obtained from 100 independent realizations of size 1000. $E(D_{j,1000})$ and $\hat{E}(D_{j,1000})$ are rounded to the nearest integer.

4. A Graphical Goodness of Fit Criterion.

The proposed goodness of fit test is based on deviations of the observed path of higher crossings from the expected path where the later is obtained under the hypothesis of an assumed model. Marked deviations of the observed path from the expected one suggest that the observed process does not oscillate as expected. The closeness of the two paths can be measured by appealing to (3) and to conditions under which the $D_{j,n}$ are asymptotically normal. It can be shown, using the technique in Cuzick (1976) that when $\{z_t\}$ is Gaussian the condition $\sum|\rho_k| < \infty$ implies the asymptotic normality of the $D_{k,n}$. It follows that approximate 95% probability limits for $D_{k,n}$ are for each k and sufficiently large n

$$(n-1)p^{(k)} \pm 1.96\{\text{Var } D_{k,n}\}^{1/2} \qquad (4)$$

where $\text{Var}(D_{k,n})$ is given by (3). When at least one observed $D_{j,n}, j=1,\ldots,6$, lies outside the limits (4) the assumed model under which (4) was derived is rejected. Before discussing the power of this test it is illustrated by a few examples.

4.1 Examples.

Annual Mean Temperature

The graph of the annual mean air temperature from 1781 to 1980 at Hohenpeissenberg, Germany, is given in Figure 2. Actually the observations for 1811 and 1812 are missing and were replaced by the mean of neighboring observations. This has only a very small effect on the sequence of higher crossings.

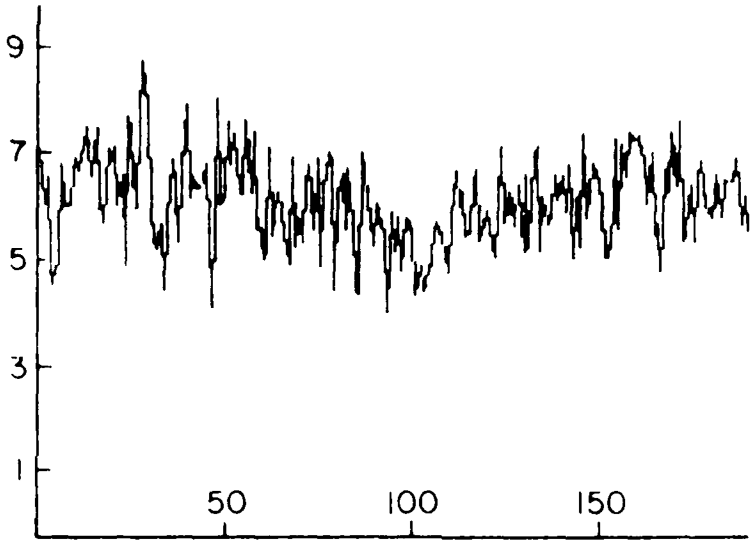


Figure 2. Annual Temperature Series. n = 195.
(Source: Report #155 of the Deutschen Wetterdienstes, West Germay (1981).)

Since annual temperature is hard to predict, we could ask the following question:

Does the series oscillate as white noise? The answer is obtained from Figure 3 where it is seen that the higher order crossings are well within the bounds (4) so that at least in this sense the series resembles white noise. For comparison, the figure portrays the higher crossings of simulated white noise which fall within the bounds too as expected.

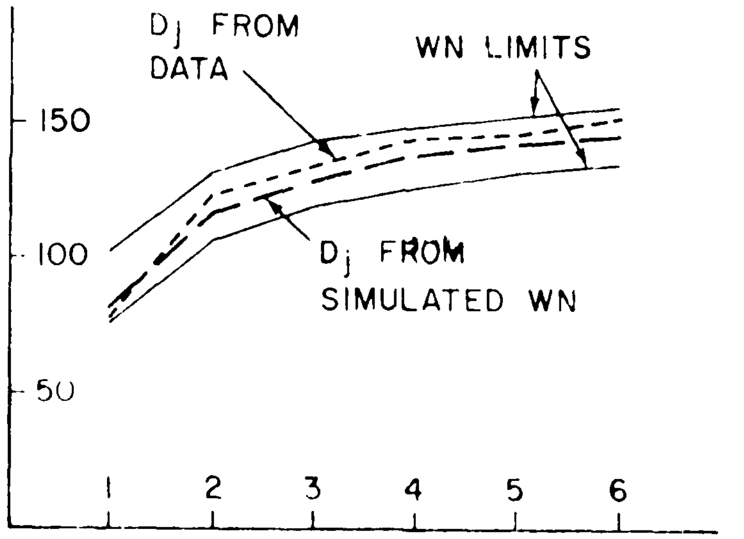


Figure 3. Probability limits for the higher order crossings from the temperature series. The series oscillates as white noise.

ARMA Models.

Figure 4 shows the probability limits (4) under various hypotheses. These are white noise, second order autoregressive process with parameters 0.4 and -0.7, and second order autoregressive moving average process with parameters $\underset{\sim}{\phi} = (-1.4, -0.5)$, $\underset{\sim}{\theta} = (0.2, 0.1)$. The actual $D_{j,450}$ were obtained from simulated data given in an appendix in Priestly (1981). The three paths fall well within their respective limits and the corresponding hypothesises are accetped.

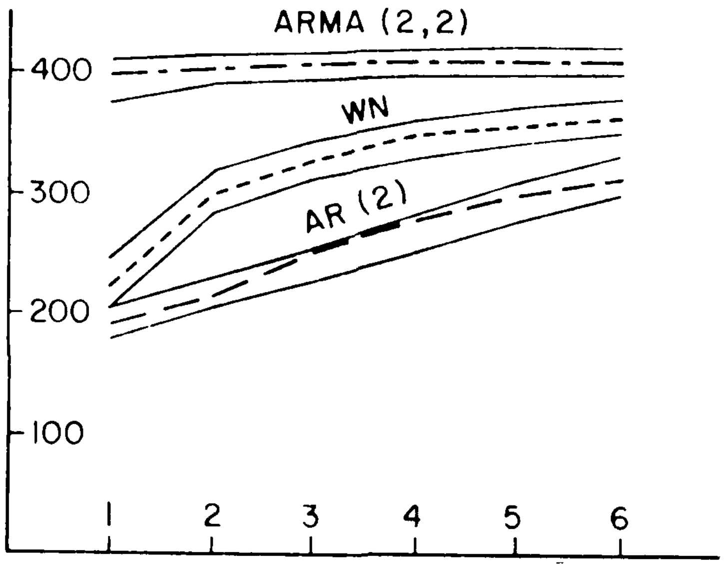


Figure 4. Sample higher order crossings paths fall within their respective limits.

It is seen that the three processes display different oscillation patterns which are captured very economically by only six higher order crossings where the ARMA (2,2) process is most oscillatory while the AR(2) is much smoother.

Signal Detection.

Figure 5 displays two series which appear to be very similar except perhaps for scale. However their higher order crossings quickly reveal that the first one oscillates as white noise while the other oscillates roughly as a low order autoregressive process. This is illustrated in Figure 6.

(a)

(b)

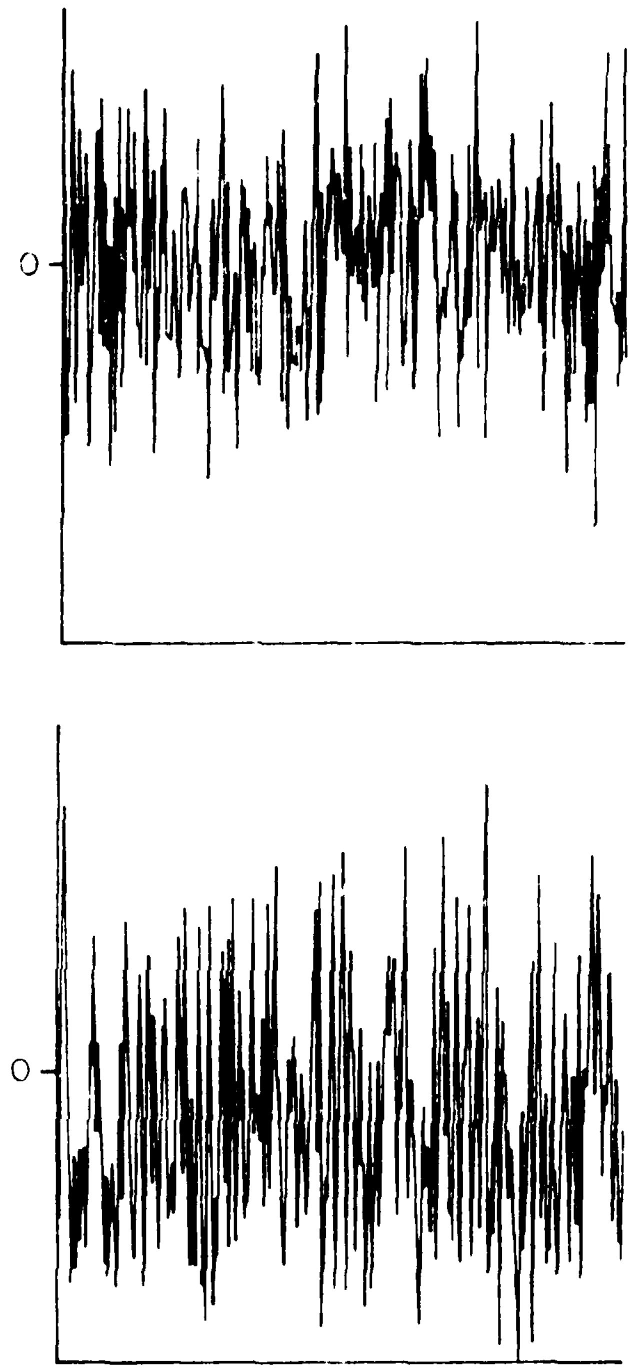


Figure 5. Two oscillating time series.

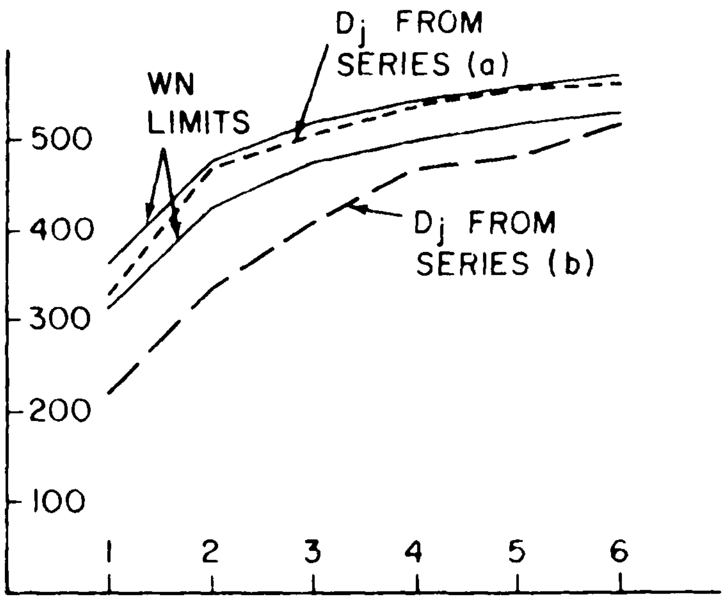


Figure 6. The higher order crossings paths of series (a), (b). The first path is within white noise bounds.

Diagonstic Check.

In testing the goodness of fit of a model one runs a residual analysis which usually tests whether the residual series constitutes white noise. Consider series A, D in Box and Jenkins (1976). The fitted models there (p. 293) are

$$\text{series A:} \quad \nabla z_t = u_t - 0.7u_{t-1}$$

$$\text{series D:} \quad z_t - 0.87z_{t-1} = 1.17 + u_t$$

where $\{u_t\}$ is the residual series. Figure 7 however reveals that the two residual series are not quite white as is signified by the axis-crossings themselves which are outside the limits (4). It is interesting to note though that the rest of the higher order crossings behave as those of white noise. Thus, except for smaller $D_{1,n}$, the two residual series oscillates as white noise.

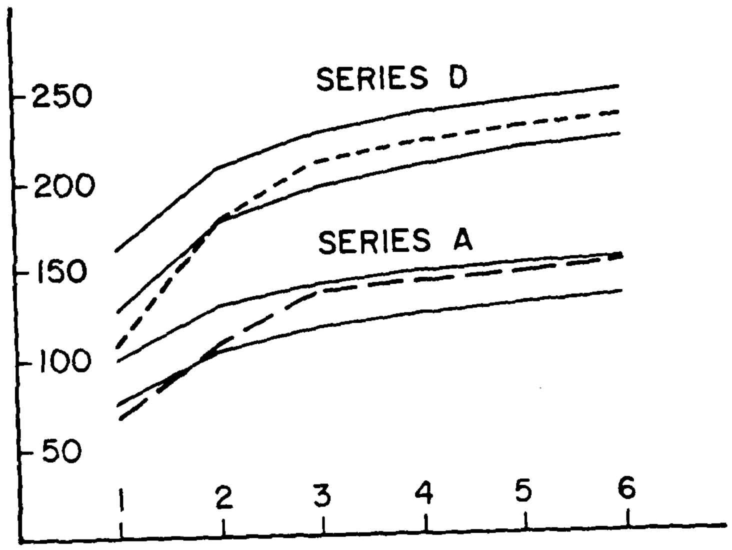


Figure 7. Diagnostic check applied to the residuals of series A (n=177) and series D (n=290). $D_{1,n}$ is outside the limits (4) for white noise.

4.2 Power Simulation.

The limits (4) provide approximate 95% bounds for each $D_{j,n}$. However our test is based on $D_{1,n},\ldots,D_{6,n}$ simultaneously and the hypothesized model is rejected if at least one $D_{j,n}$ falls outside the probability bounds. It is expected that a test which is based on more than a single $D_{j,n}$ has a higher probability of rejecting a true hypothesis than 0.05 and in fact our experience indicates that with six $D_{j,n}$ this probability is about 0.1. The exact probability is still an open problem at present.

An indication of the power is provided in Table II which gives the power for tesing the hypothesis of white noise where the alternative is the indicated process. The power is estimated from 50 independent series each of size 450. Similar results were obtained for greater series lengths.

Process	Power
White Noise	.10
AR(1), ϕ = .05	.26
MA(1), θ = .1	.40
AR(1), ϕ = .2	.90
AR(1), ϕ = .5	1.00
AR(2), ϕ = .1, ϕ_2 = −.15	.88
ARMA(1,1), ϕ_1 = .1, θ_1 = −.1	.86
ARMA(2,2), ϕ_1 = .1, ϕ_2 = −.4	1.00
θ_1 = 0, θ_2 = .3	
ARMA(2.2), ϕ_1 = .1, ϕ_2 = −.2	.88
θ_1 = .2, θ_2 = .1	

Table II. Power sinulation for testing white noise versus the indicated process.

Author's Affiliation:

Department of Mathematics
University of Maryland
College Park, Maryland 20742
USA

5. References.

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

[2] Cuzick, J.: 1976, 'A central limit theorem for the number of zeros of a stationary Gaussian process', Annals of Probability, 4, 547-556.

[3] Kedem, B.: 1980. Binary Time Series, Dekker, New York.

[4] Kedem, B.: 1985. 'A Graphical similarity measure for time series models', Report TR85-10, Mathematics Department, University of Maryland.

[5] Kedem, B. and E. Slud: 1981. 'On goodness of fit of time series models: an application of higher order crossings', Biometrika, 68, 551-556.

[6] Kedem, B. and G. Reed: 1985. 'On the asymptotic variance of higher order crossings with special reference to a fast white noise test', to appear.

[7] Priestley, M.B.: 1981. Spectal analysis and time series, Academic Press, New York

END
DTIC
8-86