Evaluation of the Time Series Analysis techniques

by F H Yung

14 Feb 1994

ABSTRACT :

Time Series Analysis (TSA) has become a popular statistical tool in many disciplines such as economics, physics and engineering just to name a few; and more recently biostatistics. However it has not been used in this department yet as reflected by the internal and external reports.

I moved to evaluate the usefulness and effectiveness of TSA to the Division's research activities by carrying out the analysis of a typical data set and discussions with research staff. Other data sets were also investigated to see whether TSA can be of use in their analysis.

To gain a perspective of our statistical analysis practice with the current development of TSA, a literature survey was conducted for relevant and interesting application examples. These were presented in a Departmental Seminar. The statistical package SYSTAT was used for demonstration. This was also aimed at raising awareness and generating interest in TSA among staff. The clarification of a few common misconceptions was used as a motivation to use TSA. Only a few important concepts were introduced such that the audience can understand the seminar intuitively. Audience appreciation was good.

TSA was suggested to be a useful tool for tree growth rate studies, wildlife management, forest disease prevention, insect outbreak control and fire control.

§1 INTRODUCTION

Time Series Analysis (TSA) is a very extensive and specialised field of statistics. It, originated from economics and since the 1920's, has influenced developments in many other fields such as physics, engineering and more recently biostatistics. Pioneering work in using TSA in biological research is represented by Diggle, 1990, done in CSIRO. In CALM, SID staff are mostly aware of its existence and understand it only superficially with the exceptions of a few namely the Research Technique, the biometrician and a botanist who did a Ph D degree in this direction. As a result the techniques have not been widely used. No report produced by CALM has mentioned this topic to-date.

I identified that this rich and very relevant discipline might have been overlooked by research scientists, either due to lack of introduction or training provided, or to simply being put off by the formidable appearance of the mathematics of the topic.

These techniques :

- are the best for analysing time based data, which is of common occurrence in CALM's activities.
- provide a parsimony (minimum number of parameters) in modelling some situations
- are the most objective way for forecasting

The specific objectives of this project are :

(1) To raise awareness among scientists in the Division by introducing them to the fundamental concepts of TSA and enable them to understand what it can do for them as application users.

(2) To assess the usefulness and effectiveness of TSA to the Division's research activities, by carrying out the analysis of a few relevant data sets and discussing with the research staff. A positive outcome may warrant further development in this direction in the Division

(3) To demonstrate these TSA techniques, as powerful tools in themselves, and which may offer shortcuts in data analysis under certain circumstances.

(4) To survey the softwares available as a spin off.

§2 METHOD

- Surveyed over the available literature to get an overview of the recent developments in TSA applications. Computer search was also done on a few databases. Selected a suitable set of examples, some of which demonstrate the fundamental techniques well, and some are more relevant and particularly appealing to the biological scientist.

- Using the statistical package SYSTAT, carry out typical TSA computer runs to demonstrate the decomposition of a series into the trend, seasonal and stochastic components by the Box-Jenkins (Auto regressive Integrated Moving Average) models.

- Cross-spectral analysis techniques were applied to some data sets to assess its suitability.

- Present a Departmental Seminar to introduce the advantages of TSA techniques. Subsequent response from the audience and staff were specially noted for awareness and interests raised.

§3 THE SEMINAR

Most of the results was presented in the Departmental Seminar on the 9 Dec 1993 using two overhead projectors simultaneously. It was based on 19 transparencies, which are included in appendix 1, numbered as T1, T2, ---.

The surprise and clarification of the common misconception of the stochastic trend as opposite to the real trend was used to arouse interest, illustrated by a few series generated by the random walk model and the speculative share market index of Hong Kong, the Hang Seng index. This also motivated the use of TSA.

Fundamental concepts in the time domain were summarized on one transparency of equations, T5, mainly based on Cryer, but only the concept that a term in a series **depends** on those before it was emphasized as the only essential one to understand the seminar intuitively. It was shown that this dependence can be described by the Moving Average and Auto Regressive models.

The diversity of TSA applications were illustrated by quoting examples like: unemployment series, airline passengers series, share indices, Einstein's mathematical explanation of the Brownian motion; anti-terrorist analysis and chronobiological applications. The latter was advocated in 1974 by Simpson as a new field resulted from the amalgamation of biology and TSA. He even speculated that in future special hospital wards for chronotherapy will have a physician, a chronobiologist and a time series analyst working as a team for diagnosis.

The Airline Passengers Series from SYSTAT was used to demonstrate the decomposition of a typical series into a trend, its seasonal and stochastic component, which was used in turn for forecasting.

The relationship between leading indicator series were illustrated by calculating the cross correlation coefficients between the Consumers Spending Series and the Money Stock Series from SYSTAT; one being a leading indicator of the other (T12). On the contrary, the number of kangaroos harvested per month peaks on the dry month as expected, (T13).

The computer literature search of the recent 24 years based on three keywords - time, series and insect - on the titles and abstracts resulted in 18 publications, but only the latest by Swetnam & Lynch was presented as it is more representative of the studies of insect outbreaks, which is of particular interest to the department.

The spectral analysis formulation in the frequency domain was presented on one transparency based on Jenkins & Watts. Its applications were illustrated by the material published by Swetnam & Lynch (T17).

§4 OUTCOMES AND DISCUSSIONS

The objectives of this project were largely achieved.

The material in 18, out of 19 transparencies, were successfully covered in the seminar and was reasonably appreciated by the audience; with interested discussions at the conclusion.

The seminar material raised has since also attracted discussion from the Economics Branch. Information was exchanged regarding software capabilities and their prices. It was preliminarily suggested that staff secondment may also lead to mutual benefits.

After all TSA demands continuous data with respect to time for quantitative modelling and forecasting. As limited by resource, not many data sets in this Department fulfil this requirement. For example tree growth rate is affected by the availability of water, its quality, temperature and other meteorological factors. They can be analysed by TSA when the data is continuous and long enough to get statistical significance.

However to identify leading indicator series and estimate their phase difference may not be too demanding. At least 4.5 years of reasonably distributed data for two series may suffice, as illustrated in the seminar by the Consumers Spending and the Money Stock Series. It was just significant. In the same spirit cross correlation techniques can also be used for studying kangaroo harvesting in parallel with the analytic models currently in use.

TSA is also a good tool for studying the sex ratio of kangaroos.

To ensure the sustainable utilisation of forest resources, it is necessary to study many factors that affect it directly or indirectly. For example, the relationship between rainfall and insect outbreaks is also of interest. Cross Spectral techniques are well suited to carry out the analysis, provided more that seven years data is collected if one wants to obtain statistical significance of annual features according to experience [Wills].

The propagation speed of Jarrah Dieback fungus is sensitive to soil temperature at the depth concerned. For example, five plots subjected to a treatment of having different degrees of vegetation removed will result in having different soil temperature behaviours. The measurement of the soil temperature at fixed time intervals over these plots yields data of the type called Repeated Measurements. We can either use the General Linear Model or the TSA approach to analyse it. The latter offer a reduction in the number of parameters to be estimated. The temperature data taken at different depths in Dwellingup Research belongs to this type.

TSA is likely to useful in our fire control research. From the set of parameters {rainfall, temperature, humidity}, the fuel moisture content can be calculated. While TSA offers a method to simulate the wind power the hourly Fire Danger Rating can be calculated.

Acknowledgment: I wish to acknowledge all my colleagues for beneficial discussions specially: I J Abbott, N D Burrows, J A Friend, Neil Gibson, S A Halse, J K Kinal, R I T Prince, T N Start, D J Ward, M R Williams A J Wills and R T Wills.

References

Cryer, J D, Time Series Analysis, 1986, Ducsbury.
Diggle, P J, Time Series, A Biostatistical Introduction, 1990, Oxford Sc. pub.
Jenkins, G M & Watts, D G, Spectral Analysis and its applications, 1968, Holden-Day.
Simpson, H W, Dept. of Pathology, Glasgow U, a paper read at a meeting of the British Region, May 7, 1974.
Swetnam, T W & Lynch, A M, Multicentury, regional-scale patterns of western spruce budworm outbreaks, Ecol. Mono., **63**(4), 1993, pp. 399-424.
Wills, R T, private communication.

APPENDIX 1 The seminar notice and the 19 transparencies presented. There title are as follows.

T1	Examples of some time series
T2	Two dimensional random walk diagrams
T3	Random walk time series generated by the first order Autoregressive model
T4	The Hong Kong share index, Hang Seng to illustrate stochastic trends
T5	The main TSA models and formulae in the time domain
T6	The most common Box-Jenkins models
T7	Autocorrelation and partial autocorrelation function for model identification, example before differencing
T8	Autocorrelation and partial autocorrelation function for model identification, example after differencing
T9	Autocorrelation function for model diagnosis after model fitting
T10	Residual plot for model diagnosis after model fitting
T11	Airline Passengers Series, model fitted and forecasted
T12	Cross correlation analysis of the Consumers Spending & Money Stock Series
T13	Cross correlation analysis of the rainfall & kangaroos harvested
T14	Basic ideas in model building
T15	Spectral analysis formulae
T16	Luteinizing homone concentration series
T17	Cross correlation analysis for the rainfall and spruce budworms outbreak series
T18	Positive signs indicating spectral analysis may be used successfully
T19	Repeated measurements formulae

DEPARTMENT OF CONSERVATION AND LAND MANAGEMENT

Please address all enquiries to:

Your Ref:
Our Ref:
Enquiries:

Seminar

Thursday 9 December 1993

TIME SERIES ANALYSIS TECHNIQUES: AN INTRODUCTION TO ITS APPLICATIONS

presented by M Yung

The basic framework of Time Series Analysis techniques (TSA) will be introduced in an intuitive style. Only elementary statistics will be assumed; no esoteric mathematical rigour will be presented.

TSA is an extensive and specialised field of statistics, owing its origin mostly to economics. Then it quickly found many important applications in physical science and engineering. To-day, there are very few disciplines that do not contemplate using TSA. In particular, it has also gained popularity in biological science during the more recent years.

One of the important turning points occurred after the publication by Box & Jenkins (B-J) in 1976, which initiated a few directions in its development and now is growing faster than ever.

This seminar intends to :

- Convince that TSA is an important technique to analyse time based data if one wants to be objective. This will be done with a discussion of smoothing, real trend, false trend, stock market and the random walk.

- Show TSA is a good method for forecasting with a discussion of the Auto regressive Integrated Moving Average (ARIMA or B-J) models, leading indicator series and Fourier transformation methods.

- Introduce a few typical examples of application in economics, physical and biological science.

- To point at future potential applications in biological science.

Time: 3:00 pm

Venue: Training Centre
CALM State Operations HQ
50 Hayman Rd
Como

Ex.

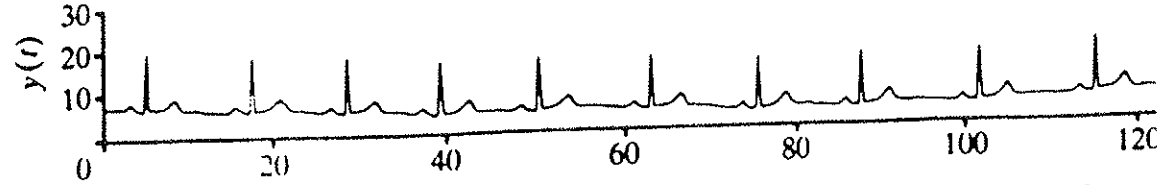


Fig. 1.1. An ECG trace from a healthy adult female. The inset shows enlargement of a single heartbeat.

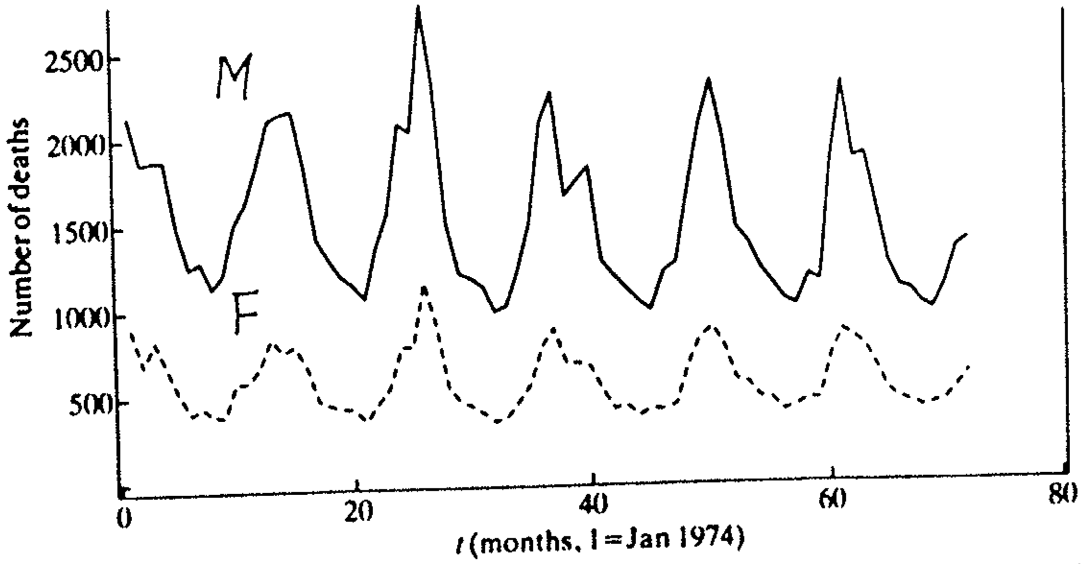


Fig. 1.6. Two time series of monthly returns of deaths in the United Kingdom attributed to bronchitis, emphysema, and asthma over the years 1974 to 1979.

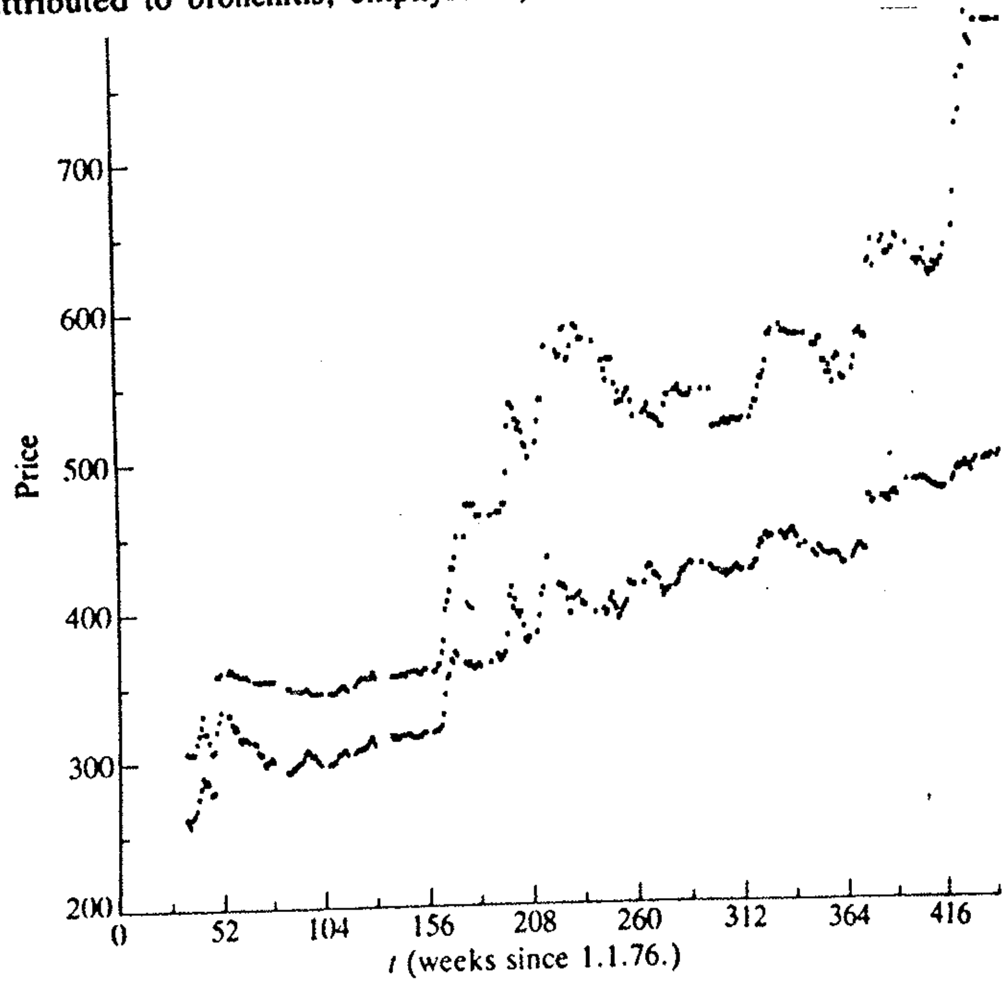
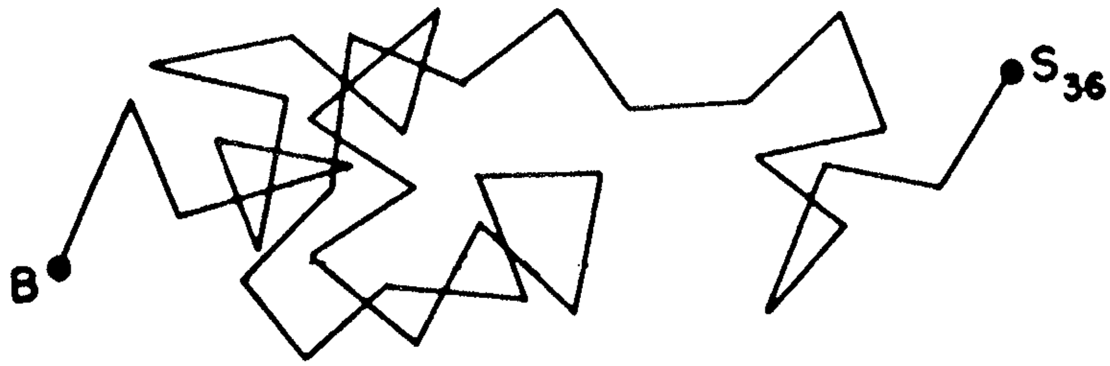
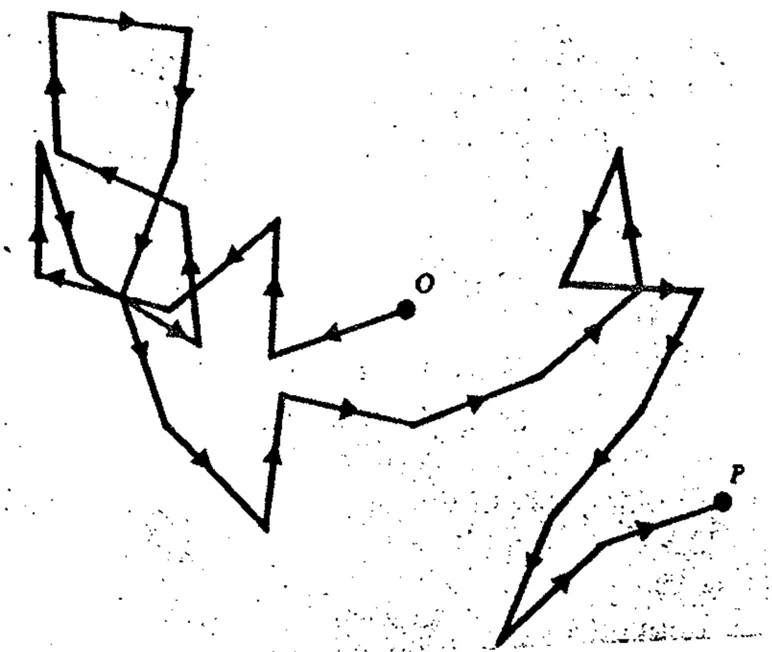


Fig. 1.4. Two time series of Australian wool prices at weekly markets during the period July 1976 to June 1984. The upper series represents the price paid for fine grade. grade wool (19 μm nominal thickness), and the lower series the floor price set by the Australian Wool Corporation.

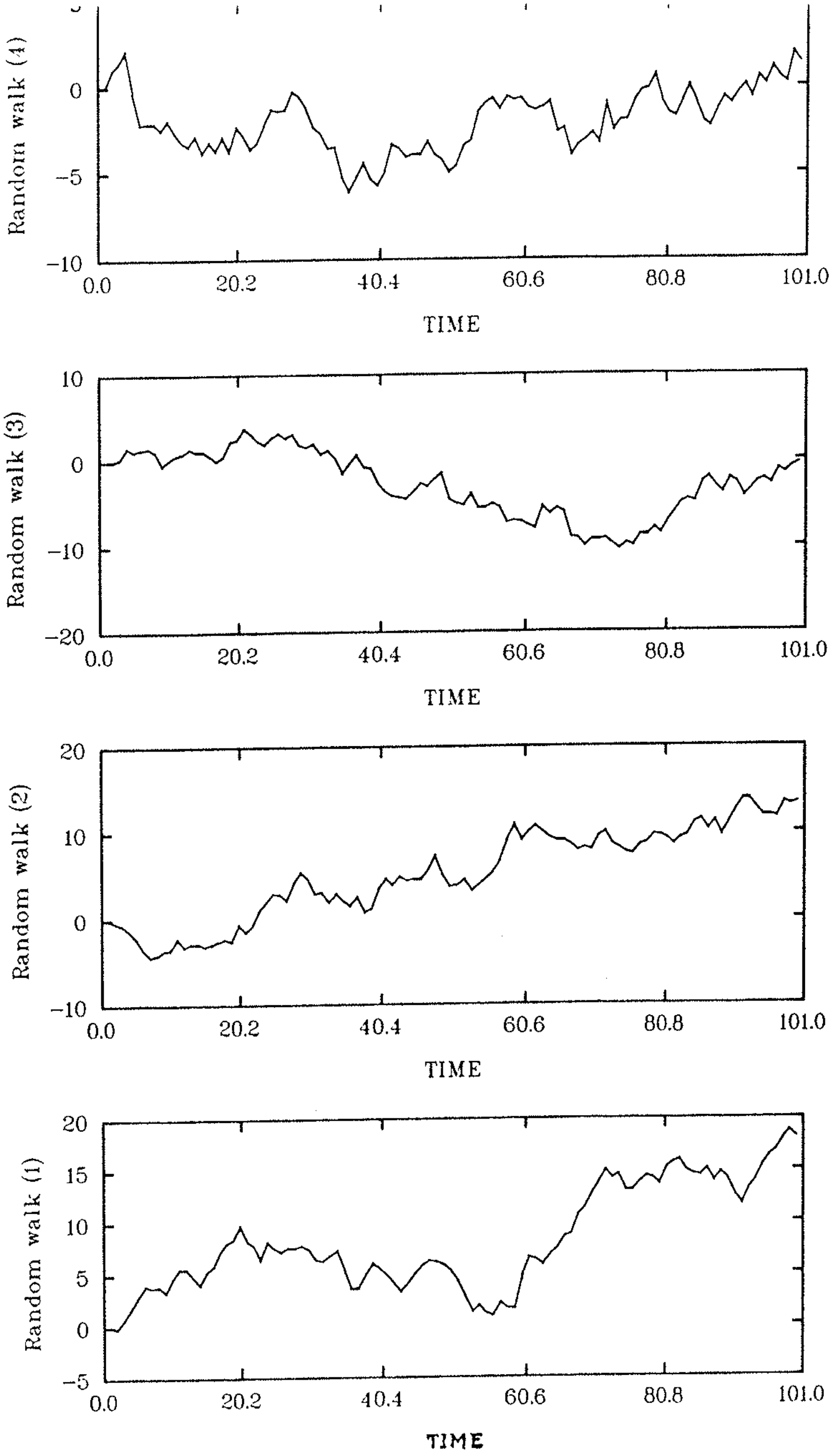
T1



A random walk of 36 steps of length l. How far is S_{36} from B? *Ans:* about $6l$ on the average.



Example of a random walk in two dimensions.



Random walk (4)
0
−5
−10
0.0
20.2
40.4
60.6
80.8
101.0
TIME
Random walk (3)
10
0
−10
−20
0.0
20.2
40.4
60.6
80.8
101.0
TIME
Random walk (2)
20
10
0
−10
0.0
20.2
40.4
60.6
80.8
101.0
TIME
Random walk (1)
20
15
10
5
0
−5
0.0
20.2
40.4
60.6
80.8
101.0
TIME

T3

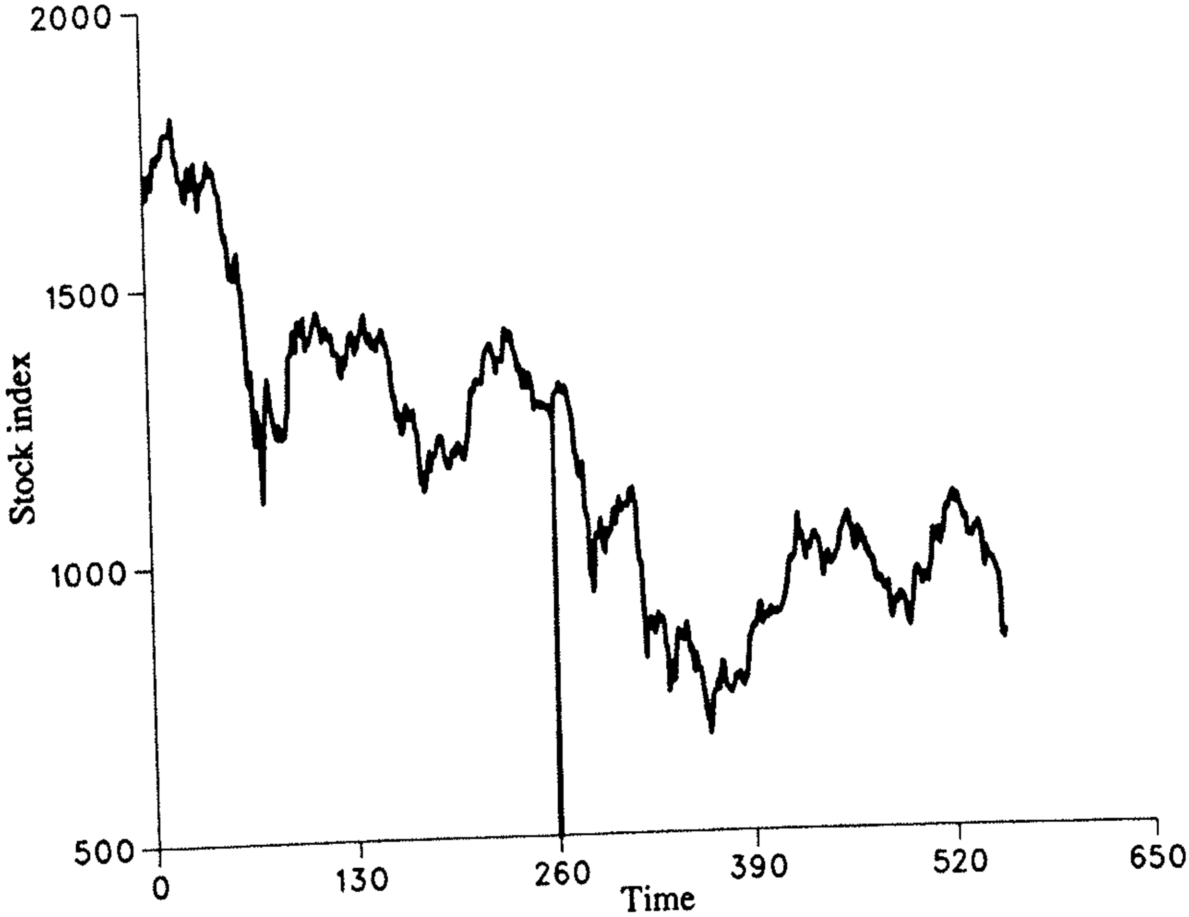


Fig. 9.3 Hang Seng index of Hong Kong stock prices from July 16, 1981, to September 31, 1983 (Series W11).

Main TSA models & formulae

Autoregressive, p th order, AR(p) :

$$Z_t = a_t + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p}$$

Autoregressive, first order, AR(1) , specially important:

$$Z_t = a_t + \phi_1 Z_{t-1}$$

Moving Average q th order, MA(q):

$$Z_t = a_t - \theta_1 Z_{t-1} - \theta_2 Z_{t-2} - \ldots\ldots - \theta_q Z_{t-q}$$

Autoregressive Moving Average, ARMA

Autoregressive Integrated Moving Average, ARIMA (B-J)
to take care of the stochastic trend

Autocorrelation function:

$$\rho_{s,t} \equiv \mathrm{Cov}(Z_t, Z_s)$$

$$\rho_k \equiv \mathrm{Cov}(Z_t, Z_{t-k}),$$ stationary series, or detrended

Partial autocorrelation function:

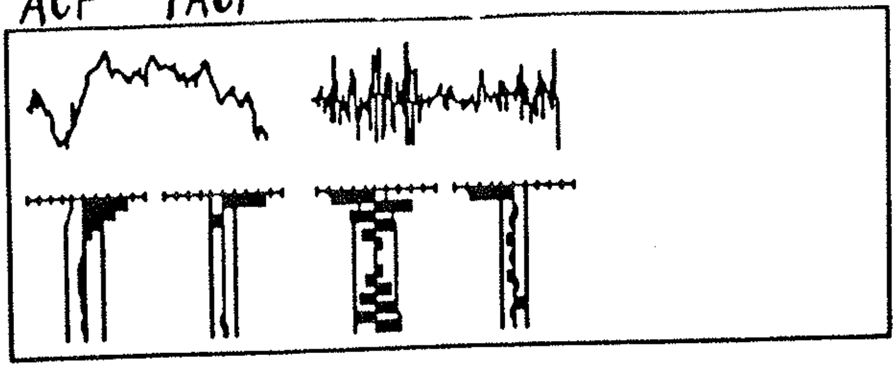
$$\rho_{kk}$$

Seasonal AR(P) :

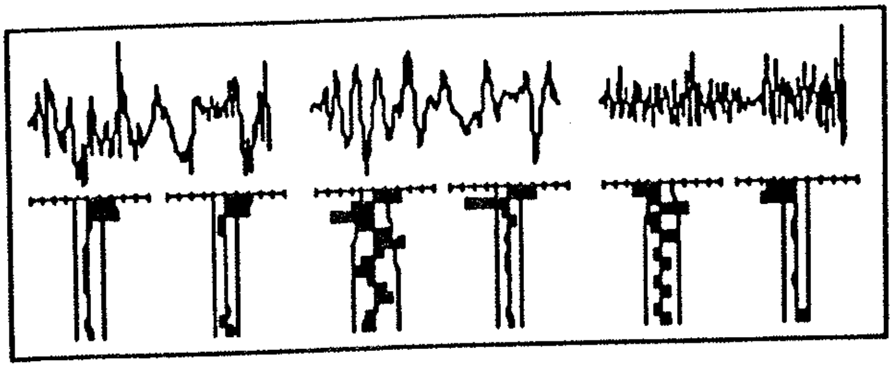
$$Z_t = a_t + \phi_1 Z_{t-s} + \phi_2 Z_{t-2s} + \ldots + \phi_p Z_{t-ps}$$

ACF PACF

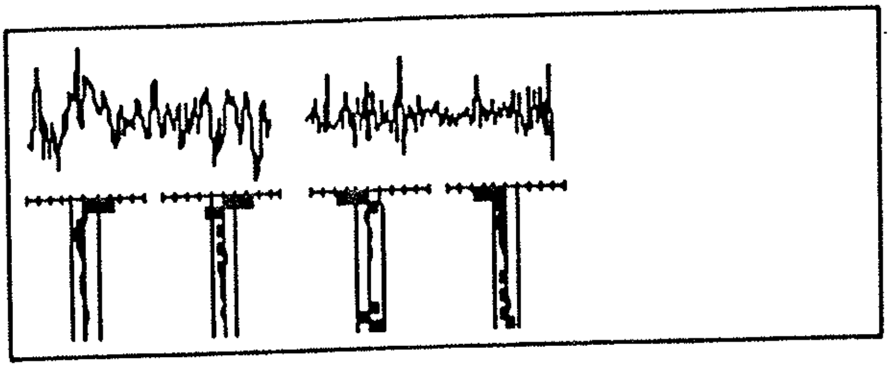
AR(1)



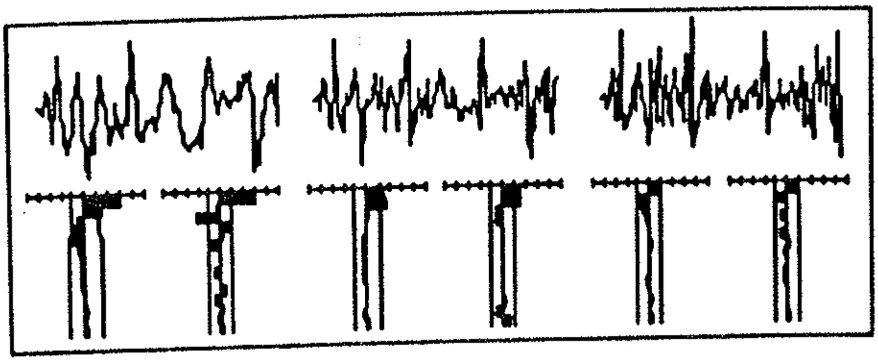
AR(2)



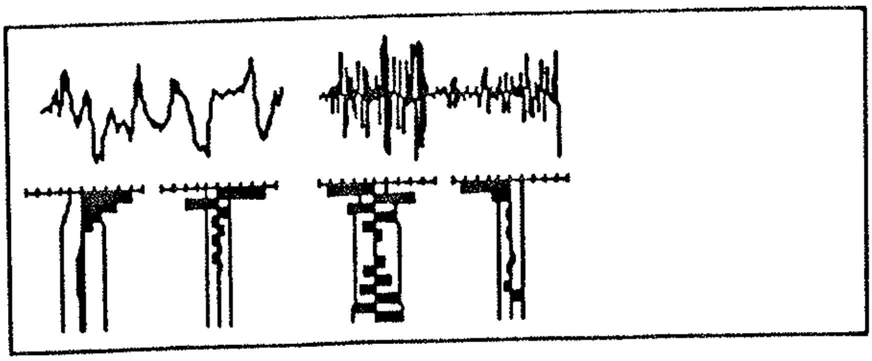
MA(1)



MA(2)



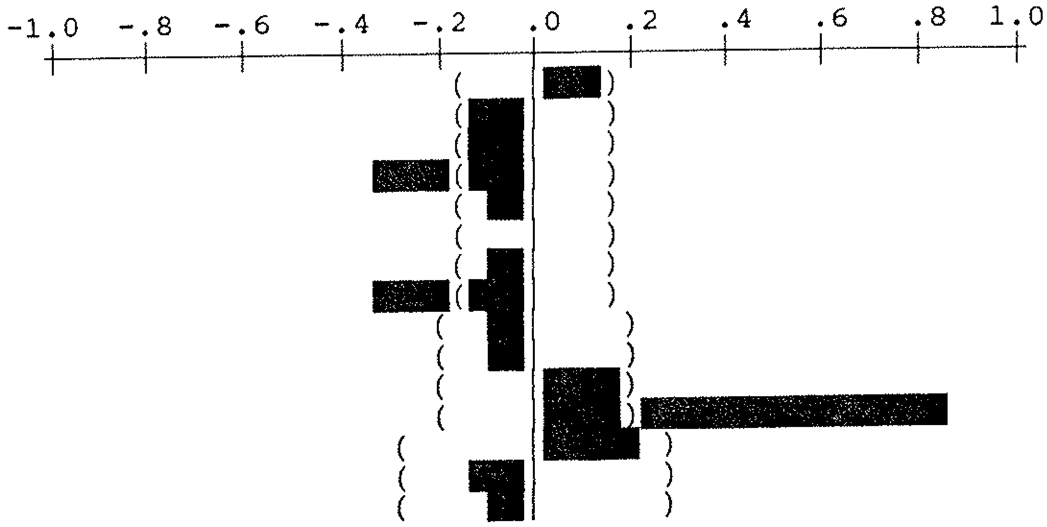
ARMA(1,1)



PLOT OF PASS (*After differencing*)
NUMBER OF CASES = 143
MEAN OF SERIES = 0.009
STANDARD DEVIATION OF SERIES = 0.106

PLOT OF AUTOCORRELATIONS

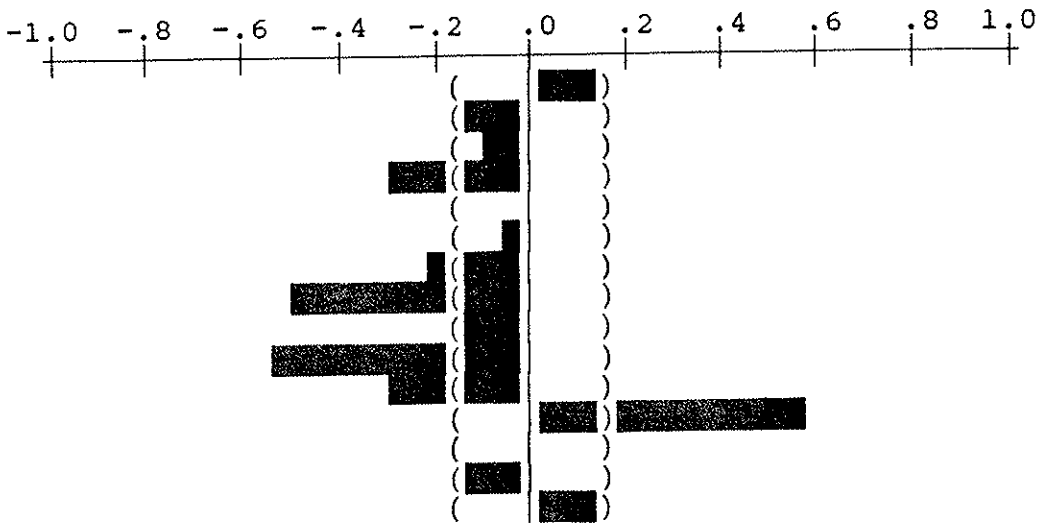
LAG	CORR	SE
1	.200	.084
2	-.120	.087
3	-.151	.088
4	-.322	.090
5	-.084	.098
6	.026	.098
7	-.111	.098
8	-.337	.099
9	-.116	.107
10	-.109	.108
11	.206	.108
12	.841	.111
13	.215	.149
14	-.140	.151
15	-.116	.152



PLOT OF PASS
NUMBER OF CASES = 143
MEAN OF SERIES = 0.009
STANDARD DEVIATION OF SERIES = 0.106

PLOT OF PARTIAL AUTOCORRELATIONS

LAG	CORR	SE
1	.200	.084
2	-.167	.084
3	-.096	.084
4	-.311	.084
5	.008	.084
6	-.075	.084
7	-.210	.084
8	-.495	.084
9	-.192	.084
10	-.532	.084
11	-.302	.084
12	.586	.084
13	.026	.084
14	-.181	.084
15	.120	.084

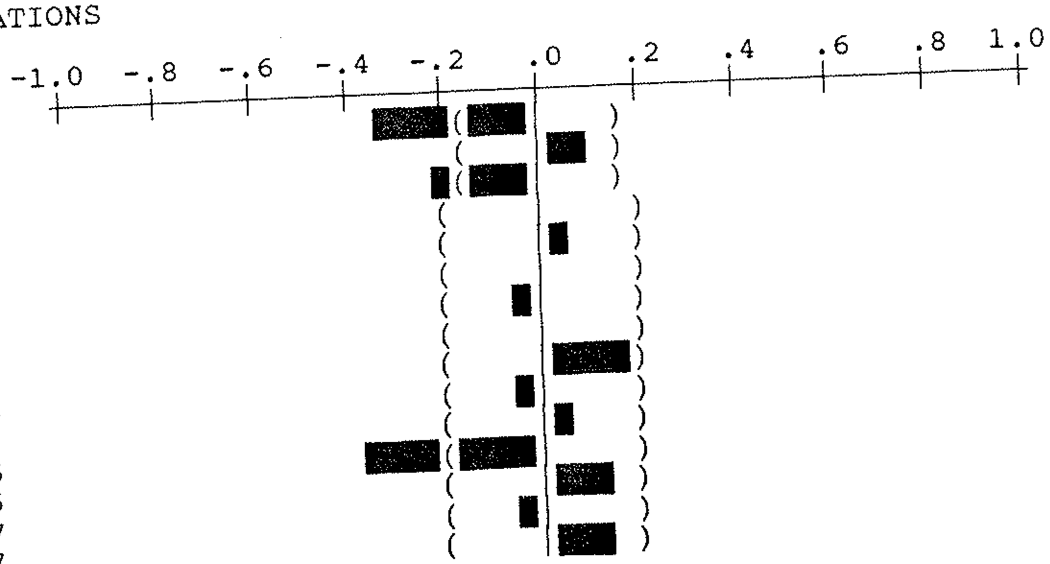


SERIES IS TRANSFORMED

PLOT OF PASS (After ∇_{12})
NUMBER OF CASES = 131
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.046

PLOT OF AUTOCORRELATIONS

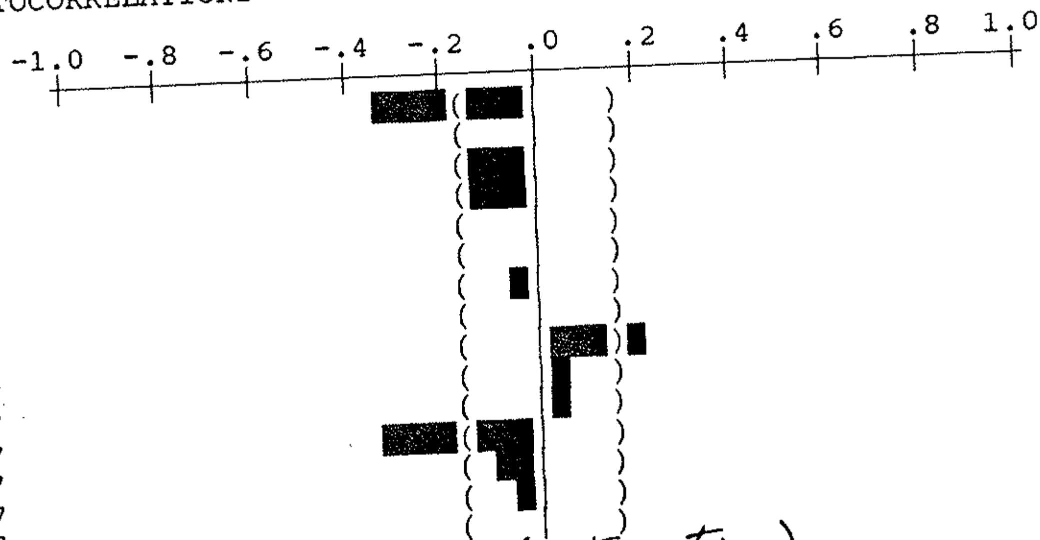
LAG	CORR	SE
1	-.341	.087
2	.105	.097
3	-.202	.098
4	.021	.101
5	.056	.101
6	.031	.101
7	-.056	.101
8	-.001	.102
9	.176	.102
10	-.076	.104
11	.064	.104
12	-.387	.105
13	.152	.115
14	-.058	.117
15	.150	.117



PLOT OF PASS
NUMBER OF CASES = 131
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.046

PLOT OF PARTIAL AUTOCORRELATIONS

LAG	CORR	SE
1	-.341	.087
2	-.013	.087
3	-.193	.087
4	-.125	.087
5	.033	.087
6	.035	.087
7	-.060	.087
8	-.020	.087
9	.226	.087
10	.043	.087
11	.047	.087
12	-.339	.087
13	-.109	.087
14	-.077	.087
15	-.022	.087



(estimation)

ITERATION	SUM OF SQUARES	PARAMETER VALUES	
0	.2392764D+00	.100	.100
1	.1835532D+00	.345	.433
2	.1764962D+00	.449	.633
3	.1759952D+00	.416	.592
4	.1758742D+00	.409	.613
5	.1758463D+00	.392	.614
6	.1758443D+00	.396	.613

7	.1758443D+00	.396	.613
8	.1758443D+00	.396	.613
9	.1758443D+00	.396	.613
10	.1758443D+00	.396	.613

FINAL VALUE OF MSE IS 0.001

INDEX	TYPE	ESTIMATE	A.S.E.	LOWER	<95%>	UPPER
1	MA	0.396	0.093	0.212		0.579
2	SMA	0.613	0.074	0.467		0.760

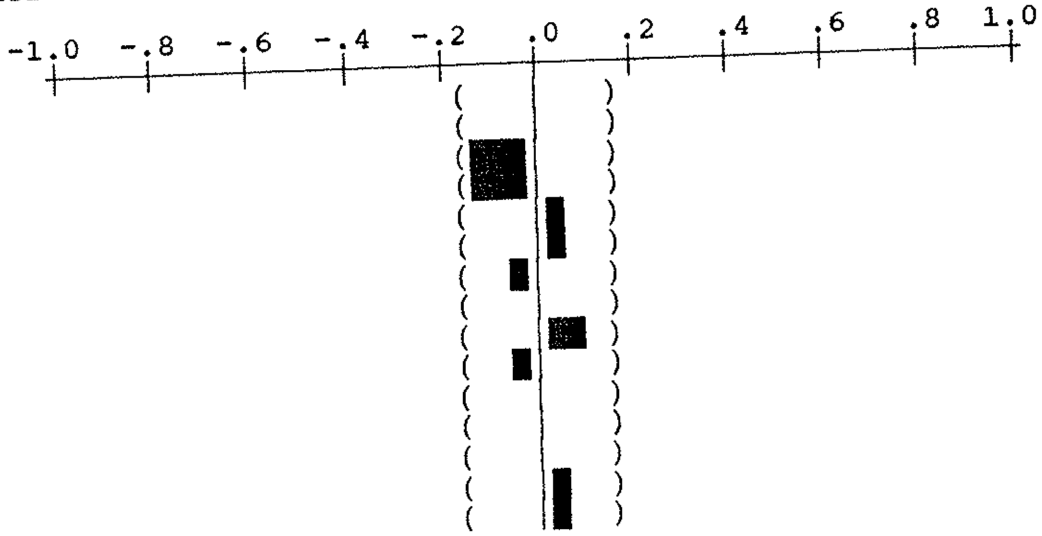
ASYMPTOTIC CORRELATION MATRIX OF PARAMETERS

	1	2
1	1.000	
2	-0.171	1.000

PLOT OF RESIDUAL
NUMBER OF CASES = 131
MEAN OF SERIES = 0.000
STANDARD DEVIATION OF SERIES = 0.036

PLOT OF AUTOCORRELATIONS

LAG	CORR	SE
1	.017	.087
2	.019	.087
3	-.126	.087
4	-.142	.089
5	.051	.091
6	.062	.091
7	-.072	.091
8	-.038	.091
9	.103	.092
10	-.079	.092
11	.024	.093
12	-.012	.093
13	.031	.093
14	.043	.093
15	.048	.093



ITERATION	SUM OF SQUARES	PARAMETER VALUES	
0	.2392764D+00	.100	.100
1	.1835532D+00	.345	.433
2	.1764962D+00	.449	.633
3	.1759952D+00	.416	.592
4	.1758742D+00	.409	.613
5	.1758463D+00	.392	.614
6	.1758443D+00	.396	.613
7	.1758443D+00	.396	.613
8	.1758443D+00	.396	.613
9	.1758443D+00	.396	.613
10	.1758443D+00	.396	.613

FINAL VALUE OF MSE IS 0.001

INDEX	TYPE	ESTIMATE	A.S.E.	LOWER	<95%>	UPPER
1	MA	0.396	0.093	0.212		0.579
2	SMA	0.613	0.074	0.467		0.760

ASYMPTOTIC CORRELATION MATRIX OF PARAMETERS

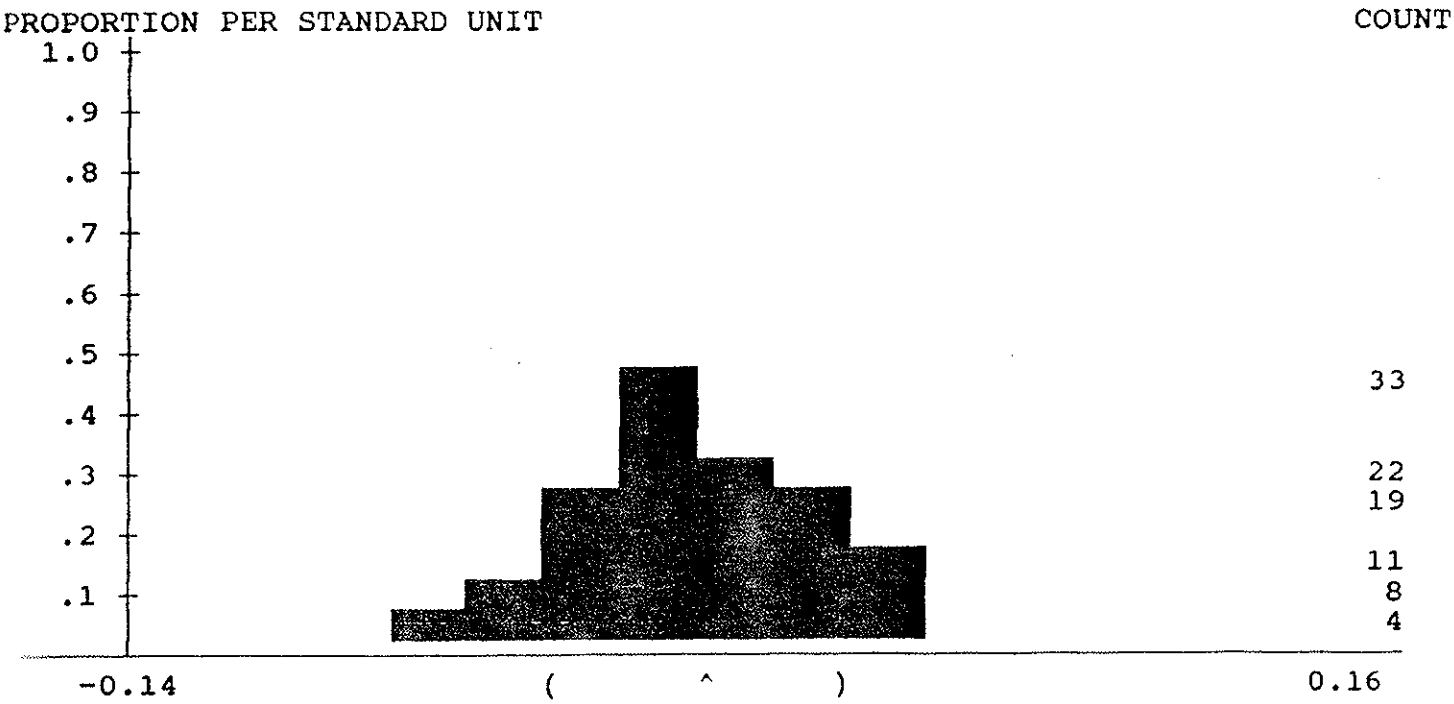
	1	2
1	1.000	
2	-0.171	1.000

FORECAST VALUES

PERIOD	LOWER95	FORECAST	UPPER95
1961.01	418.862	450.296	484.090
1961.02	391.976	426.557	464.189
1961.03	438.329	482.099	530.240
1961.04	443.296	492.246	546.602
1961.05	453.832	508.378	569.480
1961.06	516.623	583.439	658.898
1961.07	587.307	668.338	760.549
1961.08	581.136	666.091	763.464
1961.09	484.248	558.844	644.931
1961.10	427.648	496.754	577.028

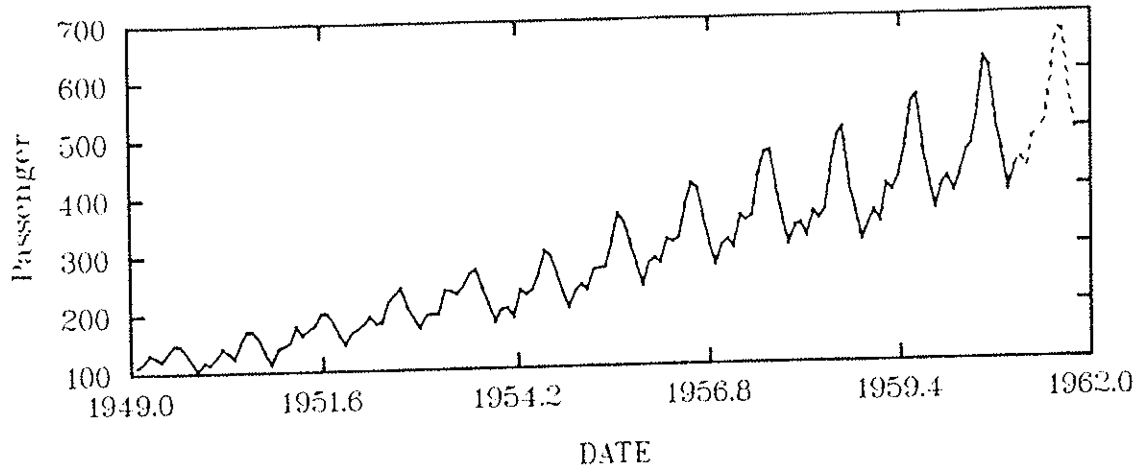
FILE HAS BEEN SAVED AND CLOSED

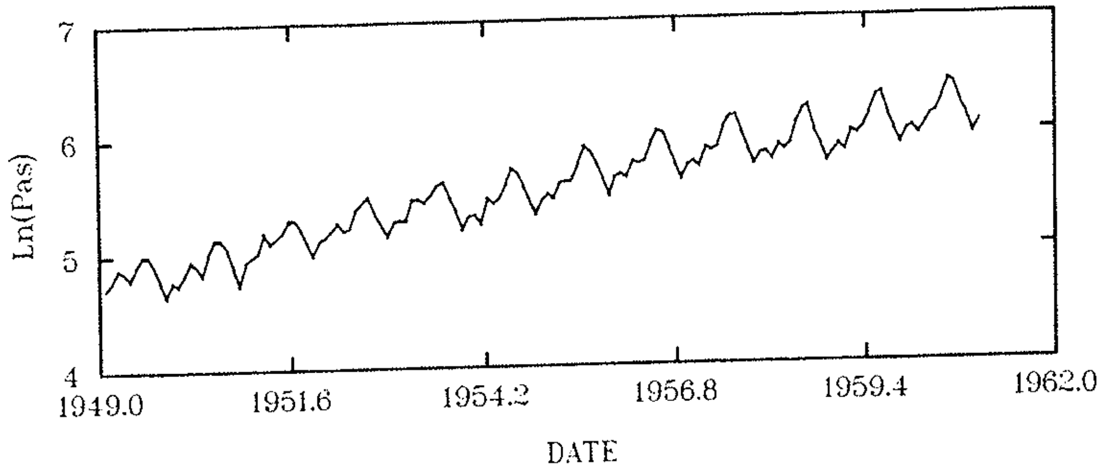
* end of prg

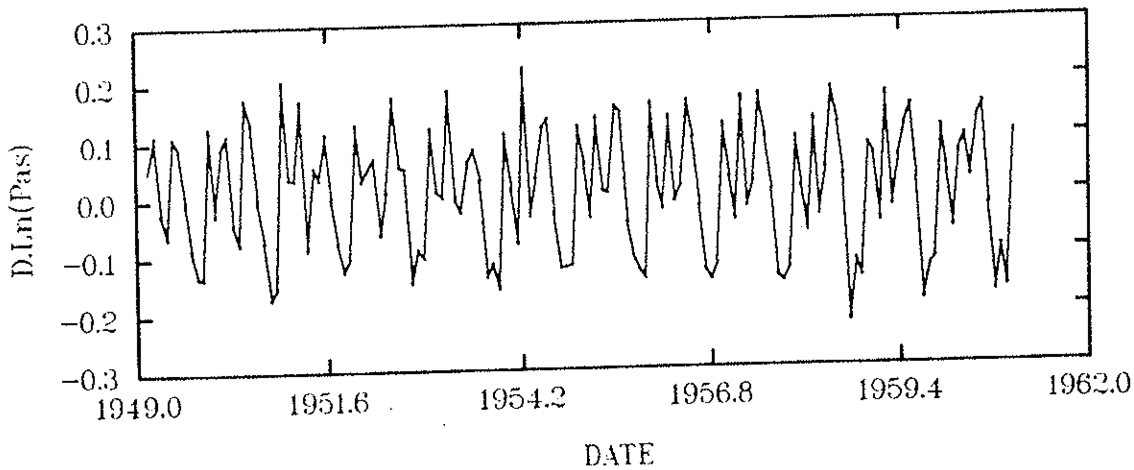


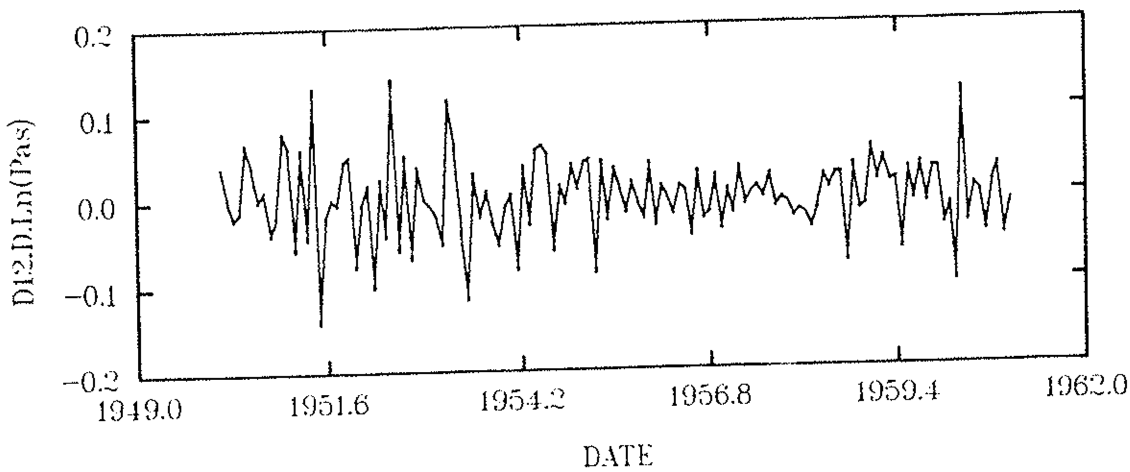
RESIDUAL

24 CASES WITH MISSING VALUES EXCLUDED FROM PLOT


700
600
500
400
300
200
100
Passenger
1949.0
1951.6
1954.2
1956.8
1959.4
1962.0
DATE

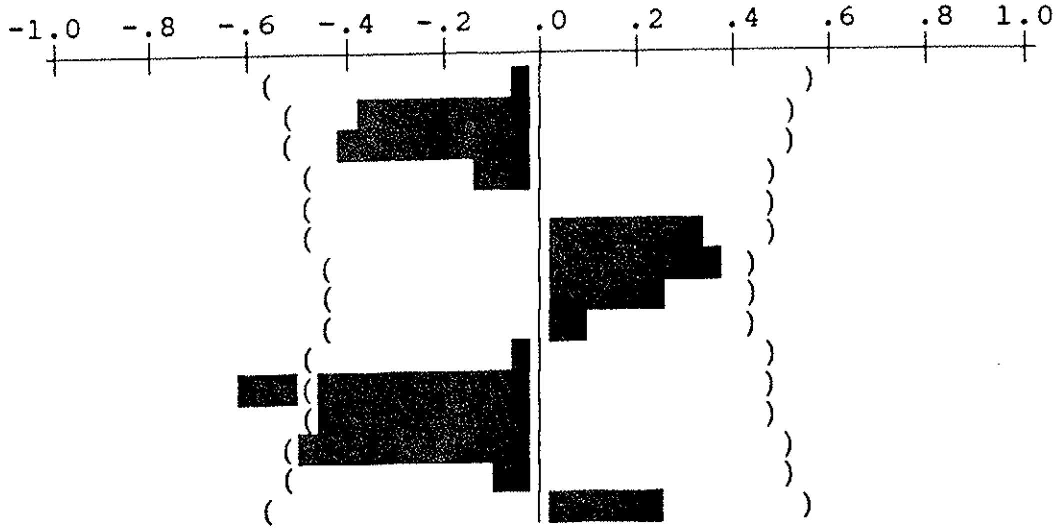

7
6
5
4
Ln(Pas)
1949.0
1951.6
1954.2
1956.8
1959.4
1962.0
DATE


0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
D.Ln(Pas)
1949.0
1951.6
1954.2
1956.8
1959.4
1962.0
DATE

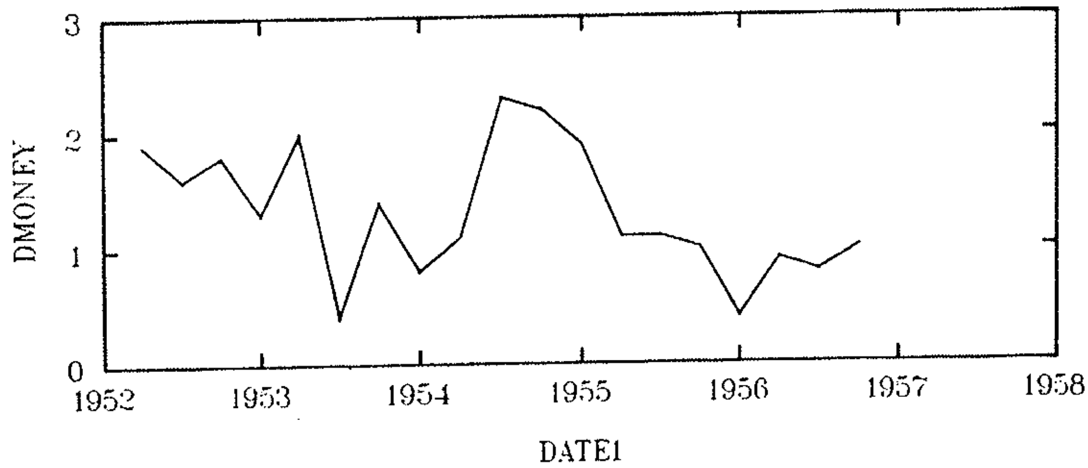

0.2
0.1
0.0
−0.1
−0.2
D12.D.Ln(Pas)
1949.0
1951.6
1954.2
1956.8
1959.4
1962.0
DATE

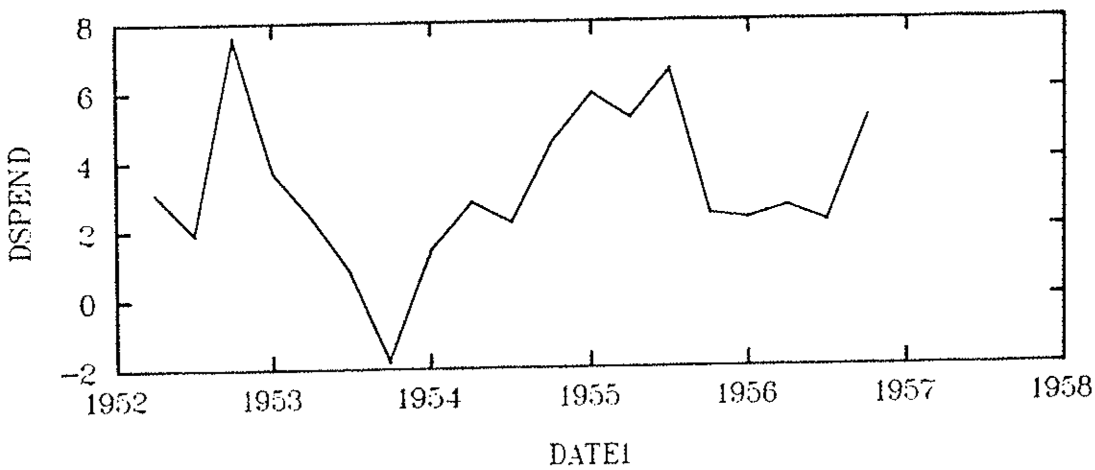
PLOT OF CROSS CORRELATIONS

LAG	CORR	SE
-7	-.048	.289
-6	-.395	.277
-5	-.405	.267
-4	-.123	.258
-3	.004	.250
-2	.342	.243
-1	.368	.236
0	.250	.229
1	.098	.236
2	-.068	.243
3	-.617	.250
4	-.494	.258
5	-.516	.267
6	-.085	.277
7	.265	.289



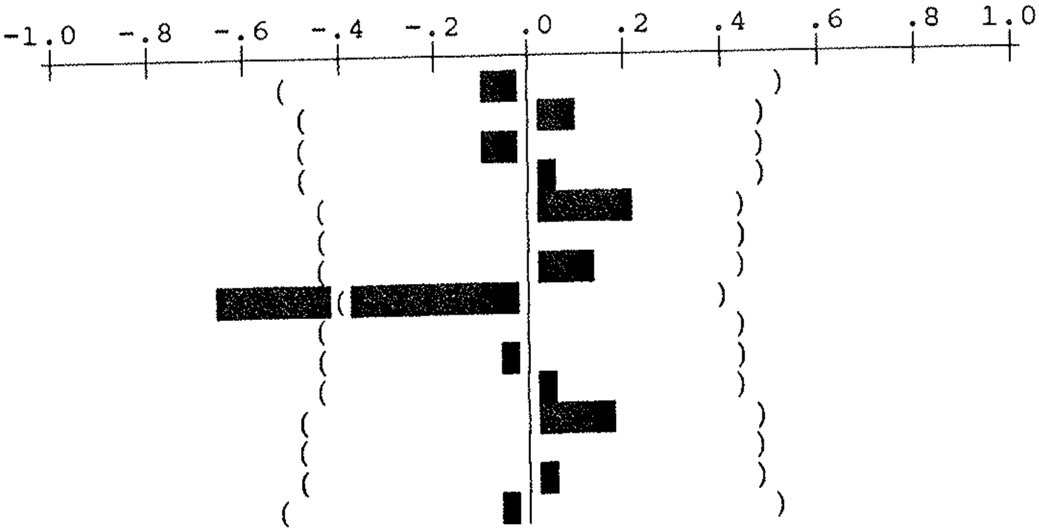
* end of prg



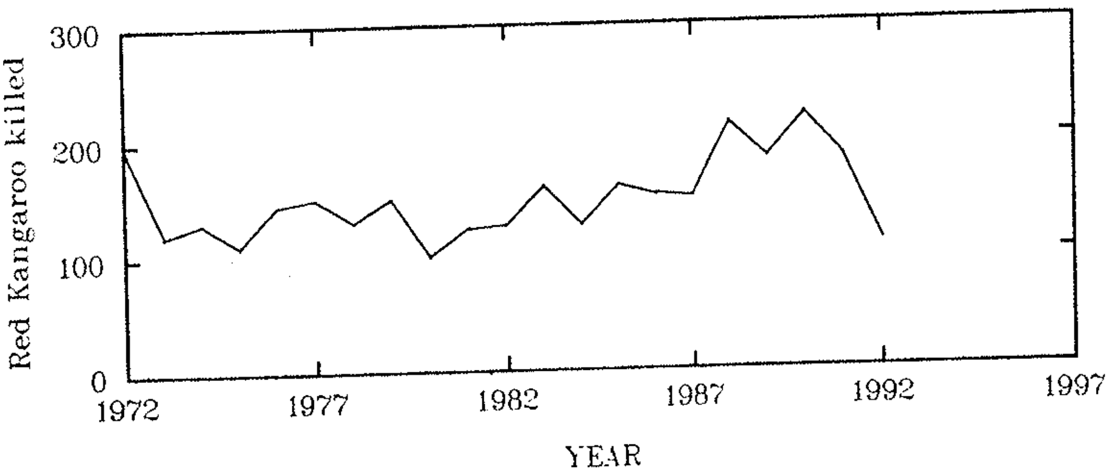


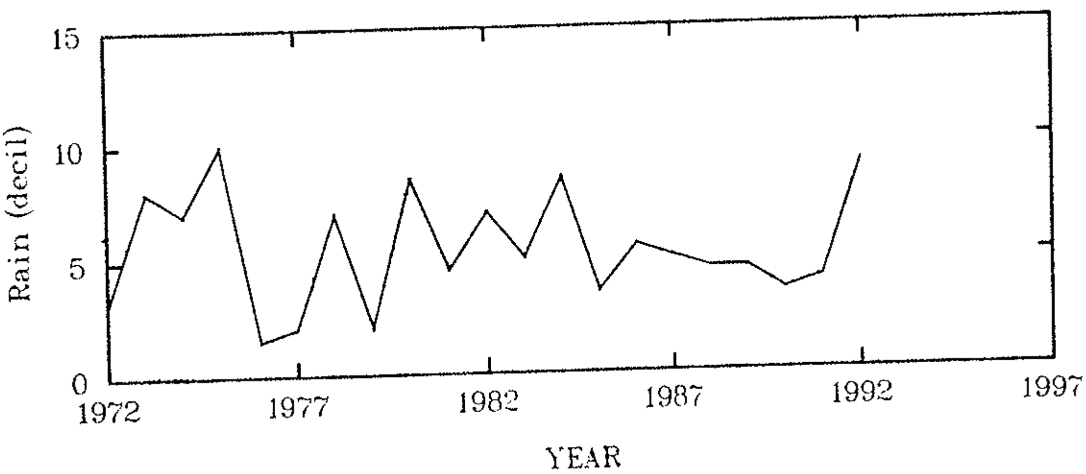
T12

LAG	CORR	SE
-7	-.089	.267
-6	.084	.258
-5	-.098	.250
-4	.053	.243
-3	.235	.236
-2	-.000	.229
-1	.133	.224
0	-.654	.218
1	.022	.224
2	-.073	.229
3	.077	.236
4	.162	.243
5	-.005	.250
6	.046	.258
7	-.048	.267

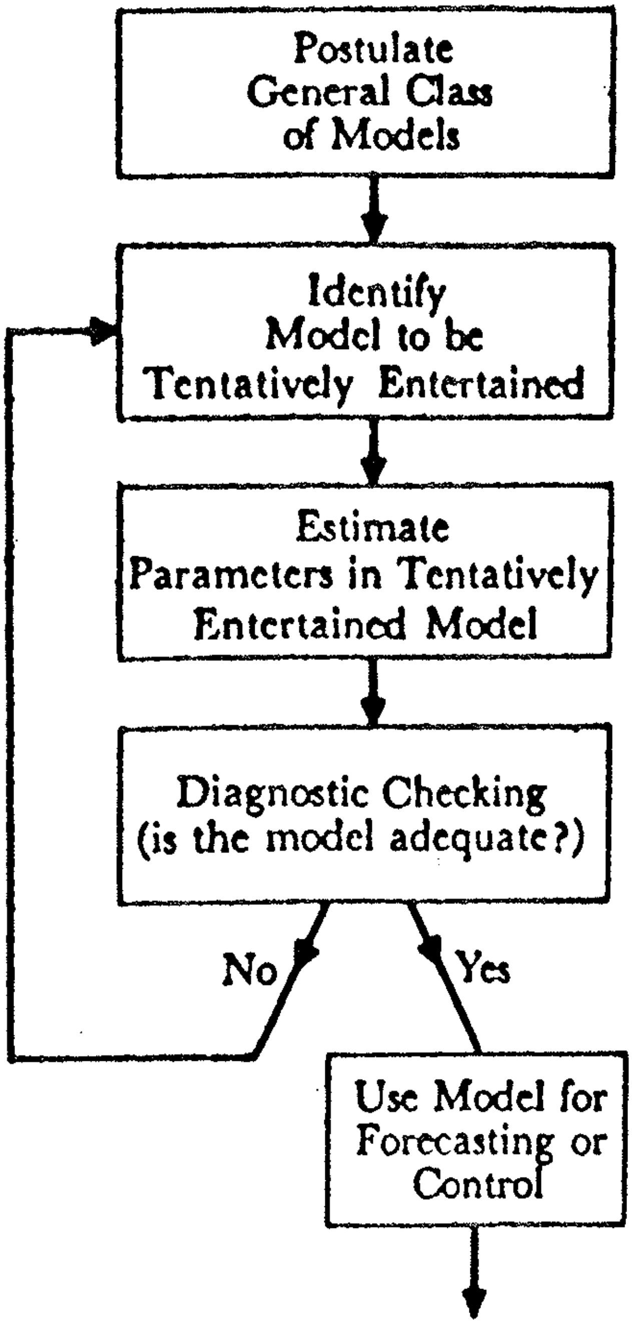


* end of prg





Basic Ideas in Model Building



Fourier Transform for TSA

acvf ≡ autocovariance function
ccvf ≡ cross covariance function

Fourier transform has inverse

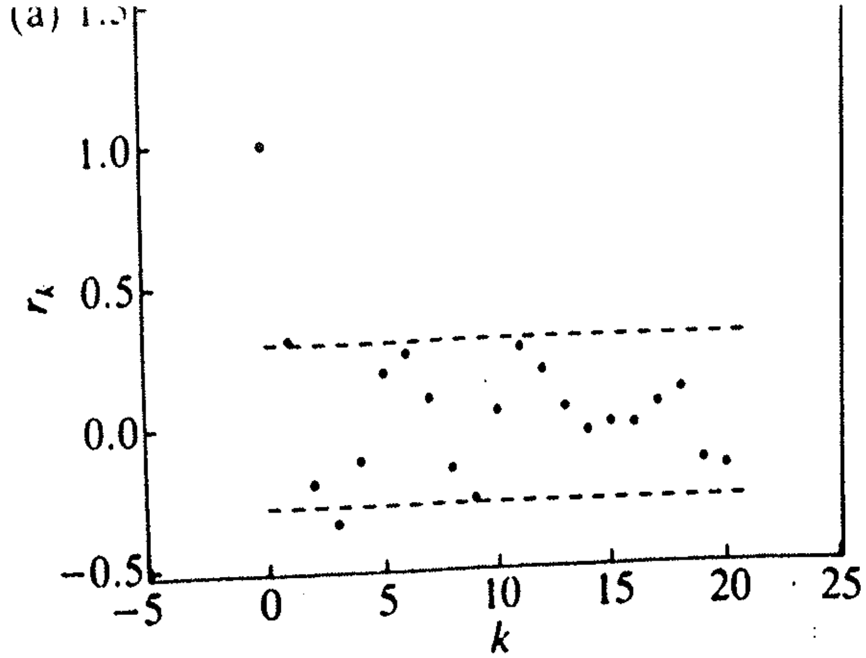
[sample] acvf	F T <--->	[sample] spectrum	--->	power spectrum $\Gamma(f)$
[sample] ccvf	F T <-->	[sample] cross spectrum	--->	squared coherency spectrum

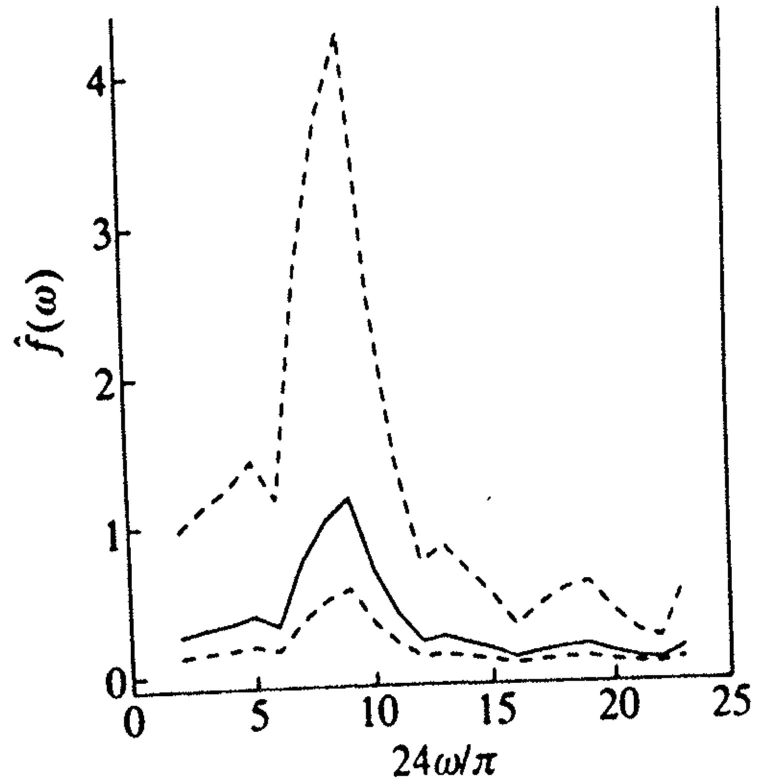
Leading indicators, done by (JW 342):

[sample] cross correlation function ρ_{xy}

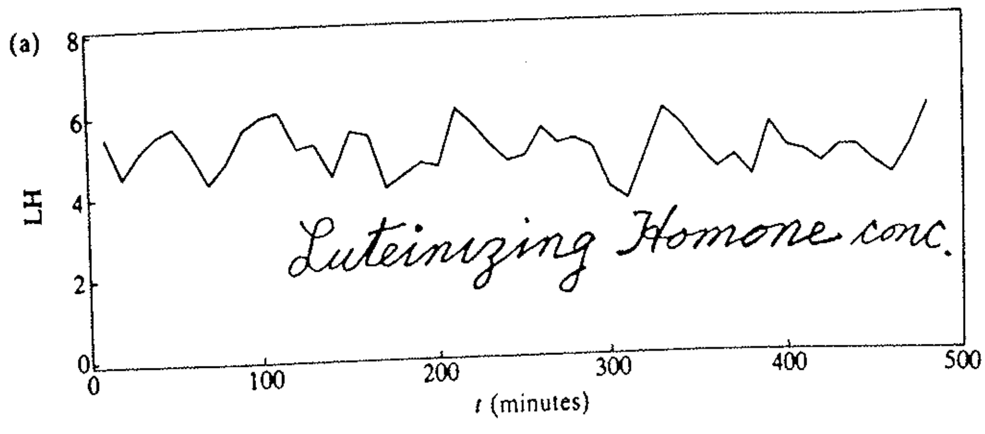
or

[sample] phase spectrum $\Psi(\omega)$





The spectral estimate



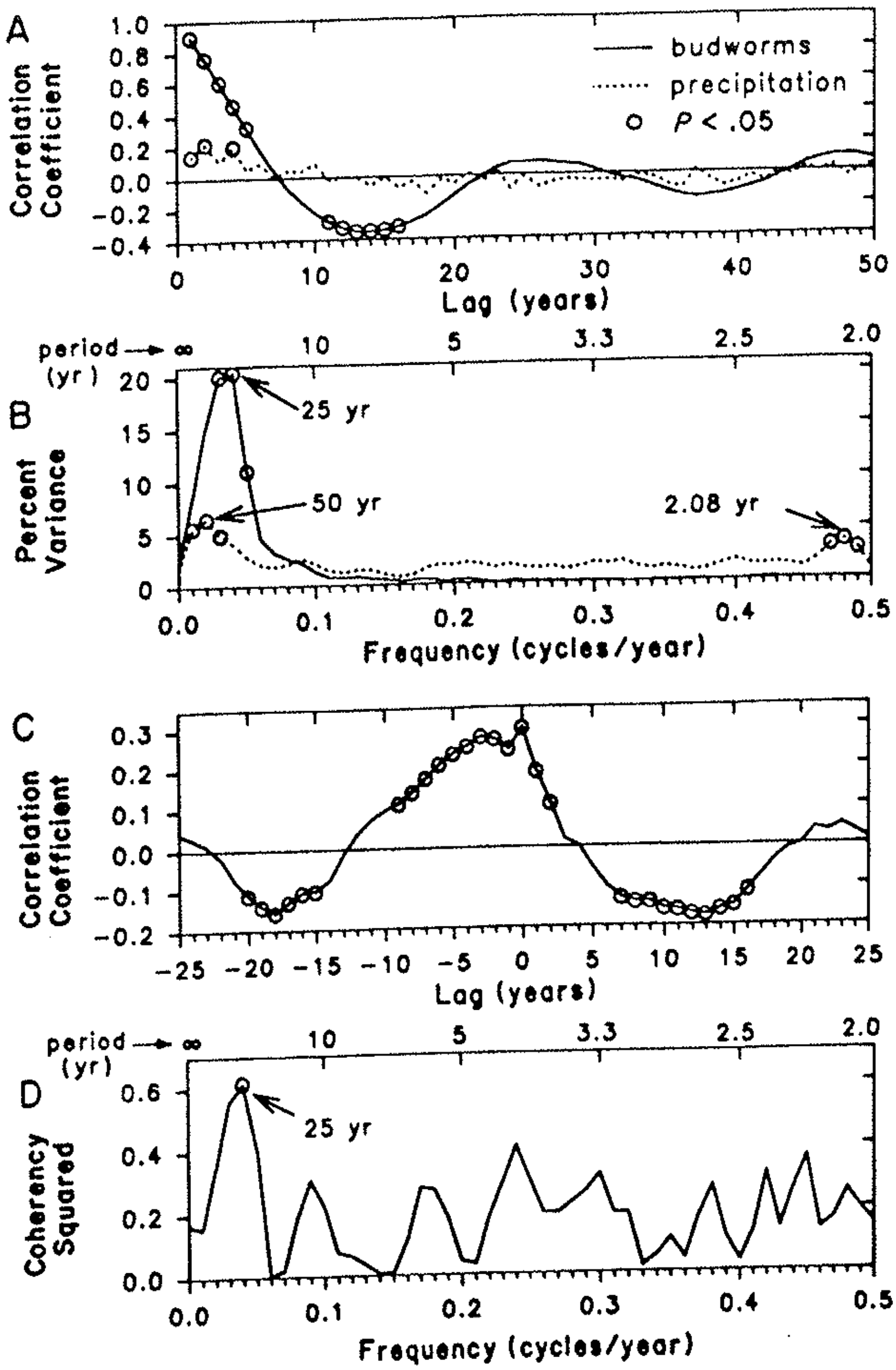


FIG. 11. (A) Autocorrelation functions of reconstructed regional budworm time series and spring precipitation (1600–1989), (B) univariate power spectra of the two time series, (C) cross-correlation function of budworms and precipitation, and (D) coherency squared spectra from the bivariate spectral analysis of budworms and precipitation. Circled values were significant at $P < .05$.

Other conclusions

- Spruce budworms outbreak cycle is significant at T= 25 yr
- The two series also associate strongly at T= 25 yr also
- they found rain leads by 3 yrs at T= 25 yr

Positive signs for the successful use of spectral analysis

1. Series are either stationary, or the trend is not of direct interest, see graph below for the opposite.

2. Series are long

3. interpretation of underlying processes in terms of cyclic patterns of variation makes sense.

4. Parametric models are difficult to justify.

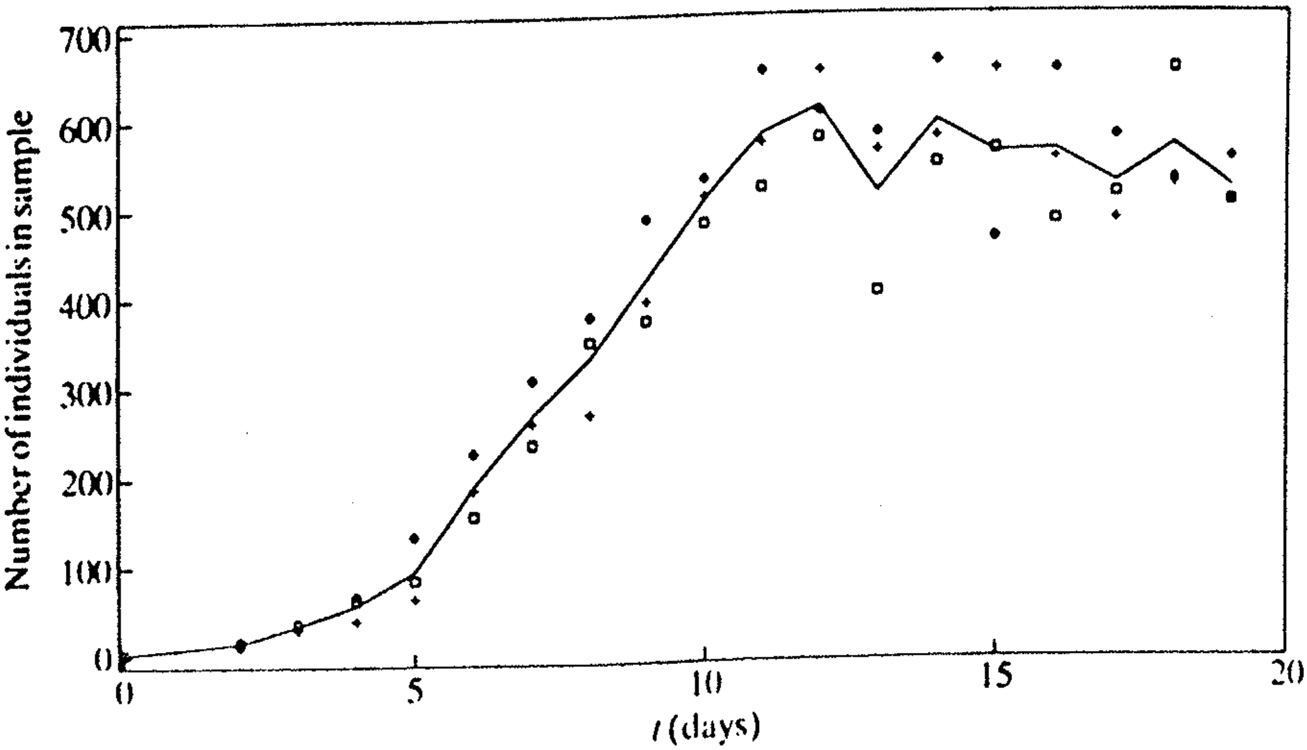


Fig. 1.7. Data on the growth of colonies of *paramecium aurelium*. +, First replicate; □, second replicate; ◇, third replicate; —, mean of three replicates.

Repeated Measurements

Let the model be $Y_i(t) = \mu_i(t) + Y_i(t)$

where

i = the label of the i th series (an experimental unit)

l = length of each series, (at the same time intervals)

$\mu_i(t)$ is the trend of direct interest

$Z_i(t)$ is stationary

Assume each Y_i Normal, then we have a multivariate Normal distribution, and we can use model:

$$\mathbf{Y}_i \sim MVN(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}), \quad i=1,2,...,m$$

where : $\boldsymbol{\Sigma}$ is a $l \times l$ covariance matrix

and there are the lower triangular & the diagonal elements to be estimated, a total of $l(l+1)/2$.

Once we assume stationary condition on

$Z_i(t) = Y_i(t) - \mu_i(t)$,

the number of elements to be estimated reduces to l, which is the possible number of $|j - k|$ values.