LECTURE NOTES IN STATISTICS provides a format for the informal publication of monographs, conference proceedings, case studies, and similar material of theoretical or empirical importance. The timeliness of a manuscript is more important than its form which may be unfinished or tentative. Thus, in some instances, proofs may be merely outlined and results presented which have recently been or will later be published elsewhere. If possible, a subject index should be included.

Publication of the Lecture Notes is intended as a service to the international statistical community, in that a commercial publisher, Springer-Verlag, can provide efficient distribution of documents which would otherwise have a restricted readership. Once published and copyrighted they can be documented and discussed in the scientific literature.

Authors receive 50 free copies of their books but no sales royalty. If necessary, Springer-Verlag can contribute toward the costs of the technical preparation of the manuscript.

Manuscripts

Manuscripts shuld be no less than 100 and preferably no more than 500 pages in length.
They are reproduced by a photographic process and therefore must be typed with extreme care. Symbols not on the typewriter should be inserted by hand in indelible black ink. Corrections to the typescript should be made by pasting in the new text or painting out errors with white correction fluid. The typescript is reduced slightly in size during reproduction; best results will not be obtained unless on each page a typing area of 18 x 26.5 cm (7 x 10½ inches) is respected. On request, the publisher can supply paper with the typing area outlined. More detailed typing instructions are also available on request.

Manuscripts generated by a word-processor or computerized typesetting are in principle acceptable. However if the quality of this output differs significantly from that of a standard typewriter, then authors should contact Springer-Verlag at an early stage.

Manuscripts should be sent to Prof. D.R. Brillinger, Department of Statistics, University of California at Berkeley, CA 94720, Dr. S. Fienberg, Department of Statistics, Carnegie-Mellon University, Pittsburgh, PA 15213, Dr. J. Gani, University of Kentucky, Department of Statistics, Lexington, KY 40506, Dr. J. Hartigan, Department of Statistics, Yale University, New Haven, CT 06520, Dr. K. Krickeberg, U.E.R. de Mathématiques, Logique Formelle et Informatique, Université de Paris V, 12 rue Cujas, F-75005 Paris, France, or to the Mathematics Editor of Springer-Verlag in New York or Heidelberg.

Springer-Verlag, 175 Fifth Avenue, New York, NY 10010/USA
Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani,
J. Hartigan, and K. Krickeberg

24

T. Subba Rao
M.M. Gabr

An Introduction to Bispectral Analysis and Bilinear Time Series Models

Lecture Notes in Statistics

Vol. 1: R. A. Fisher: An Appreciation. Edited by S. E. Fienberg and D. V. Hinkley. XI, 208 pages, 1980.

Vol. 2: Mathematical Statistics and Probability Theory. Proceedings 1978. Edited by W. Klonecki, A. Kozek, and J. Rosiński. XXIV, 373 pages, 1980.

Vol. 3: B. D. Spencer, Benefit-Cost Analysis of Data Used to Allocate Funds. VIII, 296 pages, 1980.

Vol. 4: E. A. van Doorn, Stochastic Monotonicity and Queueing Applications of Birth-Death Processes. VI, 118 pages, 1981.

Vol. 5: T. Rolski, Stationary Random Processes Associated with Point Processes. VI, 139 pages, 1981.

Vol. 6: S. S. Gupta and D.-Y. Huang, Multiple Statistical Decision Theory: Recent Developments. VIII, 104 pages, 1981.

Vol. 7: M. Akahira and K. Takeuchi, Asymptotic Efficiency of Statistical Estimators. VIII, 242 pages, 1981.

Vol. 8: The First Pannonian Symposium on Mathematical Statistics. Edited by P. Révész, L. Schmetterer, and V. M. Zolotarev. VI, 308 pages, 1981.

Vol. 9: B. Jørgensen, Statistical Properties of the Generalized Inverse Gaussian Distribution. VI, 188 pages, 1981.

Vol. 10: A. A. McIntosh, Fitting Linear Models: An Application on Conjugate Gradient Algorithms. VI, 200 pages, 1982.

Vol. 11: D. F. Nicholls and B. G. Quinn, Random Coefficient Autoregressive Models: An Introduction. V, 154 pages, 1982.

Vol. 12: M. Jacobsen, Statistical Analysis of Counting Processes. VII, 226 pages, 1982.

Vol. 13: J. Pfanzagl (with the assistance of W. Wefelmeyer), Contributions to a General Asymptotic Statistical Theory. VII, 315 pages, 1982.

Vol. 14: GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Edited by R. Gilchrist. V, 188 pages, 1982.

Vol. 15: K. R. W. Brewer and M. Hanif, Sampling with Unequal Probabilities. IX, 164 pages, 1983.

Vol. 16: Specifying Statistical Models: From Parametric to Non-Parametric, Using Bayesian or Non-Bayesian Approaches. Edited by J. P. Florens, M. Mouchart, J. P. Raoult, L. Simar, and A. F. M. Smith. XI, 204 pages, 1983.

Vol. 17: I. V. Basawa and D. J. Scott, Asymptotic Optimal Inference for Non-Ergodic Models. IX, 170 pages, 1983.

Vol. 18: W. Britton, Conjugate Duality and the Exponential Fourier Spectrum. V, 226 pages, 1983.

Vol. 19: L. Fernholz, von Mises Calculus For Statistical Functionals. VIII, 124 pages, 1983.

Vol. 20: Mathematical Learning Models – Theory and Algororithms: Proceedings of a Conference. Edited by U. Herkenrath, D. Kalin, W. Vogel. XIV, 226 pages, 1983.

Vol. 21: H. Tong, Threshold Models in Non-linear Time Series Analysis. X, 323 pages, 1983.

Vol. 22: S. Johansen, Functional Relations, Random Coefficients and Nonlinear Regression with Application to Kinetic Data. VIII, 126 pages. 1984.

Vol. 23: D. G. Saphire, Estimation of Victimization Prevalence Using Data from the National Crime Survey. V, 165 pages. 1984.

Vol. 24: T. S. Rao, M. M. Gabr, An Introduction to Bispectral Analysis and Bilinear Time Series Models. VIII, 280 pages, 1984.

Lecture Notes in Statistics

Edited by D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, and K. Krickeberg

24

T. Subba Rao
M. M. Gabr

An Introduction to Bispectral Analysis and Bilinear Time Series Models

Springer-Verlag
New York Berlin Heidelberg Tokyo 1984

Authors

Dr. T. Subba Rao
Department of Mathematics, University of Manchester
P.O. Box 88, Manchester M60 1QD, England

Dr. M.M. Gabr
Department of Mathematics, University of Alexandria
Alexandria, Egypt

AMS Subject Classification (1980): 62M15, 62L10

ISBN 3-540-96039-2 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 0-387-96039-2 Springer-Verlag New York Heidelberg Berlin Tokyo

Library of Congress Cataloging in Publication Data. Subba Rao, T. An introduction to bispectral analysis and bilinear time series models. (Lecture notes in statistics; v. 24) Bibliography: p. Includes indexes. 1. Time-series analysis. 2. Spectral theory (Mathematics) I. Gabr, M.M. II. Title. III. Lecture notes in statistics (Springer-Verlag); v. 24.
QA280.S83 1984 519.5'5 84-5501

Printing and binding: Beltz Offsetdruck, Hemsbach / Bergstr.
2146 / 3140-543210

DEDICATION

Dedicated to the memory of

Sri T. Someswara Rao,

father of T. Subba Rao.

ACKNOWLEDGEMENTS

The authors wish to thank the following publishers for giving permission to include the material published in their journals:

(1) North Holland Publishing Company, Amsterdam, Holland (Handbook of Statistics, Vol. 3)

(2) Royal Statistical Society (J. Roy. Statist. Soc. B, Vol. 43., 244-255)

(3) Tieto Limited (Journal of Time Series Analysis, 1981, Vol. 2, 153-171; Journal of Time Series Analysis, 1, 2, 145-158; Journal of Time Series Analysis, 4, 2, 95-110).

PREFACE

The theory of time series models has been well developed over the last thirty years. Both the frequency domain and time domain approaches have been widely used in the analysis of linear time series models. However, many physical phenomena cannot be adequately represented by linear models; hence the necessity of nonlinear models and higher order spectra. Recently a number of nonlinear models have been proposed.

In this monograph we restrict attention to one particular nonlinear model, known as the "bilinear model". The most interesting feature of such a model is that its second order covariance analysis is very similar to that for a linear model. This demonstrates the importance of higher order covariance analysis for nonlinear models. For bilinear models it is also possible to obtain analytic expressions for covariances, spectra, etc. which are often difficult to obtain for other proposed nonlinear models. Estimation of bispectrum and its use in the construction of tests for linearity and symmetry are also discussed. All the methods are illustrated with simulated and real data.

The first author would like to acknowledge the benefit he received in the preparation of this monograph from delivering a series of lectures on the topic of bilinear models at the University of Bielefeld, Ecole Normale Supérieure, University of Paris (South) and the Mathematisch Centrum, Amsterdam.

We are grateful to Professor M. B. Priestley for going through the manuscript and making many helpful comments. Parts of this mongraph have been read by Professor A. M. Walker, and we are also thankful to him for many suggestions. We also wish to express out sincere thanks to

Mrs. C. Rowson and Ms. V. Skwarczuk for typing the manuscript. Lastly, we are thankful to our wives Subhadra and Samia for their constant encouragement.

T. SUBBA RAO

30th January, 1984

M. M. GABR

CONTENTS

CHAPTER ONE

INTRODUCTION TO STATIONARY TIME SERIES AND SPECTRAL ANALYSIS

In this chapter we will present a brief introduction to the theory of stationary time series and spectral analysis which will be needed in later chapters. The presentation is kept very brief (for details see Priestley, 1981).

1.1 Some Basic Definitions

A stochastic (or random) process is defined as a family of random variables $\{X_t,\ t \in T\}$. The variables are indexed by the parameter t which belongs to the set T, the index set, or the parameter set. If $T = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$, then $\{X_t\}$ is said to be a discrete parameter process. If $T = \{t, -\infty < t < \infty\}$, then $\{X_t\}$ is said to be a continuous parameter process. In this book we consider only real valued discrete parameter processes. In most situations t is interpreted as a time variable, and $\{X_t\}$ is then called a time series.

The probabilistic structure of the stochastic process $\{X_t\}$ is completely specified, if for any positive integer n, and any admissable set $t_1, t_2, \ldots, t_n$, the probability distribution functions of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_n}\}$,

$$F_{t_1,\ldots,t_n}(x_1,\ldots,x_n) = \Pr(X_{t_1} \le x_1,\ldots,X_{t_n} \le x_n) \qquad (1.1.1)$$

are known.

A stochastic process $\{X_t\}$ is said to be completely (strictly) stationary if all the finite-dimensional distributions (1.1.1) defining $\{X_t\}$ remain the same if the whole set of points $\{t_1,\ldots,t_n\}$ is shifted along the t-axis, i.e. if for any admissable set $t_1,\ldots,t_n$, and all τ, the relation

$$F_{t_1+\tau,\ldots,t_n+\tau}(x_1,\ldots,x_n) = F_{t_1,\ldots,t_n}(x_1,\ldots,x_n)$$

holds.

A less restrictive requirement is called stationarity to order m. The process $\{X_t\}$ is said to be stationary up to order m, if for any admissable $t_1,\ldots,t_n$, and any τ, all the joint moments up to order m of $\{X_{t_1},\ldots,X_{t_n}\}$ exist and are equal to the corresponding joint moments up to order m of $\{X_{t_1+\tau},\ldots,X_{t_n+\tau}\}$. That is,

$$E\left\{(X_{t_1})^{K_1}(X_{t_2})^{K_2}\ldots(X_{t_n})^{K_n}\right\} = E\left\{(X_{t_1+\tau})^{K_1}(X_{t_2+\tau})^{K_2}\ldots(X_{t_n+\tau})^{K_n}\right\}$$

for all τ, and all possible non-negative integers $K_1,\ldots,K_n$ satisfying $K_1 + \ldots + K_n \le m$.

Thus, we say that the process $\{X_t\}$ is stationary to order one if

$$E[X_t] = \mu, \text{ independent of } t,$$

and is stationary to order two ("weakly stationary", or "stationary in the wide sense") if

i) $E(X_t) = \mu$, independent of t;

ii) $\mathrm{var}(X_t) = E(X_t-\mu)^2 = \sigma_X^2$, independent of t;

iii) $\mathrm{cov}(X_t,X_{t+s}) = E[(X_t-\mu)(X_{t+s}-\mu)] = R(s)$, say, $(s=0,\pm1,\pm2,\ldots)$ is a function of s only (1.1.2)

R(s) is known as the autocovariance function of lag s, and

$$\rho(s) = R(s)/R(0) \tag{1.1.3}$$

is known as the autocorrelation function of lag s. All autocovariance functions R(s) of real-valued processes possess the following properties:

$$\left.\begin{array}{lll} 1. & R(0) = \mathrm{var}(X_t) = \sigma_X^2 & \Leftrightarrow \rho(0) = 1 \\ 2. & |R(s)| \le R(0) & \Leftrightarrow |\rho(s)| \le 1, \text{ all } s \\ 3. & R(-s) = R(s) & \Leftrightarrow \rho(-s) = \rho(s) \end{array}\right\} \tag{1.1.4}$$

4. R(s) is positive semi-definite in the sense that for any set of time points $t_1, t_2,\ldots,t_n$ and all real $a_1, a_2,\ldots,a_n$,

$$\sum_{i=1}^{n}\sum_{j=1}^{n} R(t_i-t_j)\, a_i\, a_j \ge 0.$$

The stochastic process $\{X_t\}$ is a Gaussian (or normal) process if for any admissable set $t_1,\ldots,t_n$ the joint distribution of $\{X_{t_1},\ldots,X_{t_n}\}$ is an n-variate normal distribution. Since a multivariate normal distribution is completely specified by its means, variances and covariances it follows that if a Gaussian process is stationary to order two, it is completely stationary. Unless otherwise stated, throughout this book a stationary process means a second order stationary process. In most cases this will be clear from the context.

1.2 Spectral Densities and Spectral Representations

According to Wold's theorem (Wold, 1938), a necessary and sufficient condition for the sequence $\{R(s), s = 0, \pm1, \pm2, \ldots,\}$ to be the autocovariance function for some discrete parameter process $\{X_t\}$, is that there exists a function $F(\omega)$, which is bounded and non-decreasing on the interval $(-\pi,\pi)$ such that

$$R(s) = \int_{-\pi}^{\pi} e^{is\omega}\, dF(\omega) \tag{1.2.1}$$

The function $F(\omega)$ is known as the integrated (non-normalized) spectrum of the stationary process $\{X_t\}$. If $F(\omega)$ is differentiable, i.e. if we may write

$$dF(\omega) = f(\omega)\, d\omega$$

then we can write (1.2.1) as

$$R(s) = \int_{-\pi}^{\pi} e^{is\omega} f(\omega)\, d\omega \tag{1.2.2}$$

The function $f(\omega)$ is known as the (non-normalized) spectral density function (or, more simply, spectrum) of the stationary process $\{X_t\}$. From (1.2.2) we have the inverse Fourier transform

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s)\, e^{-is\omega}, -\pi \le \omega \le \pi \qquad (1.2.3)$$

The normalised spectral density function, $g(\omega)$, is then defined by

$$g(\omega) = f(\omega)/\sigma_X^2 = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \rho(s)\, e^{-is\omega}, \qquad (1.2.4)$$

The spectral density function $f(\omega)$ exists for all ω if $R(s)$ is absolutely summable, i.e. if

$$\sum_{s=-\infty}^{\infty} |R(s)| < \infty \qquad (1.2.5)$$

Hence a sufficient condition for the spectral density function to exist is that $R(s)$ damps out rapidly. The function $f(\omega)$ has the following properties:

i) $\int_{-\pi}^{\pi} f(\omega)\, d\omega = R(0) = \sigma_X^2$ (1.2.6)
(this follows by putting $s = 0$ in (1.2.2))

ii) $f(\omega) \ge 0$

iii) $f(-\omega) = f(\omega)$, all ω (since $R(-s) = R(s)$).

In an engineering context, σ_X^2 represents the total power in the process $\{X_t\}$, and the left-hand side integral of (1.2.6) represents a frequency decomposition of the total power. In other words, $f(\omega)\, d\omega$ represents the power contained in the frequency band $(\omega, \omega+d\omega)$ of the process $\{X_t\}$. Since $f(\omega)$ is the Fourier transform of the autocovariance function $R(s)$, it is natural to define $f(\omega)$ as the second order spectral density function. So, knowledge of the autocovariance function is equivalent to knowledge of the spectrum, and vice versa. If the process $\{X_t\}$ is Gaussian, all the information in the process $\{X_t\}$ is contained in the mean and the covariances

$\{R(s)\}$, and as such a second order spectral analysis on the process $\{X_t\}$ is sufficient to obtain all useful information about the process. As we shall show later, if the process $\{X_t\}$ is not Gaussian, one may have to consider a higher order spectral analysis of the process.

Sometimes it is convenient to make use of spectral representations for stationary processes to obtain spectral density functions. Let $\{X_t\}$ be a zero-mean second order stationary process; then there exists an orthogonal complex-valued process $\{dZ(\omega)\}$ on the interval $(-\pi,\pi)$ such that for all integers t,

$$X_t = \int_{-\pi}^{\pi} e^{it\omega}\, dZ(\omega) \qquad (1.2.7)$$

Equation (1.2.7) is known as the spectral representation of $\{X_t\}$. The process $\{dZ(\omega)\}$ has the following properties:

$$\left.\begin{array}{ll} \text{i)} & E[dZ(\omega)] = 0, \text{ all } \omega \\ \text{ii)} & E[|dZ(\omega)|^2] = dF(\omega) = f(\omega)\, d\omega \quad -\pi \le \omega \le \pi \\ \text{iii)} & E[dZ^*(\omega_1)\, dZ(\omega_2)] = 0, \text{ for all } \omega_1 \ne \omega_2 \end{array}\right\} \qquad (1.2.8)$$

where Z^* is the complex conjugate of Z. It will be shown later in this chapter how, using (1.2.7), one can evaluate the spectral density function of a process $\{X_t\}$ when that process $\{X_t\}$ satisfies a difference equation of finite order.

1.3 Higher Order Spectra (Polyspectra)

The usefulness of higher order spectra has been pointed out by Brillinger (1965), (1975), Brillinger and Rosenblatt (1967a, 1967b). We follow Brillinger (1965) in the definition of polyspectra.

Let $\{X_t\}$ be a real valued random process with $E(X_t^k) < \infty$ and let $m(t_1,t_2,\ldots,t_k)$ denote the k-th order product moment $E(X_{t_1} X_{t_2} \ldots, X_{t_k})$. If the process $\{X_t\}$ is stationary up to order k the product moment can

be written in the form $m(0, t_2-t_1, t_3-t_1,\ldots,t_k-t_1)$ where the arguments are the lag differences.

For the existence of higher order spectra, one needs to assume the existence of higher order moments, and as Brillinger (1975, p.23) points out, it need not cause concern, for in practice all series available for analysis appear to be bounded, i.e. $|X_t| < C$, for some finite C.

Let us now define the moment generating function

$$M(\theta_1,\theta_2,\ldots,\theta_k) = E\{\exp(\theta_1 X_{t_1}+\theta_2 X_{t_2}+\ldots+\theta_k X_{t_k}),$$

so that the product moment $m(t_1,t_2\ldots,t_k)$ is the coefficient of $(\theta_1\ \theta_2\ldots\theta_k)$ in the Taylor series expansion of the generating function $M(\theta_1,\theta_2,\ldots,\theta_k)$. Let $C(t_1,t_2,\ldots,t_k)$ denote the k-th order joint cumulant (or semi-invariant) of the set of random variables $\{X_{t_1}, X_{t_2},\ldots,X_{t_k}\}$. Then $C(t_1,t_2,\ldots,t_k)$ is the coefficient of $(\theta_1\theta_2\ldots\theta_k)$ in the expansion of the cumulant generating function $K(\theta_1,\theta_2,\ldots,\theta_k) = \log\{M(\theta_1,\theta_2,\ldots,\theta_k)\}$. The cumulant $C(t_1,t_2,\ldots,t_k)$ is a k-th order polynomial in the moments of order no higher than k. Conversely the k-th order moment $m(t_1,t_2,\ldots,t_k)$ is a k-th order polynomial in the cumulants of order no higher than k. The relations between moments and cumulants have been derived by Leonov and Shriyaev (1959) and these relations are as follows:

$$E(X_{t_1} X_{t_2}\ldots X_{t_k}) = \sum_{\nu} C(\nu_1)\ C(\nu_2)\ldots C(\nu_p) \qquad (1.3.1)$$

where the summation extends over all partitions $\nu = (\nu_1,\nu_2,\ldots,\nu_p)$ of the integers $(t_1, t_2,\ldots,t_k)$; and $C(\nu_1), C(\nu_2),\ldots,C(\nu_p)$ are the cumulants of the X's. For example, we obtain from (1.3.1), $E(X_{t_1}) = C(t_1)$, $E(X_{t_1}X_{t_2}) = C(t_1)\ C(t_2) + C(t_1,t_2)$ implying that the mean is the same as the first cumulant and the second order covariance between X_{t_1} and X_{t_2} is the same as the second order cumulant. Further, if we assume $E(X_{t_i}) = 0$ and $E(X_{t_1} X_{t_2} X_{t_3}) = 0$, we can obtain from (1.3.1) the well-known

$E(X_{t_1} X_{t_2} X_{t_3}) = C(X_{t_1})\ C(X_{t_2})\ C(X_{t_3}) + C(X_{t_1},X_{t_2},X_{t_3}) \Rightarrow$

$C(X_{t_1},t_2,X_{t_3}) = 0$

relation

$$E(X_{t_1} X_{t_2} X_{t_3} X_{t_4}) = C(t_1,t_2)\ C(t_3,t_4) + C(t_1,t_3)\ C(t_2,t_4) + C(t_1,t_4)\ C(t_2,t_3) + C(t_1,t_2,t_3,t_4)$$

and so on. Thus the moments can be written in terms of the cumulants. Equivalently we can invert the relation (1.3.1) to obtain $C(X_{t_1} X_{t_2}\ldots X_{t_k}) = \sum_{\nu} (-1)^{p-1}(p-1)!\ \mu_{\nu_1}\mu_{\nu_2}\ldots\mu_{\nu_p}$ where $\mu_{\nu_1}, \mu_{\nu_2},\ldots,\mu_{\nu_p}$ are the means of the products of X's.

Some important properties of the cumulants are as follows (see Rosenblatt, 1981):

i) $C(a_1X_{t_1}, a_2X_{t_2},\ldots a_kX_{t_k}) = a_1 a_2\ldots a_k\ C(X_{t_1},X_{t_2},\ldots,X_{t_k})$ where $(a_1,a_2,\ldots,a_k)$ are constants;

ii) If $X_{t_1},X_{t_2},\ldots,X_{t_k}$ can be partitioned into two disjoint sets which are independent of each other, then

$C(X_{t_1},X_{t_2},\ldots,X_{t_k}) = 0$;

iii) The cumulant is a symmetric and multilinear function of its arguments;

iv) If $X_{t_1}, X_{t_2},\ldots,X_{t_k}$ and $Y_{t_1}, Y_{t_2},\ldots,Y_{t_k}$ are independent

$$C(X_{t_1}+Y_{t_1}, X_{t_2}+Y_{t_2},\ldots,X_{t_k}+Y_{t_k}) = C(X_{t_1},X_{t_2},\ldots,X_{t_k}) + C(Y_{t_1},Y_{t_2},\ldots,Y_{t_k}).$$

If the k-th moment $m(t_1,t_2,\ldots,t_k)$ is absolutely summable (i.e. in L_1), then we have the representation

$$m(t_1,t_2,\ldots,t_k) = \int_{-\pi}^{\pi}\ldots\int_{-\pi}^{\pi} \exp\{i\sum_{j=1}^{k}\omega_j t_j\}\ E[\prod_{j=1}^{k} dZ(\omega_j)] \qquad (1.3.2)$$

where $E\prod_{j=1}^{k} dZ(\omega_j) = n(\omega_1+\omega_2+\ldots+\omega_k)\ dG(\omega_1,\omega_2,\ldots,\omega_k)$ and G is of bounded variation with dG zero unless $\sum_{j=1}^{k}\omega_j = 0, \pm 2\pi,\ldots$. Here $n(\lambda) = \sum_{-\infty}^{\infty}\delta(\lambda+2j\pi)$, where $\delta(\lambda)$ is the Dirac Delta function. From (1.3.2) we

see that the product moment $m(t_1,t_2,\ldots,t_k)$ can be written in terms of the moments of the product of the orthogonal process $\{dZ(\omega)\}$. Similar representations can be written for the cumulants as follows. The representation (1.3.2) leads to

$$C(t_1,t_2,\ldots,t_k) = \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi} \exp\{i\Sigma\, t_j\omega_j\}\, \mathrm{cum}(dZ(\omega_1)\, dZ(\omega_2)\ldots dZ(\omega_k)) \tag{1.3.3}$$

where

$$\mathrm{cum}(dZ(\omega_1)\, dZ(\omega_2)\ldots,dZ(\omega_j) = n(\omega_1+\omega_2+\ldots+\omega_k)dF(\omega_1,\omega_2,\ldots,\omega_k),$$

F is a function of bounded variation with dF zero unless $\sum_{j=1}^{k}\omega_j = 0$, $\pm 2\pi, \pm 4\pi, \ldots$.

If the process is k-th order stationary, we have

$$m(t_1,t_2,\ldots,t_k) = m(0,t_2-t_1,\ldots,t_k-t_1)$$

$$C(t_1,t_2,\ldots,t_k) = C(0,t_2-t_1,\ldots,t_k-t_1).$$

(From now onwards we omit 0 in this argument).

In the case of k-th order stationary process, (1.3.2) can be written as

$$m(\tau_1,\tau_2,\ldots,\tau_{k-1}) = \int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi} \exp\{i\sum_{j=1}^{k-1}\tau_j\lambda_j\}\, dG(\lambda_1,\lambda_2,\ldots,\lambda_{k-1}) \tag{1.3.4}$$

where G is a function of bounded variation. Assuming that the cumulants of $\{X_t\}$ satisfy the condition

$$\sum_{\tau_1,\tau_2,\ldots,\tau_{k-1}=-\infty}^{\infty} |\, C(\tau_1,\tau_2,\ldots,\tau_{k-1}) < \infty \tag{1.3.5}$$

the k-th order cumulant spectrum can be defined as

$$f(\omega_1,\omega_2,\ldots,\omega_{k-1}) = \frac{1}{(2\pi)^{k-1}} \sum_{\tau_1=-\infty}^{\infty}\cdots\sum_{\tau_{k-1}}^{\infty} C(\tau_1,\tau_2,\ldots,\tau_{k-1})$$

$$\exp\{-i(\omega_1\tau_1+\ldots+\omega_{k-1}\tau_{k-1})\} \tag{1.3.6}$$

In view of (1.3.5), the function $F(\omega_1,\omega_2,\ldots,\omega_k)$ defined in (1.3.3) is

differentiable on the manifold $\sum_{j=1}^{k}\omega_j = 0 \pm 2\pi,\ldots$, and hence

$$dF(\omega_1,\omega_2,\ldots,\omega_k) = f(\omega_1,\omega_2,\ldots,\omega_k)\, d\omega_1\, d\omega_2\ldots d\omega_k \tag{1.3.7}$$

In writing (1.3.7) we have written the function f as a function of k variables $\omega_1,\omega_2,\ldots,\omega_k$, though in view of the condition $\sum_{j=1}^{k}\omega_j = 0$, it can be written as a function of only (k-1) variables, as done in (1.3.6).

In studying the properties of estimates of the function $f(\omega_1,\omega_2,\ldots,\omega_{k-1})$ given by (1.3.6), Brillinger and Rosenblatt (1967a) make use of the following condition on the cumulants

$$\sum_{\tau_1,\tau_2,\ldots,\tau_{k-1}=-\infty}^{\infty} |\tau_j C(\tau_1,\tau_2,\ldots,\tau_{k-1})\,| < \infty \tag{1.3.8}$$

(j=1,2,...,k-1) which is stronger than (1.3.5). This condition ensures faster decay to zero of the cumulants than (1.3.5).

When k=2, the cumulant spectrum reduces to the second order spectral density function (1.2.3), since $C(\tau_1) = R(\tau_1)$ for all τ_1.

Brillinger and Rosenblatt (1967b) have defined the k-th order normalized polyspectrum (or standardized k-th order spectrum $k \geq 3$) as

$$g(\omega_1,\ldots,\omega_{k-1}) = \frac{f(\omega_1,\ldots,\omega_{k-1})}{\{f(\omega_1)\ldots f(\omega_{k-1}).f(\omega_1+\ldots+\omega_{k-1})\}^{\frac{1}{2}}} \tag{1.3.9}$$

where $f(\omega)$ is the second order spectral density function. It is well known that all joint cumulants of higher order than the second vanish for the multivariate normal distribution. This leads to the fact that if the process $\{X_t\}$ is Gaussian, all the polyspectra of higher order than the second vanish. A process $\{X_t\}$ which is non-linear is always non-Gaussian; hence higher order spectra can be used to study departures from linearity.

Time Reversibility The higher order cumulant spectra can be used not only for studying departure from linearity (and Gaussianity), but can be used also for studying "time reversibility" of the process. (See Brillinger and Rosenblatt, 1967a, p.210). A process X_t is said to be time-reversible if the probability structure of X_{-t} is the same as that of X_t. In this case we have $C(\tau_1,\tau_2,\ldots,\tau_{k-1}) = C(-\tau_1,-\tau_2,\ldots,-\tau_k)$, and therefore the imaginary part of $f(\omega_1,\omega_2,\ldots,\omega_{k-1})$ is identically zero. Therefore, if the process is determined by its moments, and if the imaginary part of all the cumulant spectra are zero, the process is time reversible.

The simplest higher order spectra one can consider as a part of non-linear analysis are the bispectra, i.e., the Fourier transforms of the third order moment. In view of their importance, we will consider in the following section the properties of bispectra in detail.

1.4 Bispectral Density Functions

Let $\{X_t\}$ be a real valued third order stationary process; then the third order moment.

$$E[X_t\, X_{t+t_1}\, X_{t+t_2}] = E[X_0\, X_{t_1}\, X_{t_2}] = m(t_1,t_2) \tag{1.4.1}$$

depends only on t_1,t_2 for all admissable integers t, t_1 and t_2. The third order central moment $C(t_1,t_2)$ is defined as

$$C(t_1,t_2) = E[(X_t-\mu)\,(X_{t+t_1}-\mu)\,(X_{t+t_2}-\mu)] \tag{1.4.2}$$

Since the process is real valued, the following symmetry relations hold;

$$C(t_1,t_2) = C(t_2,t_1) = C(-t_1,t_2-t_1) = C(t_1-t_2,-t_2) \tag{1.4.3}$$

The relations (1.4.3) imply that $C(t_1,t_2)$ is completely specified over the entire plane by its values in any one of the six sectors shown in Figure 1.1 and Figure 1.2. The bispectral density function, $f(\omega_1,\omega_2)$, is now defined as

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} C(t_1,t_2)\, e^{-it_1\omega_1 - it_2\omega_2} \quad \pi \le \omega_1,\omega_2 \le \pi \tag{1.4.4}$$

so that by inversion,

$$C(t_1,t_2) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} e^{it_1\omega_1+it_2\omega_2} f(\omega_1,\omega_2)\, d\omega_1\, d\omega_2 \tag{1.4.5}$$

(Not that $f(\omega_1,\omega_2)$ is a complex valued function). The bispectral density function exists for all ω_1,ω_2 if

$$\sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} |C(t_1,t_2)| < \infty \tag{1.4.6}$$

By putting $t_1 = t_2 = 0$ in (1.4.5) we get

$$C(0,0) = E(X_t-\mu)^3 = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(\omega_1,\omega_2)\, d\omega_1\, d\omega_2 \tag{1.4.7}$$

A comparison of (1.4.7) with (1.2.3) shows that the right-hand side of (1.4.7) is a frequency decomposition of the third order central moments. If the process $\{X_t\}$ is Gaussian, then $C(t_1,t_2) = 0$, all t_1,t_2. Hence, $f(\omega_1,\omega_2)$ can be considered as a measure of departure from Gaussianity of the process $\{X_t\}$. This will be further considered later. Since $C(t_1,t_2)$ is symmetric and satisfies the relations (1.4.3), we have the following symmetry relations for $f(\omega_1,\omega_2)$,

$$f(\omega_1,\omega_2) = f(\omega_2,\omega_1) = f(\omega_1,-\omega_1-\omega_2) = f(-\omega_1-\omega_2,\omega_2) = f^*(-\omega_1,-\omega_2) \tag{1.4.8}$$

In view of these relations, the values of $f(\omega_1,\omega_2)$ over the entire plane are completely specified by the values defined over any one of the twelve sectors shown in Figure 1.3.

1.5 Standard Linear Models - their spectra and bispectra

We now consider the evaluation of the spectra and bispectra of the process $\{X_t\}$ when the process satisfies some standard linear time series models.

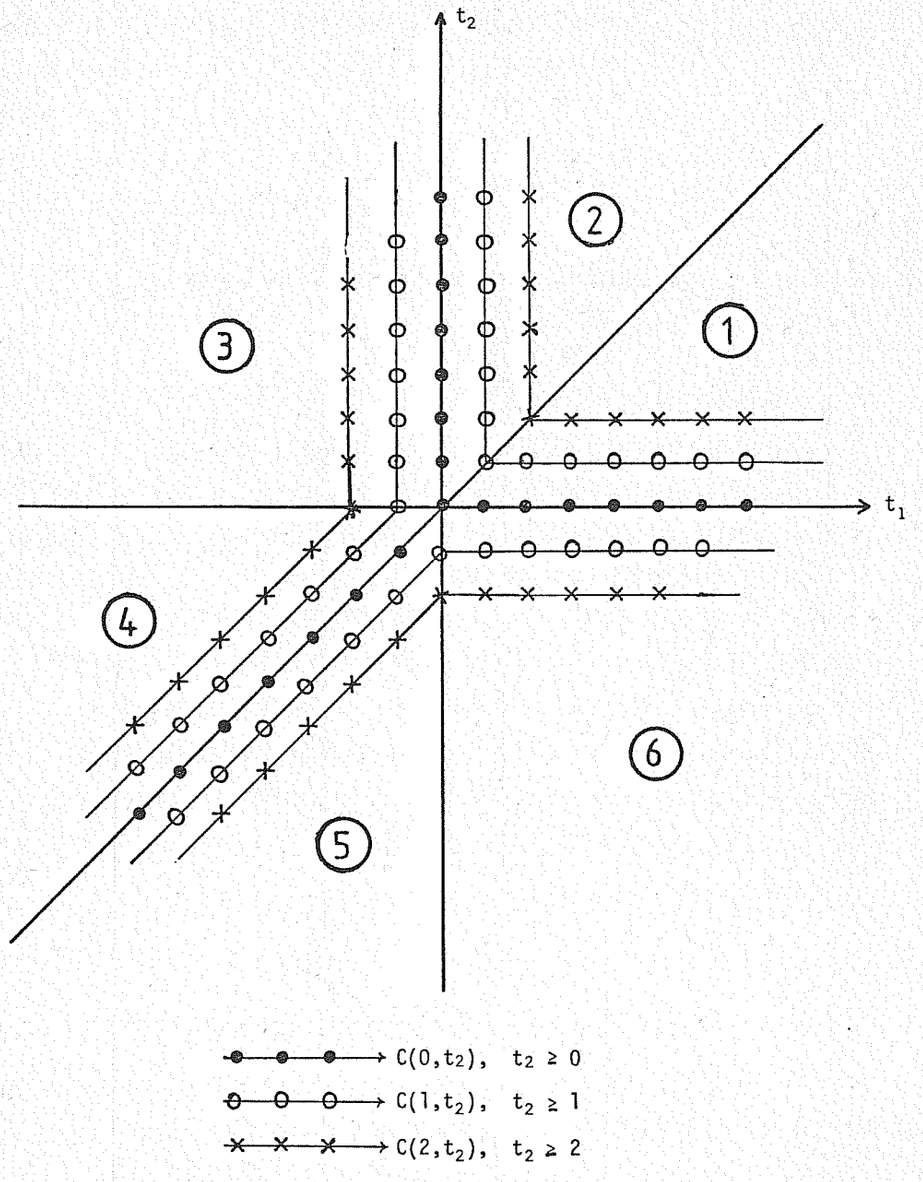


Figure 1.1

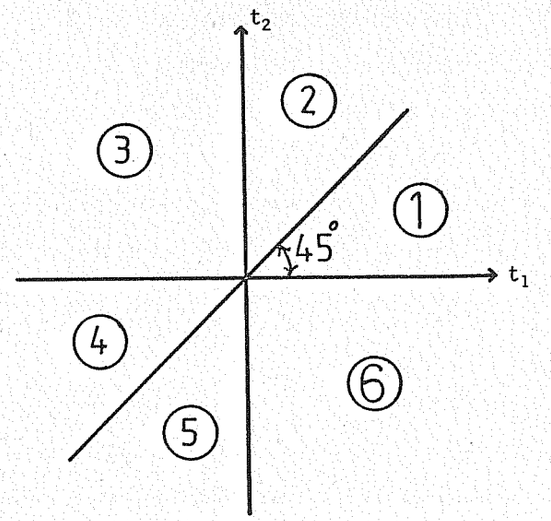


Figure 1.2

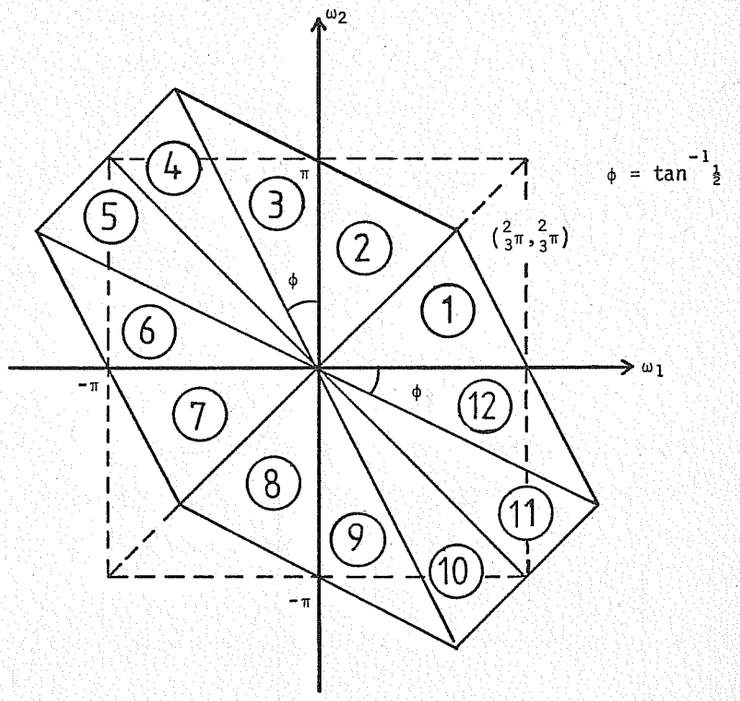


Figure 1.3

The simplest type of time series which may arise is that generated by a "purely random process", and we denote this special type of series by $\{e_t\}$. Such a series consists of a sequence of uncorrelated random variables, and, if it is stationary in the wide sense (or second order stationary), its mean μ_e and variance σ_e^2 are the same at all time points. The covariances at different time points are, of course, zero and we may therefore write

$$E(e_t) = \mu_e, \text{ all } t$$

$$R_e(s) = \text{cov}(e_t, e_{t+s}) = \begin{cases} \sigma_e^2 , & s = 0 \\ 0 , & s \neq 0 \end{cases}$$

By virtue of (1.2.3), the spectral density function of the stationary purely random process $\{e_t\}$ is given by

$$f_e(\omega) = \frac{\sigma_e^2}{2\pi} = \text{a constant, all } \omega.$$

From now on, when we use the term "purely random processes", we mean a stationary purely random process. The purely random process is often called "white noise", particularly in the engineering literature.

If the variables $\{e_t\}$ are just mutually uncorrelated, the spectral density is constant, whereas the bispectral density $f(\omega_1, \omega_2)$ is not. However, if the variables $\{e_t\}$ are strictly independent, it can easily be shown that the bispectral density of the process $\{e_t\}$ is given by

$$f(\omega_1, \omega_2) = \frac{\mu_3}{(2\pi)^2} \quad , \text{ where } \mu_3 = E[(e_t - \mu_e)^3].$$

In other words, the bispectrum is constant when the process is independent. This important result suggests that higher order moment analysis (or spectral analysis) may be used to distinguish between independent random "errors" and uncorrelated random "errors".

(i) Let us now consider a time series satisfying an autoregressive model.

The process $\{X_t\}$ is said to be an autoregressive process of order p (denoted by AR(p)) if it satisfies the difference equation

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} = e_t \qquad (1.5.1)$$

where $a_1, \ldots, a_p$ are constants and $\{e_t\}$ is a purely random process. Defining the shift operator B by $B^k X_t = X_{t-k}$ equation (1.5.1) may be rewritten as $\Phi(B) X_t = e_t$ where

$$\Phi(z) = 1 + a_1 z + \ldots + a_p z^p \qquad (1.5.2)$$

It can be shown that if all the roots of $\Phi(z)$ lie outside the unit circle, then the process $\{X_t\}$ is asymptotically stationary (see, for example, Box and Jenkins, 1970). Under this condition, an important recurrence relation for the autocorrelation function can be found. Without loss of any generality, we may assume $E(e_t) = 0 \Rightarrow E(X_t) = 0$. Multiplying both sides of (1.5.1) by X_{t-r}, taking the expectations and dividing by σ_X^2, we get

$$\rho(r) + a_1 \rho(r-1) + \ldots + a_p \rho(r-p) = 0, \quad r = 1, 2, \ldots, p \qquad (1.5.3)$$

This set of equations is called the Yule-Walker equations. We can now obtain the second order spectral density function of X_t (when X_t satisfies the relation (1.5.1)) as follows.

Substituting the spectral representations for X_t and e_t from (1.2.7) in (1.5.1), we obtain

$$[1 + a_1 e^{-i\omega} + a_2 e^{-2i\omega} + \ldots + a_p e^{-pi\omega}]\, dZ_X(\omega) = dZ_e(\omega) \qquad (1.5.4)$$

where $dZ_X(\omega)$ is the orthogonal process corresponding to $\{X_t\}$, and $\{dZ_e(\omega)\}$ is the orthogonal process corresponding to $\{e_t\}$.

From (1.5.4) we obtain, by taking the expectations of the square of the modulus both sides,

$$f(\omega) = \frac{\sigma_e^2}{2\pi} \frac{1}{|1+a_1 e^{-i\omega}+\ldots+a_p e^{-ip\omega}|^2}$$

$$= \frac{\sigma_e^2}{2\pi} |\phi^{-1}(e^{-i\omega})|^2 \qquad (1.5.5)$$

where $\phi(e^{-i\omega}) = 1 + a_1 e^{-i\omega}+\ldots+a_p e^{-ip\omega}$.

(ii) Now consider a moving average process. The process $\{X_t\}$ is said to be a moving average process of order q (denoted by MA(q)) if

$$X_t = b_0 e_t + b_1 e_{t-1} +\ldots + b_q e_{t-q} \qquad (1.5.6)$$

where $b_0, b_1, \ldots, b_q$ are constants. (The e_t are usually scaled so that $b_0 = 1$.) From the relation (1.5.6) we find that

$$E[X_t] = 0,$$

$$\text{var}(X_t) = \sigma_e^2 \left(\sum_{i=0}^{q} b_i^2 \right)$$

$$R(s) = E[X_t X_{t+s}] = \begin{cases} 0 & , s > q \\ \sigma_e^2 \sum_{i=0}^{q-s} b_i b_{i+s}, & s = 0,1,\ldots,q \end{cases}$$

As R(s) does not depend on t, and the mean is constant, the process is stationary irrespective of the values of $\{b_i\}$. However, the process is <u>invertible</u> only under certain restrictions on the parameters $\{b_i\}$. The imposition of the invertibility condition ensures that there is a unique MA process for a given autocorrelation function. Rewriting (1.5.6) using the shift operator B we have, $X_t = \Psi(B)\, e_t$ where $\Psi(B) = b_0 + b_1 z +\ldots+ b_q z^q$.

Then an MA(q) process is invertible if all the roots of $\Psi(z)$ lie outside the unit circle. In this case $\Psi^{-1}(B)$ exists, and we can write $e_t = \Psi^{-1}(B)\, X_t$ (see Box and Jenkins, 1970).

Proceeding as in the case of the AR model, we can show that the spectral density function of MA(q) process is given by

$$f(\omega) = \frac{\sigma_e^2}{2\pi} |b_0 + b_1 e^{-i\omega} +\ldots+ b_q e^{-iq\omega}|^2 = \frac{\sigma_e^2}{2\pi} |\Psi(e^{-i\omega})|^2$$

where $\Psi(e^{-i\omega}) = b_0 + b_1 e^{-i\omega}+\ldots+b_q e^{-iq\omega}$

In general, an MA process of finite order can be expressed as an AR process of infinite order, while an AR process of finite order can be expressed as an MA process of infinite order.

iii) <u>Mixed Autoregressive Moving Average Process</u>

A useful class of models for time series is that formed from a combination of AR and MA processes. A mixed autoregressive moving average process of order (p,q) (denoted by ARMA (p,q)) is given by

$$X_t + a_1 X_{t-1} +\ldots+ a_p X_{t-p} = b_0 e_t + b_1 e_{t-1} +\ldots+ b_q e_{t-q} \qquad (1.5.7)$$

where again $\{e_t\}$ is a purely random process with zero mean and $\{a_1,\ldots,a_p, b_0, b_1,\ldots,b_q\}$ are constants. (We can, without loss of generality, take $b_0 = 1$). Using the same operator notation given before, (1.5.7) may be rewritten as $\Phi(B)\, X_t = \Psi(B)\, e_t$. Again, the process is asymptotically stationary if all the roots of $\Phi(z)$ lie outside the unit circle and the model is invertible if all the roots of $\Psi(z)$ lie outside the unit circle.

By multiplying both sides of (1.5.7) by X_{t-r} and taking expectations both sides, we obtain the recursive equations for the autocorrelations, namely,

$$\rho(r) + a_1\rho(r-1) +\ldots+ a_p\rho(r-p) = 0 \qquad \text{if } r > q \qquad (1.5.8)$$

These equations are known as the Yule-Walker equations for the ARMA (p,q) process. It should be noted that the AR, MA and ARMA models are particular cases of the general linear process considered below.

iv) <u>The General Linear Process</u>

The general linear process may be regarded as an MA process with infinite order, and is given by

$$X_t = \sum_{u=0}^{\infty} h_u e_{t-u} = H(B)\, e_t \text{ , say} \qquad (1.5.9)$$

where $\{e_t\}$ is a purely random process, and $\{h_u\}$ is a given sequence of constants satisfying

$$\sum_{u=0}^{\infty} h_u^2 < \infty, \quad \text{and} \quad H(z) = \sum_{u=0}^{\infty} h_u z^u . \qquad (1.5.10)$$

From (1.5.9) it follows that

$$\text{var}(X_t) = \sigma_X^2 = \sigma_e^2 \left(\sum_{u=0}^{\infty} h_u^2 \right)$$

so condition (5.1.10) is clearly necessary in order that σ_X^2 be finite. The spectral density function of the general linear process (1.5.9) is given by

$$f(\omega) = \frac{\sigma_e^2}{2\pi} |H(e^{-i\omega})|^2 = \frac{\sigma_e^2}{2\pi} |\Gamma(\omega)|^2 \qquad (1.5.11)$$

where

$$\Gamma(\omega) = \sum_{u=0}^{\infty} h_u e^{-i\omega u} \qquad (1.5.12)$$

and is sometimes known as the "transfer function". In fact both MA and AR processes are special cases of the general linear process. To express X_t in the form (1.5.9), we have to represent the function $f(\omega)$ in the form (1.5.11). The method of representing the function $f(\omega)$ is given by the well known theorem on "Spectral Factorization" (Doob, 1953, p.160), which may be stated as follows:

If the spectral density function $f(\omega)$ is absolutely continuous and satisfies the Paley-Wiener condition

$$\int_{-\pi}^{\pi} \log f(\omega)\, d\omega > -\infty$$

then there exists a uniquely determined sequence $h_0, h_1, \ldots$, with h_0 real and positive, such that

(i) $\sum_{0}^{\infty} |h_u|^2 < \infty$

(ii) $H(z) = \sum_{0}^{\infty} h_u z^u \neq 0, \quad |z| < 1$

(i.e. the zeros of $H(z)$ lie outside the unit circle)

(iii) $f(\omega) \propto |H(e^{-i\omega})|^2$

This theorem is extremely important in the theory of linear prediction (for more details, see, for example, Priestley, 1981, Chapter 11).

To find the bispectral density function of the process which satisfies (1.5.9), let us consider the linear relation

$$X_t = \sum_{u=0}^{\infty} h_u Y_{t-u} \qquad (1.5.13)$$

where $\{Y_t\}$ and $\{X_t\}$ are assumed to be stationary at least to the third order. By obtaining the third order moments and taking Fourier transforms, we can show that

$$f_X(\omega_1,\omega_2) = H(-\omega_1-\omega_2)\, H(\omega_1)\, H(\omega_2)\, f_Y(\omega_1,\omega_2) \qquad (1.5.14)$$

where $f_X(\omega_1,\omega_2)$ is the bispectral density function of $\{X_t\}$ and $f_Y(\omega_1,\omega_2)$ is the bispectral density function of $\{Y_t\}$.

If the process $\{Y_t\}$ is mutually independent and identically distributed with $E(Y_t) = 0$, $E(Y_t^2) = \sigma_y^2$, $E(Y_t^3) = \mu_{3,y}$ then the relation (1.5.14) reduces to

$$f_X(\omega_1,\omega_2) = \frac{\mu_{3,y}}{(2\pi)^2} H(-\omega_1-\omega_2)\, H(\omega_1)\, H(\omega_2) \qquad (1.5.15)$$

From the relation (1.5.15) one can easily see that if the process $\{Y_t\}$ is Gaussian then the bispectral density $f_X(\omega_1,\omega_2)$ is zero for all ω_1 and ω_2. We make use of this fact to construct statistical tests for testing linearity and symmetry (Gaussianity) of time series in later sections.

We now consider a harmonic process. The process $\{X_t\}$ is called a harmonic process if

$$X_t = \sum_{i=1}^{p} A_i \cos(\omega_i t+\phi_i) \qquad (1.5.16)$$

where p, $\{A_i\}$, $\{\omega_i\}$, (i=1,...,p) are constants, and the $\phi_1,\ldots,\phi_p$ are independent random variables, each having a rectangular distribution on the interval $(-\pi,\pi)$ with probability density function

$$f_\phi(\rho) = \begin{cases} \frac{1}{2\pi} & -\pi \le \rho \le \pi \\ 0 & \text{otherwise} \end{cases}$$

From (1.5.16) one may show that

$$E[X_t] = 0, \quad \text{all } t$$
$$R(s) = E[X_t\, X_{t+s}] = \frac{1}{2}\sum_{i=1}^{p} A_i^2 \cos\omega_i s \qquad (1.5.17)$$

Since R(s) is independent of the time t, the process is stationary. The essential feature to be noted for this process is that the autocovariance function R(s) consists of a sum of cosine terms, and hence never dies out. This feature contrasts sharply with the behaviour of R(s) for the linear models discussed before. In these linear models R(s) may oscillate, but they eventually decay to zero.

In view of the relation (1.5.18), the bispectral density function $f(\omega_1,\omega_2)$ of the process $\{X_t\}$ is zero for all ω_1 and ω_2.

Now we suppose the phases $\{\phi_i\}$ are deterministic, let p = 3, and let the frequencies ω_1,ω_2 and ω_3 satisfy the relation $\omega_1+\omega_2+\omega_3 = 0 \pmod{2\pi}$. Then Brillinger (1980) has shown that the third order periodogram, which is an estimate of the bispectral density function, can be used to estimate the "bi-frequencies" ω_1,ω_2 and ω_3.

1.6 State Space Representation of Linear Time Series Models

It is often convenient to represent the linear time series models

discussed in section 1.5 in the so-called state space form.

Consider, for example, the autoregressive model, AP(p) given by the representation (1.5.1). Let $\underline{x}_t' = (X_t, X_{t-1},\ldots,X_{t-p+1})$,

$$\underline{A} = \begin{bmatrix} -a & -a & \ldots & -a_p \\ 1 & 0 & \ldots & 0 \\ & & \ldots & \\ 0 & 0 & \ldots 1 & 0 \end{bmatrix} \qquad \underline{C}' = (1,0,\ldots,0) \qquad \underline{H}' = (1,0,\ldots,0) \qquad (1.6.1)$$

Then the autoregressive model (1.5.1) can be written in the form

$$\underline{x}_t = \underline{A}\,\underline{x}_{t-1} + \underline{C}\, e_t$$
$$X_t = \underline{H}'\underline{x}_t \qquad (1.6.2)$$

Here, $\underline{x}_t$ is known as the state vector, and $\underline{A}$ is known as the transition matrix. Note that the process $\{\underline{x}_t\}$ is a Markov process even though, in general, the process $\{X_t\}$ is not. The condition for stationarity of the process X_t can now be stated in terms of the eigenvalues of the matrix $\underline{A}$. The process is second order stationary if all the eigenvalues of $\underline{A}$ have modulus less than one. It is, however, possible to obtain an alternative form of state space representation. As an illustration let us consider an ARMA (p,p-1) model of the form

$$X_t + a_1X_{t-1} +\ldots+ a_pX_{t-p} = b_0e_t + b_1e_{t-1} +\ldots+b_{p-1}e_{t-p+1} \qquad (1.6.3)$$

Define the state vector $\underline{x}_t = (x_1(t), x_2(t),\ldots,x_p(t))$, the matrix

$$\underline{F} = \begin{bmatrix} 0 & 0 \ldots & 0 & -a_p \\ 1 & 0 \ldots & 0 & -a_{t-1} \\ & & \vdots & \\ 0 & 0 \ldots & 1 & -a_1 \end{bmatrix}$$

$$\underline{G} = \begin{bmatrix} b_{p-1} \\ b_{p-2} \\ \vdots \\ b_0 \end{bmatrix} , \quad \underline{H}' = (0, 0, \ldots, 1).$$

Then the model (1.6.3) can be written in the form

$$\underline{x}_t = \underline{F}\, \underline{x}_{t-1} + \underline{G}\, e_t$$

$$X_t = \underline{H}'\, \underline{x}_t \qquad (1.6.4)$$

Explicit expressions for $\{x_k(t)\}$ of the state vector $\underline{x}_t$ can be obtained recursively. The process $\{X_t\}$ is stationary if all the eigenvalues of the matrix $\underline{F}$ are less than one in modulus. (One can obtain similar state space forms for the time series models with time dependent coefficients. The main advantages of using the state space forms arise when these models are used for forecasting and when one is dealing with multiple time series (see Mehra, 1977).

1.7 Bispectra and linear processes

In a recent paper, Lii and Rosenblatt (1982) have shown how bispectral density function of a non-Gaussian, but linear process, can be used for estimating the phase relationships, and this in turn can be applied to the problem of deconvolution of e.g. seismic traces. In this section, we briefly outline their approach.

Suppose $\{X_t\}$ satisfies the model

$$\sum_{j=0}^{p} b_j X_{t-j} = \sum_{k=0}^{q} a_k V_{t-k} , \quad a_0, b_0 \neq 0 \qquad (1.7.1)$$

Define the polynomials $B(z) = \sum_{j=0}^{p} b_j z^j$, $A(z) = \sum_{j=0}^{p} a_j z^j$. If the

roots of the polynomials A(z) and B(z) are greater than one, then one can write (1.7.1) as

$$X_t = \sum_{j=0}^{\infty} \beta_j V_{t-j} , \quad \sum|\beta_j| < \infty$$

or

$$V_t = \sum \gamma_j X_{t-j} , \quad \sum|\gamma_j| < \infty \qquad (1.7.2)$$

If the roots of A(z), B(z) are greater than one, whether or not the process X_t is Gaussian, the optimal predictor of X_{t+s} given $(X_t, X_{t-1} \ldots)$ is always linear in $(X_t, X_{t-1} \ldots)$. Even if the roots of A(z), B(z) are less than one, but the process is Gaussian, the predictors are still linear. (This is because of the fact that in the Gaussian case, there is an equivalent linear representation with the roots outside the unit circle). However, in the non-Gaussian case, when the roots are less than one, the optimal predictors may be non-linear. In such situations different specifications generally correspond to different probability structures and differenct stationary processes. To illustrate this, Lii and Rosenblatt (1982), have considered the two representations, namely,

$$X_t = 6V_t - 5V_{t-1} + V_{t-2} , \qquad (1.7.3)$$

$$Y_t = 3V_t - 7V_{t-1} + 2V_{t-2} \qquad (1.7.4)$$

where $\{V_t\}$ are independent, with zero mean and variance one. The polynomial associated with the representation (1.7.3) have roots 2 and 3, the polynomial associated with (1.7.4) have roots $\frac{1}{2}$ and 3, and the two processes have the same spectral density function. If the independent random variables $\{V_t\}$ are exponentially distributed, then the marginal distributions of X_t and Y_t are different.

Now consider the linear representation

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j V_{t-j} , \quad \sum \alpha_j^2 < \infty \qquad (1.7.5)$$

where the random variables $\{V_t\}$ are independent, identically distributed with mean zero, and variance one. The second order spectral density function of X_t is given by

$$f(\lambda) = \frac{1}{2\pi} |\alpha(e^{-i\lambda})|^2 \qquad (1.7.6)$$

where $\alpha(e^{-i\lambda}) = \sum \alpha_j e^{-ij\lambda}$, and hence we have $|\alpha(e^{-i\lambda})| = \sqrt{2\pi f(\lambda)}$.

If the process X_t is Gaussian, the full probability structure of $\{X_t\}$ is determined by $f(\lambda)$ or equivalently by the modulus $|\alpha(e^{-i\lambda})|$, and therefore the phase information in $\alpha(e^{-i\lambda})$ is not identifiable. This phase information can be recovered using bispectra as shown by Rosenblatt (1980) and Lii and Rosenblatt (1982) if the random variables $\{V_t\}$ are non-Gaussian.

Let $\gamma = E(v_t^3) \neq 0$. Then the bispectrum of the process $\{X_t\}$ is given by

$$b(\lambda,\mu) = \frac{\gamma}{(2\pi)^2} \alpha(e^{-i\lambda}) \alpha(e^{-i\mu}) \alpha(e^{i(\lambda+\mu)}) \qquad (1.7.7)$$

Let $\alpha(e^{-i\lambda}) = |2\pi f(\lambda)|^{\frac{1}{2}} e^{ih(\lambda)}$, where we define

$$h(\lambda) = \text{Arg}\{\alpha(e^{-i\lambda}) \frac{\alpha(1)}{|\alpha(1)|}\}$$

Since $\{\alpha_u\}$ are real, $h(-\lambda) = -h(\lambda)$, and also $h(0) = 0$.

From the relation (1.7.7), and the definition of $h(\lambda)$, we have

$$\gamma \frac{\alpha^3(1)}{|\alpha(1)|^3} = \frac{(2\pi)^{5/2} b(0,0)}{[f(0)]^{3/2}},$$

$$h(\lambda) + h(\mu) + h(-\lambda-\mu) = \text{Arg}\left\{\alpha(e^{-i\lambda}) \alpha(e^{-i\mu}) (e^{i\lambda+i\mu}) \left(\gamma^{-1} \frac{\sqrt{2\pi}\, b(0,0)}{[f(0)]^{3/2}}\right)\right\} \qquad (1.7.8)$$

$$= \text{Arg}\left\{b(\lambda,\mu) \cdot \gamma^{-1} (2\pi)^{5/2} \frac{b(0,0)}{[f(0)]^{3/2}}\right\}$$

From this expression, it is clear that $h(\lambda)$ cannot be explicitly

evaluated. We have, since $h(0) = 0$,

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} \{h(\lambda) + h(\Delta) - h(\lambda + \Delta)\} = h'(0) - h'(\lambda)$$

By setting $C = h'(0)$, we can write

$$h(\lambda) = \int_0^\lambda (h'(\mu) - h'(0))\, d\mu + C\lambda.$$

$$= h_1(\lambda) + C\lambda ,$$

where

$$h_1(\lambda) = \int_0^\lambda (h'(\mu) - h'(0))\, d\lambda = h(\lambda) - \lambda h'(0).$$

To evaluate $h_1(\lambda)$, we use the expression which is valid up to a sign change,

$$h(\lambda) + h(\Delta) + h(-\lambda-\Delta) = \text{Arg}\{b(\lambda,\Delta)\}$$

Set $\Delta = \Delta(n)$, $k\Delta = \lambda$, and let $\Delta = \Delta(n) \to 0$ as $n \to \infty$. Assume for convenience $b(0,0)$ is positive. (The case when $b(0,0)$ is negative requires only a slight modification). Then

$$h_1(\lambda) = h(k\Delta) - \frac{h(\Delta)}{\Delta} k\Delta.$$

$$= -\sum_{j=1}^{k-1} \{h(j\Delta) + h(\Delta) - h((j+1)\Delta)\}$$

$$= -\sum_{j=1}^{k-1} \text{arg}\{b(j\Delta,\Delta)\}.$$

In other words, to estimate $h_1(\lambda)$ and thus $h(\lambda)$, the bispectral density has to be estimated.

Lii and Rosenblatt (1982) have proposed a method of estimation for $h_1(\lambda)$ and studied its sampling properties.

Since

$$\alpha_k = \frac{1}{2\pi} \int_0^{2\pi} \alpha(e^{-i\lambda}) e^{ik\lambda} d\lambda,$$

an estimate of α_k can be obtained from

$$\hat{\alpha}_k = \frac{1}{2\pi} \int_0^{2\pi} \hat{\alpha}(e^{-i\lambda}) e^{ik\lambda} d\lambda,$$

where $\hat{\alpha}(e^{-i\lambda})$ is an estimate of $\alpha(e^{-i\lambda})$.

Suppose the linear process X_t has a finite one sided representation,

$$X_t = \sum_{j=0}^{q} \alpha_j V_{t-j} = \alpha(B)V_t ,$$

where $\alpha(B) = \sum_{j=0}^{q} \alpha_j B^j$, B denoting, as previously, the backward shift operator, so that we can write $V_t = \alpha^{-1}(B)X_t$.

Quite a number of seismic records are observed to be non-Gaussian, and in many geophysical problems it is often required to estimate the coefficients α_j's, so that one can recover the process $\{V_t\}$. This operation is called deconvolution, and details are given by, e.g. Lii and Rosenblatt (1982).

If it happens that $E(V_t^3) = 0$, and $E(V_t^4) \neq 0$, the above procedure can be extended to the fourth order cumulants (see Lii and Rosenblatt, 1983).

1.8 Invertibility of Time Series Models

For the models of the form (1.5.7) to be useful for forecasting purposes, it is necessary to be able to estimate the unobservable random variables $\{e_t\}$ when only $\{X_t\}$ is available. In other words, it must be possible to invert the model (1.5.7) so that one can write e_t in terms of the past and the present $\{X_s, s \leq t\}$. The condition for the invertibility for MA and ARMA models are discussed in section 1.5. Recently Granger and Andersen (1978c) have provided another definition of invertibility which can be applied to both linear and non-linear time series.

A model is said to be invertible if it is possible to estimate the e_t sequence from the given X_t values together with an exact knowledge of the generating model (Granger and Andersen, 1978a, p.69). We apply their definition of invertibility to the model (1.5.7) and derive the required condition. Their definition is as follows:

Let X_t be a discrete parameter time series satisfying the model

$$X_t = f(X_{t-j}, e_{t-j}; j = 1,2,\ldots,p) + e_t \qquad (1.8.1)$$

where the $\{e_t\}$ are independent random variables. The random variables $\{e_t\}$ are unobservable. Starting with some initial values $\hat{e}_0 = \hat{e}_{-1} = \ldots = \hat{e}_{-p} = 0$, and $x_0 = x_{-1} = \ldots = x_{-p} = 0$, we obtain a sequence of estimates $\hat{e}_t$ (t=1,2,...) using the model (1.8.1). Thus the model (1.8.1) is said to be invertible if

$$\lim_{t\to\infty} E(e_t - \hat{e}_t)^2 \to 0 \qquad (1.8.2)$$

where the model and the parameters are known completely.

We now consider the model (1.5.7) with $b_0 = 1$. From the model we obtain

$$\eta_t = -b_1\eta_{t-1} - b_2\eta_{t-2} \cdots - b_q\eta_{t-q} \qquad (1.8.3)$$

where $\eta_t = \hat{e}_t - e_t$. Let $\underline{\eta}'_t = (\eta_t, \eta_{t-1}, \ldots, \eta_{t-q+1})$. Then (1.8.3) can be written in the form

$$\eta_t = \underline{B}\, \underline{\eta}_{t-1},$$

$$\eta_t = \underline{H}'\underline{\eta}_t \qquad (1.8.4)$$

where

$$\underline{B} = \begin{bmatrix} -b_1 & -b_2 & -b_3 & & b_q \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

and $\underline{H}$ is as defined earlier. From (1.8.4), we get $E(n_t^2) = \underline{H}'\underline{V}_t\,\underline{H}$, where $\underline{V}_t = E[\underline{n}_t\,\underline{n}'_t]$. Let $V_t(1,1)$ be the element corresponding to the first row and the first column of $\underline{V}_t$. Then $E(n_t^2) = V_t(1,1)$. From the relation (1.8.4), and the definition of V_t, we get the relation

$$\underline{V}_t = \underline{B}\,\underline{V}_{t-1}\underline{B}'$$

and from the results given in Appendix A, we obtain

$$\text{vec}\,(\underline{V}_t) = (\underline{B} \otimes \underline{B})\,\text{vec}(\underline{V}_{t-1}) \qquad (1.8.5)$$

where $(\underline{B} \otimes \underline{B})$ is a Kronecker product matrix. As $t\to\infty$, $V_t(1,1)$ tends to zero if the highest eigenvalue of $\underline{B} \otimes \underline{B}$ is less than unity. From the theory of Kronecker product matrices, we know that if the highest eigenvalue of $\underline{B}$ is less than unity, the highest eigenvalue of $\underline{B} \otimes \underline{B}$ is also less than unity. This condition is equivalent to the condition that the roots of the polynomial $\psi(z) = 1 + b_1 z + b_2 z^2 + \ldots + b_q z^q$ must be outside the unit circle. In other words, the condition of invertibility discussed by Granger and Andersen (1978c) leads to the usual condition in the case of linear time series models.

CHAPTER TWO

THE ESTIMATION OF SPECTRAL AND BISPECTRAL DENSITY FUNCTIONS

2.1 Introduction

As noted previously, if a stationary time series is Gaussian, and therefore linear, the second order spectra contain all the useful information present in the series, including, for example, information about the possible presence of harmonic components. If the series is non-linear the second order spectra will not adequately characterise the series. For instance, for some types of non-linear time series (e.g. bilinear models which will be considered later) one can show that the second order properties are similar to those of a linear time series model. As such, second order spectral analysis will not necessarily show up any features of non-linearity (or non-Gaussianity) present in the series. It may be necessary, therefore, to perform higher order spectral analysis on the series in order to detect departures from linearity and Gaussianity.

The simplest type of higher order spectral analysis one can perform is "bispectral analysis". (There is an extensive literature on higher order spectral analysis, notably by Brillinger (1964), Brillinger and Rosenblatt (1967a, b), Rosenblatt (1966), Rosenblatt and Van Ness (1965) and Van Ness (1966)).

The bispectrum has been used in a number of investigations as a data analytic tool; we mention in particular the work of Hasselman, Munk and MacDonald (1963) on ocean waves, the papers of Lii, Rosenblatt and Van Atta (1976) on turbulence, and Helland, Lii and Rosenblatt (1979) on the energy transfer in grid-generated turbulence.

In this chapter we consider the estimation of the bispectral density function using the spectral window approach. The bispectral density function can also be estimated using complex demodulation techniques (see Godfrey, 1965) and also using Fast Fourier transforms (see Huber et al, 1971). The method considered here is similar to the estimation of the second order spectral density function and an extensive treatment of this subject has been given by Priestley (1981). In the following section, therefore, we describe only briefly the method of estimation of the second order spectral density function, and illustrate it with some simulated examples.

2.2 Estimation of the Spectral Density Function

Let $X_1, X_2, \ldots, X_N$ be a realization of a real valued second order stationary process $\{X_t\}$ with mean μ and covariance function $R(s)$. The natural estimates of μ and $R(s)$, respectively, are

$$\bar{X} = \frac{1}{N} \sum_{t=1}^{N} X_t,$$

$$\hat{R}(s) = \frac{1}{N} \sum_{t=1}^{N-|s|} (X_t - \bar{X})(X_{t+|s|} - \bar{X}), \quad s = 0, \pm 1, \pm 2, \ldots, \pm(N-1) \tag{2.2.1}$$

A natural estimate of $f(\omega)$ is the periodogram $I(\omega)$ defined by

$$I(\omega) = \frac{I}{2\pi N} \left| \sum (X_t - \bar{X}) e^{-it\omega} \right|^2 = \frac{1}{2\pi} \sum_{s=-N+1}^{N-1} \hat{R}(s) e^{-is\omega} \tag{2.2.2}$$

However, it is well known that although $I(\omega)$ is an asymptotically unbiased estimate of $f(\omega)$ it is not a consistent estimate of $f(\omega)$. To obtain a consistent estimate of $f(\omega)$ we "smooth" the periodogram using a weight function, $\{K_M(\theta)\}$, which is then known as a "spectral window" (see Parzen, 1957; Priestley, 1981). The spectral window is usually derived from a "spectral window generator", $K_0(\theta)$, by setting $K_M(\theta) =$

$MK_0(M\theta)$, with $K_0(\theta)$ satisfying the following conditions

$$\text{(i)} \quad \int_{-\infty}^{\infty} K_0(\theta)\, d\theta = 1, \quad \int_{-\infty}^{\infty} K_0^2(\theta)\, d\theta < \infty,$$

$$\text{(ii)} \quad K_0(-\theta) = K_0(\theta) \tag{2.2.3}$$

$$\text{(iii)} \quad K_0(\theta) \to 0 \text{ as } |\theta| \to \infty.$$

The "smoothed" estimate of $f(\omega)$ is then given by

$$\hat{f}(\omega) = \int_{-\pi}^{\pi} K_M(\omega - \theta)\, I(\theta)\, d\theta$$

$$= \frac{1}{2\pi} \sum_{\tau=-(N-1)}^{(N-1)} \lambda\left(\frac{\tau}{M}\right) R(\tau) \cos \omega\tau\,, \tag{2.2.4}$$

$$\lambda(s) = \int_{-\infty}^{\infty} K_0(\theta)\, e^{-\theta s}\, d\theta\,, \tag{2.2.5}$$

where $M = M(N)$ is an integer chosen so that, as $N \to \infty$, $M \to \infty$ but $M/N \to 0$. (If $\lambda(s) = 0$, $|s| > 1$, M corresponds to the "truncation point"). The function $\lambda(s)$ is known as a lag window.

Parzen defined the "characteristic exponent" of the function $K_0(\theta)$ as the largest integer r, such that

$$k(r) = \lim_{\theta \to 0} \left\{ \frac{1 - \lambda(\theta)}{|\theta|^r} \right\}$$

exists, is finite and non-zero.

Parzen (1957) has shown that the most useful class of windows to consider is that for which $r = 2$. We have the well known asymptotic results (see, e.g. Priestley, 1981),

$$E[\hat{f}(\omega)] \sim f(\omega)$$

$$\text{var}[\hat{f}(\omega)] \sim \frac{2\pi}{N} f^2(\omega) \int_{-\pi}^{\pi} K_M^2(\theta)\, d\theta \simeq \frac{2\pi M}{N} f^2(\omega) \int_{-\pi}^{\pi} K_0^2(\theta) d\theta \qquad (2.2.6)$$

from which it follows that, under the above conditions, $\hat{f}(\omega)$ is a consistent estimate of $f(\omega)$. The basic problem in the estimation of $f(\omega)$ is to find a suitable weight function $K_M(\theta)$. During the decade 1955-1965 several authors, notably Lominicki and Zaremba (1957), Parzen (1957, 1958), Whittle (1957), Tukey (1959), Bartlett (1948, 1950, 1966), and Priestley (1962) have suggested various windows, some of which depend on unknown parameters of the spectral density function $f(\omega)$. A comparison of these windows has been made by Neave (1972). Some of the windows which are currently used are given in Table 2.1. All the windows given in Table 2.1 have characteristic exponent 2.

Daniell window	$\lambda_D(s) = \frac{\sin(s\pi)}{s\pi}$
Tukey-Hamming window	$\lambda_T(s) = \begin{cases} 0.54 + 0.46 \cos \pi s & \lvert s\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Parzen window	$\lambda_P(s) = \begin{cases} 1 - 6s^2 + 6\lvert s\rvert^3 & \lvert s\rvert \le \frac{1}{2} \\ 2(1-\lvert s\rvert)^3 & \frac{1}{2} \le \lvert s\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Bartlett-Priestley window	$\lambda_{BP}(s) = \frac{3}{(\pi s)^2} \left\{ \frac{\sin \pi s}{\pi s} - \cos \pi s \right\}$

TABLE 2.1: Lag windows $\lambda(s)$

Taking the relative mean square error as the optimality criterion, Priestley (1981) has shown that the Bartlett-Priestley windown is optimal amongst all non-negative windows with characteristic exponent 2. (Bartlett (1966, p.316) suggested the same window independently). In fact Priestley (1981) defines an "Efficiency Index" of a window which is proportional to the relative mean square error. He shows that the Bartlett-Priestley window has the smallest Index value amongst all the non-negative windows with characteristic exponent 2.

The actual estimation of $f(\omega)$ depends on several factors, some of which are (i) the degree of smoothness required of the spectral estimate and (ii) the resolvability of the peaks in the estimate. These in turn influence the choice of the number of terms used in the right-hand side summation of (2.2.4) and the value of M. For an excellent account of these design relations, the reader is referred to the book by Priestley (1981, Ch.7) and hence we do not go further into these details.

To estimate $f(\omega)$, we write (2.2.4) as

$$\hat{f}(\omega) = \frac{1}{2\pi} \sum_{\tau=-n}^{n} \lambda\left(\frac{\tau}{M}\right) \hat{R}(\tau) \cos \omega\tau \qquad (2.2.7)$$

For the lag windows of the truncated type, i.e. of the form $\lambda(s) = 0$, $|s| < 1$, $n = M$, otherwise $n = N-1$. In practice, even when using windows of the non-truncated type, we often choose a value of n smaller than $(N-1)$. This is because the covariances tend to zero as lags increase. For all our illustrations in this section we have used the Parzen window.

To illustrate the estimation procedure, a set of 500 independent variables $\{e_t\}$, each distributed normally with mean zero and variance unity, are generated. Using these e_t's the series $\{X_t\}$ is generated from the following models.

Series A

The series $\{X_t\}$ is generated from

$$X_t - 0.8X_{t-1} + 0.4X_{t-2} = e_t \qquad (2.2.8)$$

The spectral density function $f(\omega)$ of the series generated from (2.2.8) can be shown to be

$$f(\omega) = \frac{1}{2\pi|1-0.8e^{-i\omega}+0.4e^{-2i\omega}|^2} \qquad (2.2.9)$$

The function $f(\omega)$ is plotted against ω (by continuous lines) for $\omega=0.0(.01)\pi$ and these plots are given in Fig. 2.1. The estimated values of $f(\omega)$ are obtained by choosing n = M = 12, and the plot of this graph is given in Fig. 2.1 (broken line).

Series B

A non-linear time series X_t is generated from the bilinear time series model

$$X_t - 0.8X_{t-1} + 0.4X_{t-2} = e_t + 0.2X_{t-1}\, e_{t-1} \qquad (2.2.10)$$

The parametric spectral density function of a bilinear process satisfying the difference equation

$$X_t + a_1X_{t-1} + a_2X_{t-2} = e_t + b_1X_{t-1}\, e_{t-1} \qquad (2.2.11)$$

(where the parameters a_1, a_2 and b_1 satisfy the conditions of stationarity) can be shown to be

$$f(\omega) = \frac{1}{2\pi}\left[\frac{1-a_2{}^2+a_1(a_1+2\rho_1)+2[\rho_1(1+a_2)+a_1]\cos\omega}{(1-a_2)^2+a_1{}^2+2a_1(1+a_2)\cos\omega+4a_2\cos^2\omega}\right] \qquad (2.2.12)$$

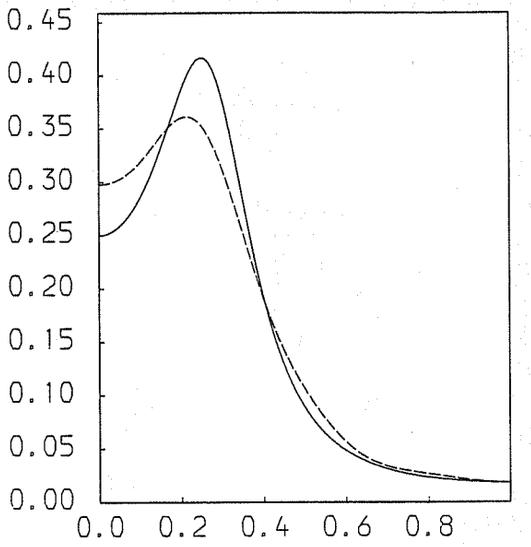


Fig. 2.1: Series A

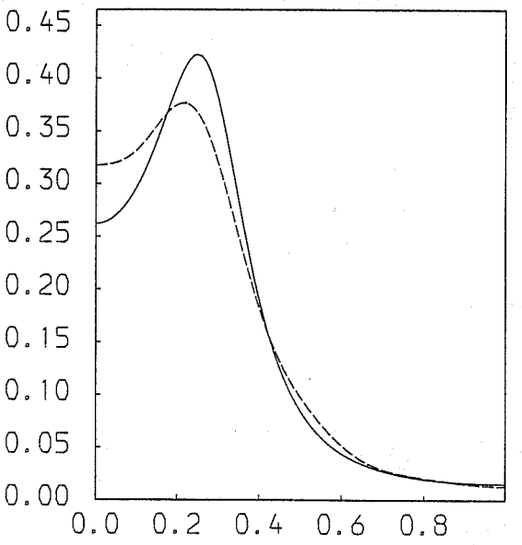


Fig. 2.2: Series B

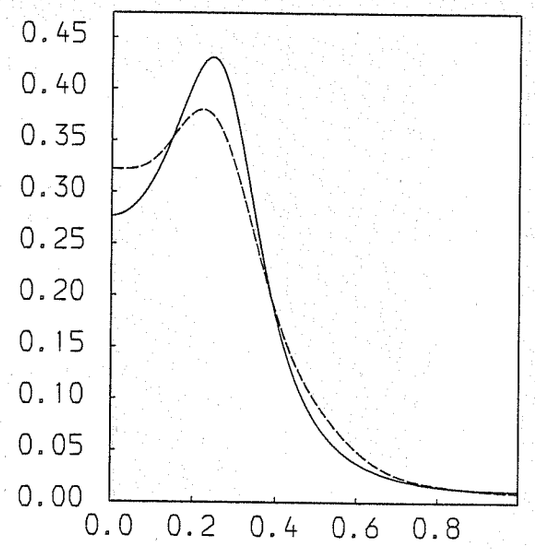


Fig. 2.3: Series C

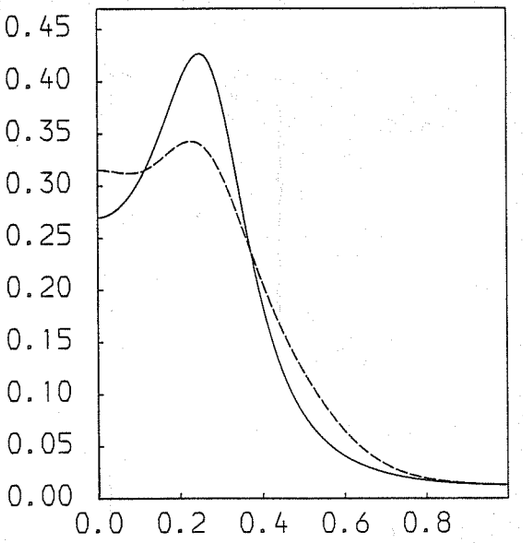


Fig. 2.4: Series D

where

$$\rho_1 = \frac{R(1)}{\sigma_X^2}, \quad R(1) = \frac{2b_1\mu_1 - a_1\mu_2}{1+a_2} - \mu_1^2,$$

$$\sigma_X^2 = \mu_2 - \mu_1^2, \quad \mu_2 = E(X_t^2) = \frac{(1+a_2)(1+2b_1^2) - 2b_1\mu(2a_1+a_2(1+a_2))}{(1+a_2)(1-a_1^2-a_2^2-b_1^2)+2a_1^2a_2},$$

$$\mu_1 = E(X_t) = \frac{b_1}{1+a_1+a_2} .$$

The estimated spectral density (n =12 , M = 12), together with the parametric spectral density, is plotted in Fig. 2.2.

Series C and D

Non-linear series C and D were also generated from the difference equation (2.2.11). To obtain Series C, we put $a_1 = -0.8$, $a_2 = 0.4$, $b_1 = 0.4$, and Series D is obtained by putting $a_1 = -0.8$, $a_2 = 0.4$, and $b_1 = 0.6$. The number of observations generated for all the four series is 500. The theoretical and the estimated spectral densities for Series C and D are plotted in Figs. 2.3 and 2.4 respectively.

We note that the spectral shapes of the non-linear series B, C and D are similar to the spectral shape of Series A.

2.3 Estimation of the Bispectral Density Function

Let $X_1, X_2, \ldots, X_N$ be a realization from a third order stationary process $\{X_t\}$ with third order central moments $C(\tau_1,\tau_2)$, defined in section 1.4. The natural estimate of $C(\tau_1,\tau_2)$ is given by

$$\hat{C}(\tau_1,\tau_2) = \frac{1}{N}\sum_{t=1}^{N-\gamma} (X_t-\bar{X})\,(X_{t+\tau_1}-\bar{X})\,(X_{t+\tau_2}-\bar{X}), \quad \tau_1,\tau_2 \geq 0 \qquad (2.3.1)$$

where $\gamma = \max(0,\tau_1,\tau_2)$.

Following the approach used to estimate the spectral density function, we introduce the two-dimensional function $I(\omega_1,\omega_2)$, defined by

$$I(\omega_1,\omega_2) = \frac{1}{(2\pi)^2} \sum_{\tau_1=-(N-1)}^{N-1} \sum_{\tau_2=-(N-1)}^{N-1} \hat{C}(\tau_1,\tau_2)\, e^{-i\tau_1\omega_1 - i\tau_2\omega_2}$$

$$= \frac{1}{N(2\pi)^2} \left[\sum_{t=1}^{N} (X_t-\bar{X})\, e^{-i\omega_1 t}\right] \left[\sum_{t=1}^{N} (X_t-\bar{X})\, e^{-i\omega_2 t}\right] \qquad (2.3.2)$$

$$\left[\sum_{t=1}^{N} (X_t-\bar{X})\, e^{i(\omega_1+\omega_2)t}\right]$$

The function $I(\omega_1,\omega_2)$ is called the "third order periodogram". It can be shown (Brillinger and Rosenblatt, 1967a, 1967b; Van Ness, 1966) that although $I(\omega_1,\omega_2)$ is an asymptotically unbiased estimate of $f(\omega_1,\omega_2)$, it is not a consistent estimate of $f(\omega_1,\omega_2)$. To obtain a consistent estimate, $I(\omega_1,\omega_2)$ has to be "smoothed", as in the second order case. Let $K_0(\theta_1,\theta_2)$ be a bounded and non-negative function satisfying

$$\text{(i)} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_0(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = 1$$

$$\text{(ii)} \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_0^2(\theta_1.\theta_2)\, d\theta_1\, d\theta_2 < \infty. \qquad (2.3.3)$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta_i^2\, K_0(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 < \infty,\ i = 1,2$$

$$\text{(iii)} \quad K_0(\theta_1,\theta_2) = K_0(\theta_2,\theta_1) = K_0(\theta_1,-\theta_1-\theta_2) = K_0(-\theta_1-\theta_2,\theta_2)$$

Let $\lambda(\tau_1,\tau_2)$ be its inverse Fourier transform, viz,

$$\lambda(\tau_1,\tau_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i\tau_1\theta_1 + i\tau_2\theta_2} K_0(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$

Then $\lambda(\tau_1,\tau_2)$ also satisfies the symmetry conditions,

$$\lambda(\tau_1,\tau_2) = \lambda(\tau_2,\tau_1) = \lambda(-\tau_1,\tau_2-\tau_1) = \lambda(\tau_1-\tau_2,-\tau_2).$$

The bispectral estimate $f(\omega_1,\omega_2)$ is then given by

$$\hat{f}(\omega_1,\omega_2) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} K_M(\theta_1-\omega_1,\theta_2-\omega_2)\, I(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$

$$= \frac{1}{(2\pi)^2} \sum_{\tau_1=-(N-1)}^{(N-1)} \sum_{\tau_2=-(N-1)}^{(N-1)} \lambda\left(\frac{\tau_1}{M}, \frac{\tau_2}{M}\right) \hat{C}(\tau_1,\tau_2)\, e^{-i\tau_1\omega_1 - i\tau_2\omega_2} \qquad (2.3.4)$$

where $K_M(\theta_1,\theta_2) = M^2\, K_0(M\theta_1, M\theta_2)$, and M, the window parameter, is chosen such that $M^2/N \to 0$ as $M \to \infty$, $N \to \infty$.

Since $f(\omega_1,\omega_2)$ and $\hat{f}(\omega_1,\omega_2)$ are complex valued functions, we can write

$$f(\omega_1,\omega_2) = r(\omega_1,\omega_2) + i\, q(\omega_1,\omega_2),$$
$$\hat{f}(\omega_1,\omega_2) = \hat{r}(\omega_1,\omega_2) + i\, \hat{q}(\omega_1,\omega_2)$$

Then the mean and the variance of $\hat{f}(\omega_1,\omega_2)$ can be defined by

$$E(\hat{f}(\omega_1,\omega_2)) = E(\hat{r}(\omega_1,\omega_2)) + i\, E(\hat{q}(\omega_1,\omega_2)),$$

$$\text{var}(\hat{f}(\omega_1,\omega_2)) = E\left|\hat{f}(\omega_1,\omega_2) - E\,\hat{f}(\omega_1,\omega_2)\right|^2$$
$$= E\left[\hat{r}(\omega_1,\omega_2) - E(\hat{r}(\omega_1,\omega_2)\right]^2$$
$$+ E\left[\hat{q}(\omega_1,\omega_2) - E(\hat{q}(\omega_1,\omega_2)\right]^2$$

The mean square error is defined by

$$\text{MSE}\ (\hat{f}(\omega_1,\omega_2)) = E\left|\hat{f}(\omega_1,\omega_2) - f(\omega_1,\omega_2)\right|^2$$
$$= \text{var}(\hat{f}(\omega_1,\omega_2)) + \left|b(\omega_1,\omega_2)\right|^2$$

where $b(\omega_1,\omega_2)$, the bias, is given by

$$b(\omega_1,\omega_2) = E(\hat{f}(\omega_1,\omega_2)) - f(\omega_1,\omega_2).$$

We need the following definition to obtain an expression for the bias.

Definition

Consider the class of all functions $K_0(\theta_1,\theta_2)$ satisfying the conditions (2.3.3) and let the partial derivatives of the two-dimensional lag window generators $\lambda(s_1,s_2)$ satisfy the conditions

$$\left.\frac{\partial}{\partial s_j}\lambda(s_1,s_2)\right|_{s_1=s_2=0} = i\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_j K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 = 0 \qquad (j = 1, 2) \tag{2.3.5}$$

and

$$\left.\frac{\partial^2}{\partial s_j^2}\lambda(s_1,s_2)\right|_{s_1=s_2=0} = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_j^2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 \neq 0 \qquad (j = 1, 2) \tag{2.3.6}$$

We will say that the functions satisfying (2.3.3), (2.3.5) and (2.3.6) belong to the "$C_K^{(2)}$ class". (Here the condition (2.3.6) is similar to the requirement that a one-dimensional window has characteristic exponent 2; all non-negative windows necessarily have characteristic exponent 2 (Parzen, 1957). We now obtain an expression for the bias of the bispectral estimate, the derivation of which is essentially similar to the one given by Brillinger and Rosenblatt (1967a) for the k-th order case.

Theorem 1 Let the second order partial derivatives of $f(\omega_1,\omega_2)$ exist and be finite, and let the third order moment $C(\tau_1,\tau_2)$ satisfy condition (1.3.8). Further let $K_0(\theta_1,\theta_2) \in C_K^{(2)}$. Then the bias of $f(\omega_1,\omega_2)$ is given by

$$b(\omega_1,\omega_2) = E[\hat{f}(\omega_1,\omega_2)] - f(\omega_1,\omega_2)$$

$$= \frac{B_K}{M^2} D^{(2)}(\omega_1,\omega_2) + O(M^{-3}) \tag{2.3.7}$$

where $D^{(2)}(\omega_1,\omega_2) = \left(\frac{\partial^2}{\partial\omega_1^2} - \frac{\partial^2}{\partial\omega_1\partial\omega_2} + \frac{\partial^2}{\partial\omega_2^2}\right) f(\omega_1,\omega_2)$,

$$2B_K = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1^2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_2^2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2$$

$$= -2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1\theta_2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 \tag{2.3.8}$$

Proof

It can be shown that (Brillinger and Rosenblatt, 1967a,p.163):

$$b(\omega_1,\omega_2) = \int_{-\pi}^{\pi}\int \{f(\omega_1-\theta_1,\omega_2-\theta_2) - f(\omega_1,\omega_2)\} K_M(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 + O\left(\frac{M}{N}\right) \tag{2.3.9}$$

Expanding $f(\omega_1-\theta_1,\omega_2-\theta_2)$ in Taylor series and making use of the conditions (2.3.3) and (2.3.5) we obtain

$$b(\omega_1,\omega_2) = \frac{1}{2!}\sum_{\substack{r_1,r_2=0\\ r_1+r_2=2}}^{2}\binom{2}{r_1,r_2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\theta_1^{r_1}\theta_2^{r_2} K_M(\theta_1,\theta_2)\,d\theta_1\,d\theta_2$$

$$\frac{\partial^2 f(\omega_1,\omega_2)}{\partial\omega_1^{r_1}\partial\omega_2^{r_2}} + O(M^{-3}) + O\left(\frac{M}{N}\right) \tag{2.3.10}$$

In obtaining (2.3.10) from (2.3.9) we have used the fact that $K_0(\theta_1,\theta_2)$ is symmetric and (after change of variables) we can write

$$2B_K = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1^2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\theta_1+\theta_2)^2 K_0(-\theta_1-\theta_2,\theta_2)\,d\theta_1\,d\theta_2$$

$$= 4B_K + 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1\theta_2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2,$$

which implies

$$B_K = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta_1\theta_2 K_0(\theta_1,\theta_2)\,d\theta_1\,d\theta_2.$$

The variance of the bispectral estimate, when (ω_1,ω_2) are defined inside one of the twelve sectors, is given by (see Van Ness, 1966; Brillinger and Rosenblatt, 1967):

$$\mathrm{var}(\hat{f}(\omega_1,\omega_2)) \simeq \frac{M^2}{N}\frac{V_2}{2\pi} f(\omega_1)f(\omega_2)f(\omega_1+\omega_2), \quad 0 < \omega_2 < \omega_1 \tag{2.3.11}$$

where

$$V_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \lambda^2(u_1,u_2)\, du_1\, du_2 = (2\pi)^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} K_0^2\,(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 \ .$$

If the parameter M is chosen such that as $N \to \infty$, $M \to \infty, \frac{M^2}{N} \to 0$, then the bispectral estimate is a consistent estimate of $f(\omega_1,\omega_2)$.

The mean square error of the bispectral estimate which is a function of $f(\omega_1)$, $f(\omega_2)$ and $f(\omega_1+\omega_2)$ can now be written as

$$M(\omega_1,\omega_2) \simeq \frac{M^2}{N} \frac{V_2}{2\pi} f(\omega_1)\, f(\omega_2)\, f(\omega_1+\omega_2) + \frac{1}{M^4} B_K^2 \left[D^{(2)}(\omega_1,\omega_2)\right]^2 \qquad (2.3.12)$$

The mean square error is a function of V_2 and B_K only when the second order spectral density functions $f(\omega_1)$, $f(\omega_2)$ and $f(\omega_1+\omega_2)$ are given.

We assume that the spectral windows which are non-negative and belong to $C_K^{(2)}$ class give rise effectively to the same bias when used for estimating the bispectral density function. With this assumption and for fixed $f(\omega_1)$, $f(\omega_2)$ and $f(\omega_1+\omega_2)$, the minimization of $M(\omega_1,\omega_2)$ is the same as the minimization of V_2. In the following section we find an optimum weight function for which V_2 is minimum (we note that the assumption that the non-negative windows belong to $C_K^{(2)}$ class implies that the windows are invariant under scale transformation).

2.4 Optimum Bispectral Window

Let $K^*(\theta_1,\theta_2)$ be a bispectral window given by

$$K^*(\theta_1,\theta_2) = \begin{cases} \frac{\sqrt{3}}{\pi^3} \left[1 - \frac{1}{\pi^2}(\theta_1^2 + \theta_2^2 + \theta_1\theta_2)\right] & \text{if } (\theta_1,\theta_2) \in G_1 \\ 0 & \text{otherwise} \end{cases} \qquad (2.4.1)$$

where the region G_1 is given by the set

$$\{(\theta_1,\theta_2);\ \theta_1^2 + \theta_2^2 + \theta_1\theta_2 \le \pi^2\}.$$

(G_1' denotes the complementary set). Schematically, this region can be represented by the ellipse area in Fig. 2.5.

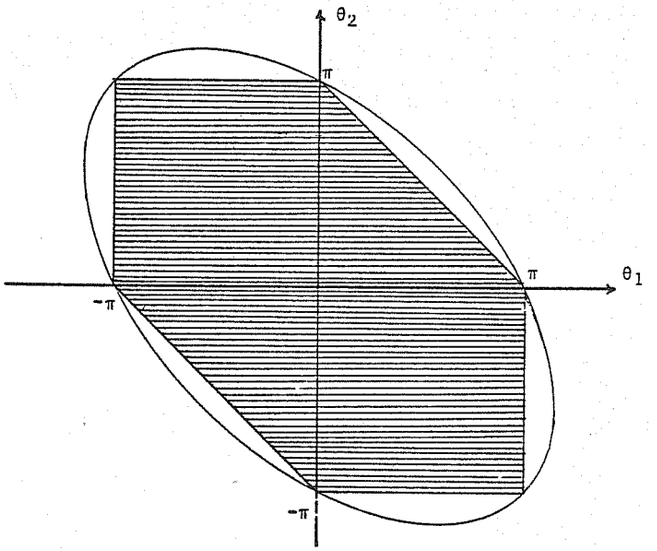


Fig. 2.5: G_1 is the ellipse $\{(\theta_1,\theta_2) : \theta_1^2 + \theta_2^2 + \theta_1\theta_2 \le \pi^2\}$
G is the shaded area $\{(\theta_1,\theta_2) : |\theta_1| + |\theta_2| + |\theta_1+\theta_2| \le 2\pi\}$
$G \subset G_1$

It is simple to check that $K^*(\theta_1,\theta_2)$ satisfies the conditions (2.3.3), (2.3.5) and (2.3.6) and is non-negative and it belongs to $C_K^{(2)}$ class. The optimum property is given in the following theorem.

Theorem 2:- Let $K_a(\theta_1,\theta_2)$ be any other bispectral window belonging to the $C_K^{(2)}$ class, which is non-negative, and satisfies

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_j^2\, K_a(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta_j^2\, K^*(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 \qquad (2.4.2)$$
$$j = 1 \text{ or } 2$$

Then

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_a^2(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 \geq \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^{*2}(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 \qquad (2.4.3)$$

Proof:- Let $K_a(\theta_1,\theta_2) = K^*(\theta_1,\theta_2) + \varepsilon(\theta_1,\theta_2)$. Since $K_a(\theta_1,\theta_2) \in C_K^{(2)}$ and is non-negative, we have

$$\varepsilon(\theta_1,\theta_2) \geq 0 \quad \text{for all } (\theta_1,\theta_2) \in G_1'$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = 0 \qquad (2.4.4)$$

and because of the assumption (2.4.2), $\varepsilon(\theta_1,\theta_2)$ should satisfy the condition

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta_1^2\, \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta_2^2 \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$
$$= -2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \theta_1\theta_2 \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = 0 \qquad (2.4.5)$$

Hence,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_a^2(\theta_1,\theta_2)\, d\theta_1\, d\theta_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K^{*2}(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$
$$+ 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon(\theta_1,\theta_2)\, K^*(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$
$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon^2(\theta_1,\theta_2)\, d\theta_1\, d\theta_2.$$

Consider the term

$$I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varepsilon(\theta_1,\theta_2)\, K^*(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$
$$= \frac{\sqrt{3}}{\pi^2}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[1 - \frac{1}{\pi^2}(\theta_1^2 + \theta_2^2 + \theta_1\theta_2)\right] \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2\right.$$
$$\left. - \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[1 - \frac{1}{\pi^2}(\theta_1^2 + \theta_2^2 + \theta_1\theta_2)\right] \varepsilon(\theta_1,\theta_2)\, d\theta_1\, d\theta_2\right\}.$$

In view of (2.4.4) and (2.4.5), the first term is zero, and since

$$\varepsilon(\theta_1,\theta_2) \geq 0 \quad \text{for all } (\theta_1,\theta_2) \in G_1'$$

and in G_1', $(\theta_1^2 + \theta_2^2 + \theta_1\theta_2) \geq \pi^2$, the second term is positive. Hence the result (2.4.3), which shows that $K^*(\theta_1,\theta_2)$ is an optimum bispectral window.

The inverse Fourier transform of $K^*(\theta_1,\theta_2)$ which may be called the optimum two-dimensional lag window, is given by

$$\lambda^*(s_1,s_2) = \int_{G_1}\int e^{is_1\theta_1 + is_2\theta_2}\, K^*(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$

where the double integration has to be performed over the region G_1. This integration over the region G_1 is complicated and has to be done numerically. To obtain an analytic expression, however, we can approximate the set G_1 by G, the set G given by $\{(\theta_1,\theta_2);\ |\theta_1| + |\theta_2| + |\theta_1+\theta_2| \leq 2\pi\}$ which is the shaded area of Fig. 2.5. This approximation of the set G_1 by G alters slightly the constant of proportionality from $\frac{\sqrt{3}}{\pi^3}$ to $\frac{4}{7\pi^2}$. Now the approximate two-dimensional lag window is given by

$$\lambda^*(s_1,s_2) = \int_{G_1}\int e^{is_1\theta_1 + is_2\theta_2}\, K^*(\theta_1,\theta_2)\, d\theta_1\, d\theta_2$$
$$\simeq \frac{4}{7\pi^2}\int_{G_1}\int e^{is_1\theta_1 + is_2\theta_2}\left[1 - \frac{1}{\pi^2}(\theta_1^2+\theta_2^2+\theta_1\theta_2)\right] d\theta_1\, d\theta_2$$
$$= \frac{4}{7\pi^2}\{J_0 - \frac{1}{\pi^2}(J_1+J_2+J_3)\}$$

where $J_0 = \int_{G_1}\int e^{is_1\theta_1 + is_2\theta_2}\, d\theta_1\, d\theta_2 = 2\mathrm{Re}\left[\int_0^{\pi} e^{is_2\theta_2}\left\{\int_{-\pi}^{\pi-\theta_2} e^{is_1\theta_1}\, d\theta_1\right\}\, d\theta_2\right]$

$$= 2\left\{\frac{(s_2-s_1)\cos(s_1-s_2)\pi + s_1\cos s_1\pi - s_2\cos s_2\pi}{s_1 s_2 (s_2-s_1)}\right\}$$

$$J_1 = \int_{G_1}\int \theta_1^2\, e^{is_1^2 1 + is_2\theta_2}\, d\theta_1\, d\theta_2$$
$$= 2\mathrm{Re}\left[\int_0^{\pi} \theta_1^2\, e^{is_1\theta_1} d\theta_1\left\{\int_{-\pi}^{\pi-\theta_1} e^{is_2\theta_2} d\theta_2\right\}\right]$$

$$= 2 \left\{ \frac{1}{s_1 s_2} (\pi^2 - \frac{2}{s_1^2}) \cos(s_1-s_2)\pi - \frac{1}{s_2(s_1-s_2)} (\pi^2 - \frac{2}{(s_1-s_2)} \cos s_1\pi \right.$$

$$- 2 \frac{(3s_1^2+s_2^2-2s_1s_2)}{s_1^3(s_1-s_2)^3} \cos s_2\pi - \frac{2\pi}{s_1^2 s_2} \sin(s_1-s_2)\pi$$

$$\left. + \frac{2\pi}{s_2(s_1-s_2)} \sin s_1\pi \right\}.$$

$$J_2 = \int_{G_1} \int \theta_2^2 \, e^{is_1\theta_1+is_2\theta_2} \, d\theta_1 \, d\theta_2$$

which can be obtained from J_1 by replacing s_1 by s_2 and s_2 by s_1. Finally,

$$J_3 = \int_{G_1} \int \theta_1\theta_2 e^{is_1\theta_1+is_2\theta_2} \, d\theta_1 \, d\theta_2$$

$$= 2 \, \mathrm{Re} \left[\int_0^{\pi} \theta_2 \, e^{is_2\theta_2} \left\{ \int_{-\pi}^{\pi-\theta_2} \theta_1 e^{is_1\theta_1} \, d\theta_1 \right\} \right]$$

$$= 2 \left\{ \frac{(s_2-3s_1)}{s_1^2(s_2-s_1)^3} \cos s_2\pi + \frac{(3s_2-s_1)}{s_2^2(s_2-s_1)^3} \cos s_1\pi \right.$$

$$- \left(\frac{\pi^2}{s_1 s_2} + \frac{1}{s_1^2 s_2^2} \right) \cos(s_2-s_1)\pi - \frac{(s_2-s_1)\pi}{s_1^2 s_2^2} \sin(s_2-s_1)\pi$$

$$\left. + \frac{(s_2-2s_1)\pi}{s_1^2(s_2-s_1)^2} \sin s_2\pi - \frac{(2s_2-s_1)\pi}{s_2^2(s_2-s_1)^2} \sin s_1\pi \right\}.$$

Combining all these results and simplifying, the optimal two-dimensional lag window can be shown to simplify to

$$\lambda^*(s_1,s_2) = \frac{8}{7\pi^3} \left\{ \frac{1}{\pi} \left[\frac{(2s_1^2+2s_2^2+s_1s_2)}{s_1^3 s_2^3} \cos(s_2-s_1)\pi \right. \right.$$

$$\left. + \frac{(2s_1^2+5s_2^2-5s_1s_2) \cos s_1\pi}{s_2^3 (s_2-s_1)^3} + \frac{(5s_1^2+2s_2^2-5s_1s_2)\cos s_2\pi}{(s_1-s_2)^3 s_1^3} \right]$$

$$- \left[\frac{(s_2-s_1)}{s_1^2 s_2^2} \sin(s_2-s_1)\pi + \frac{s_1}{s_2^2(s_2-s_1)} \sin(s_1\pi) \right.$$

$$\left. \left. + \frac{s_2}{s_1^2(s_2-s_1)^2} \sin(s_2\pi) \right] \right\}. \qquad (2.4.6)$$

This function may also be written in the form

$$\lambda^*(s_1,s_2) = \frac{8}{7\pi^3} \left\{ \nu(s_1,s_2) + \nu(-s_1,s_2-s_1) + \nu(s_1-s_2,-s_2) \right\}.$$

where

$$\nu(s_1,s_2) = \frac{(2s_1^2+2s_2^2+s_1s_2)}{\pi s_1^3 s_2^3} \cos(s_2-s_1)\pi - \frac{(s_2-s_1)}{s_1^2 s_2^2} \sin(s_2-s_1)\pi .$$

2.5 Comparison of Bispectral Lag Windows

A natural choice of the two-dimensional lag window for estimating $f(\omega_1,\omega_2)$ is to take the product of one dimensional lag windows. Brillinger (1965) has suggested that the lag window be chosen of the form

$$\lambda_B(s_1,s_2) = \lambda(s_1) \, \lambda(s_2) \qquad (2.5.1)$$

where $\lambda(s)$ is a one-dimensional lag window; some of the lag windows $\lambda(s)$ are given in Table 2.1. This choice of $\lambda_B(s_1,s_2)$ does not satisfy the symmetry conditions (iii) of (2.3.3) and because of this we consider the windows of the form

$$\lambda_{SG}(s_1,s_2) = \lambda(s_1) \, \lambda(s_2) \, \lambda(s_1-s_2) \qquad (2.5.2)$$

where $\lambda(s)$ are lag windows, defined above, each with characteristic exponent 2, so it satisfies

$$\lambda(0) = 1, \; \lambda'(0) = 0 \text{ and } \lambda''(0) \neq 0.$$

By virtue of equations (2.3.6), (2.3.8) and (2.5.2) we get

$$B_K = -\tfrac{1}{2} \frac{\partial^2}{\partial s_1^2} \lambda_{SG}(s_1,s_2) \Big|_{s_1 = s_2 = 0}$$

$$= -\tfrac{1}{2} \frac{\partial^2}{\partial s_1^2} \{\lambda(s_1) \, \lambda(s_2) \, \lambda(s_1-s_2)\} \Big|_{s_1 = s_2 = 0}$$

$$= -\lambda''(0) \qquad (2.5.3)$$

hence $\lambda_{SG}(s_1,s_2)$ satisfies the conditions (2.3.5) and (2.3.6) and thus belongs to $C_K^{(2)}$ class.

To compare the various lag windows we define an index of efficiency as in the case of spectral density functions. For fixed $f(\omega_1)$ $f(\omega_2)$ $f(\omega_1+\omega_2)$, the mean square error is minimum with respect to M, when

$$M = \left[2 B_K D^{(2)}(\omega_1,\omega_2) \left[\frac{\pi N}{V_2 f(\omega_1) f(\omega_2) f(\omega_1+\omega_2)} \right]^{\frac{1}{2}} \right]^{\frac{1}{3}} \qquad (2.5.4)$$

The minimum mean square error with respect to M is now given by

$$\min_M \left\{ M(\omega_1,\omega_2) \right\} = 2 \left[\frac{1}{2\pi N} f(\omega_1) f(\omega_2) f(\omega_1+\omega_2) \cdot D^{(2)}(\omega_1,\omega_2) V_2 B_K \right]^{\frac{2}{3}} \qquad (2.5.5)$$

In view of the relationship (2.5.5) we can define $E_B = V_2 B_K$ as an index of efficiency of a bispectral window. The smaller E_B, the smaller the mean square error. In Table 2.2, the values of E_B are given for the optimal window, and other product windows.

TABLE 2.2: Values of E_B

Bispectral window $\lambda(s)$		B_K	VV_2	E_B
$\lambda_{SG}(s_1,s_2)$	Daniell $\lambda_D(s)$	3.29	0.5	1.65
	Tukey-Hamming $\lambda_T(s)$	4.54	0.36	1.63
	Parzen $\lambda_p(s)$	12	0.17	2.04
	Bartlett-Priestley $\lambda_{BP}(s)$	1.97	0.84	1.65
optimal $\lambda^*(s_1,s_2)$	-	1.03	1.43	1.47

The optimal lag window has the smallest E_B value and the next smallest is the the product of Daniell windows. The lag windows, $\lambda(s_1,s_2)$ are plotted in Figures 2.6, 2.7, 2.8, 2.9 and 2.10. The optimal lag window has a much flatter surface when compared to the other windows, and the rate of decay of $\lambda(s_1,s_2)$ as $s_1 \to \infty$, $s_2 \to \infty$ is much slower than for other windows. This means that the Fourier transform of this window will be

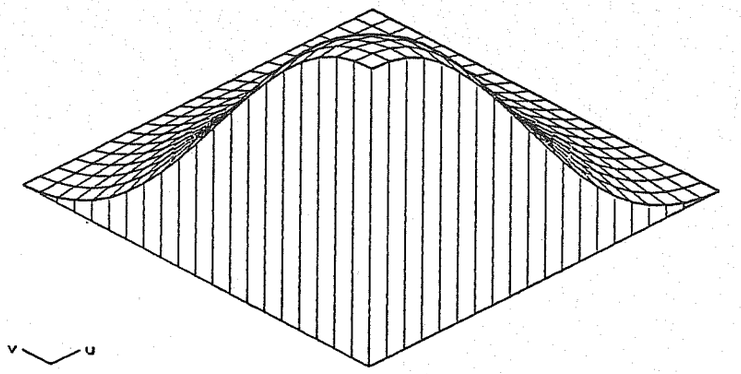


Fig. 2.6: The 2-dimensional Daniell lag-window

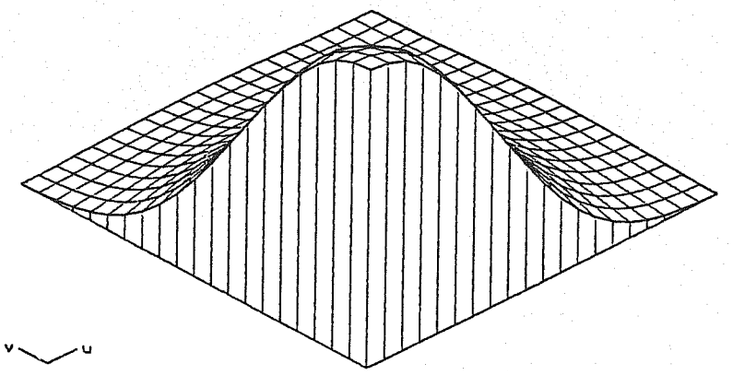


Fig. 2.7: The 2-dimensional Tukey lag-window

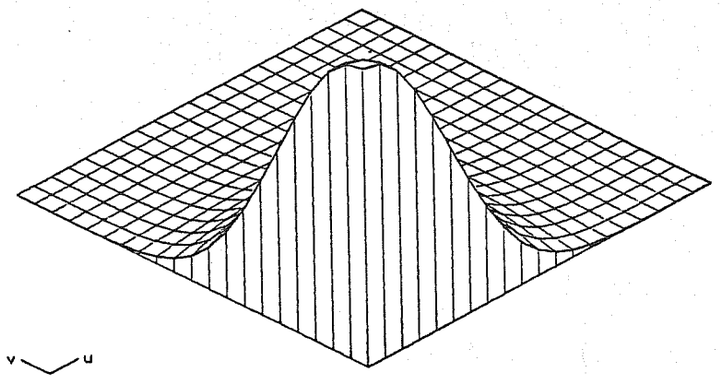


Fig. 2.8: The 2-dimensional Parzen lag-window

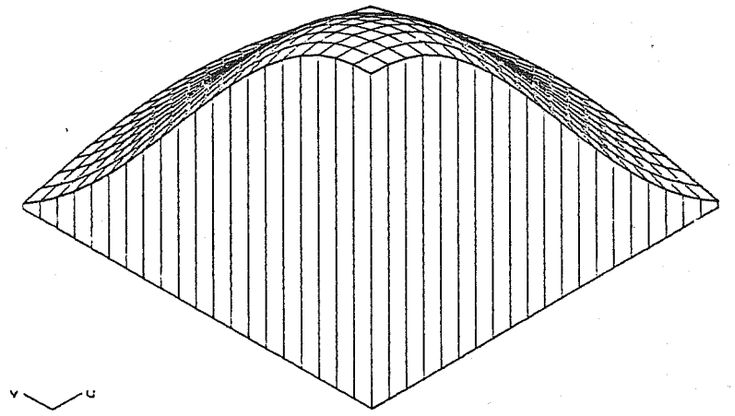


Fig. 2.9: The 2-dimensional Bartlett-Priestley lag-window

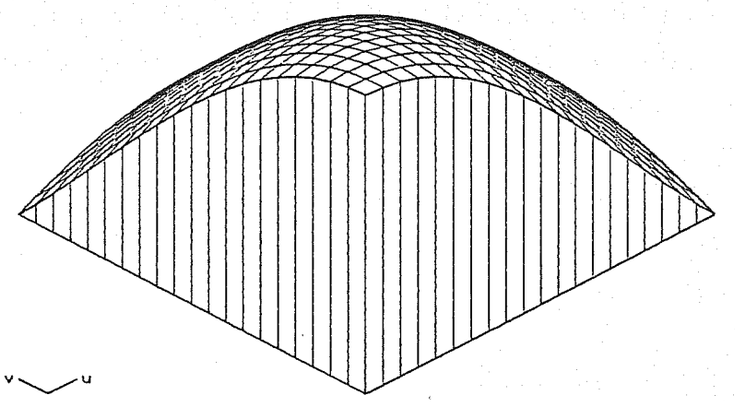


Fig. 2.10: The 2-dimensional optimum lag-window

like a two-dimensional Dirac delta function concentrating all its mass around the origin (0,0). Of course, as is well known, this is a desirable property. We now consider an illustration.

Numerical Illustration

A set of 1000 independent variables $\{e_t\}$, each e_t distributed as an N(0,1) variable, are generated.

Series E: The time series $\{X_t\}$ $(t = 1,2,3,\ldots,1000)$ is generated from the model

$$X_t - 0.8X_{t-1} + 0.4X_{t-2} = e_t + 0.4X_{t-1}\, e_{t-1} \qquad (2.5.6)$$

The bispectral density function is then estimated using the relation (2.3.4) and the Parzen window

$$\hat{f}(\omega_1,\omega_2) = \frac{1}{(2\pi)^2} \sum_{\tau_1=-(n-1)}^{n-1} \sum_{\tau_2=-(n-1)}^{n-1} \lambda\left(\frac{\tau_1}{M},\frac{\tau_2}{M}\right) \hat{C}(\tau_1,\tau_2)\; e^{-i\tau_1\omega_1 - i\tau_2\omega_2} \qquad (2.5.7)$$

where we have chosen n = 10, M = 10, Since $\hat{f}(\omega_1,\omega_2)$ is complex valued, we can write

$$\hat{f}(\omega_1,\omega_2) = \hat{r}(\omega_1,\omega_2) + i\,\hat{q}(\omega_1,\omega_2),$$

and hence the modulus and the phase are given by

$$\text{Modulus} = \left[\hat{r}^2(\omega_1,\omega_2) + \hat{q}^2(\omega_1,\omega_2)\right]^{\frac{1}{2}}$$

$$\text{Phase} = \tan^{-1}\frac{\hat{q}(\omega_1,\omega_2)}{\hat{r}(\omega_1,\omega_2)}.$$

The modulus of the bispectrum is calculated at the frequencies $\omega_1 = 0.00(.05)0.65\pi$, $\omega_2 = .00(.05)\pi$ and the graph is plotted in Fig. 2.11.

We now recall that the spectral density function of the series A and C (the series C and E are one and the same except for sample sizes) are similar in shape, and hence point out that the second order spectral analysis cannot distinguish between the linear model A and the bilinear model C (or E). However, we note that the bispectral density function of the series A is zero for all frequencies whereas these are not zero in the case of series E. This suggests the use of higher order spectral analysis for discrimination purposes.

The graphs of the modulus of the normalized bispectral density, i.e.

$$|\hat{g}(\omega_1,\omega_2)| = \frac{\left[\hat{r}^2(\omega_1,\omega_2)+\hat{q}^2(\omega_1,\omega_2)\right]^{\frac{1}{2}}}{\left\{\hat{f}(\omega_1)\,\hat{f}(\omega_2)\,\hat{f}(\omega_1+\omega_2)\right\}^{1/2}}$$

for the series E is plotted in Fig. 2.12. There is a dominant peak in the neighbourhood of the origin and there are several prominent peaks over the entire (ω_1,ω_2) plane which suggests that the process is non-linear.

In the following section we shall find the parametric bispectral density function of the time series $\{X_t\}$ satisfying a bilinear model A comparison between the parametric and non-parametric bispectral estimates will be considered in the next chapter.

2.6 Bispectral Density Function of BL(1,0,1,1) Model

Let the time series $\{X_t\}$ satify the model

$$X_t = a\,X_{t-1} + b\,X_{t-1}\,e_{t-1} + e_t \qquad (2.6.1)$$

where $\{e_t\}$ is a sequence of independent identically distributed N(0,1) variables. The model (2.6.1) will be defined as a bilinear model BL(1,0,1,1).

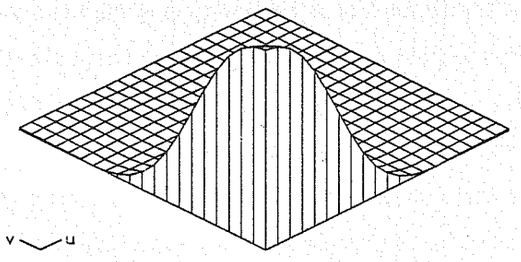


Fig. 2.11: Modulus of the Bispectrum

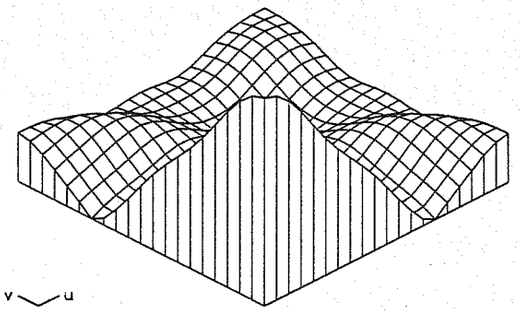


Fig. 2.12: Normalised Bispectrum

The conditions of stationarity, together with the expression for the mean and second order moments of the process $\{X_t\}$ generated from (3.2.9) are given in Section 2.2 . (see Subba Rao, 1981, Granger and Andersen, 1978b). Here we need to find all the third order moments. Although it is complicated, it is still possible to evaluate explicit expressions for them. We will state some of these results which will be needed later. Under the conditions

$$|a| < 1, \ |a^2 + b^2| < 1$$

we can show that

$$E[X_{t-1}\, e_{t-1}] = 1,$$

$$\mu = E[X_t] = \frac{b}{1-a} \tag{2.6.2}$$

$$\mu_2 = E[X_t^2] = \frac{1+2b^2+4ab\mu}{1-a^2-b^2} \tag{2.6.3}$$

$$E[X_{t-1}^2\, e_{t-1}^2] = \frac{1}{1-b^2}\,(a^2\,\mu_2 + 3 + 4ab\mu)$$

$$E[X_{t-1}^2\, e_{t-1}] = 2\mu,$$

$$\left.\begin{aligned} \gamma_1 &= E[X_t\, X_{t-1}] = a\mu_2 + 2b\mu \\ \gamma_s &= E\,(X_t X_{t-s}) = a^{s-1}\,\{a\mu_2 + (1-2a)\mu^2\} + \mu^2 \end{aligned}\right\} \tag{2.6.4}$$

Hence

$$R(s) = \gamma_s - \mu^2 = a^{s-1}\,\{a\mu_2 + (1-2a)\mu^2\},\ s \geq 1 \tag{2.6.5}$$

The spectral density function $f(\omega)$ is then given by

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s)\, e^{-is\omega} \\ &= \frac{1}{2\pi}\left\{(\mu_2-\mu^2) + 2\,\frac{[a\mu_2+(1-2a)\mu^2]\,(\cos\omega - a)}{(1+a^2-2a\cos\omega)}\right\} \end{aligned} \tag{2.6.6}$$

and the normalized spectral density function is

$$g(\omega) = f(\omega)/(\mu_2-\mu^2)\,. \tag{2.6.7}$$

Multiply both sides of the expression for X_{t-1}^3 which can be obtained from (2.6.1) by e_{t-1}, e_{t-1}^2 and e_{t-1}^3 and then by taking expectation both sides we obtain

$$\begin{aligned} Q_1 &= E[X_{t-1}^3\, e_{t-1}] = \frac{3}{1-b^2}\ (1 + a^2\mu_2 + 2b^2 + 4ab\mu), \\ Q_3 &= E[X_{t-1}^3\, e_{t-1}^3] = \frac{3}{1-b^2}\ (5 + 4b^2 + 3a^2\mu_2 + 12ab\mu), \\ Q_2 &= E[X_{t-1}^3\, e_{t-1}^2] = \frac{1}{1-3ab^2}\ (a^3\mu_2 + b^3Q_3 + 3a^2bQ_1 + 9\mu) \end{aligned} \tag{2.6.8}$$

and therefore

$$\mu_3 = E[X_t^3] + \frac{1}{1-3ab^2-a^3}\ \{b^3Q_3 + 3a^2bQ_1 + 3\mu(1+6ab^2)\} \tag{2.6.9}$$

so a sufficient condition for μ_3 to be finite is that

$$a(a^2+3b^2) \neq 1 \tag{2.6.10}$$

Using (2.6.1), we have, for $\tau > 0$,

$$\begin{aligned} \xi_\tau &= E[X_t\, X_{t-\tau}^2] \\ &= E[\{aX_{t-1} + bX_{t-1}\, e_{t-1} + e_t\}\, X_{t-\tau}^2] \end{aligned}$$

Hence,

$$\xi_\tau = \begin{cases} a\ \xi_{\tau-1} + b\ \mu_2, & \tau \geq 2 \\ a\ \mu_3 + b\ Q_1, & \tau = 1 \end{cases}\,. \tag{2.6.11}$$

Now, let

$$\begin{aligned} \nu_\tau &= E[X_t^2\, X_{t-\tau}], \\ \eta_\tau &= E[X_t^2\, e_t^2\, X_{t-\tau}], \\ \zeta_\tau &= E[X_t^2\, e_t\, X_{t-\tau}]. \end{aligned}$$

Then for $\tau > 0$,

$$\begin{aligned} \zeta_\tau &= E[(a\ X_{t-1} + b\ X_{t-1}\, e_{t-1} + e_t)^2\, e_t\, X_{t-\tau}] \\ &= 2E[(a\ X_{t-1} + b\ X_{t-1}\, e_{t-1})\, e_t^2\, X_{t-\tau}] \end{aligned}$$

Hence,

$$= \begin{cases} 2a\mu_2 + 4b\mu & , \tau = 1 \\ 2a\gamma_{\tau-1} + 2b\mu, & , \tau \geq 2 \end{cases} \tag{2.6.12}$$

$$\eta_\tau = E[\{(a^2+b^2e_{t-1}^2+2ab\ e_{t-1})\ X_{t-1}^2 + e_t^2\}e_t^2\ X_{t-\tau}] \tag{2.6.13}$$

$$= \begin{cases} a^2\nu_{\tau-1} + b^2\eta_{\tau-1} + 2ab\zeta_{\tau-1} + 3\mu, & \tau \geq 2 \\ a^2\mu_3 + b^2Q_2 + 2ab\ Q_1 + 3\mu & , \tau = 1 \end{cases} \tag{2.6.14}$$

Hence

$$\begin{aligned} \nu_\tau &= E[\{(a^2+b^2e_{t-1}^2 + 2ab\ e_{t-1})\ X_{t-1}^2 + e_t^2\}\ X_{t-\tau}] \\ &= \begin{cases} a^2\nu_{\tau-1} + b^2\eta_{\tau-1} + 2ab\zeta_{\tau-1} + \mu, & \tau \geq 2 \\ a^2\mu_3 + b^2Q_2 + 2abQ_1 + \mu & , \tau = 1 \end{cases} \end{aligned} \tag{2.6.15}$$

Subtracting (2.6.14) from (2.6.13), we get

$$\eta_\tau = \nu_\tau + 2\mu \tag{2.6.16}$$

Substituting for $\eta_{\tau-1}$ from (2.6.16) into (2.6.15), we get

$$\nu_\tau = (a^2+b^2)\ \nu_{\tau-1} + 2ab\ \zeta_{\tau-1} + \mu(1+2b^2), \quad \tau \geq 2 \tag{2.6.17}$$

By successive substitutions for X_{t-1}, X_{t-2},... from (2.6.1), X_t may be written as

$$X_t = X_{t-\tau} \prod_{j=1}^{\tau} (a+b\ e_{t-j}) + e_t + \sum_{i=1}^{\tau-1} e_{t-i} \prod_{j=1}^{i} (a+b\ e_{t-j}) \tag{2.6.18}$$

Using this expression together with (2.6.1) for $\tau \geq 1$, we get

$$E[X_t X_{t-\tau} X_{t-\tau-1}] = a^{\tau-1} E[(a+b e_{t-\tau}) X^2_{t-\tau} X_{t-\tau-1}]$$
$$+ b \left(\frac{1-a^{\tau-1}}{1-a}\right) . E(X_{t-\tau} X_{t-\tau-1})$$
$$= a^{\tau} \nu_1 + b a^{\tau-1} \zeta_1 + \mu(1-a^{\tau-1})\gamma_1$$
$$= a^{\tau-1}\{a\nu_1 + \mu(1-2a).(a\mu_2 + 2b\mu)\}$$
$$+ a\mu\mu_2 + 2b\mu^2 \qquad (2.6.19)$$

Similarly, for $\tau \geq 1$ and $s \geq 2$,

$$E[X_t X_{t-\tau} X_{t-\tau-s}] = a^{\tau-1} E[(a+b e_{t-\tau}) X^2_{t-\tau} X_{t-\tau-s}]$$
$$+ b \left(\frac{1-a^{\tau-1}}{1-a}\right) E[X_{t-\tau} X_{t-\tau-s}]$$
$$= a^{\tau-1} (a\nu_s + b\zeta_s) + \mu(1-a^{\tau-1})\gamma_s \qquad (2.6.20)$$
$$\tau \geq 1, \; s \geq 2$$

Now, the third order central moments for all lags t_1, t_2 are given by

$$C(t_1,t_2) = E[(X_t-\mu)(X_{t+t_1}-\mu)(X_{t+t_2}-\mu)]$$
$$= E[X_t X_{t+t_1} X_{t+t_2}] - \mu\{E(X_t X_{t+t_1})$$
$$+ E(X_t X_{t+t_2}) + E(X_{t+t_1} X_{t+t_2})\} + 2\mu^3 \qquad (2.6.21)$$

Using all the above results, we obtain the following

(i) for $t_1 = t_2 = 0$, (2.6.21) reduces to

$$C(0,0) = \mu_3 - 3\mu\mu_2 + 2\mu^3.$$

(ii) for $t_1 = t_2 = -\tau$, we get

$$C(-\tau,-\tau) = E[X_t X^2_{t-\tau}] - 2\mu\gamma_\tau - \mu\mu_2 + 2\mu^3$$
$$= (\xi_\tau - \mu\mu_2) - 2\mu(\gamma_\tau - \mu_2)$$

(iii) for $t_1 = 0$, $t_2 = -\tau$, we get

$$C(0,-\tau) = E[X^2_t X_{t-\tau}] - \mu(2\gamma_\tau + \mu_2) + 2\mu^3$$
$$= (\nu_\tau - \mu\mu_2) - 2\mu(\gamma_\tau - \mu^2)$$

(iv) for $t_1 = -\tau$, $t_2 = -\tau-1$, we get

$$C(-\tau,-\tau-1) = E[X_t X_{t-\tau} X_{t-\tau-1}] - \mu[\gamma_\tau + \gamma_{\tau+1} + \gamma_1] + 2\mu^3.$$

Substituting for $E[X_t X_{t-\tau} X_{t-\tau-1}]$ from (2.6.19) and for $\gamma_\tau, \gamma_{\tau-1}$ from (2.6.4) and simplifying, we get

$$C(-\tau,-\tau-1) = a^{\tau-1}[a\nu_1 - 3a^2\mu\mu_2 + (1-5a+6a^2)\mu^3], \; \tau \geq 1 \qquad (2.6.22)$$

(v) for $t_1 = -\tau$, $t_2 = -\tau-s$ ($\tau \geq 1$, $s \geq 2$), we get

$$C(-\tau,-\tau-s) = E[X_t X_{t-\tau} X_{t-\tau-s}] - \mu[\gamma_\tau + \gamma_s + \gamma_{\tau+s}] + 2\mu^3,$$
$$= a^{\tau-1}[a\nu_s + b\zeta_s] + \mu(1-a^{\tau-1})\gamma_s - \mu[\gamma_\tau + \gamma_s + \gamma_{\tau+s}] + 2\mu^3$$
$$= a^{\tau}(\nu_s - \mu\mu_2) + 2aba^{\tau-1}(\gamma_{s-1} - \mu^2)$$
$$- \mu a^{\tau-1}(\gamma_s - \mu^2) - \mu(\gamma_{\tau+s} - \mu^2) \qquad (2.6.23)$$

All the other third order moments can be obtained using the following symmetry relations

$$C(0,\tau) = C(\tau,0) = C(-\tau,-\tau),$$
$$C(\tau,\tau) = C(-\tau,0) = C(0,-\tau),$$
$$C(s,s+\tau) = C(s+\tau,s) = C(-s,\tau) = C(\tau,-s) = C(-s-\tau,-\tau)$$
$$= C(-\tau,-\tau-s). \qquad (2.6.24)$$

The domains represented by the third order moments can be schematically represented as in Figure 2.13.

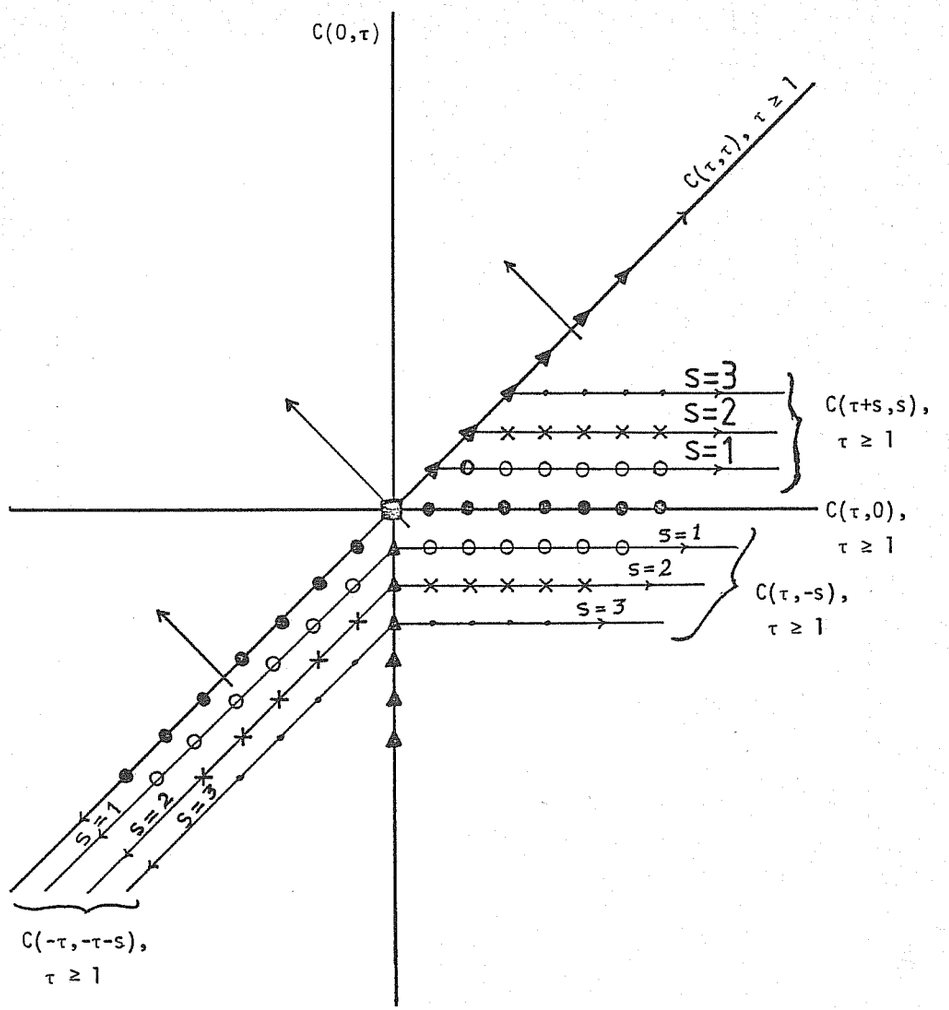


Figure 2.13

The bispectral density function, using the symmetry relations (2.6.24) is given by

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} C(t_1,t_2)\, e^{-i(t_1\omega_1+t_2\omega_2)}$$

$$= \frac{1}{(2\pi)^2} \{C(0,0) + \sum_{\tau=1}^{\infty} C(-\tau,-\tau)\,(e^{+i\tau(\omega_1+\omega_2)} + e^{-i\tau\omega_1} + e^{-i\tau\omega_2})$$

$$+ \sum_{\tau=1}^{\infty} C(0,-\tau)\,(e^{i\tau\omega_1}+e^{i\tau\omega_2}+e^{-i\tau(\omega_1+\omega_2)})$$

$$+ \sum_{\tau=1}^{\infty}\sum_{s=1}^{\infty} C(-\tau,-\tau-s)\{e^{i\tau\omega_1+i(\tau+s)\omega_2}+e^{i(\tau+s)\omega_1+i\tau\omega_2}$$

$$+ e^{-i\tau\omega_1+is\omega_2}+e^{is\omega_1-i\tau\omega_2}+e^{-is\omega_1-i(s+\tau)\omega_2}$$

$$+ e^{-i(s+\tau)\omega_1-is\omega_2}\} \qquad (2.6.25)$$

All these summations can be evaluated explicitly by defining the generating functions. Let $z = e^{i\theta}$, then

$$g_1(\theta) = \sum_{\tau=1}^{\infty} a^{\tau-1} z^{\tau} = \frac{z}{1-az}$$

From the equation (2.6.5) we obtain

$$g_2(\theta) = \sum_{\tau=1}^{\infty} (\gamma_\tau-\mu^2) z^{\tau} = [a\,\mu_2+(1-2a)\,\mu^2]g_1(\theta).$$

Note that the equation (2.6.17) may be rewritten in the form

$$(\nu_\tau-\mu\,\mu_2) = (a^2+b^2)\,(\nu_{\tau-1}-\mu\,\mu_2) + 4a^2b(\gamma_{\tau-2}-\mu^2).$$

Hence,

$$g_3(\theta) = \sum_{\tau=2}^{\infty} (\nu_\tau-\mu\,\mu_2) z^{\tau} = (\nu_2-\mu\,\mu_2) z^2 + \sum_{\tau=3}^{\infty}(\nu_\tau-\mu\,\mu_3)z^{\tau}$$

$$= (\nu_2-\mu\,\mu_2) z^2 + \sum_{\tau=3}^{\infty} \{(a^2+b^2)\,(\nu_{\tau-1} - \mu\,\mu_2)$$

$$+ 4a^2b(\gamma_{\tau-2}-\mu^2)\}\; z^{\tau}$$

$$= (\nu_2-\mu\,\mu_2) z^2 + (a^2+b^2)\, z\, g_3(\theta) + 4a^2b\, z^2\, g_2(\theta).$$

Therefore,

$$g_3(\theta) = \frac{1}{1-(a^2+b^2)z}\{(\nu_2-\mu\ \mu_2)\ z^2 + 4a^2b\ z^2 g_2(\theta)\}.$$

The first and second summations in (2.6.25) are then given by

$$F_1(\theta) = \sum_{\tau=1}^{\infty} C(-\tau,-\tau)\ z^\tau$$

$$= (a\mu_3+bQ_1-\mu\ \mu_2)\ g_1(\theta) - 2\mu\ g_2(\theta),$$

$$F_2(\theta) = \sum_{\tau=1}^{\infty} C(0,-\tau)\ z^\tau$$

$$= C(0,-1)\ z + \sum_{\tau=2}^{\infty} C(0,-\tau)\ z^\tau$$

$$= \{\nu_1-(1+2a)\mu\ \mu_2 \quad - 2\ (1-2a)\mu^3\}\ z$$

$$+ \sum_{\tau=2}^{\infty} \{(\gamma_\tau-\mu\ \mu_2) - 2\mu\ (\gamma_\tau-\mu^2)\}\ z^\tau$$

$$= \{\nu_1-(1+2a)\mu\ \mu_2 - 2\ (1-2a)\mu^3\}\ z + g_3(\theta)$$

$$- 2\mu\ g_2(\theta).$$

To find the fourth term of (2.6.25) define

$$z_1 = e^{i\theta_1},\quad z_2 = e^{i\theta_2}.$$

Then,

$$\sum_{\tau=1}^{\infty}\sum_{s=1}^{\infty} C(-\tau,-\tau-s)\ {z_1}^\tau\ {z_2}^s = z_2 \sum_{\tau=1}^{\infty} C(-\tau,-\tau-1)\ {z_1}^\tau + \sum_{\tau=1}^{\infty}\sum_{s=2}^{\infty} C(-\tau,-\tau-s)\ {z_1}^\tau\ {z_2}^s \qquad (2.6.26)$$

and by virtue of equations (2.6.22) and (2.6.23), we get

$$\sum_{\tau=1}^{\infty} C(-\tau,-\tau-1)\ {z_1}^\tau = \{a\nu_1-3a^2\mu\ \mu_2 + (1-5a+6a^2)\ \mu_3\}\ g_1(\theta_1)$$

$$\sum_{\tau=1}^{\infty}\sum_{s=2}^{\infty} C(-\tau,-\tau-s)\ {z_1}^\tau\ {z_2}^s = a\ g_1(\theta_1)\ g_3(\theta_2) + 2ab\ g_1(\theta_1).z_2 g_2(\theta_2) - \mu\ g_1(\theta_1)\ \{g_2(\theta_2) - (\gamma_1-\mu^2)\ z_2\} - \mu \sum_{\tau=1}^{\infty}\sum_{s=2}^{\infty} (\gamma_{\tau+s}-\mu^2)\ {z_1}^\tau\ {z_2}^s$$

and using (2.6.4), the final term is given by

$$\sum_{\tau=1}^{\infty}\sum_{s=2}^{\infty} (\gamma_{\tau+s}-\mu^2)\ {z_1}^\tau\ {z_2}^s = \{a\mu_2 + (1-2a)\ \mu^2\} \sum_{\tau=1}^{\infty}\sum_{s=2}^{\infty} a^{\tau+s-1}\ {z_1}^\tau\ {z_2}^s$$

$$= \{a\mu_2 + (1-2a)\ \mu^2\} g_1(\theta_1)\ a^2\ z_2\ g_1(\theta_2).$$

Therefore, (2.6.26), after some simplifications, reduces to

$$\sum_{\tau=1}^{\infty}\sum_{s=1}^{\infty} C(-\tau,-\tau-s)\ {z_1}^\tau\ {z_2}^s = g_1(\theta_1)\ F_3(\theta_2)$$

where

$$F_3(\theta_2) = \{a\nu_1 + (1-3a)\ a\mu\mu_2 + (2-7a+6a^2)\mu^3\}\ z_2 + a\ g_3(\theta_2) - \mu\ g_2(\theta_2) + \{2abg_2(\theta_2) - a^2\mu[a\mu_2 + (1-2a)\ \mu^2]\ g_1(\theta_2)\}\ z_2.$$

The above results, together with (2.6.25) lead to the final expression for the bispectral density function, namely,

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2}\ \{C(0,0) + F_1(\omega_1+\omega_2) + F_1(-\omega_1) + F_1(-\omega_2) + F_2(\omega_1) + F_2(\omega_2) + F_2(-\omega_1-\omega_2) + g_1(\omega_1+\omega_2)\ F_3(\omega_2) + g_1(\omega_1+\omega_2)\ F_3(\omega_1) + g_1(-\omega_1)\ F_3(\omega_2) + g_1(-\omega_2)\ F_3(\omega_1) + g_1(-\omega_2)\ F_3(-\omega_1-\omega_2) + g_1(-\omega_1)\ F_3(-\omega_1-\omega_2)\} \qquad (2.6.27)$$

From this formula, the parametric bispectral density function of the time series $\{X_t\}$ satisfying (2.6.1) can be calculated for any values of a, b satisfying the conditions

(i) $|a| < 1$,

(ii) $a^2 + b^2 < 1$,

(iii) $a(a^2+3b^2) \neq 1$,

and at any frequencies ω_1,ω_2. In the following chapter a comparison between the parametric and non-parametric estimates of $f(\omega_1,\omega_2)$ will be made in detail.

CHAPTER 3

PRACTICAL BISPECTRAL ANALYSIS

In the previous chapter, we considered the estimation of the bispectral density function of a stationary time series, and we obtained an optimum two-dimensional "weight function" which leads to an estimate with minimum mean square error. In this chapter, the practical aspects of computation of the bispectral density function from a sample is considered. For illustration purposes, we have simulated time series from a bilinear model and considered the estimation of the spectral density function and the bispectral density function of these series. A comparison with the theoretical (parametric) spectral density and the bispectral density is then made.

3.1 The choice of truncation point (M)

From Chapter 2 we know that the bias and the variance of the bispectrum estimate are, respectively, var $\{\hat{f}(\omega_1,\omega_2)\} = O(M^2/N)$, bias $\{\hat{f}(\omega_1,\omega_2)\} = O(1/M^2)$. In general, as in the case of spectral density estimation, the effect of increasing the parameter M is to increase the variance and decrease the bias. But, as mentioned by Brillinger and Rosenblatt (1967b), and others, the value of M should decrease as the order of the spectra increases in order to maintain reasonable stability of the estimate. They suggest that the value of the window bandwidth in estimating the bispectrum

should be the square root of the window bandwidth used in estimating the spectral density. Therefore, the value of M in estimating $f(\omega_1,\omega_2)$ should be smaller than the value of M used in estimating $f(\omega)$. Also, we have to choose a reasonable value of M corresponding to each window used. To be more specific, the value of M which minimizes the mean square error, given by (2.5.4), can be rewritten as

where

$$\left.\begin{aligned} M &= M_R \cdot G \\ M_R &= (B_K/V_2)^{\frac{1}{3}}, \\ G &= \{2D^{(2)}(\omega_1,\omega_2).\sqrt{\pi N/f(\omega_1)\ f(\omega_2)\ f(\omega_1+\omega_2)}\}^{\frac{1}{3}} \end{aligned}\right\} \qquad (3.1.1)$$

and B_K, V_2 are defined by (2.3.8) and (2.3.11) respectively. For given spectral density and bispectral density functions at known frequencies, G is the same for all windows. Hence, the ratios between the values of M, in using different windows, is given by the ratios of M_R values. Using the values of B_K and V_2, given in Table 2.2, the values of M_R for various windows are calculated and are given in Table 3.1. From the relation (3.1.1) it is clear that M is proportional to M_R and hence when the optimal window is used we need a smaller value of M when compared with other windows as can be seen from Table 3.1.

TABLE 3.1: The values of M_R for different windows and ratios to the optimum window

Window	M_R	Ratios
Daniell	1.7	1.79
Tukey	1.86	1.96
Parzen	3.1	3.26
B.P.	1.23	1.3
Optimum	.95	1

The choice of M can be made on the basis of the following considerations:-

(i) if it is possible, M should be less than the square root of the sample size, N;

(ii) it should be smaller than that value of M used in estimating the spectral density;

(iii) from the relation (3.1.1), we can calculate the ratio of M values; for example, when Parzen and optimum windows are used, we have

$$\frac{M_{Par}}{M_{opt}} = \frac{M_{R,Par}}{M_{R,opt}} ,$$

and this implies

$$M_{Par} = \frac{M_{R,Par}}{M_{R,opt}} \cdot M_{opt}, \text{ and so on.}$$

The ratios $M_{R,Par}/M_{R,opt}$ are given in Table 3.1. For example, if M_{opt} is 5, then

$$M_{Par} = 3.26 \times 5 = 16.3 \simeq 16.$$

3.2 Comparison of Parametric and Non-parametric Bispectral Estimates

In section 2.6, the parametric spectral and bispectral density functions of the BL(1,0;1,1) model (given by (2.6.1)) are obtained explicitly and these can be evaluated for any values of a and b using (2.6.4) and (2.6.27). For our illustration we have chosen $a = .4$ and $b = .4$.

For this choice of a and b, the normalized parametric spectral density function, given by (2.6.7), is plotted in Fig. 3.1. It is interesting to note that the spectral density function decreases as ω increases and the overall shape is similar to the shape of the spectrum of an AR(1) model, although in this case the rate of damping is slow.

The parametric bispectral density $f(\omega_1,\omega_2)$ is calculated from (2.6.77) at frequencies $\omega_1,\omega_2 = 0.0(0.05\pi)\pi$, such that (ω_1,ω_2) lie inside the sector (2) of Fig. 1.3, and the results are given in Tables 3.2 and 3.3. The numbers presented are in fact the modulus and arguments of $f(\omega_1,\omega_2)$ at varying frequencies. The modulus of $f(\omega_1,\omega_2)$ is plotted in Fig. 3.2.

Also, the modulus of the normalized bispectral density is calculated from

$$\mathcal{g}(\omega_1,\omega_2) = \frac{f(\omega_1,\omega_2)}{\{f(\omega_1)\ f(\omega_2)\ f(\omega_1+\omega_2)\}^{\frac{1}{2}}}$$

where $f(\omega)$ and $f(\omega_1,\omega_2)$ are given by (2.6.6) and (2.6.27) respectively. The results are plotted in Fig. 3.3.

A set of 1000 mutually independent random variables $\{e_t\}$, each distributed N(0,1), is generated. Then a time series $(X_t,\ t = 1,2,\ldots,1000)$ is generated from (2.6.1) with a = 0.4 and b = 0.4. The graph of $\{X_t\}$ thus generated is plotted in Fig. 3.10. The bispectral density is estimated using the optimum two-dimensional lag window (given by (2.46) and also the two-dimensional lag window $\lambda_{SG}(s_1,s_2)$ (defined by (2.5.2) with $\lambda(s)$ chosen to be Daniell, Tukey, Parzen and Bartlett-Priestley windows.

Also, the normalized bispectral density function is estimated by

$$\hat{g}(\omega_1,\omega_2) = \frac{\hat{f}(\omega_1,\omega_2)}{\{\hat{f}(\omega_1)\ \hat{f}(\omega_2)\ \hat{f}(\omega_1+\omega_2)\}^{\frac{1}{2}}}$$

where $\hat{f}(\omega)$ and $\hat{f}(\omega_1,\omega_2)$ are the spectral and bispectral density estimates and are given by (2.2.4) and (2.3.4) respectively. The bispectral density $f(\omega_1,\omega_2)$, used in estimating $g(\omega_1,\omega_2)$, is estimated using $\lambda_{SG}(s_1,s_2)$ windows, which are in fact obtained by multiplying the one-dimensional lag windows $\lambda(s)$. We have used the same one-dimensional window $\lambda(s)$ in estimating the spectral density function $f(\omega)$. Also, as pointed out in Brillinger and Rosenblatt (1967b), we have used the same value of M for estimating both $f(\omega)$ and $f(\omega_1,\omega_2)$. The values of the modulus of the non-normalized bispectral density estimate, using different windows and different values of M, are given in Tables 3.5 - 3.13. These numbers, together with the modulus of the normalized bispectral estimates, are shown in Figs. 3.4 - 3.9. The number M has been chosen according to the ratios tabulated in the last column of Table 3.1 and the criteria (ii) and (iii) of section 3.1.

To compare the bispectral estimates using different windows, we used the sample mean square error criterion for measuring the accuracy of $\hat{f}(\omega_1,\omega_2)$ as an estimate of $f(\omega_1,\omega_2)$. This sample mean square error (M.S.E.) is defined as

$$\text{M.S.E.} = \frac{1}{K}\sum\sum\{|\hat{f}(\omega_i,\omega_j)| - |f(\omega_i,\omega_j)|\}^2$$

where $|\hat{f}(\omega_i,\omega_j)|$ is the modulus of the bispectral density estimate and $|\hat{f}(\omega_i,\omega_j)|$ is the theoretical (parametric) bispectral modulus. The summations are taken over all the frequencies (ω_i,ω_j), $\omega_i,\omega_j = 0.0(0.05\pi)\pi$ and are such that (ω_i,ω_j) lie inside the sector (2) of Fig. 1.3., K is the total number of these frequencies (ω_i,ω_j). The root mean square error (R.M.S.E.) is then defined as

$$\text{R.M.S.E.} = \sqrt{\text{M.S.E.}}.$$

For example, when the optimum window is used with M = 5, the M.S.E. is calculated using the values given in Table 3.2 and Table 3.4, i.e.

$$\text{M.S.E.} = \frac{1}{154}\{(.914-1.279)^2 + (.876-1.203)^2 + \ldots + (.022-.007)^2 + (.805-1.071)^2 + \ldots + (.036-.019)^2\}$$

$$\simeq .0032,$$

where the total number of frequencies considered is K = 154. Similarly, for other windows, the values of M.S.E. and R.M.S.E. are calculated for different M values. These are tabulated in Table 3.14.

From the results tabulated in Table 3.14 and the figures 3.4 - 3.9 we can draw the following conclusions:-

(i) when $M = M_1$, the shapes of the modulus of the bispectral estimates are smooth for all the windows used. However, the root mean square error for the optimum window is smaller than the R.M.S.E. obtained by other windows;

(ii) when M_2 values are used, the shapes of the graphs are not as smooth as the shape of the bispectral modulus when M_1 values are used. Though, in this case, the shapes are not smooth, the R.M.S.E. are always smaller than the corresponding R.M.S.E.'s obtained from M_1 and M_3 values;

(iii) the mean square error obtained when $M_2 = 7$ for the optimum window is the smallest.

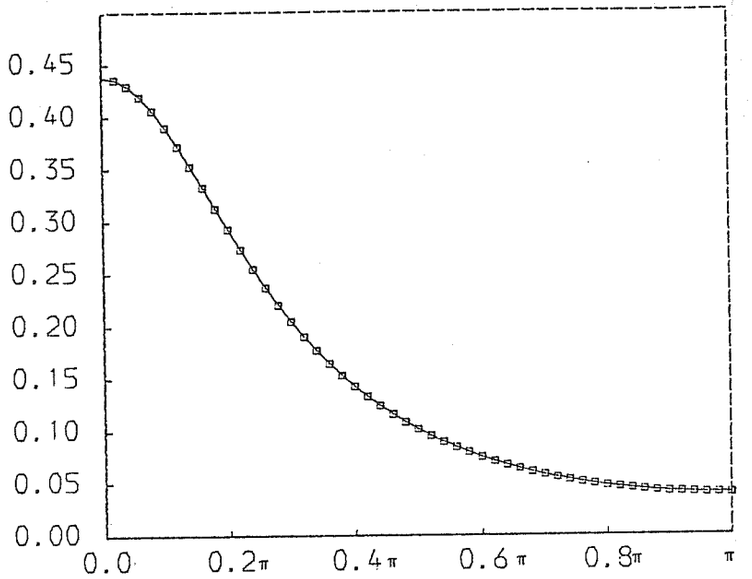


Figure 3.1: Parametric normalized spectral density function of the series $\{X_t\}$ satisfying the bilinear model $X_t + .4X_{t-1} = .4X_{t-1}e_{t-1} + e_t$

TABLE 3.2 PARAMETRIC BISPECTRAL MODULUS

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.007													
.95	.007	.008	.010											
.90	.006	.007	.009	.012	.014									
.85	.003	.006	.008	.011	.013	.015	.016							
.80	.000	.003	.006	.010	.013	.015	.016	.017	.018					
.75	.005	.002	.004	.008	.011	.014	.016	.017	.018	.019	.019			
.70	.011	.007	.003	.005	.009	.012	.015	.016	.017	.018	.019	.019	.019	
.65	.020	.014	.008	.004	.006	.010	.013	.015	.017	.018	.018	.019	.019	.019
.60	.032	.024	.016	.008	.005	.008	.011	.014	.016	.017	.018	.019	.019	
.55	.048	.037	.026	.016	.008	.006	.008	.012	.014	.016	.017	.018		
.50	.069	.055	.040	.027	.016	.008	.006	.009	.012	.014	.016			
.45	.099	.080	.060	.041	.026	.015	.008	.007	.009	.012				
.40	.141	.114	.087	.062	.041	.026	.015	.008	.007					
.35	.201	.163	.125	.090	.062	.041	.025	.014						
.30	.287	.233	.178	.130	.092	.062	.041							
.25	.408	.331	.254	.187	.133	.092								
.20	.573	.469	.361	.265	.189									
.15	.783	.650	.503	.371										
.10	1.013	.864	.678											
.05	1.203	1.071												
0.00	1.279													

TABLE 3.3 PARAMETRIC BISPECTRAL ARGUMENT

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	3.14													
.95	3.14	3.06	3.06											
.90	3.14	3.04	3.05	3.10	-3.12									
.85	-3.14	2.97	3.00	3.08	-3.13	-3.06	-3.00							
.80	0.00	2.71	2.88	3.02	3.13	-3.07	-3.01	-2.96	-2.92					
.75	.00	.92	2.51	2.89	3.07	-3.11	-3.03	-2.97	-2.93	-2.90	-2.87			
.70	0.00	.32	1.24	2.55	2.94	3.11	-3.06	-2.99	-2.94	-2.90	-2.88	-2.86	-2.85	
.65	-.00	.21	.57	1.57	2.65	3.00	-3.13	-3.03	-2.97	-2.92	-2.89	-2.87	-2.85	-2.85
.60	.00	.16	.38	.80	1.87	2.74	3.04	-3.10	-3.01	-2.95	-2.91	-2.88	-2.86	
.55	.00	.14	.30	.54	1.01	2.09	2.80	3.07	-3.08	-3.00	-2.94	-2.91		
.50	0.00	.12	.25	.42	.67	1.19	2.23	2.85	3.09	-3.07	-2.99			
.45	-.00	.10	.22	.35	.51	.77	1.33	2.31	2.87	3.10				
.40	-.00	.09	.19	.30	.42	.58	.84	1.39	2.33					
.35	.00	.08	.17	.26	.35	.47	.61	.86						
.30	.00	.07	.14	.22	.30	.38	.48							
.25	.00	.05	.11	.18	.25	.31								
.20	.00	.04	.09	.14	.19									
.15	-.00	.03	.06	.10										
.10	.00	.01	.04											
.05	.00	.01												
0.00	0.00													

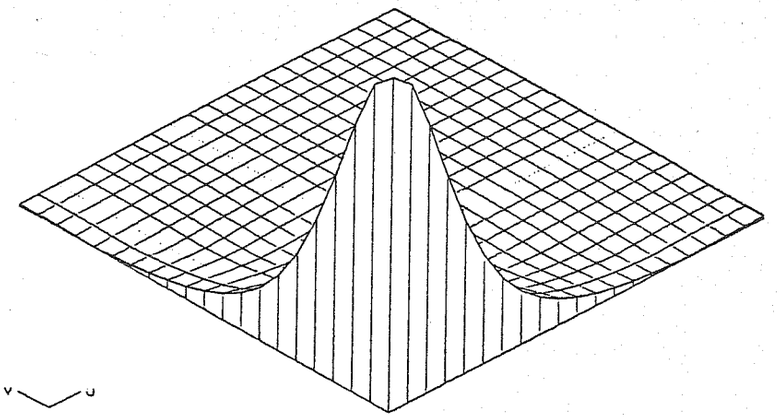


Fig. 3.2: The parametric (non-normalized) bispectral modulus

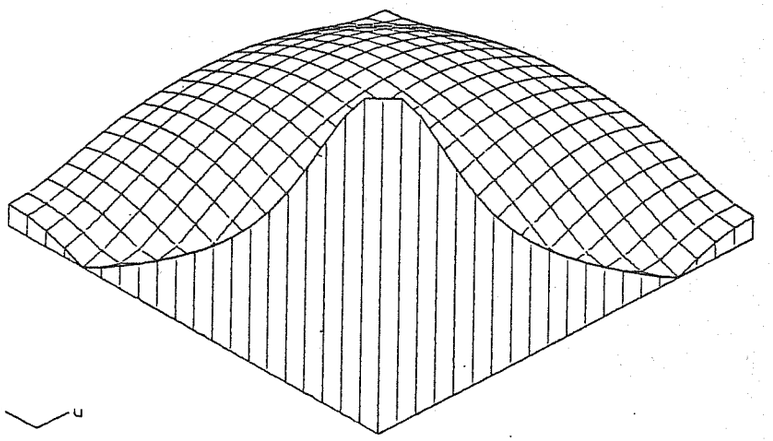


Fig. 3.3: The parametric (normalized) bispectral modulus

TABLE 3.4 ESTIMATED BISPECTRAL MODULUS
USING OPTIMUM WINDOW, M= 5

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.022													
.95	.020	.022	.022											
.90	.016	.018	.021	.022	.022									
.85	.015	.016	.018	.021	.022	.021	.020							
.80	.016	.016	.016	.019	.022	.023	.021	.020	.022					
.75	.013	.015	.016	.017	.020	.022	.023	.022	.023	.027	.031			
.70	.003	.009	.013	.015	.017	.019	.021	.022	.024	.027	.031	.034	.035	
.65	.011	.003	.005	.011	.015	.016	.017	.019	.024	.028	.031	.033	.035	.036
.60	.027	.018	.007	.005	.011	.014	.014	.014	.020	.026	.031	.032	.033	
.55	.048	.035	.022	.008	.005	.010	.011	.011	.014	.022	.029	.032		
.50	.081	.060	.041	.023	.007	.003	.007	.008	.009	.015	.024			
.45	.130	.099	.068	.042	.022	.008	.001	.004	.005	.008				
.40	.186	.150	.106	.067	.039	.021	.010	.003	.002					
.35	.242	.204	.152	.099	.059	.034	.020	.011						
.30	.302	.260	.202	.140	.086	.050	.031							
.25	.381	.327	.262	.192	.127	.079								
.20	.493	.420	.340	.259	.186									
.15	.632	.545	.442	.343										
.10	.772	.683	.564											
.05	.876	.805												
0.00	.914													

W1(AS A FRACTION OF PI)

TABLE 3.5 ESTIMATED BISPECTRAL MODULUS
USING OPTIMUM WINDOW, M= 7

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.023													
.95	.020	.022	.024											
.90	.014	.018	.022	.024	.023									
.85	.012	.015	.019	.022	.023	.022	.023							
.80	.012	.014	.018	.021	.023	.023	.022	.022	.021					
.75	.011	.014	.016	.020	.023	.023	.021	.021	.021	.022	.025			
.70	.006	.010	.014	.019	.023	.023	.021	.020	.021	.024	.028	.031	.032	
.65	.006	.003	.009	.017	.022	.021	.018	.018	.021	.027	.035	.039	.036	.032
.60	.028	.014	.002	.011	.019	.019	.014	.014	.019	.027	.037	.045	.044	
.55	.062	.039	.018	.002	.013	.016	.011	.009	.014	.022	.032	.042		
.50	.099	.073	.041	.017	.004	.010	.009	.005	.008	.015	.023			
.45	.136	.107	.069	.035	.013	.003	.004	.003	.002	.008				
.40	.190	.147	.099	.059	.027	.010	.004	.003	.002					
.35	.280	.209	.139	.090	.054	.025	.008	.005						
.30	.390	.300	.191	.123	.087	.056	.025							
.25	.475	.390	.253	.154	.111	.088								
.20	.532	.456	.318	.196	.135									
.15	.632	.532	.398	.275										
.10	.833	.680	.518											
.05	1.069	.897												
0.00	1.176													

W1(AS A FRACTION OF PI)

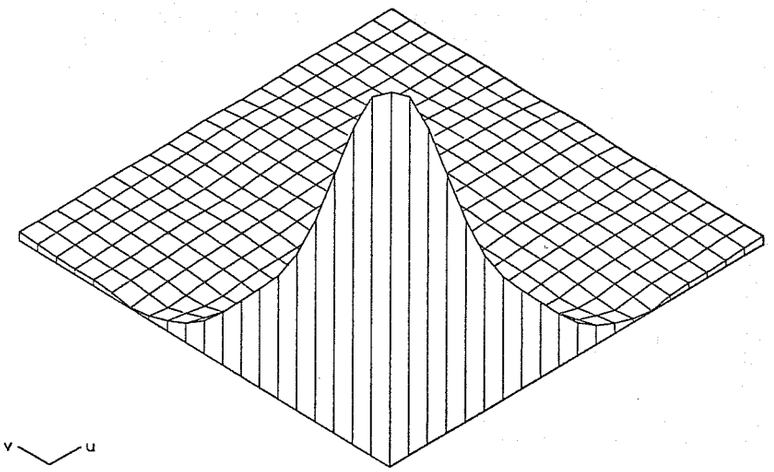


Fig. 3.4: Bispectral modulus estimate using optimum window with M = 5

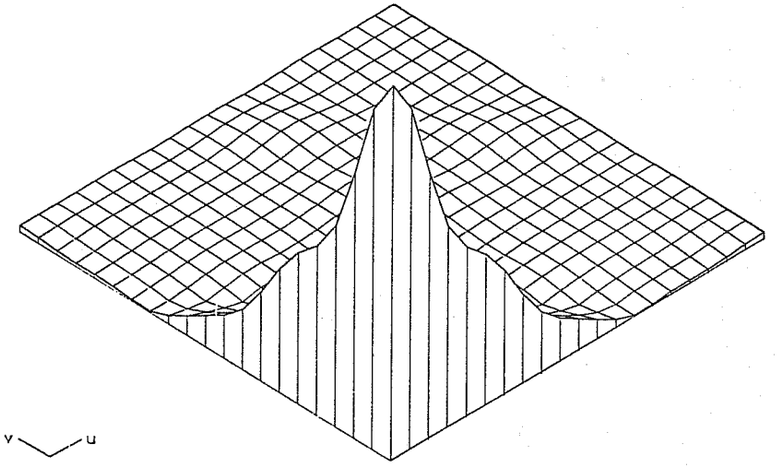


Fig. 3.5: Bispectral modulus estimate using optimum window with M = 7

TABLE 3.6 ESTIMATED BISPECTRAL MODULUS
USING DANIELL WINDOW, M=9

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.024													
.95	.020	.022	.021											
.90	.018	.019	.020	.019	.020									
.85	.019	.019	.019	.019	.020	.021	.020							
.80	.016	.018	.017	.016	.018	.020	.019	.019	.020					
.75	.013	.014	.015	.015	.016	.018	.018	.018	.020	.022	.025			
.70	.008	.011	.010	.013	.017	.018	.018	.019	.020	.023	.027	.029	.030	
.65	.005	.003	.006	.009	.014	.016	.016	.017	.019	.024	.028	.030	.030	.030
.60	.022	.012	.003	.005	.011	.013	.012	.013	.016	.022	.028	.031	.031	
.55	.047	.032	.018	.007	.005	.010	.010	.009	.012	.018	.025	.030		
.50	.077	.057	.037	.021	.009	.005	.007	.006	.007	.013	.020			
.45	.115	.088	.060	.036	.018	.006	.004	.002	.002	.005				
.40	.170	.131	.091	.060	.033	.014	.006	.002	.001					
.35	.230	.186	.130	.087	.057	.031	.013	.007						
.30	.291	.244	.180	.120	.079	.053	.030							
.25	.363	.308	.236	.166	.111	.074								
.20	.449	.386	.304	.225	.163									
.15	.565	.491	.398	.299										
.10	.712	.620	.511											
.05	.831	.751												
0.00	.872													

W1(AS A FRACTION OF PI)

TABLE 3.7 ESTIMATED BISPECTRAL MODULUS
USING DANIELL WINDOW, M= 13

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.025													
.95	.022	.023	.020											
.90	.017	.019	.019	.020	.020									
.85	.012	.017	.017	.020	.020	.020	.020							
.80	.012	.016	.016	.018	.024	.022	.021	.020	.020					
.75	.016	.017	.016	.017	.021	.023	.021	.019	.018	.018	.021			
.70	.010	.012	.013	.018	.023	.021	.019	.017	.019	.021	.024	.028	.027	
.65	.003	.009	.007	.016	.022	.020	.015	.014	.015	.025	.031	.036	.032	.027
.60	.024	.015	.003	.012	.021	.017	.013	.014	.015	.022	.035	.043	.042	
.55	.055	.035	.019	.003	.015	.013	.008	.008	.013	.019	.030	.040		
.50	.090	.066	.039	.016	.008	.010	.008	.003	.009	.012	.021			
.45	.125	.096	.063	.032	.010	.001	.005	.006	.001	.008				
.40	.179	.136	.090	.054	.025	.009	.003	.003	.006					
.35	.250	.186	.122	.079	.049	.021	.008	.008						
.30	.372	.271	.166	.114	.080	.051	.020							
.25	.462	.375	.224	.136	.105	.077								
.20	.513	.436	.287	.165	.125									
.15	.574	.486	.338	.218										
.10	.759	.605	.453											
.05	1.018	.821												
0.00	1.160													

W1(AS A FRACTION OF PI)

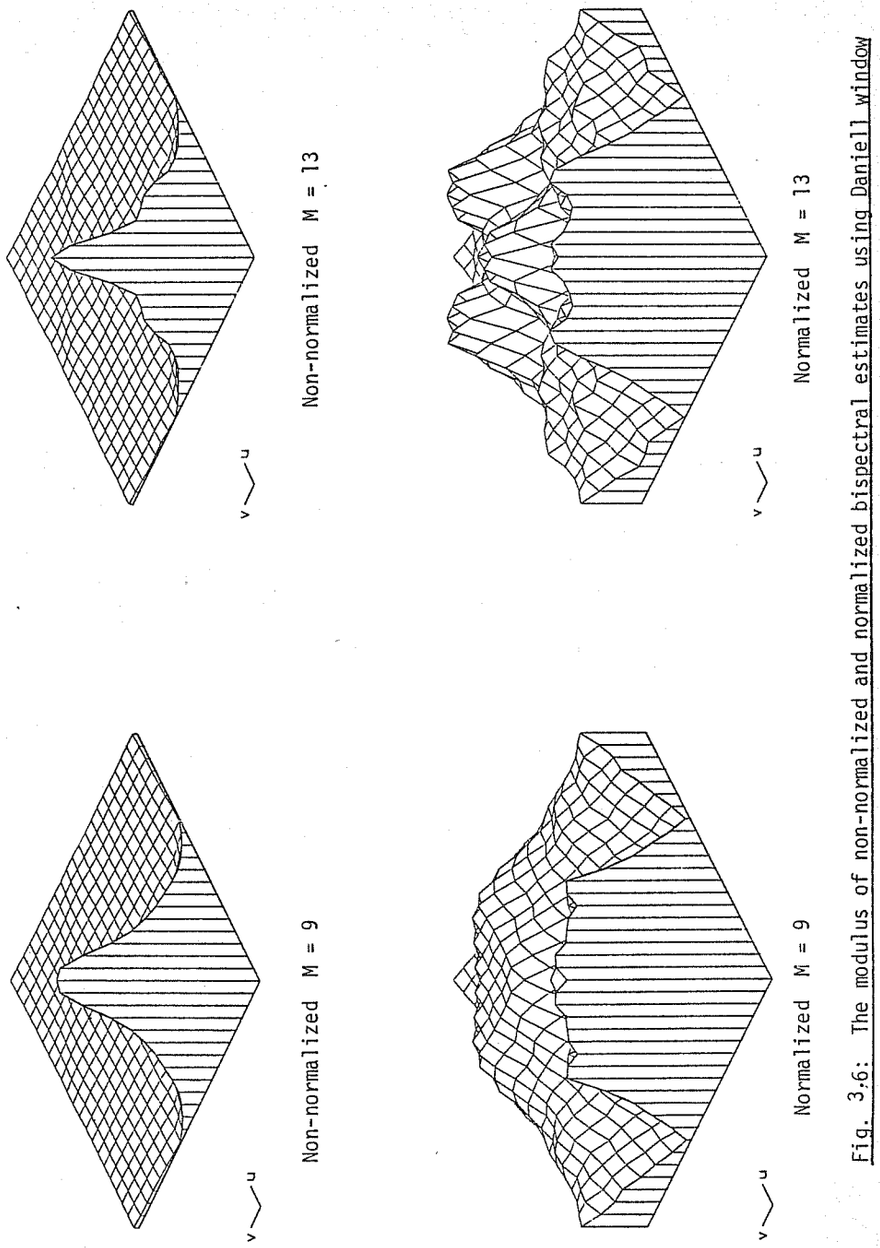


Fig. 3.6: The modulus of non-normalized and normalized bispectral estimates using Daniell window with M = 9, M = 13

TABLE 3.8 ESTIMATED BISPECTRAL MODULUS
USING TUKEY WINDOW, M= 12

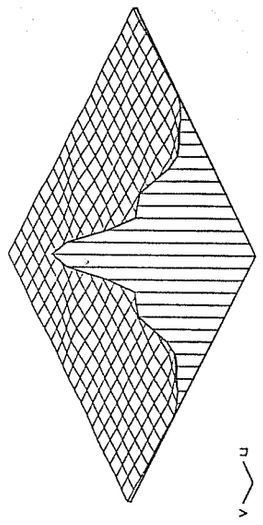
W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.022													
.95	.020	.020	.019											
.90	.018	.019	.018	.018	.018									
.85	.016	.016	.016	.017	.019	.019	.018							
.80	.015	.015	.014	.016	.018	.019	.019	.018	.018					
.75	.013	.013	.013	.014	.017	.019	.018	.018	.018	.020	.023			
.70	.007	.010	.010	.013	.016	.017	.017	.017	.018	.021	.024	.026	.027	
.65	.004	.003	.005	.011	.015	.016	.015	.015	.017	.022	.027	.029	.029	.027
.60	.023	.013	.004	.006	.012	.014	.012	.012	.015	.021	.028	.032	.032	
.55	.047	.033	.019	.007	.008	.010	.009	.008	.011	.017	.024	.031		
.50	.078	.058	.038	.019	.007	.007	.006	.004	.006	.011	.018			
.45	.115	.090	.062	.036	.017	.006	.004	.003	.002	.006				
.40	.164	.130	.092	.059	.032	.014	.006	.003	.003					
.35	.228	.181	.129	.086	.054	.030	.014	.006						
.30	.304	.246	.176	.119	.079	.051	.029							
.25	.380	.318	.231	.156	.110	.077								
.20	.457	.391	.296	.208	.149									
.15	.562	.480	.376	.280										
.10	.707	.604	.482											
.05	.850	.748												
0.00	.913													

W1(AS A FRACTION OF PI)

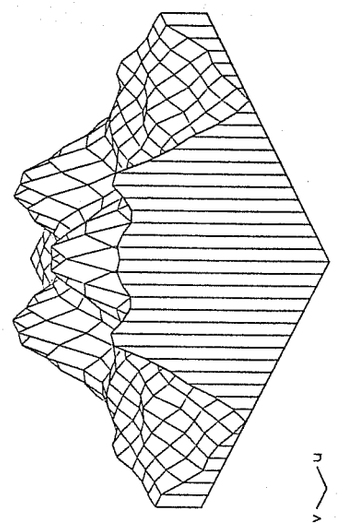
TABLE 3.9 ESTIMATED BISPECTRAL MODULUS
USING TUKEY WINDOW, M= 16

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.022													
.95	.021	.021	.019											
.90	.015	.019	.019	.018	.017									
.85	.011	.015	.016	.019	.019	.018	.020							
.80	.012	.014	.014	.018	.023	.020	.019	.019	.017					
.75	.012	.014	.014	.015	.022	.022	.019	.018	.017	.016	.020			
.70	.007	.012	.015	.017	.020	.020	.017	.016	.017	.018	.023	.026	.026	
.65	.004	.006	.009	.018	.022	.019	.014	.013	.016	.021	.029	.034	.030	.026
.60	.022	.016	.005	.011	.020	.017	.013	.012	.013	.021	.033	.043	.039	
.55	.053	.036	.021	.004	.015	.015	.008	.009	.012	.017	.027	.040		
.50	.092	.066	.039	.013	.008	.010	.006	.003	.007	.012	.019			
.45	.124	.099	.063	.032	.010	.005	.005	.006	.001	.006				
.40	.169	.131	.088	.053	.028	.011	.004	.004	.006					
.35	.250	.186	.123	.078	.048	.023	.008	.005						
.30	.361	.266	.168	.110	.077	.052	.022							
.25	.472	.361	.216	.139	.105	.080								
.20	.519	.436	.272	.164	.128									
.15	.562	.474	.335	.219										
.10	.746	.585	.429											
.05	1.000	.813												
0.00	1.113													

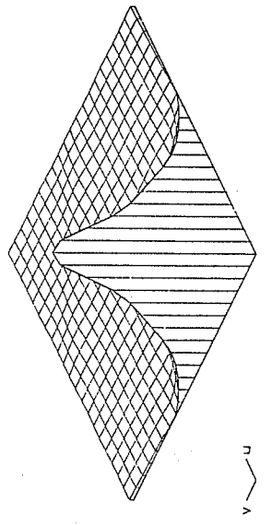
W1(AS A FRACTION OF PI)



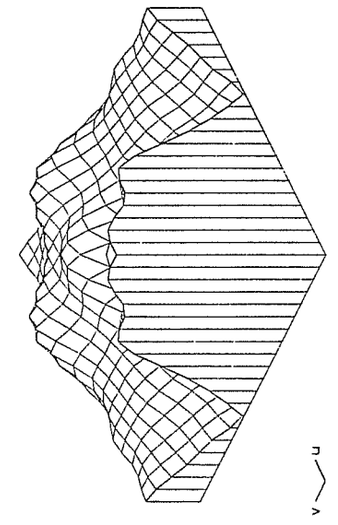
Non-normalized M = 16



Normalized M = 16



Non-normalized M = 12



Normalized M = 12

Fig. 3.7: The modulus of non-normalized and normalized bispectral estimates using the Tukey window with M = 12, M = 16

TABLE 3.10 ESTIMATED BISPECTRAL MODULUS
USING PARZEN WINDOW, M= 15

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.023													
.95	.022	.022	.020											
.90	.020	.020	.020	.019	.019									
.85	.018	.018	.018	.018	.019	.019	.019							
.80	.016	.016	.016	.017	.018	.019	.019	.019	.020					
.75	.013	.014	.014	.015	.017	.018	.018	.019	.020	.022	.024			
.70	.006	.009	.011	.012	.015	.016	.017	.018	.020	.022	.025	.027	.028	
.65	.005	.001	.005	.009	.013	.014	.015	.016	.018	.023	.026	.029	.029	.028
.60	.022	.013	.005	.004	.009	.012	.012	.013	.016	.021	.026	.030	.030	
.55	.046	.032	.019	.008	.006	.008	.009	.009	.012	.017	.023	.028		
.50	.076	.057	.038	.021	.010	.006	.005	.005	.007	.012	.018			
.45	.113	.089	.062	.039	.021	.009	.005	.003	.003	.007				
.40	.160	.129	.093	.062	.037	.019	.009	.004	.002					
.35	.219	.179	.132	.091	.058	.034	.018	.009						
.30	.287	.239	.180	.127	.086	.056	.033							
.25	.362	.308	.237	.171	.120	.084								
.20	.448	.386	.306	.227	.165									
.15	.555	.481	.390	.301										
.10	.680	.596	.491											
.05	.788	.712												
0.00	.832													

W1(AS A FRACTION OF PI)

TABLE 3.11 ESTIMATED BISPECTRAL MODULUS
USING PARZEN WINDOW, M= 22

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.024													
.95	.022	.022	.020											
.90	.017	.020	.020	.019	.018									
.85	.013	.017	.018	.021	.021	.019	.021							
.80	.013	.015	.015	.019	.024	.021	.020	.021	.019					
.75	.013	.016	.015	.016	.023	.023	.020	.020	.019	.018	.021			
.70	.007	.013	.016	.017	.021	.021	.018	.018	.018	.020	.024	.027	.027	
.65	.003	.006	.010	.018	.022	.019	.015	.014	.017	.023	.031	.035	.031	.027
.60	.021	.016	.005	.011	.020	.017	.013	.012	.015	.022	.034	.044	.040	
.55	.053	.036	.020	.003	.015	.013	.008	.010	.013	.018	.029	.041		
.50	.091	.067	.039	.013	.008	.010	.005	.003	.008	.013	.020			
.45	.125	.098	.063	.032	.010	.003	.005	.005	.001	.007				
.40	.171	.132	.089	.054	.029	.011	.004	.004	.006					
.35	.251	.187	.124	.079	.049	.024	.008	.006						
.30	.364	.267	.170	.112	.079	.053	.023							
.25	.479	.366	.219	.142	.108	.082								
.20	.528	.443	.277	.168	.131									
.15	.569	.482	.340	.224										
.10	.753	.592	.435											
.05	1.009	.820												
0.00	1.122													

W1(AS A FRACTION OF PI)

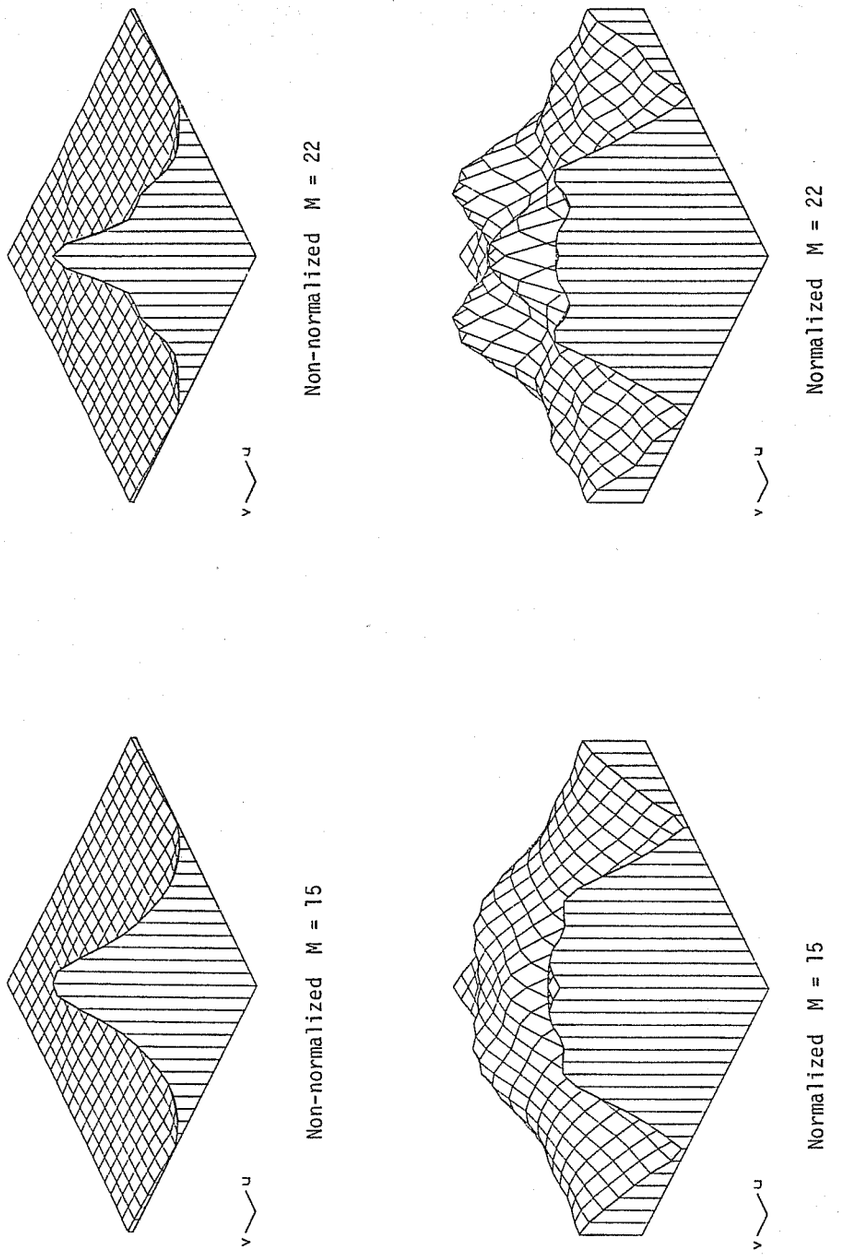


Fig. 3.8: The modulus of non-normalized and normalized bispectral estimates using the Parzen window with M = 15, M = 22.

TABLE 3.12 ESTIMATED BISPECTRAL MODULUS
USING BARTLETT PRIESTLEY WINDOW, M= 8

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.024													
.95	.022	.022	.020											
.90	.019	.020	.019	.019	.020									
.85	.016	.017	.017	.019	.020	.021	.020							
.80	.016	.016	.016	.017	.020	.021	.020	.019	.019					
.75	.014	.015	.014	.016	.019	.021	.019	.019	.019	.021	.024			
.70	.009	.012	.012	.015	.019	.020	.018	.017	.019	.022	.026	.028	.029	
.65	.002	.005	.007	.012	.017	.018	.016	.016	.018	.024	.030	.032	.031	.029
.60	.022	.012	.003	.007	.015	.015	.013	.013	.016	.023	.031	.036	.036	
.55	.049	.033	.018	.005	.010	.012	.009	.008	.012	.019	.027	.034		
.50	.081	.060	.038	.018	.005	.007	.007	.004	.007	.013	.020			
.45	.118	.092	.062	.035	.015	.004	.003	.003	.002	.007				
.40	.169	.132	.091	.057	.030	.012	.004	.003	.002					
.35	.241	.187	.129	.085	.052	.026	.010	.005						
.30	.325	.259	.177	.116	.079	.050	.026							
.25	.401	.334	.233	.152	.106	.077								
.20	.471	.403	.297	.199	.140									
.15	.574	.487	.376	.273										
.10	.737	.619	.485											
.05	.909	.785												
0.00	.984													

W1(AS A FRACTION OF PI)

TABLE 3.13 ESTIMATED BISPECTRAL MODULUS
USING BARTLETT PRIESTLEY WINDOW, M= 12

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.023													
.95	.022	.023	.021											
.90	.016	.020	.021	.020	.016									
.85	.009	.017	.019	.023	.021	.018	.022							
.80	.010	.016	.016	.020	.028	.022	.021	.023	.019					
.75	.011	.018	.018	.016	.026	.025	.020	.021	.018	.015	.020			
.70	.005	.015	.022	.020	.024	.021	.017	.018	.018	.018	.023	.028	.027	
.65	.004	.006	.013	.023	.026	.020	.014	.013	.017	.022	.032	.038	.031	.025
.60	.019	.020	.008	.015	.025	.018	.014	.013	.014	.021	.037	.051	.045	
.55	.057	.040	.025	.003	.019	.013	.007	.013	.013	.017	.030	.048		
.50	.102	.073	.040	.008	.014	.012	.005	.004	.009	.013	.020			
.45	.131	.104	.064	.031	.005	.005	.007	.007	.002	.007				
.40	.175	.133	.085	.051	.031	.013	.006	.005	.009					
.35	.262	.190	.120	.072	.047	.022	.011	.011						
.30	.397	.273	.164	.109	.077	.057	.021							
.25	.555	.393	.203	.135	.109	.086								
.20	.589	.486	.257	.144	.130									
.15	.566	.480	.311	.180										
.10	.773	.569	.388											
.05	1.105	.860												
0.00	1.249													

W1(AS A FRACTION OF PI)

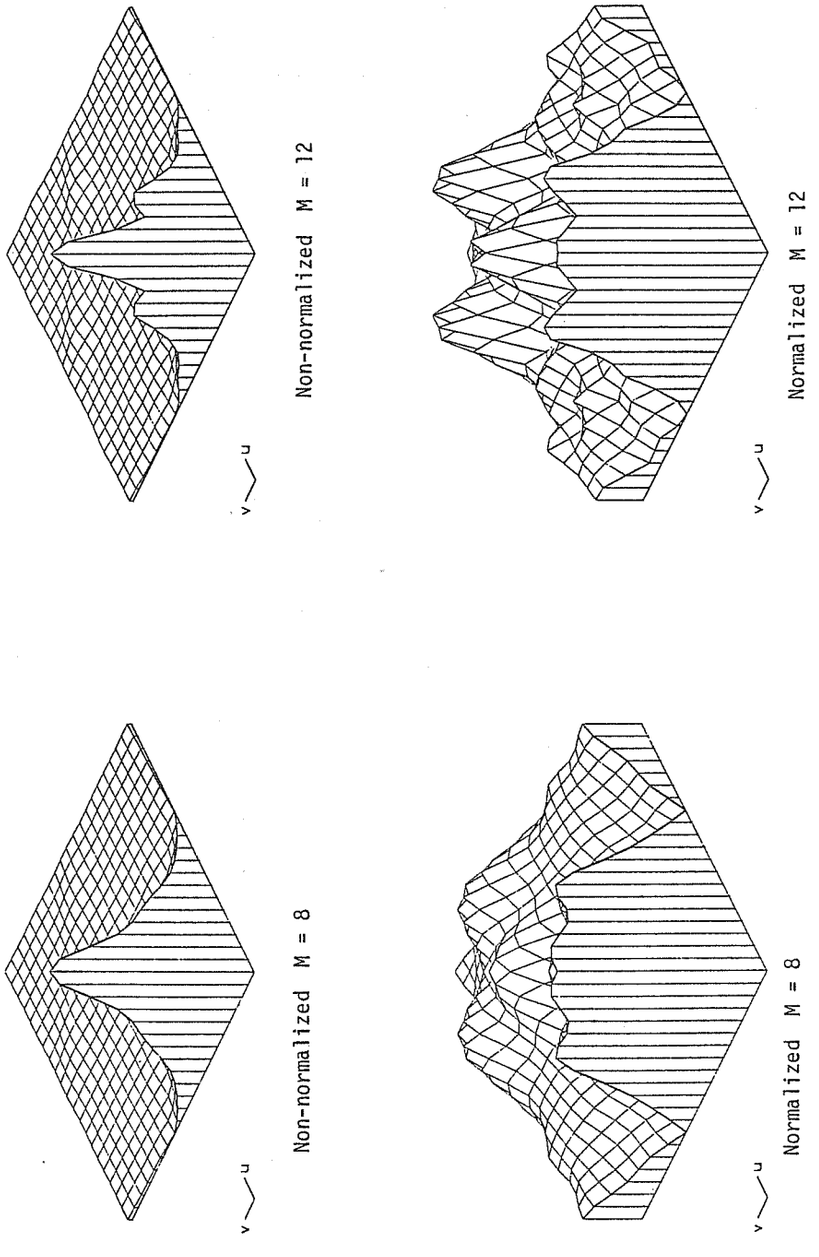


Fig. 3.9: The modulus of non-normalized and normalized bispectral estimates using the Bartlett-Priestley window with M = 8, M = 12.

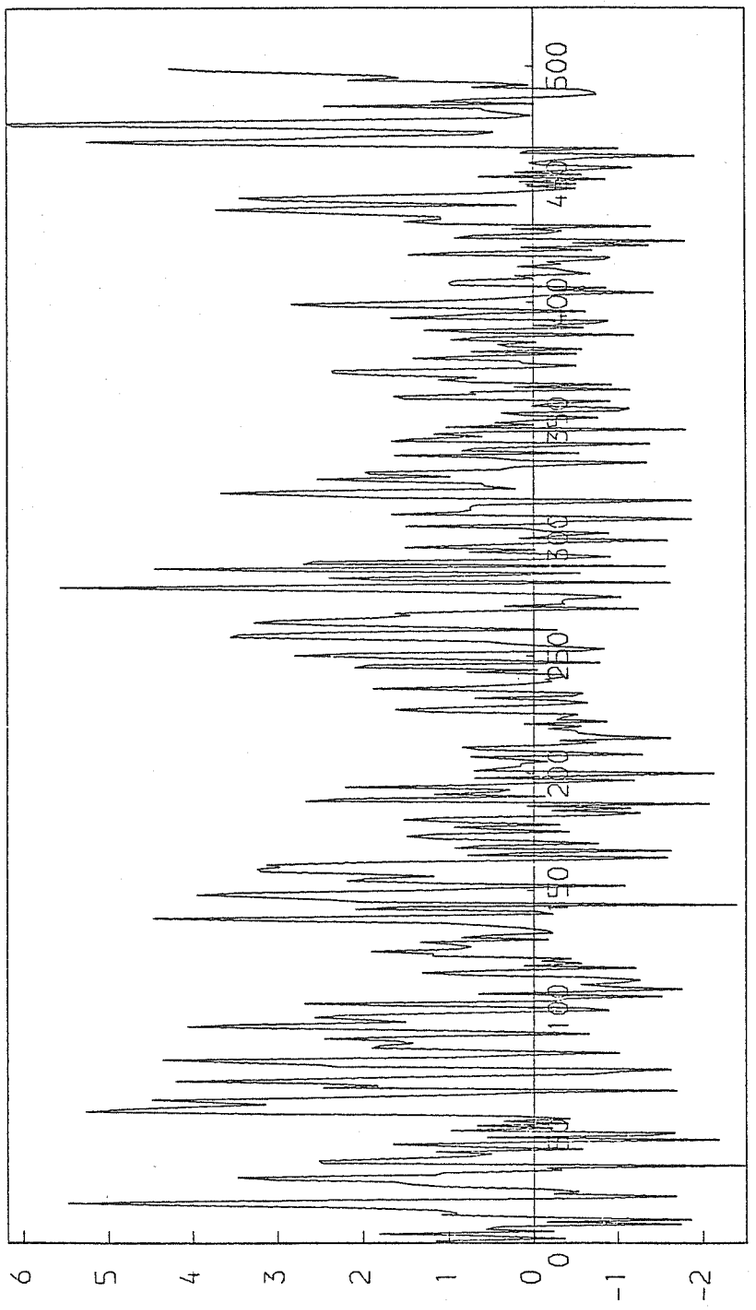


Fig. 3.10: Simulated series $\{x_t, t = 1, \ldots, 500\}$ from the BL(1,0;1,1), $X_t + .4X_{t-1} = .4X_{t-1}e_{t-1} + e_t$

TABLE 3.14: The root mean square error values of the bispectral estimates

Window	M_1	R.M.S.E.	M_2	R.M.S.E.	M_3	R.M.S.E.
Optimum	5	.0566	7	.0421	8	.0450
Daniell	9	.0682	13	.0556	15	.0574
Tukey	12	.0678	16	.0590	18	.0609
Parzen	15	.0745	22	.0574	26	.0578
B.P.	8	.0608	12	.0588	14	.0644

3.3 Bispectral Analysis of Some Time Series data

We have pointed out earlier that when the time series is non-linear, the second order spectral analysis cannot necessarily show the non-linear effects contained in the time series. In view of this one has to perform higher order spectral analysis on the time series. Therefore, it is interesting to reconsider the well known time series data from the higher order spectral analysis point of view. The series considered are

(i) Wölfer sunspot numbers;

(ii) The Canadian lynx data;

(iii) The number of people registered as unemployed in W. Germany.

The three time series (mean deleted) are strictly bounded, so definitely the third sample moments for these series does exist. The third order sample moments $C(\tau_1,\tau_2)$ do decrease as $\tau_1 \to \pm\infty$, $\tau_2 \to \pm\infty$ hence the absolute summability condition is also satisfied.

(i) Sunspot Numbers (annual data)

The data we consider is the Wölfer sunspot for the years 1700-1955 (Waldmeirer, 1961), giving 256 observations. These numbers are

given in Appendix D and plotted in Figure 3.11. This series has a certain historic interest for statisticians, see, e.g. Yule (1927), Bartlett (1950), Whittle (1954), Brillinger and Rosenblatt (1967b). It is believed by many scientists that this series has an eleven year cycle. We find from the data

Sample mean = $\bar{X}$ = 44.782,

Sample variance = S^2 = 1248.62, and

Sample third order central moment = $\hat{C}(0,0) = -.1036$,

$$\hat{C}(0,0)/S^3 = .235 \times 10^{-7}$$

The bispectral density function of the series is estimated using the optimum window with M = 20. The modulus and the arguments of the bispectral estimate are calculated for several frequencies (ω_1,ω_2) as shown in Tables 3.15, 3.16, and Figure 3.12. In Table 3.17 and Figure 3.13, however, we have given the values of the modulus for frequencies ω_1,ω_2 = 0.0 (0.01π) 0.27π. There is some evidence of a hill (ridge) along $\omega_2 = 0$; however, this ridge may simply be a manifestation of the increased sampling variability (see Brillinger and Rosenblatt, 1967b, p.217). In fact, the variance of the bispectral density on the boundary line $\omega_2 = 0$ is greater than the variance inside the boundary (see Rosenblatt and Van Ness, 1966). Turning to Table 3.16 specifically, we now note the possibility of a peak (hill) in the neighbourhood of

$$\omega_1 = \omega_2 = 0.18\pi = \frac{2\pi}{11.1}$$

which corresponds to the periodicity 11 years (approximately). This confirms the general belief drawn on the basis of second order spectral analysis. Also, the non-normalized and normalized bispectral density functions are estimated using the Daniell window with M = 24. The results are shown in Figures 3.14 and 3.15.

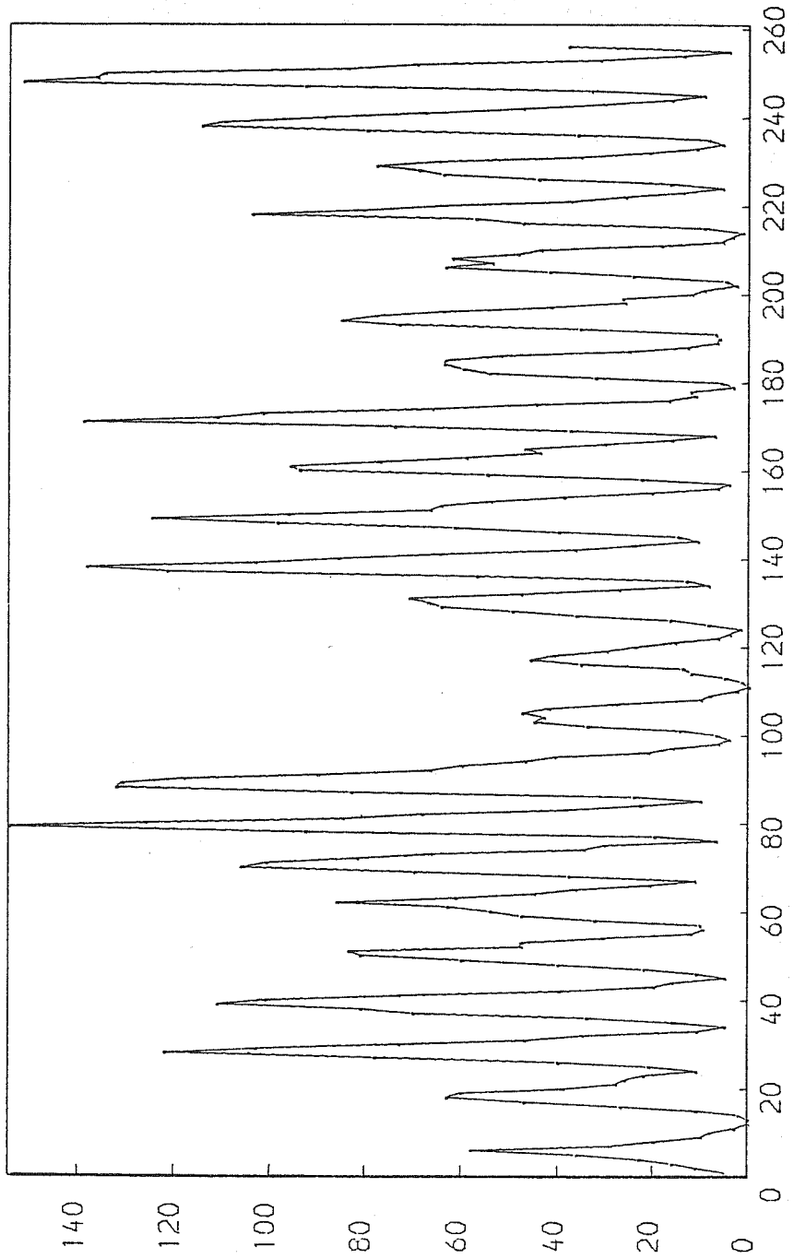


Fig. 3.11: Sunspot Numbers 1700-1955

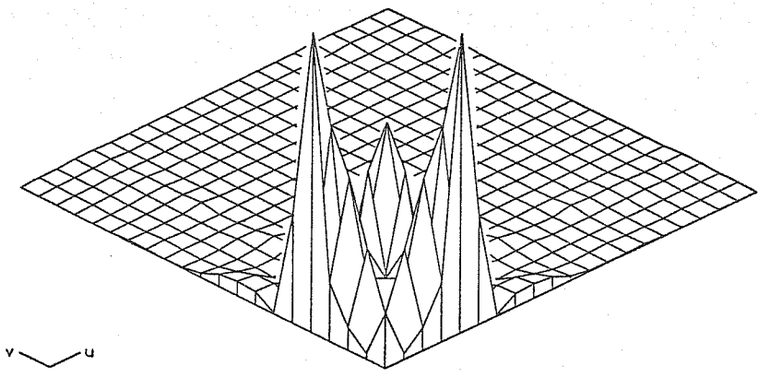


Fig. 3.12: Estimated bispectral modulus of sunspot data using the optimum window, with M = 20, at frequencies (ω_1,ω_2) = $0.0(0.05\pi)\pi$

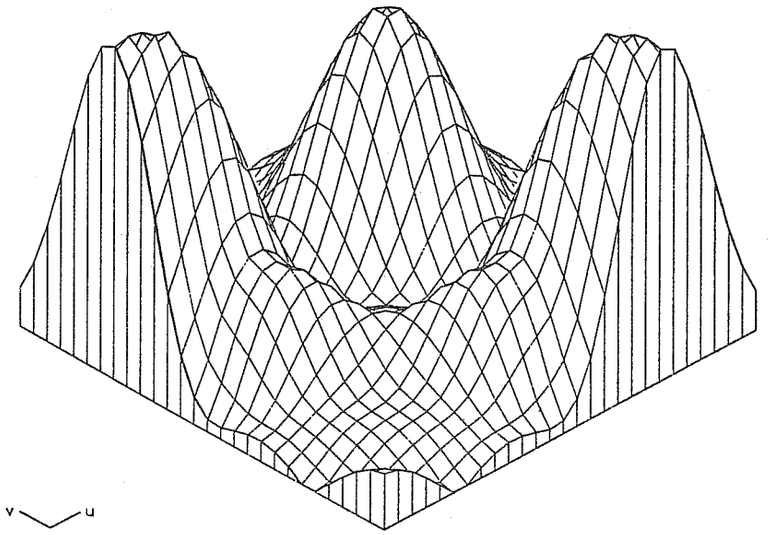


Fig. 3.13: Estimated bispectral modulus of sunspot data at the frequencies $(\omega_1,\omega_2 = 0.0(0.01\pi).2\pi)$

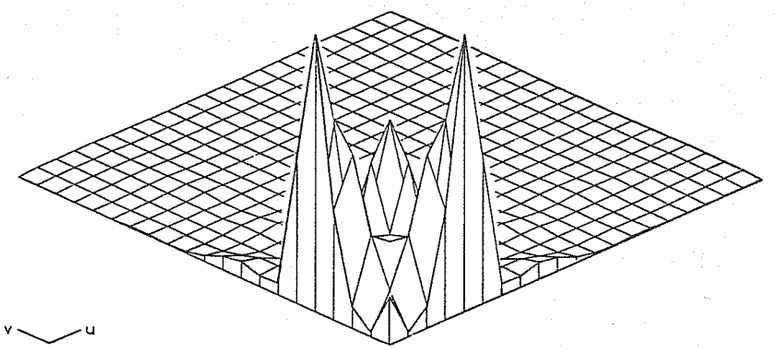


Fig. 3.14: The modulus of the non-normalized bispectral estimate using Daniell window with M = 24

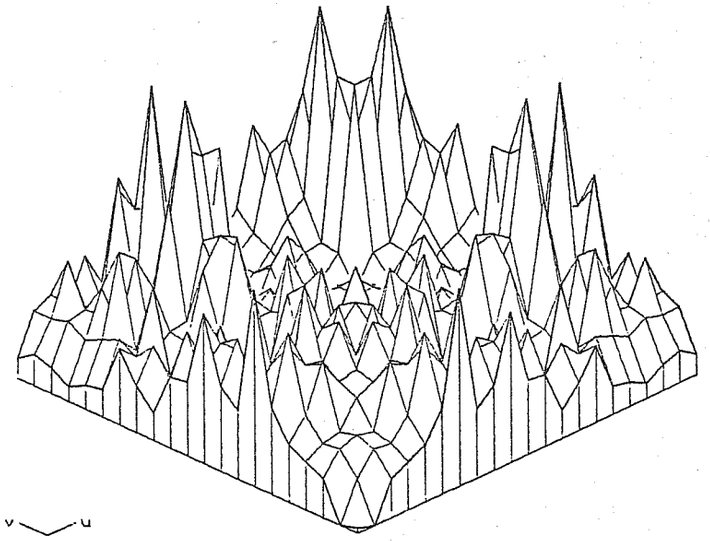


Fig. 3.15 The modulus of the normalized bispectral estimate using Daniell window with M = 24

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	68.0													
.95	58.1	141.3	142.1											
.90	85.0	110.5	108.9	200.5	159.9									
.85	14.4	116.5	102.8	251.9	275.5	38.7	36.7							
.80	27.5	2.6	48.0	193.0	264.2	27.8	30.8	37.9	21.3					
.75	71.7	40.8	56.4	147.7	266.4	72.9	53.6	55.4	18.8	16.4	11.9			
.70	193.5	69.6	63.3	59.9	415.2	127.5	31.6	36.9	36.7	2.1	19.7	21.0	31.5	
.65	156.2	218.3	40.3	101.5	31.3	118.7	44.3	54.9	17.2	35.5	19.2	1.1	3.1	9.1
.60	83.2	86.8	56.7	107.6	126.8	40.6	21.9	58.3	13.5	8.3	16.9	9.8	12.9	
.55	186.3	27.9	79.1	397.7	318.6	37.1	40.3	79.1	60.6	22.1	7.7	13.0		
.50	576.6	220.5	33.1	258.5	797.1	153.4	38.8	41.2	68.7	59.3	34.4			
.45	1650.7	819.8	164.3	199.9	793.2	587.1	46.0	26.8	34.1	163.3				
.40	1896.1	1566.5	59.9	669.0	695.2	687.4	111.4	100.4	22.1					
.35	1745.7	1624.0	395.0	811.8	1712.6	216.1	90.5	308.1						
.30	168.8	687.0	694.8	1001.2	1323.1	331.4	159.7							
.25	14051.9	3595.7	2112.5	4248.3	7835.3	2983.4								
.20	39184.3	25856.1	2706.5	17466.4	22771.9									
.15	7744.8	20044.4	5914.1	5148.5										
.10	2996.7	11986.0	7076.6											
.05	769.0	4673.0												
0.00	7163.9													

W1(AS A FRACTION OF PI)

TABLE 3.15: The modulus of the bispectral estimate of the sunspot data using the optimum window, with M = 20

W2 \ W1	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
.90	-.000													
.85	-.000	.420	.890											
.80	.000	-1.147	.889	-.364	-.108									
.75	.000	1.249	-.922	-.376	-.079	-.228	-.877							
.70	-.000	-.073	-.613	-.650	.082	.069	-.580	-.441	-1.100					
.65	.000	.095	-1.335	1.304	.373	.887	.724	-.278	-1.218	.136	-1.310			
.60	.000	-.558	-1.304	-.743	.005	-.983	1.056	-.665	.508	1.065	-.116	1.520	-1.130	
.65	.000	.095	-1.335	1.304	.373	.887	.724	-.278	-1.218	.136	-1.310	-1.421	.012	1.277
.60	.000	-.558	-1.304	-.743	.005	-.983	1.056	-.665	.508	1.065	-.116	1.520	-1.130	
.55	.000	.974	.462	-.501	-.692	.575	-1.060	-.149	-.252	-.150	-1.143	.429		
.50	-.000	.760	-.758	-.173	-.563	-.633	-.369	1.234	1.413	.708	-.698			
.45	.000	-.497	-.940	.080	-1.147	-.818	1.488	1.404	-.106	.909				
.40	.000	-.378	1.522	-.162	-1.182	-1.385	-.667	-.136	-.863					
.35	-.000	-.258	-.905	-.437	-.617	-.666	-.937	.235						
.30	.000	.279	-.341	1.529	-1.462	-.999	-.738							
.25	.000	.186	.713	-.921	-1.267	-1.321								
.20	-.000	-.481	-.030	-1.002	-1.276									
.15	-.000	-.000	-.969	-.859										
.10	.000	1.335	.845											
.05	.000	.855												
0.00	0.000													

TABLE 3.16: The argument of the bispectral estimate of the sunspot data using the optimum window, with M = 20

USING OPTIMUM WINDOW

W1 \ W2 (FRACTION OF PI)	0.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.11	.12	.13	.14	.15	.16	.17	.18	.19	.20	.21	.22	.23	.24	.25	.26	.27
0.00	7164.																											
.01	6777.	6046.																										
.02	5701.	4747.	3423.																									
.03	4161.	3156.	2167.	2037.																								
.04	2436.	1640.	1769.	2780.	3738.																							
.05	769.	1046.	2362.	3544.	4325.	4673.																						
.06	685.	1220.	3077.	4045.	4594.	4811.	5032.																					
.07	1850.	2649.	3605.	4357.	4835.	5199.	5794.	6969.																				
.08	2669.	3202.	3961.	4752.	5557.	6457.	7535.	8777.	10027.																			
.09	3071.	3426.	4346.	5723.	7329.	8884.	10324.	10860.	10987.	10464.																		
.10	2997.	3452.	5211.	7676.	10131.	11986.	12844.	12573.	11299.	9324.	7077.																	
.11	2525.	3696.	6801.	10261.	13122.	14717.	14736.	13273.	10752.	7800.	5191.	3912.																
.12	2017.	4443.	8448.	12537.	15199.	16101.	15181.	12810.	9656.	6561.	4531.	4196.	4874.															
.13	2163.	5523.	10121.	13829.	15908.	16129.	14697.	12129.	9139.	6554.	5115.	4755.	4566.	3987.														
.14	3837.	7154.	11671.	14895.	16563.	16452.	15354.	13030.	10233.	7648.	5823.	4755.	3942.	3085.	2408.													
.15	7745.	11335.	15195.	18103.	19861.	20044.	18599.	15765.	12174.	8667.	5914.	4105.	2961.	2375.	3002.	5149.												
.16	14006.	17918.	21653.	24565.	26051.	25531.	22847.	18499.	13469.	8817.	5291.	3126.	2187.	2664.	4811.	8397.	12960.											
.17	21688.	25912.	29403.	31923.	32964.	30259.	25678.	19599.	13324.	7998.	4288.	2382.	2263.	3977.	7346.	12129.	17434.	22733.										
.18	29842.	33556.	36148.	37434.	36201.	32099.	25768.	18530.	11799.	6600.	3414.	2307.	3094.	5668.	9936.	15451.	21208.	26181.	28961.									
.19	36152.	38510.	39679.	39238.	36118.	30400.	23123.	15736.	9500.	5163.	2945.	2632.	4013.	7140.	11816.	17424.	22934.	27115.	29015.	28326.								
.20	39184.	39567.	39121.	36892.	32112.	25854.	18718.	12176.	7157.	4000.	2707.	2920.	4626.	7867.	12368.	17466.	22159.	25404.	26529.	25475.	22772.							
.21	38540.	36090.	34824.	31283.	26094.	19944.	13872.	8768.	5131.	3092.	2467.	2995.	4694.	7580.	11416.	15607.	19301.	21698.	22371.	21392.	19251.	16556.						
.22	34054.	31274.	28137.	24093.	19232.	14170.	9608.	5988.	3551.	2368.	2207.	2862.	4243.	6625.	9323.	12497.	15280.	17113.	17695.	17111.	15691.	13832.	11844.					
.23	27590.	24208.	20784.	17040.	13160.	9485.	6317.	3877.	2346.	1793.	2033.	2697.	3602.	4918.	6619.	9055.	11210.	12759.	13469.	13149.	12568.	11349.	9876.	8263.				
.24	20698.	17228.	14188.	11271.	8530.	6061.	3914.	2346.	1524.	1502.	2034.	2663.	3133.	3637.	4441.	6204.	7913.	9329.	10181.	10381.	9993.	9138.	7962.	6608.	5201.			
.25	14052.	11362.	9063.	7038.	5247.	3590.	2233.	1357.	1092.	1453.	2113.	2643.	2819.	2799.	3165.	4248.	5673.	6963.	7815.	8100.	7835.	7129.	6134.	5013.	3923.	2983.		
.26	8929.	6967.	5418.	4127.	2958.	1931.	1194.	851.	968.	1477.	2053.	2405.	2405.	2198.	2314.	3138.	4315.	5384.	6042.	6237.	5943.	5308.	4494.	3449.	2881.	2228.	1667.	
.27	5223.	3901.	2920.	2124.	1447.	951.	693.	693.	976.	1397.	1758.	1972.	1853.	1693.	1844.	2528.	3431.	4195.	4608.	4614.	4284.	3758.	3166.	2598.	2083.	1610.	1160.	

TABLE 3.17: The modulus of the bispectral estimate of the sunspot data at frequencies $(\omega_1,\omega_2 = 0.0(0.01\pi)0.27\pi)$

(ii) Canadian Lynx Data

The second series we consider is the annual number of Canadian lynx trapped in the Mackenzie River district of North-west Canada for the years 1821-1934, giving a total of 114 observations. These numbers are given in Appendix D, and plotted in Figure 3.16. There is an obvious cycle of approximately ten years with varying amplitude and it is non-linear (and non-Gaussian). We find that

$$\bar{X} = 1537.89$$

$$S^2 = 2{,}492{,}516.0$$

$$\hat{C}(0,0) = .13593 \times 10^9. \quad \hat{C}(0,0)/S^3 = .0345$$

Campbell and Walker (1977) have analysed the logarithm (to the base 10) of this data and observe that the logarithm of the series has a harmonic component corresponding to the periodicity 9.63 years. The graph of the transformed data $(\log_{10}X_t)$ is given in Figure 3.17. The sample measures of skewness (b_1) and kurtosis (b_2) have been calculated by Campbell and Walker, and these are found to be -.36 and 2.27 respectively. On the basis of these measures they have concluded that the logarithm of the Canadian lynx data is nearly Gaussian.

The bispectral density function is estimated using the optimal window, with M = 16, and at frequencies $\omega_1,\omega_2 = 0.0(0.05\pi)\pi$. The results are shown in Tables 3.18 and 3.19 and Figure 3.18. In Table 3.20 we have given the values of the modulus at the frequencies $\omega_1,\omega_2 = 0.0(0.01\pi)\ 0.26\pi$ (see also Figure 3.19). From this table, it is clear that there is a dominant peak in the neighbourhood of the frequency $\omega_1 = \omega_2 = 0.20\pi = \frac{2\pi}{10}$ which corresponds to the periodicity of ten years.

Also, the non-normalized and normalized bispectral density functions are estimated using the Daniell window with M = 20, and at the same frequencies as before. The results are shown in Figs. 3.20 and 3.21.

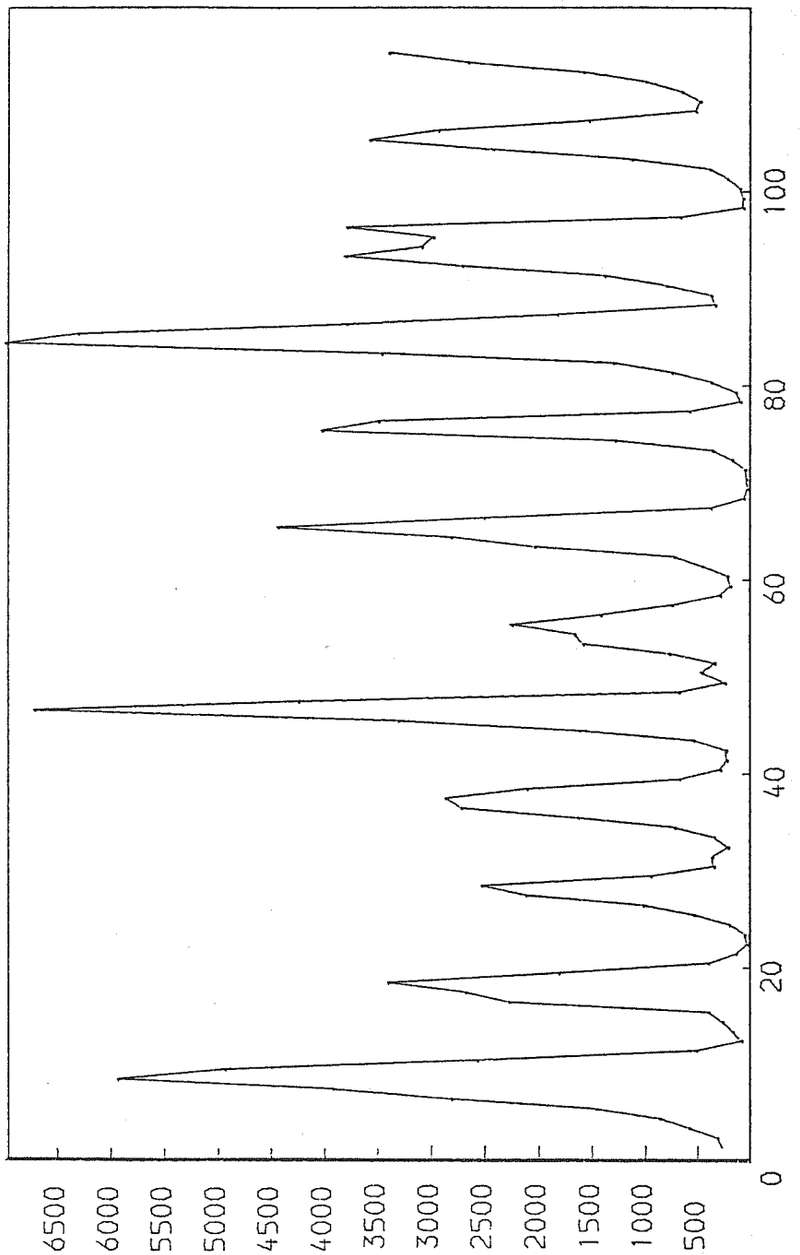


Fig. 3.16: The original Canadian lynx data

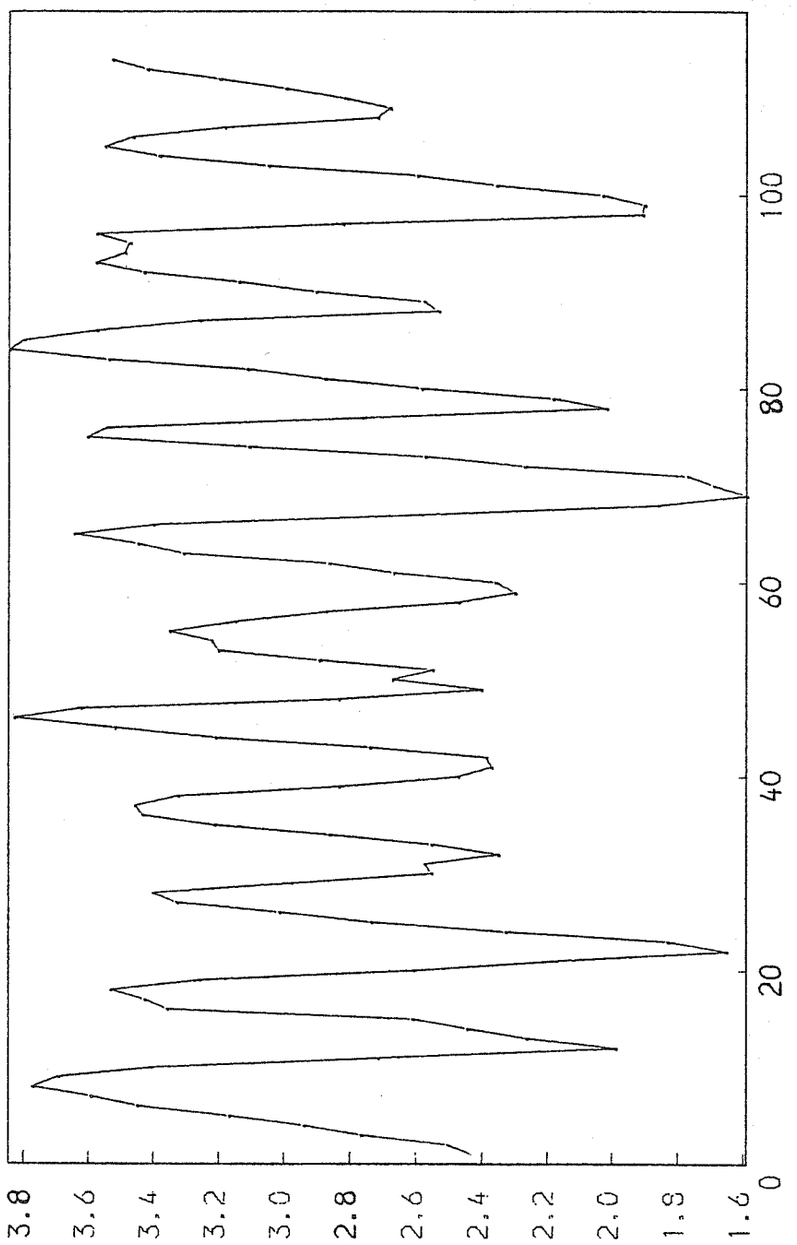


Fig. 3.17: $\log_{10}$ (Canadian Lynx data)

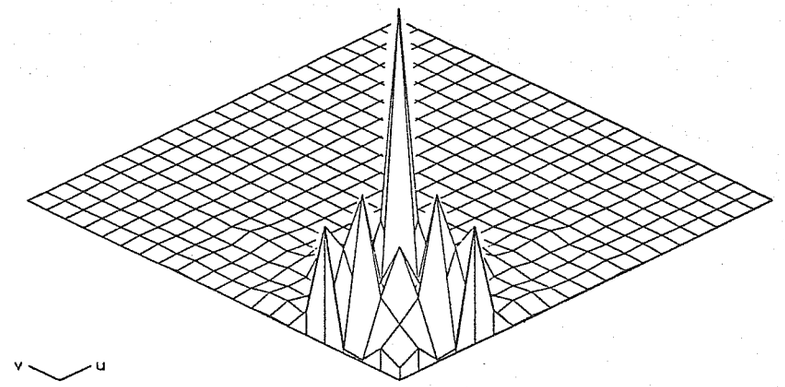


Fig. 3.18: The modulus of the bispectral estimate of log Canadian lynx data, using the optimum window, with M = 16, at frequencies $(\omega_1,\omega_2 = 0.0(0.05\pi)\pi)$.

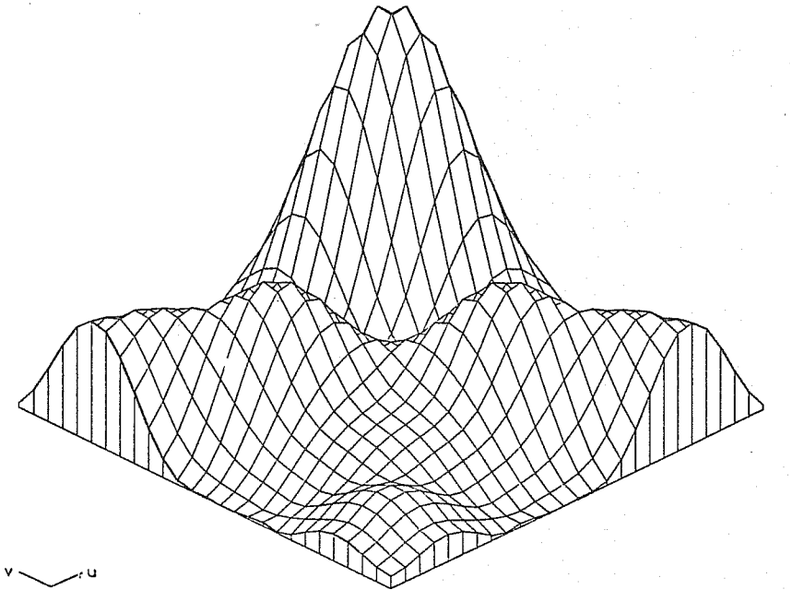


Fig. 3.19: The modulus of the bispectral estimate of log Canadian lynx data at frequencies $(\omega_1,\omega_2 = 0.0(0.01\pi).26\pi)$

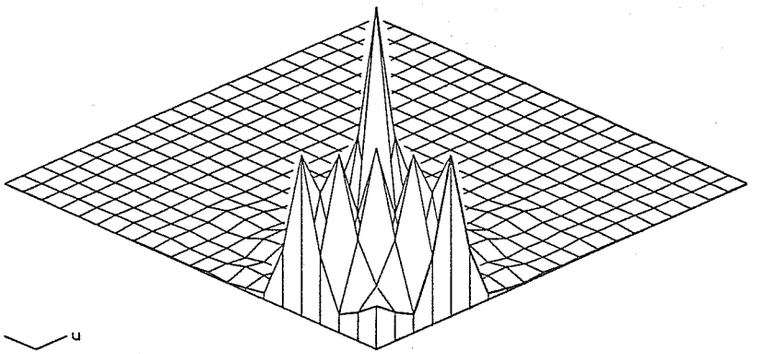


Fig. 3.20: The modulus of the non-normalized bispectral estimate of log Canadian lynx data, using the Daniell window, with M = 20

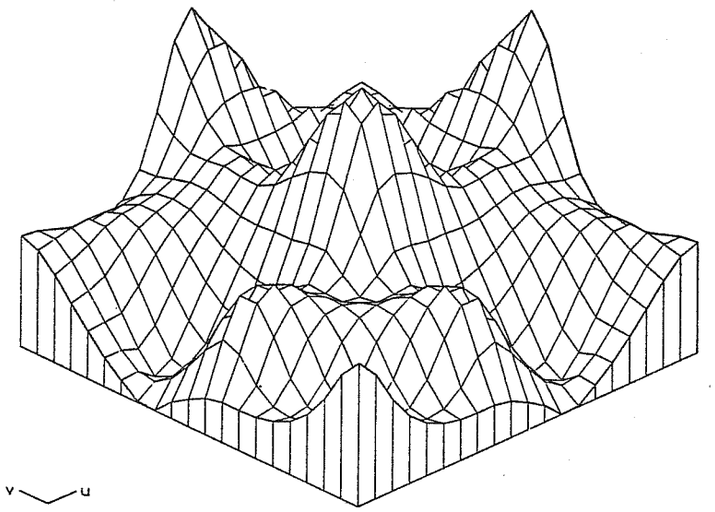


Fig. 3.21: The modulus of the normalized bispectral estimate of log Canadian lynx data, using the Daniell window, with M = 20

W1 \ W2	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.0002													
.95	.0000	.0000	.0001											
.90	.0002	.0001	.0001	.0004	.0004									
.85	.0000	.0002	.0002	.0001	.0006	.0001	.0000							
.80	.0001	.0002	.0005	.0003	.0003	.0004	.0001	.0001	.0003					
.75	.0001	.0005	.0002	.0002	.0004	.0001	.0001	.0000	.0001	.0001	.0000			
.70	.0001	.0008	.0003	.0002	.0008	.0003	.0000	.0001	.0002	.0001	.0001	.0001	.0001	
.65	.0001	.0003	.0004	.0001	.0005	.0002	.0001	.0001	.0001	.0001	.0001	.0002	.0001	.0002
.60	.0008	.0013	.0011	.0003	.0025	.0009	.0001	.0002	.0002	.0001	.0001	.0001	.0002	
.55	.0001	.0007	.0004	.0008	.0023	.0014	.0002	.0001	.0002	.0001	.0001	.0001		
.50	.0003	.0005	.0006	.0006	.0016	.0005	.0001	.0001	.0002	.0001	.0001			
.45	.0003	.0011	.0006	.0023	.0034	.0010	.0003	.0002	.0002	.0001				
.40	.0002	.0033	.0019	.0020	.0139	.0019	.0003	.0004	.0007					
.35	.0007	.0036	.0016	.0003	.0061	.0036	.0002	.0001						
.30	.0002	.0009	.0016	.0027	.0039	.0012	.0006							
.25	.0112	.0022	.0039	.0164	.0361	.0025								
.20	.0539	.0399	.0079	.0120	.1371									
.15	.0020	.0684	.0245	.0027										
.10	.0010	.0254	.0447											
.05	.0105	.0173												
0.00	.0057													

W1 (AS A FRACTION OF PI)

TABLE 3.18: The modulus of the bispectral estimate of the log Canadian lynx data using the optimum window with M = 16.

ω_2	W1 0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
.90	-.000													
.85	.000	1.564	-1.226											
.80	.000	-.775	-1.349	1.543	-.819									
.75	.000	1.385	-.824	-1.112	-.362	.529	.566							
.70	-.000	.979	-1.363	.990	.756	-1.455	-.720	.419	.647					
.65	-.000	1.096	-1.407	-1.387	1.552	.777	-.562	-1.346	-.678	-1.312	-.217			
.60	.000	.611	1.500	.693	-.523	-.306	.263	1.439	-.435	-.893	1.339	-.701	-.165	
.65	-.000	1.096	-1.407	-1.387	1.552	.777	-.562	-1.346	-.678	-1.312	-.217	-.999	-.085	.602
.60	.000	.611	1.500	.693	-.523	-.306	.263	1.439	-.435	-.893	1.339	-.701	-.165	
.55	-.000	1.233	.669	.620	-.351	-.710	.769	1.510	1.345	-.868	1.107	1.352		
.50	.000	1.117	-.121	.206	.533	-.895	.481	1.329	1.432	1.413	-.000			
.45	.000	1.417	.720	-.979	.921	.661	-1.349	-1.108	.522	-.719				
.40	-.000	.274	.019	1.447	-1.121	1.487	1.462	-.517	.041					
.35	.000	-1.363	.117	.069	-1.491	-1.216	-1.240	-.266						
.30	-.000	-.890	.348	-.497	1.036	1.455	-1.444							
.25	.000	-.833	.368	-1.565	1.390	1.033								
.20	.000	.504	.876	-1.335	-1.488									
.15	-.000	-.643	-.683	.855										
.10	.000	-.249	-.747											
.05	.000	.720												
0.00	0.000													

TABLE 3.19: The argument of the bispectral estimate of log Canadian lynx data, using the optimum window, with M = 16

ω_2	W1 0.00	.01	.02	.03	.04	.05	.06	.07	.08	.09	.10	.11	.12	.13	.14	.15	.16	.17	.18	.19	.20	.21	.22	.23	.24	.25	.26
.26	.0023	.0069	.0023	.0029	.0028	.0023	.0019	.0021	.0028	.0033	.0034	.0034	.0035	.0029	.0042	.0093	.0153	.0201	.0222	.0212	.0176	.0129	.0084	.0050	.0027	.0017	.0015
.25	.0112	.0066	.0038	.0028	.0026	.0022	.0014	.0012	.0024	.0035	.0039	.0037	.0031	.0019	.0069	.0164	.0276	.0372	.0425	.0419	.0361	.0271	.0177	.0100	.0050	.0025	
.24	.0240	.0178	.0130	.0094	.0070	.0053	.0037	.0017	.0007	.0028	.0040	.0042	.0033	.0021	.0096	.0230	.0399	.0560	.0668	.0693	.0630	.0501	.0347	.0207	.0106		
.23	.0379	.0317	.0259	.0205	.0159	.0124	.0094	.0063	.0028	.0007	.0033	.0045	.0040	.0024	.0103	.0264	.0481	.0711	.0893	.0976	.0937	.0792	.0566	.0376			
.22	.0496	.0450	.0399	.0337	.0273	.0216	.0170	.0127	.0082	.0034	.0013	.0040	.0046	.0030	.0086	.0250	.0495	.0778	.1034	.1193	.1210	.1083	.0853				
.21	.0555	.0541	.0514	.0461	.0388	.0313	.0249	.0196	.0143	.0087	.0032	.0023	.0044	.0039	.0055	.0195	.0433	.0737	.1046	.1279	.1371	.1297					
.20	.0539	.0565	.0575	.0548	.0483	.0399	.0317	.0252	.0196	.0140	.0079	.0027	.0031	.0041	.0032	.0120	.0317	.0601	.0923	.1206	.1371						
.19	.0455	.0516	.0570	.0585	.0549	.0472	.0379	.0295	.0230	.0177	.0122	.0064	.0022	.0031	.0030	.0053	.0187	.0414	.0703	.0993							
.18	.0327	.0411	.0505	.0572	.0587	.0545	.0458	.0354	.0260	.0194	.0146	.0097	.0048	.0018	.0024	.0021	.0081	.0230	.0451								
.17	.0193	.0283	.0406	.0526	.0605	.0620	.0568	.0461	.0333	.0221	.0152	.0113	.0075	.0036	.0013	.0016	.0022	.0093									
.16	.0085	.0168	.0304	.0460	.0597	.0677	.0678	.0597	.0457	.0300	.0174	.0111	.0086	.0060	.0029	.0010	.0013										
.15	.0020	.0091	.0219	.0385	.0554	.0684	.0740	.0704	.0585	.0416	.0245	.0122	.0078	.0071	.0052	.0027											
.14	.0005	.0055	.0157	.0306	.0475	.0628	.0724	.0738	.0663	.0519	.0343	.0182	.0079	.0063	.0066												
.13	.0005	.0038	.0114	.0253	.0378	.0521	.0633	.0684	.0662	.0569	.0428	.0273	.0140	.0062													
.12	.0001	.0028	.0087	.0177	.0287	.0402	.0501	.0567	.0586	.0553	.0474	.0364	.0245														
.11	.0002	.0021	.0071	.0143	.0225	.0307	.0379	.0435	.0472	.0487	.0476	.0435															
.10	.0010	.0010	.0058	.0124	.0193	.0254	.0302	.0340	.0372	.0407	.0447																
.09	.0016	.0017	.0046	.0107	.0174	.0229	.0269	.0294	.0315	.0348																	
.08	.0006	.0038	.0058	.0097	.0156	.0212	.0252	.0276	.0292																		
.07	.0024	.0063	.0094	.0112	.0144	.0191	.0233	.0259																			
.06	.0066	.0094	.0137	.0153	.0156	.0173	.0204																				
.05	.0105	.0125	.0173	.0197	.0190	.0173																					
.04	.0124	.0144	.0189	.0221	.0220																						
.03	.0118	.0141	.0182	.0215																							
.02	.0094	.0117	.0153																								
.01	.0067	.0087																									
0.00	.0057																										

W1 W1(AS A FRACTION OF PI)

TABLE 3.20: The modulus of the bispectral estimate of log Canadian lynx data, using the optimum window, with M = 16, at frequencies $(\omega_1,\omega_2 = 0.0(0.01\pi).26\pi$.

(iii) Unemployment Data of W. Germany

For our third illustration we consider the number of people registered as unemployed in W. Germany for the period January 1948 - May 1980 (inclusive). The data is monthly, and the total number of observations are 389 (see Appendix D). The graph of this data is plotted in Fig. 3.22, and its sample autocorrelation up to 100 lags are plotted in Fig. 3.24. The steep increase in the number of unemployed in the beginning can be due to war effect, and the sudden increase from the 320th observation onwards coincides with the world energy crisis and the world recession in general. The sample autocorrelation function shows that there is a seasonality in the series and possibly a linear trend. In their book, Granger and Newbold (1977) have analysed unemployment figures in Belgium and it is interesting to note that features exhibited in these two series are similar. They have suggested differencing the series to remove the trend and seasonality on the series. Suppose Y_t denotes the number of people registered as unemployed at time t. In order to remove the trend and seasonality (twelve month period), we have differenced the series and obtained $X_t = (1-B)(1-B^{12})Y_t$.

The series $\{X_t\}$ and its sample autocorrelations are plotted in Fig. 3.23 and Fig. 3.25 respectively. The series $\{X_t\}$ seems to be free from trend and seasonality; and the bispectral density is estimated for this series. The mean, variance and third order central moments for the original series $\{Y_t\}$ and for the differenced series $\{X_t\}$ are, respectively,

$$\bar{Y} = .71103 \times 10^6 , \qquad \bar{X} = -1301.2$$

$$S_Y^2 = .3223 \times 10^{12} , \qquad S_X^2 = .12558 \times 10^{11},$$

$$\hat{C}_Y(0,0) = -.1029 \times 10^{15}, \quad \hat{C}_X(0,0) = .47325 \times 10^{11},$$

$$\hat{C}_Y(0,0)/S^3 = -.00056 \qquad \hat{C}_X(0,0)/S^3 = .000034$$

The bispectral density function of the transformed series is estimated using the optimum window with M = 10. The results are given in Tables 3.21 and 3.22, and Fig. 3.26. Also, the non-normalized and normalized bispectral modulus are estimated using the Daniell window with M = 16; the results are shown in Figs. 3.27 and 3.28.

From these results, we note that there is no obvious "peak" in the modulus of the bispectral estimate similar to the "peaks" which appeared in the sunspot numbers and the Canadian lynx data. Therefore, the conclusion is that there is no periodicity in the differenced unemployment data. From Table 3.21 it is clear that the values of the estimated bispectral modulus at the frequencies ($\omega_1, \omega_2 \geq .35$) are larger than the values at the other frequencies and this may be due to some non-linearity effect in the series.

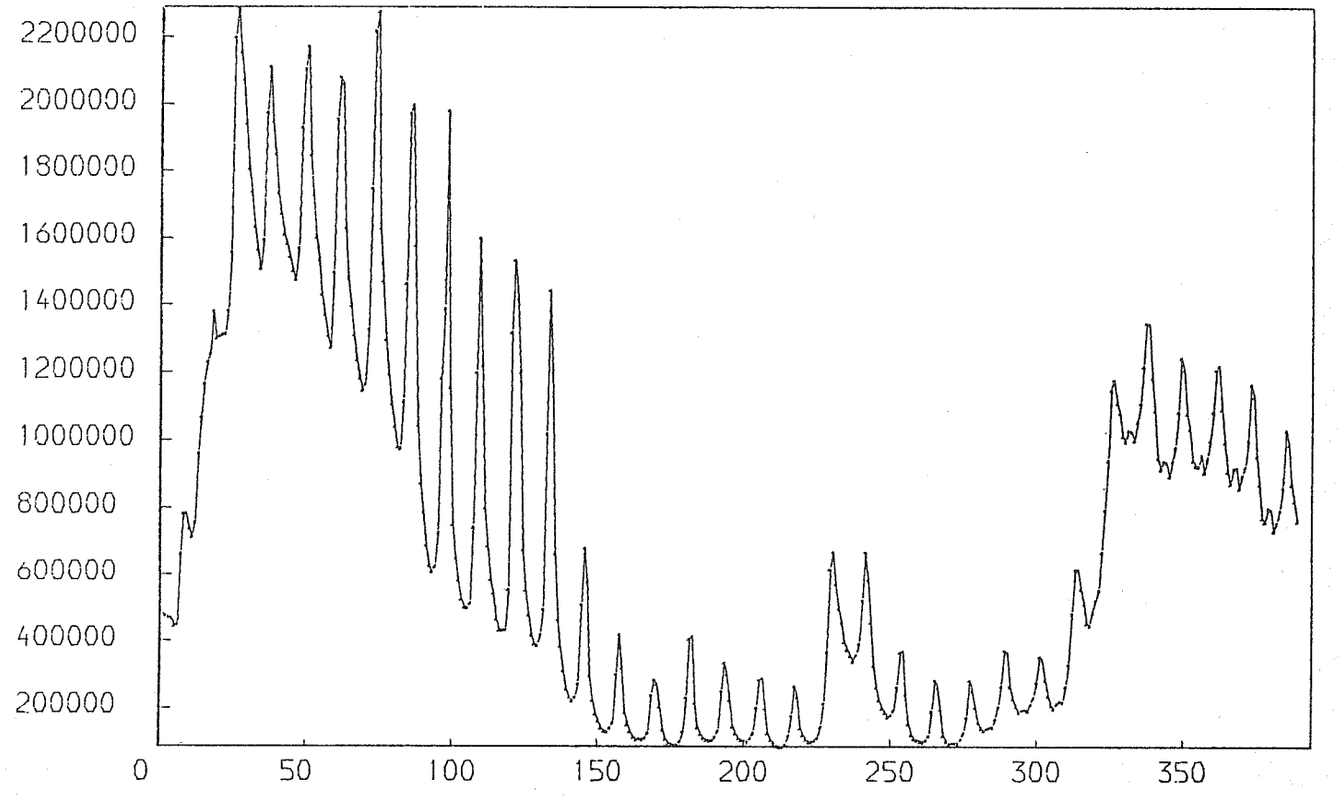


Fig. 3.22: Unemployment Figures in W. Germany $\{Y_t\}$ (monthly) Jan. 1948 - May 1980

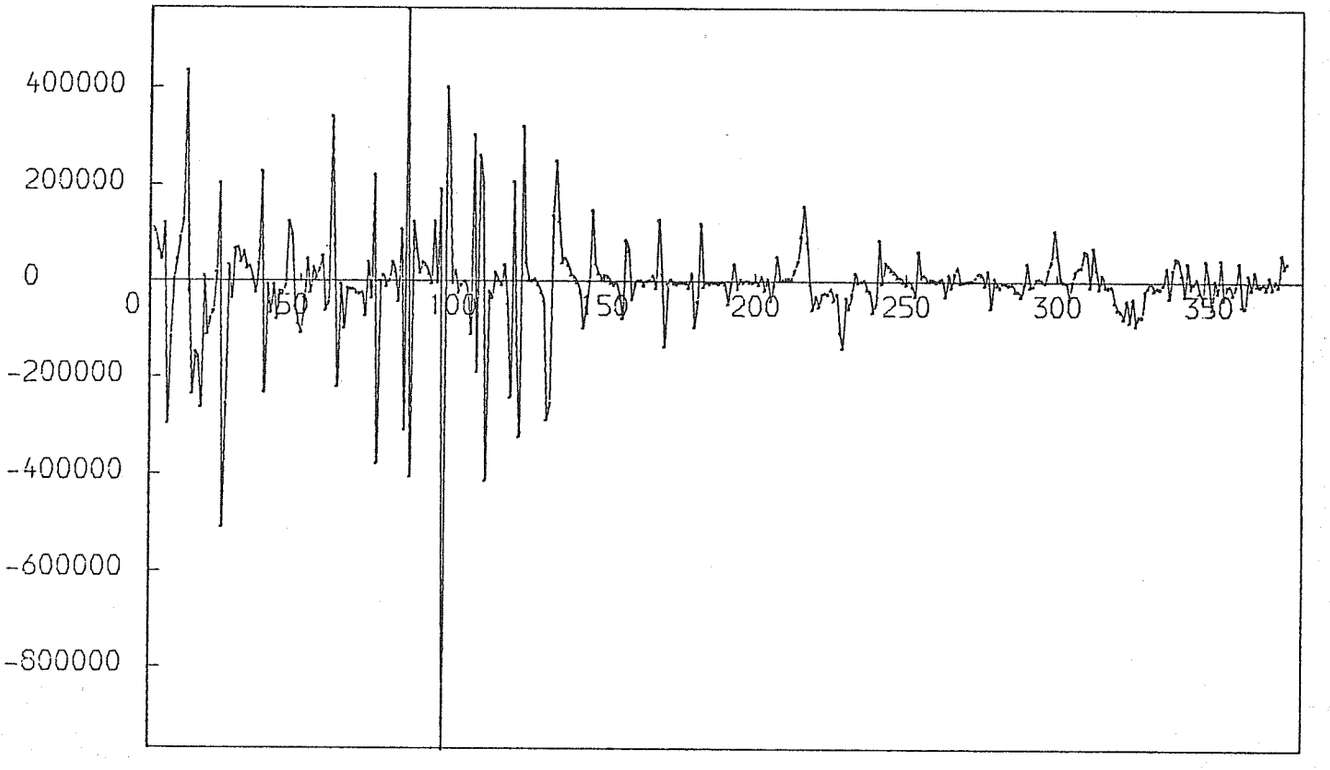


Fig. 3.23: The differenced unemployment figures in W. Germany, $\{X_t\}$.

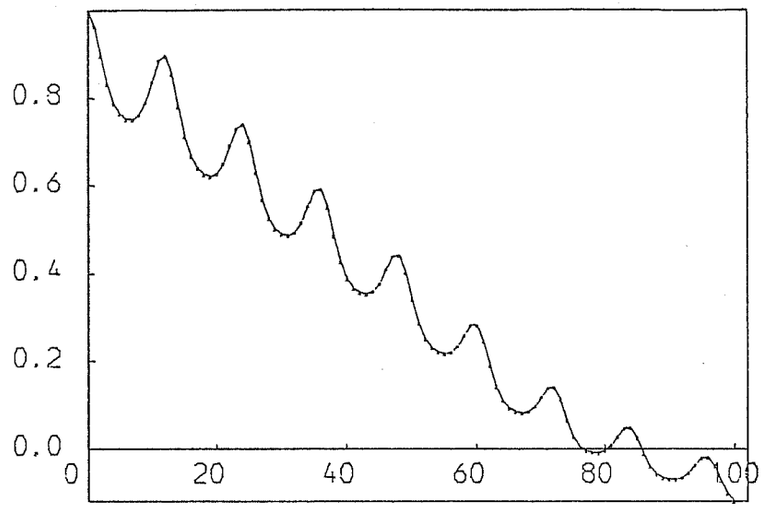


Fig. 3.24: Sample autocorrelations of the original unemployment figures in West Germany

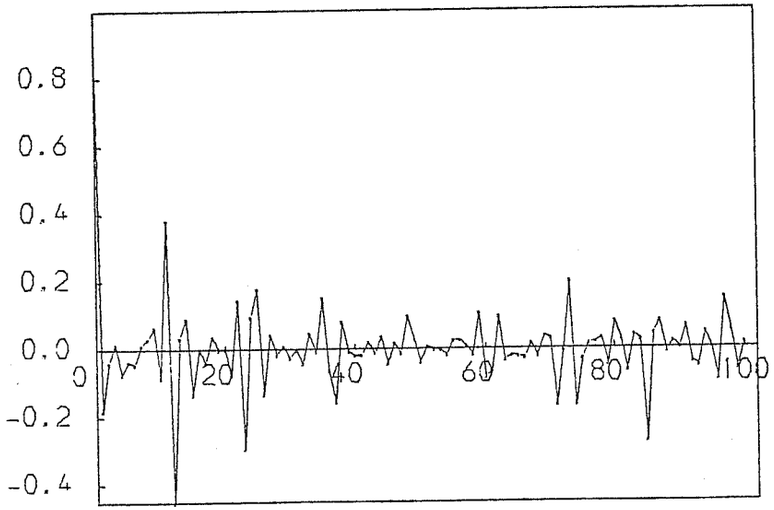


Fig. 3.25: Sample autocorrelations of the differenced unemployment figures in West Germany

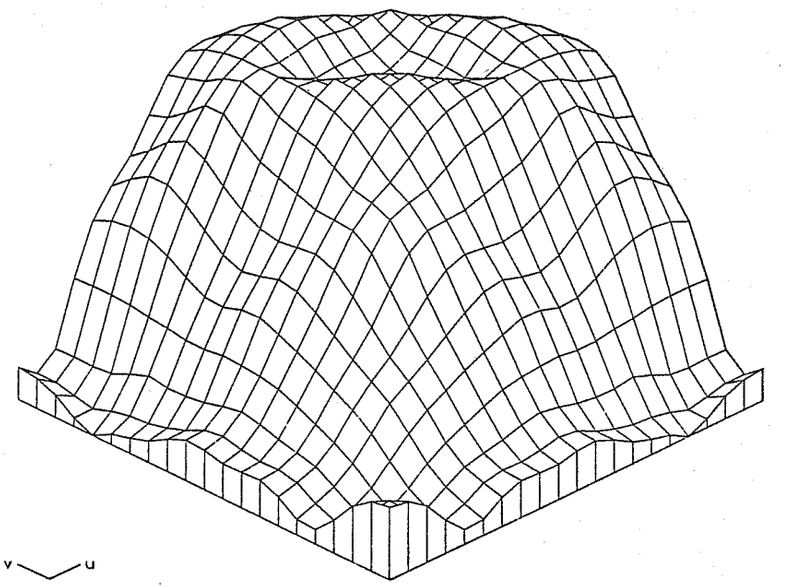


Fig. 3.26: Estimated bispectral modulus of the differenced unemployment figures in West Germany using the optimum window with M = 10.

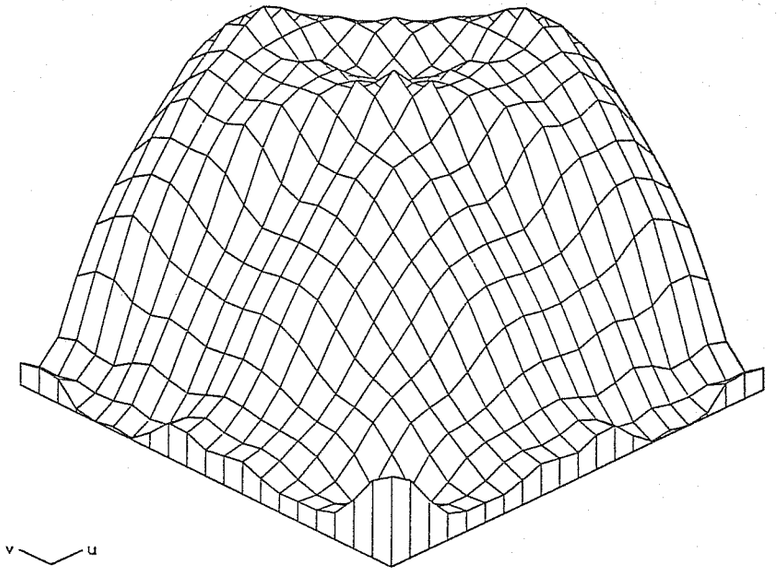


Fig. 3.27: The modulus of the non-normalized bispectral estimate of the differenced unemployment figures in West Germany using the Daniell window with M = 16.

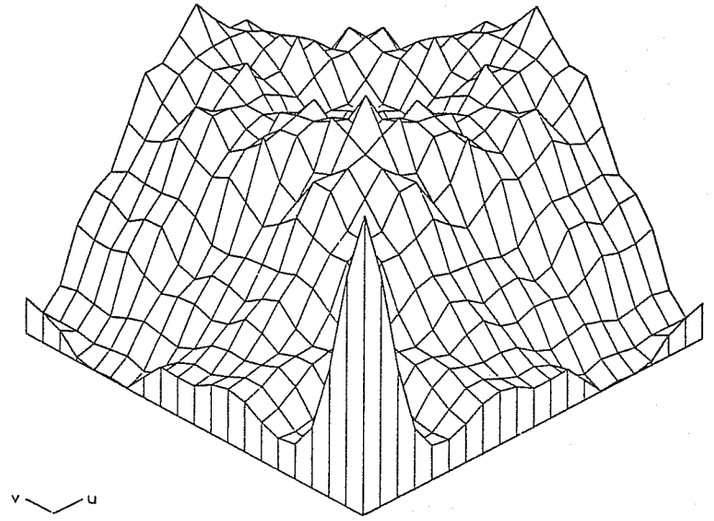


Fig. 3.28: The modulus of the normalized bispectral estimate of the differenced unemployment data

W1 \ W2	0.00	.05	.10	.15	.20	.25	.30	.35	.40	.45	.50	.55	.60	.65
1.00	.168E+17													
.95	.158E+17	.977E+16	.174E+17											
.90	.123E+17	.762E+16	.171E+17	.469E+17	.749E+17									
.85	.572E+16	.576E+16	.181E+17	.463E+17	.736E+17	.948E+17	.111E+18							
.80	.162E+16	.883E+16	.222E+17	.449E+17	.696E+17	.893E+17	.107E+18	.124E+18	.136E+18					
.75	.654E+16	.108E+17	.255E+17	.438E+17	.635E+17	.817E+17	.983E+17	.116E+18	.133E+18	.137E+18	.133E+18			
.70	.887E+16	.115E+17	.241E+17	.424E+17	.586E+17	.743E+17	.916E+17	.110E+18	.127E+18	.137E+18	.136E+18	.129E+18	.124E+18	
.65	.119E+17	.136E+17	.244E+17	.424E+17	.583E+17	.709E+17	.869E+17	.107E+18	.124E+18	.136E+18	.141E+18	.137E+18	.130E+18	.124E
.60	.170E+17	.169E+17	.258E+17	.438E+17	.612E+17	.728E+17	.851E+17	.103E+18	.121E+18	.134E+18	.141E+18	.144E+18	.139E+18	
.55	.208E+17	.213E+17	.282E+17	.433E+17	.617E+17	.756E+17	.865E+17	.100E+18	.116E+18	.129E+18	.136E+18	.142E+18		
.50	.204E+17	.225E+17	.298E+17	.424E+17	.581E+17	.741E+17	.876E+17	.995E+17	.110E+18	.120E+18	.130E+18			
.45	.182E+17	.205E+17	.280E+17	.406E+17	.539E+17	.683E+17	.842E+17	.988E+17	.108E+18	.113E+18				
.40	.184E+17	.192E+17	.252E+17	.367E+17	.496E+17	.617E+17	.760E+17	.922E+17	.104E+18					
.35	.191E+17	.195E+17	.233E+17	.325E+17	.442E+17	.546E+17	.657E+17	.795E+17						
.30	.153E+17	.177E+17	.208E+17	.281E+17	.391E+17	.486E+17	.559E+17							
.25	.904E+16	.121E+17	.158E+17	.225E+17	.324E+17	.425E+17								
.20	.856E+16	.859E+16	.102E+17	.152E+17	.236E+17									
.15	.172E+17	.125E+17	.987E+16	.961E+16										
.10	.282E+17	.221E+17	.149E+17											
.05	.345E+17	.307E+17												
0.00	.360E+17													

W1 (AS A FRACTION OF PI)

TABLE 3.21: The modulus of the normalized bispectral estimate of the differenced unemployment data using the optimum window, with M = 10.

w2 \ w1	0.000	.050	.100	.150	.200	.250	.300	.350	.400	.450	.500	.550	.600	.650
.90	.000													
.85	-.000	-.971	.672											
.80	.000	1.319	.866	.530	.326									
.75	.000	.786	.885	.676	.452	.307	.189							
.70	.000	.171	.585	.628	.496	.371	.291	.172	.008					
.65	.000	-.063	.209	.403	.429	.388	.348	.275	.131	-.027	-.116			
.60	.000	.153	.217	.278	.335	.389	.415	.374	.267	.125	-.009	-.094	-.137	
.65	.000	-.063	.209	.403	.429	.388	.348	.275	.131	-.027	-.116	-.140	-.168	-.216
.60	.000	.153	.217	.278	.335	.389	.415	.374	.267	.125	-.009	-.094	-.137	
.55	.000	.312	.462	.378	.316	.368	.460	.469	.384	.270	.146	.013		
.50	.000	.293	.565	.552	.404	.363	.448	.513	.470	.373	.274			
.45	.000	.248	.504	.596	.520	.433	.448	.509	.517	.458				
.40	.000	.259	.428	.491	.507	.508	.510	.519	.527					
.35	.000	.219	.378	.338	.320	.420	.536	.555						
.30	.000	.069	.254	.215	.097	.161	.386							
.25	.000	-.188	.015	.115	.005	-.023								
.20	.000	-.552	-.407	-.045	.028									
.15	.000	-.504	-.806	-.466										
.10	.000	-.284	-.626											
.05	.000	-.114												
0.00	0.000													

Table 3.2 The arguments of the normalized bispectral estimate of the differenced unemployment data using the optimum window with M = 10

3.4 Some nonlinear phenomena

In chapter 1, we have shown that the bispectral density function of a Gaussian process (and hence a linear process) is zero for all frequencies. The process under consideration may be linear, but may be non-Gaussian, in which case the normalised bispectral density function is constant for all frequencies. In other words, a study of bispectral density function is likely to reveal whether the process we are dealing with is linear or nonlinear, and this question is pursued further in chapter 4. If the bispectral density function is zero, it does not necessarily mean that the process is linear and Gaussian. For some nonlinear processes, which are symmetric, the bispectral density function can be zero. A well known process, known as Duffing process, is an example of such a process.

On the basis of bispectral analysis carried out on the three time series in section 3.3, one can confirm that the three time series are all nonlinear.

Besides the bispectral density function, one can look for other phenomena which may also characterise nonlinear time series. In this connection it is useful to summarise some properties of (deterministic) nonlinear differential equations used in the study of nonlinear vibrations (see Cunningham, 1958, Stoker, 1950).

A typical nonlinear differential equation considered is

$$\ddot{X}(t) + \phi(\dot{X}(t)) + f(X(t)) = 0 \ , \qquad (3.4.1)$$

$$\ddot{X}(t) + \phi(\dot{X}(t)) + f(X(t)) = A \cos \omega t \ ,$$

where $X(t)$ is the response, $\dot{X}(t)$ is the first derivative of $X(t)$ with respect to t etc. Here $\phi(\dot{X}(t))$ is the 'damping force' and $f(X(t))$ is the

restoring force of the oscillator. If $\phi(\ddot{X}(t)) = 0$ and $f(X(t))$ is linear, we get the equation of a simple harmonic motion.

If $\phi(\cdot)$ and $f(\cdot)$ are nonlinear functions, then the response $X(t)$ exhibits two typical nonlinear phenomena. They are (i) jump phenomena (ii) limit cycle behaviour.

In particular, let us consider the Duffing (1918) equation,

$$\ddot{X}(t) + C\dot{X}(t) + \alpha\dot{X}(t) + \beta\dot{X}^3(t) = A \cos \omega t \qquad (3.4.3)$$

which is a special case of (3.4.2). In this case the damping force is linear, and the restoring force is nonlinear. If $\beta>0$, the equation (3.4.3) is known as a 'hard spring' oscillator, and if $\beta<0$ it is called a 'soft spring' oscillator. As the frequency ω of the input changes, the amplitude of the response function of the system exhibits a "jump phenomena". Basically, in the response function, which is a function of ω, there will be a jump and hence it is non differentiable (see Haggan and Ozaki, 1979).

We say a process $\{X(t)\}$is a Duffing process if $X(t)$ satisfies the nonlinear differential equation

$$\ddot{X}(t) + C\dot{X}(t) + \alpha\dot{X}(t) + \beta\dot{X}^3(t) = n(t) \qquad (3.4.4)$$

where the input $n(t)$ is a Guassian white noise. A typical nonlinear phenomenen associated with this process is that of amplitude dependent frequency, i.e., the amplitude increases as frequency increases and amplitude decreases as frequency decreases.

In view of the symmetric nature of the process $\{X(t)\}$, one can show that the bispectral density function of the process $\{X(t)\}$ is zero (see Perrochaud, 1982). As stated earlier, the Duffing process is an example of a nonlinear process with bispectral density function equal to zero.

Other statistical measures used in the study of nonlinear processes are (i) zero crossings, (ii) threshold crossings, (iii) distribution of peaks. The peak and envelope probability density functions of the process $\{X(t)\}$ when $\beta=0$ in (3.4.4) are Gaussian. Crandall (1963) has shown that in the case of Duffing process (when $\beta>0$) these probability density functions, though symmetric, are very narrow (when compared to normal) with very small tail probability. In this case, clearly the fourth order cumulant is non zero.

On the basis of these probability density functions, Crandall (1963, p.1698) has concluded that the Duffing process is one for which large amplitude cycles have small periods (this is an amplitude dependent frequency property).

Another statistical measure used by Crandall (1963, p.1702) to study the nonlinear phenomena with reference to Duffing process is the "mean square" property. Suppose we denote the variance of $X(t)$, when $X(t)$ satisfies (3.4.4) with $\beta>0$, by $\text{var}_{non}(X(t))$, and by $\text{var}_{lin}(X(t))$ when $\beta=0$. Crandall (1963) has shown that $\text{var}_{non}(X(t)) < \text{var}_{lin}(X(t)$ in the case of hard spring oscillator. i.e. $\beta>0$.

To study the other typical nonlinear phenomena known as limit cycles, consider the van Der Pol equation

$$\ddot{X}(t) + \phi(\dot{X}(t)) + \alpha X(t) = 0 \qquad (3.4.5)$$

where the damping force $\phi(\dot{X}(t))$ is nonlinear.

If $\quad \dot{X}(t)\phi(\dot{X}(t)) < 0 \quad$ for small $|\dot{X}(t)|$,

$\quad\quad \dot{X}(t)\phi(\dot{X}(t)) > 0 \quad$ for large $|\dot{X}(t)|$,

then the asymptotic solution of the equation (3.4.5) is a <u>periodic function</u>. This is called the limit cycle behaviour. From our analysis we observe that these phenomena are discussed for continuous time parameter

observations only. In recent years, Haggan and Ozaki (1979), Tong and Lim (1980) have shown that similar phenomena can be realised by exponential autoregressive models and threshold autoregressive models, and these models are discrete parameter time series models. The two time series, namely sunspot numbers and Canadian lynx data, exhibit limit cycle behaviour, and this, of course, implies that these series are nonlinear.

A close examination of the plots of the three series, namely, sunspot numbers, Canadian lynx data and German unemployment data reveals that these three series are time irreversible (see section 1.3 , and Brillinger and Rosenblatt, 1967a). This implies that the series are non-Gaussian, and may be nonlinear too.

An examination of the differenced unemployment figures in W. Germany as shown in Fig. 3.23 shows that there is a clear evidence of "nonlinearity" in the first 150 observations. In this portion of the plot, we observe high amplitude, high frequency oscillation when compared to the later portion of the data where there are low amplitude, low frequency oscillations.

In the foregoing discussion, we have discussed various nonlinear phenomena one can look for in the analysis of time series. In this monograph we are mainly interested in the study of nonlinearity via higher order spectral analysis (bispectral analysis).

Just as we can use bispectral density function to study "nonlinearity" of a time series, we can use cross bispectral analysis for studying the nonlinear relationships, if any, between two time series, say $\{Y_t\}$ and $\{X_t\}$. Recently, cross bispectral analysis have been used to model human pupillary system (see Hung, Brillinger and Stark, 1979). Other interesting applications are modelling of (i) human body subjected to seat to head vibrations, and (ii) loudspeaker excited with band limited white noise in the low frequency systems (see Perrochaud, 1982). We do not pursue cross bispectral analysis in this monograph.

In the following chapter we consider the topic of nonlinearity by examining the bispectral density function.

CHAPTER 4

TESTS FOR LINEARITY AND GAUSSIANITY OF STATIONARY TIME SERIES

4.1 General Introduction

The assumptions that are commonly made in time series analysis are

(i) that the process is stationary, and

(ii) that the process can be described by a linear model.

It is probably no exaggeration to say that the majority of time series in the real world do not satisfy these assumptions. So the important problem is how one can test these assumptions when a realization of a random process is given. Based on the theory of "oscillatory" stochastic processes, as developed by Priestley (1965), Priestley and Subba Rao (1969) have proposed a test for non-stationarity of time series. In this chapter we consider statistical tests for testing the assumption of linearity (and Gaussianity) of the time series, and apply these tests to real and simulated time series data (Subba Rao and Gabr, 1980).

A considerable amount of work has been reported in the engineering

literature on the effects of non-linearity in various physical systems. (For details, refer to Cunningham, 1958; Lyon, 1963; and Stoker, 1950). In this context, the effects of non-linearity are studied by obtaining the distribution of zero crossings, threshold crossings and peaks of the response functions (see Crandall, 1963). However, in recent years several research workers, notably Brillinger (1965), Brillinger and Rosenblatt (1967a, 1967b), Rosenblatt and Van Ness (1966), Godfrey (1965), Huber et al (1971), Hasselman et al (1963), have pointed out the importance of higher order spectra in the analysis of non-linear time series is further strengthened by the fact that no specific assumption on the underlying model governing the time series is necessary.

Brillinger (1965) pointed out that the bispectral density function could, in principle, be used for testing linearity. Here, we construct some specific test procedures and apply these to real and simulated data.

4.2 Spectrum and Bispectrum of Linear Processes

Let $\{X_t\}$ have the linear representation

$$X_t = \sum_{r=-\infty}^{\infty} a_r \, e_{t-r} \qquad (4.2.1)$$

where $\{e_t\}$ is a sequence of independent, identically distributed random variables with

$$E(e_t) = 0,$$
$$E(e^2_t) = \sigma_e^2, \quad \text{and}$$
$$E(e^3_t) = \mu_3.$$

Then, the autocovariance function of $\{X_t\}$ is

$$R(s) = E[X_t X_{t+s}]$$
$$= E\left[\left\{\sum_{r=-\infty}^{\infty} a_r e_{t-r}\right\}\left\{\sum_{r'=-\infty}^{\infty} a_{r'} e_{t+s-r'}\right\}\right]$$
$$= \sum_r \sum_{r'} a_r a_{r'} E[e_{t-r} e_{t+s-r'}]$$

Hence, because of the independence of the e_t's, we get

$$R(s) = \sigma_e^2 \left(\sum_{r=-\infty}^{\infty} a_r a_{r+s}\right)$$

Also, the third-order central moment of $\{X_t\}$ is

$$C(t_1,t_2) = E[X_t X_{t+t_1} X_{t+t_2}]$$
$$= \sum_{r_1} \sum_{r_2} \sum_{r_3} a_{r_1} a_{r_2} a_{r_3} E[e_{t-r_1} e_{t+t_1-r_2} e_{t+t_2-r_3}]$$
$$= \mu_3 \left(\sum_r a_r a_{r+t_1} a_{r+t_2}\right)$$

Let

$$H(\omega) = \sum_{r=-\infty}^{\infty} a_r e^{-ir\omega}$$

be the transfer function of the process (4.2.1); then the spectral density and bispectral density functions of $\{X_t\}$ are given by

$$f(\omega) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} R(s) e^{-is\omega}$$
$$= \frac{\sigma_e^2}{2\pi} \sum_{s=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} a_r a_{r+s} e^{-is\omega}$$

$$= \frac{1}{2\pi}\left\{\sum_{r=-\infty}^{\infty} a_r e^{ir\omega} \cdot \sum_{s=-\infty}^{\infty} a_{r+s} e^{-i(r+s)\omega}\right\}$$
$$= \frac{\sigma_e^2}{2\pi} H(-\omega) H(\omega) = \frac{\sigma_e^2}{2\pi} |H(\omega)|^2. \qquad (4.2.2)$$

Similarly,

$$f(\omega_1,\omega_2) = \frac{1}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} C(t_1,t_2) e^{-it_1\omega_1 - it_2\omega_2}$$
$$= \frac{\mu_3}{(2\pi)^2} \sum_{t_1=-\infty}^{\infty} \sum_{t_2=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} a_r a_{r+t_1} a_{r+t_2} e^{-t_1\omega_1 - it_2\omega_2}$$
$$= \frac{\mu_3}{(2\pi)^2} \cdot H(\omega_1) H(\omega_2) H(-\omega_1-\omega_2) \qquad (4.2.3)$$

Thus, if we write

$$X_{ij} = \frac{|f(\omega_i,\omega_j)|^2}{f(\omega_i) f(\omega_j) f(\omega_i+\omega_j)} = |g(\omega_i,\omega_j)|^2$$

then from (4.2.3) we obtain

$$X_{ij} = \frac{\mu_3^2}{2\pi\sigma_e^2}, \text{ all } i, j \qquad (4.2.4)$$

i.e. X_{ij} does not depend on ω_i and ω_j.

The relation (4.2.3) shows that if $\mu_3 = 0$, then the bispectral density function $f(\omega_i,\omega_j) = 0$ for all values of ω_i and ω_j. Of course, if the random variables $\{e_t\}$ are Gaussian, then $\mu_3 = 0$ and $f(\omega_i,\omega_j) = 0$. Now under suitable conditions on the sequence $\{a_r\}$, one can show that the normality of $\{e_t\}$ implies the normality of $\{X_t\}$. It is then clear that if the process $\{X_t\}$ is Gaussian then the bispectral density function $f(\omega_i,\omega_j) = 0$ for all ω_i and ω_j. However, it may happen that the process $\{X_t\}$ is linear but not necessarily Gaussian. For example, consider the

model used to describe the Poisson triggered activity (Huber, 1971.p.79). Here, the process is non-Gaussian, but still has a linear representation. We observe that in this situation the ratio X_{ij} given by (4.2.4) is a constant for all ω_i and ω_j. In other words, the constancy of X_{ij} is a test for adequacy of linearity of the process (see Brillinger, 1965).

In this chapter we propose tests for testing

(1) H_0: $f(\omega_i,\omega_j) = 0$, for all ω_i and ω_j; and

(2) H_1: X_{ij} is constant for all ω_i and ω_j.

Acceptance of H_0 is only consistent with linearity and $\mu_3 = 0$. Acceptance of H_1 and rejection of H_0 implies that the process is not Gaussian but consistent with being linear.

4.3 Test for Symmetry and Linearity

To test the hypothesis $H_0: f(\omega_i,\omega_j) = 0$, all ω_i, ω_j, we proceed in two stages. The first stage consists of testing the hypothesis

$$f(\omega_i,\omega_j) = 0 \quad (\text{all } \omega_i \text{ and } \omega_j)$$

when ω_i and ω_j are restricted to the range

$$\left.\begin{aligned} &0 < \omega_i < \frac{2\pi}{3} \\ &\omega_{i1} < \omega_j < \frac{2\pi}{3}, \text{ and} \\ &\frac{2\pi}{3} < \omega_j < \pi - \tfrac{1}{2}\omega_i \end{aligned}\right\} \qquad (4.3.1)$$

(excluding the boundaries and the origin of the region (2) of Fig. 1.3).

Within this region the bispectral estimate is approximately complex normal (see Miller, 1974, for a definition of complex normal), and the test we describe below depends on the complex analogue of Hotelling T^2 statistic (see Giri, 1965; Khatri, 1965). If we accept the null hypothesis at the first stage, then in order to confirm that the time series $\{X_t\}$ is linear and $\mu_3 = 0$, we must also test that $f(\omega_i,\omega_j) = 0$ when ω_i and ω_j are defined at the origin and on the boundaries. This is our second stage of the test. In this context, we use Hotelling T^2 for real random variables (see Anderson, 1958; Kshirsagar,1972). The theoretical details are omitted here. However, the section on numerical illustrations does include the testing procedure of the second stage.

We now consider the first stage of the procedure. We construct a column vector from the set of bispectral densities $f(\omega_i,\omega_j)$ defined on the region given by (4.3.1) as follows. We select a 'coarse' grid of frequencies

$$(\omega_i,\omega_j),\ \omega_i = \frac{i\pi}{K},\ \omega_j = \frac{j\pi}{K},\ i = 1,2,\ldots,L;$$
$$j = i+1, i+2, \ldots, \gamma(i) \qquad (4.3.2)$$

where $L = \left[\frac{2K}{3}\right]$, $\gamma(i) = K - \left[\frac{i}{2}\right] - 1$ and $K << N$

In view of the symmetry relations (1.4.8), it is sufficient to consider only frequencies (ω_i,ω_j) lying in the region OAB of Fig. 4.1 (the equations of the lines OA, OB and AB which define the region OAB are, respectively,

$$\omega_2 - \omega_1 = 0,\ \omega_2 = -\tfrac{1}{2}\omega_1 + \pi \quad \text{and } \omega_1 = 0).$$

It is clear from Fig. 4.1 that the above restrictions on L, $\gamma(i)$ ensure that the points (ω_i,ω_j) lie within the required region.

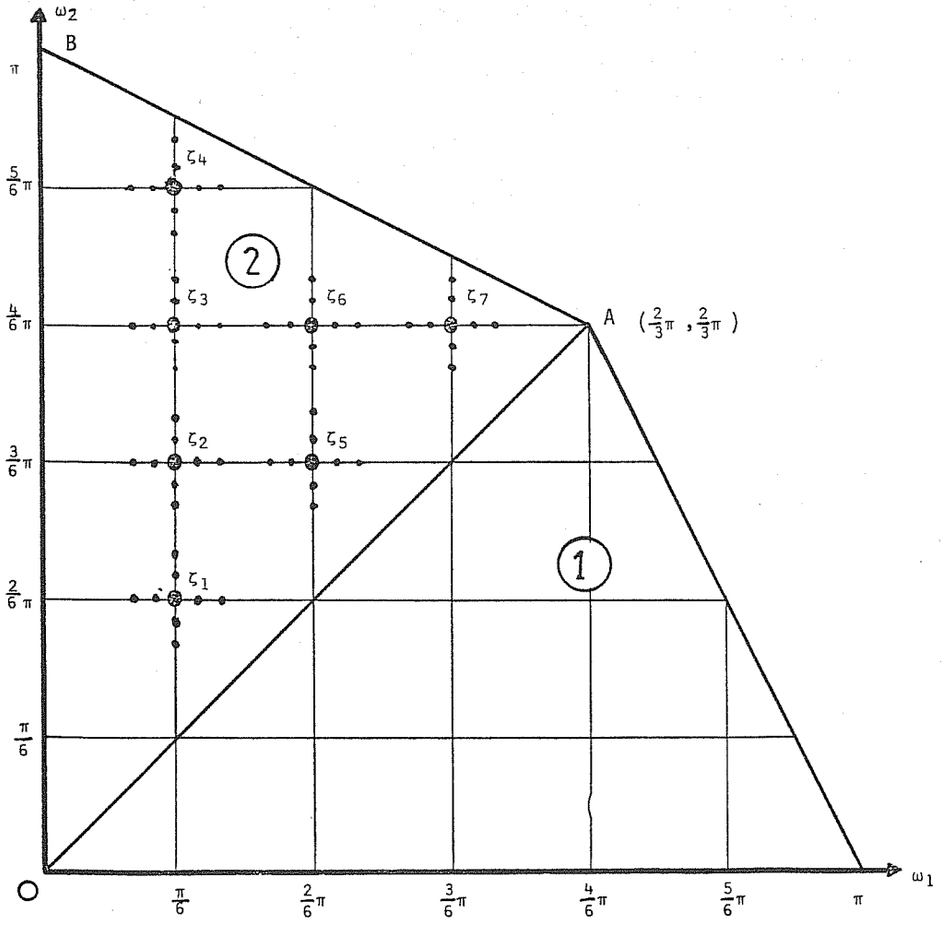


Fig. 4.1: Bispectrum sample with K = 6 and r = 2 which imply L = 4, P = 7, n = 9

Write $\eta_{ij} = f(\omega_i, \omega_j)$, and for each i (i = 1, 2, ..., L) define the vector

$$\underline{\eta}_i' = (\eta_{i,i+1}, \eta_{i,i+2}, \cdots, \eta_{i,\gamma(i)}), \tag{4.3.3}$$

and let the vector $\underline{\eta}$ be defined by

$$\underline{\eta}' = (\underline{\eta}_1', \underline{\eta}_2', \ldots, \underline{\eta}_L'). \tag{4.3.4}$$

We now relabel the elements of $\underline{\eta}$ as

$$\underline{\eta}' = (\zeta_1, \zeta_2, \ldots, \zeta_P) \tag{4.3.5}$$

where $P = \sum_{i=1}^{L} (\gamma(i) - i)$, so that for each ℓ ($1 \le \ell \le P$), $\zeta_\ell = \eta_{ij}$ for some i, j, satisfying $1 \le i \le L$; $i+1 \le j \le \gamma(i)$. (See Fig. 4.1 for an illustration).

We now form a set of (approximately) uncorrelated estimates of each ζ_i by constructing a "fine" frequency grid around each (ω_i, ω_j) point. Specifically, for each (ω_i, ω_j), let

$$\left.\begin{aligned} \omega_{i_p} &= \omega_i + \frac{pd\pi}{N}, \; p = -r, -r+1, \ldots, 0, 1, \ldots, r \\ \omega_{j_q} &= \omega_j + \frac{qd\pi}{N}, \; q = -r, -r+1, \ldots, -1, 1, \ldots, r, \; (q \neq 0) \end{aligned}\right\} \tag{4.3.6}$$

where the distance 'd' is chosen so that the bispectral estimates at neighbouring points on this fine grid are approximately uncorrelated. (In effect, this means that d must be chosen so that $\frac{\pi d}{N}$ must be greater than the band width of the spectral window corresponding to the lag window $\lambda(s)$). Now let $\hat{f}(\omega_{i_p}, \omega_{j_q})$ denote the estimated bispectral density function (as given by(2.3.4)) at the points $(\omega_{i_p}, \omega_{j_q})$. Assuming that the true bispectral density function is sufficiently smooth

so as to be effectively constant over the 'fine' frequency grid, we may write

$$E[\hat{f}(\omega_{i_p},\omega_{j_q})] = f(\omega_i,\omega_j) \quad \text{all } p, q.$$

We may thus regard the set of estimators $\{\hat{f}(\omega_{i_p},\omega_{i_q})\}$ as $n = 4r + 1$ approximately uncorrelated and unbiased estimates of $f(\omega_i,\omega_j)$.
To facilitate the analogy with standard multivariate tests, we now form the bispectral estimates

$$\{\hat{f}(\omega_i - \tfrac{rd\pi}{N},\omega_j),\ \hat{f}(\omega_i - \tfrac{(r-1)d\pi}{N},\omega_j),\ \ldots,\ \hat{f}(\omega_i,\omega_j),\ \ldots,\ \hat{f}(\omega_i + \tfrac{rd\pi}{N},\omega_j),$$

$$\hat{f}(\omega_i,\omega_j - \tfrac{rd\pi}{N}),\ \hat{f}(\omega_i,\omega_j - \tfrac{(r-1)d\pi}{N}),\ \ldots,\ \hat{f}(\omega_i,\omega_j + \tfrac{rd\pi}{N})\} \qquad (4.3.7)$$

into a n x 1 vector, which after relabelling, can be denoted by $\underline{\xi} = (\xi_1,\xi_2,\ldots,\xi_n)'$. Using this device, we obtain a n x 1 vector of estimates for each element $\zeta_\ell (1 \le \ell \le P)$ of the vector $\underline{\eta}$ defined by (4.3.4). We may thus form the complete set of bispectral estimates into a "data matrix", $\underset{\sim}{D}$,

$$D = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots & \xi_{1n} \\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n} \\ \vdots & \vdots & & \vdots \\ \xi_{p1} & \xi_{p2} & \cdots & \xi_{pn} \end{bmatrix} = \begin{bmatrix} \underline{\xi}_1' \\ \underline{\xi}_2' \\ \vdots \\ \underline{\xi}_p' \end{bmatrix} = \left[\underline{\xi}_{(1)}\ \underline{\xi}_{(2)} \cdots \underline{\xi}_{(n)}\right] \qquad (4.3.8)$$

where $\underline{\xi}'_{(i)} = (\xi_{1i}, \xi_{2i}, \ldots, \xi_{Pi})$ $(i = 1, 2, \ldots, n)$.
For large N, $\underline{\xi}_{(i)}$ $(i = 1, 2, \ldots, n)$ is distributed as complex normal with mean $\underline{\eta}$ and variance covariance matrix $\underline{\Sigma}_\xi$. Under the null hypothesis that the process is linear and $\mu_3 = 0$, the mean vector $\underline{\eta} = \underline{0}$. The maximum likelihood estimates of $\underline{\eta}$ and $\underline{\Sigma}_\xi$ are, respectively,

$$\hat{\underline{\eta}} = \frac{1}{n}\sum_{\ell=1}^{n} \underline{\xi}_{(\ell)},\quad \hat{\underline{\Sigma}}_\xi = \frac{\underline{A}}{n},\quad \underline{A} = \sum_{\ell=1}^{n} (\underline{\xi}_{(\ell)} - \hat{\underline{\eta}})(\underline{\xi}_{(\ell)} - \hat{\underline{\eta}})^*. \qquad (4.3.9)$$

The likelihood ratio test for testing the hypothesis $\underline{\eta} = \underline{0}$ against the

alternative $\underline{\eta}^* \underline{\Sigma}_\xi^{-1} \underline{\eta} > 0$ leads to the rejection of the hypothesis if the statistic (see Giri, 1965; Khatri, 1965)

$$T^2 = n\, \hat{\underline{\eta}}^*\, \underline{A}^{-1} \hat{\underline{\eta}}$$

is greater than a constant λ, where λ is determined by the significance level α. Under the null hypothesis the statistic

$$F_1 = \frac{2(n-P)}{2P}\, T^2$$

is distributed as a central F with (2P, 2(n-P)) degrees of freedom. If the null hypothesis is accepted, we proceed to the second stage of the testing procedure.

4.4 Test for Linearity

To test whether $\{X_t\}$ is linear, but allowing for $\mu_3 \neq 0$, we now use the property that the ratio X_{ij} given by (4.2.4) is constant for all ω_i and ω_j. As before, we form a column vector of order P x 1 of the ratios $\{X_{ij}\}$ (ω_i and ω_j are restricted by (4.3.1). Let us denote this column vector by $\underline{Y} = (Y_1, Y_2, \ldots, Y_P)'$, where, for each ℓ, $Y_\ell = X_{ij}$ for some pairs of integers (i,j). (It may be noted that in this column vector we may include the elements X_{ij} defined at the origin and all the points on the line $\omega_1 = 0$ since X_{ij}'s defined at these frequencies are always real.) We can now construct n estimates of each Y_ℓ from the bispectral estimates and spectral estimates at the n points in the "fine grid" $\{\omega_{i_p},\omega_{i_q}\}$. (These estimates of Y_ℓ are asymptotically normally distributed, Brillinger, 1965, p.1368). We thus obtain a random sample of n estimates of $\underline{Y}$, which we may denote by $\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_n$. If the null hypothesis is true then all the elements of the mean vector $\{Y_i\}$ are identical. This corresponds to a classical problem of symmetry

in multivariate analysis (Anderson, 1958; Kshirsagar, 1972).

$$\text{Let } \bar{\underline{Y}} = \frac{1}{n}\sum_{i=1}^{n} \underline{Y}_i,\ \hat{\underline{S}}_Y = \sum_{i=1}^{n} (\underline{Y}_i - \bar{\underline{Y}})(\underline{Y}_i - \bar{\underline{Y}})' \text{ and } \hat{\underline{\Sigma}}_Y = \frac{1}{n}\hat{\underline{S}}_Y. \qquad (4.4.1)$$

Define a column vector $\underline{\beta}$ of order Q x 1, where Q = P-1, such that $\underline{\beta} = \underline{B}\,\underline{Y}$, where $\underline{B}$ is a matrix of order Q x P and it is of the form

$$\underline{B} = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ 0 & 1 & -1 & \dots & 0 \\ \vdots & & \vdots & & \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

Under the null hypothesis $\underline{\beta}$ is asymptotically jointly normally distributed with mean vector $\underline{0}$ and variance covariance matrix $\underline{B}\,\underline{\Sigma}_Y\,\underline{B}'$. The likelihood ratio test for testing the null hypothesis leads to the rejection of the hypothesis if the statistic

$$T^2 = n\,\bar{\beta}'\,\hat{\underline{S}}^{-1}\,\bar{\beta}, \text{ where } \bar{\underline{\beta}} = \underline{B}\,\bar{\underline{Y}},\ \hat{S} = \underline{B}\,\hat{\underline{\Sigma}}_Y\underline{B}',$$

is greater than a constant λ_0, where λ_0 is determined by the significance level α. The statistic

$$F_2 = \frac{n-Q}{Q}\,T^2$$

has, under the null hypothesis, an F distribution with (Q,n-Q) degrees of freedom. The statistic T^2 is invariant and is independent of the choice of $\underline{B}$ (Anderson, 1958, p.111). The test is illustrated in Section 4.6.

4.5 Choice of the Parameters

We first choose K equally spaced points in the interval $(0,\pi)$ to form the "coarse" grid. Here, K must be substantially smaller than N in order to provide sufficient 'space' around each pair of frequencies (ω_i,ω_j) to enable spectra and bispectra to be estimated at the neighbouring points $(\omega_{i_p},\omega_{j_q})$ on the "fine" grid. The total number of points in each "fine" grid is (4r + 1) and there are $\frac{K^2}{3}$ grids. We thus require at least that $(4r + 1)\frac{K^2}{3} < N$. In addition, for the spectral and bispectral estimates at different points on the "fine" grid to be effectively uncorrelated, we require d to be chosen so that $\frac{\pi d}{N}$ is larger than the bandwidth of the spectral window. Also in order to be able to assume that $f(\omega)$, $f(\omega_1,\omega_2)$ are roughly constant over the points in the "fine" grid, r must be chosen so that $\frac{2\pi r}{N}$ is less than the "band width" of $f(\omega)$ and $f(\omega_1,\omega_2)$. Finally, to ensure that points in different "fine" grids do not overlap, we require $d \le \frac{N}{K(2r+1)}$.

A priori information on the bandwidths of $f(\omega)$, $f(\omega_1,\omega_2)$ would, of course, rarely be available, and in practice we would have to exercise discretion in the choice of the parameters. However, this situation is common to virtually all types of frequency domain analysis of time series, and the tests described in this chapter are certainly not unique in this respect. However, Table 4.1 (constructed taking into account the above considerations) shows some typical values of K, L and P, together with the corresponding constraints on the associated values of r and n. Also, Figs. 4.2 - 4.6 show clearly how the choice of K implies the values of L, P and $\gamma(i)$'s.

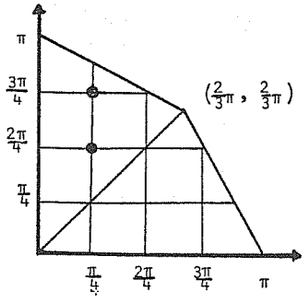


Figure 4.2

K = 4 ⇒ L = 2

γ(1) = 4 - 1 = 3 ⇒ 2 points

γ(2) = 4 - 1 - 1 = 2 —

∴ P = 2

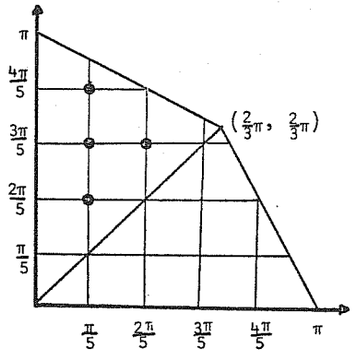


Figure 4.3

K = 5 ⇒ L = 3

γ(1) = 5 - 1 = 4 ⇒ 3 points

γ(2) = 5 - 1 - 1 = 3 ⇒ 1 point

γ(3) = 5 - 1 - 1 = 3 —

∴ P = 4

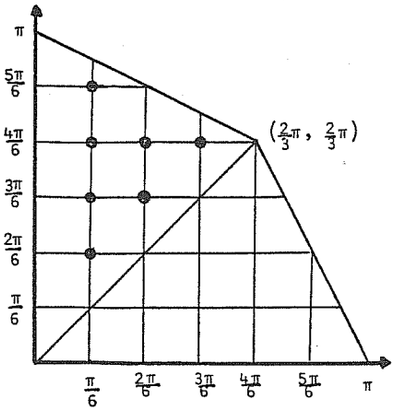


Figure 4.4

K = 6 ⇒ L = 4

γ(1) = 6 - 1 = 5 ⇒ 4 points

γ(2) = 6 - 1 - 1 = 4 ⇒ 2 points

γ(3) = 6 - 1 - 1 = 4 ⇒ 1 point

γ(4) = 6 - 2 - 1 = 3 —

∴ P = 7

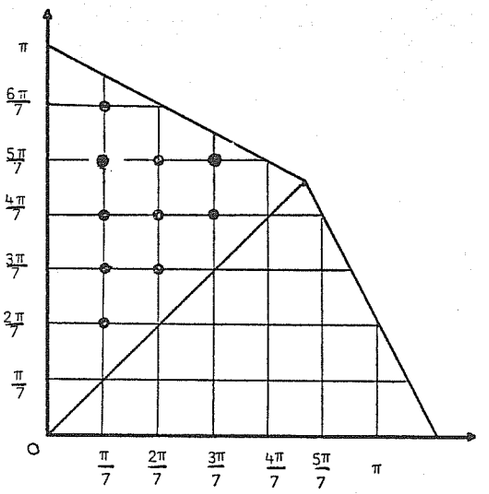


Figure 4.5

K = 7 ⇒ L = 4

γ(1) = 7 - 1 = 6 ⇒ 5 points

γ(2) = 7 - 1 - 1 = 5 ⇒ 3 points

γ(3) = 7 - 1 - 1 = 5 ⇒ 2 points

γ(4) = 7 - 1 - 2 = 4 —

∴ P = 9

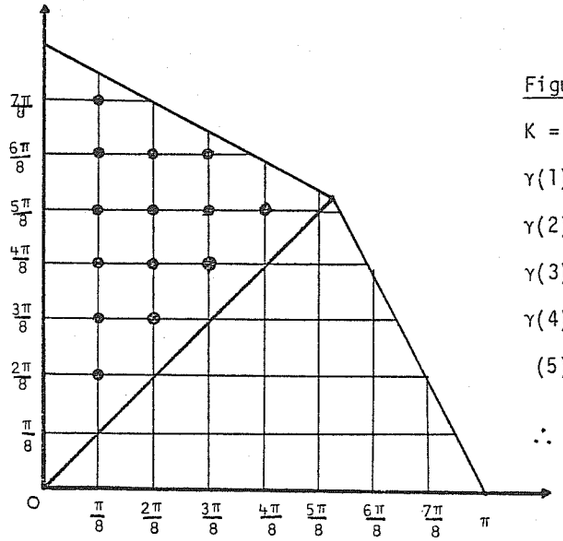


Figure 4.6

K = 8 ⇒ L = 5

γ(1) = 8 - 1 = 7 ⇒ 6 points

γ(2) = 8 - 1 - 1 = 6 ⇒ 4 points

γ(3) = 8 - 1 - 1 = 6 ⇒ 3 points

γ(4) = 8 - 1 - 2 = 5 ⇒ 1 point

(5) = 8 - 1 - 2 = 5 —

∴ P = 14

TABLE 4.1

Values of (K, L, P, r, n)

K	L	P	r	n
4	2	2	≥1	≥1
5	3	4	≥1	≥5
6	4	7	≥2	≥9
7	4	10	≥3	≥13
8	5	14	≥4	≥17

$L = \left[\frac{2K}{3}\right]$

$P = \Sigma\{\gamma(i)-i\}$

$n = 4r + 1$

4.6 Numerical Illustrations

In order to illustrate the above tests, we have generated two linear time series and two non-linear time series, each one with size N = 500.

Linear Time Series

Two stationary time series are generated as follows:-

Series A: The series $\{X_t\}$ is generated from the AR(2) model,

$$X_t - 1.3\, X_{t-1} + .4\, X_{t-2} = e_t$$

where $\{e_t\}$ is a sequence of independent and identically distributed N(0,1) random variables. This series is plotted in Figure 4.7.

Series B: This series is generated from the MA(1) model,

$$X_t = .9\, e_{t-1} + e_t$$

where $\{e_t\}$ is generated as above. This series is plotted in Figure 4.8.

Non-Linear Time Series

Two non-linear stationary time series are generated from the BL(1,0,1,1) model,

$$X_t = \alpha\, X_{t-1} + \beta\, X_{t-1}\, e_{t-1} + e_t \qquad (4.6.1)$$

where $\{e_t\}$ is generated as above.

Series C: The series $\{X_t\}$ is generated from (4.6.1) with $\alpha = .4$, $\beta = .4$. This series is shown in Figure 3.10.

Series D: The series $\{X_t\}$ is generated from (4.6.1) with $\alpha = .4$, $\beta = .6$, and is plotted in Figure 4.9.

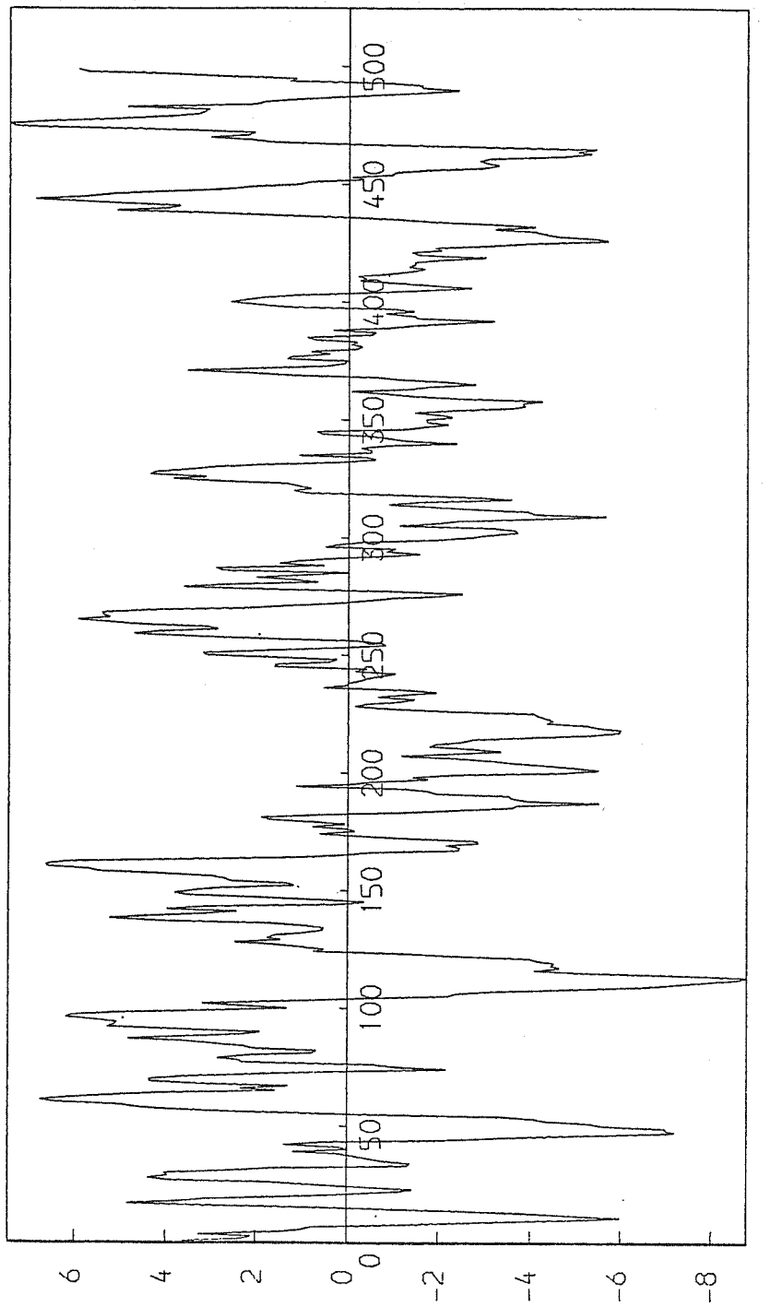


Figure 4.7: Series A, simulated from the AR(2) model, $X_t - 1.3\ X_{t-1} + .4\ X_{t-2} = e_t$

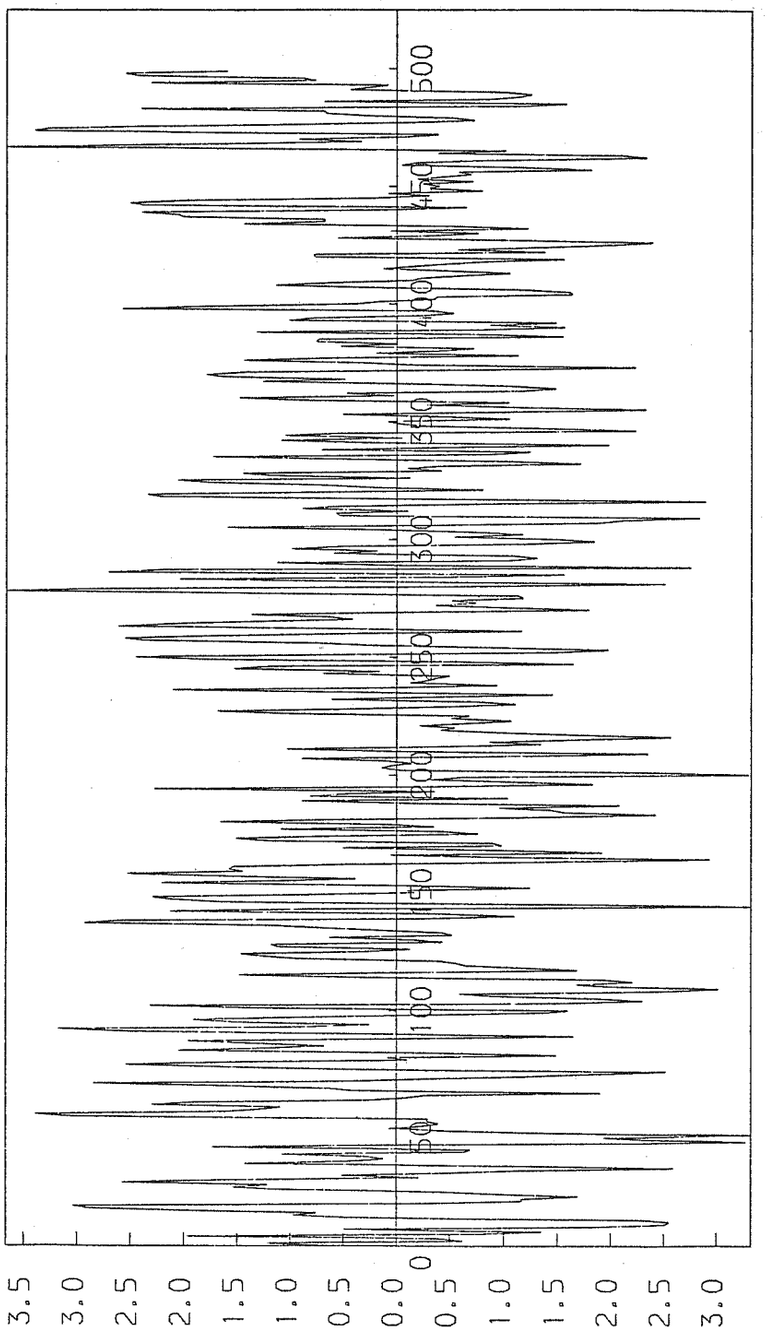


Figure 4.8: Series B, simulated from the MA(1) model, $X_t = .9\ e_{t-1} + e_t$

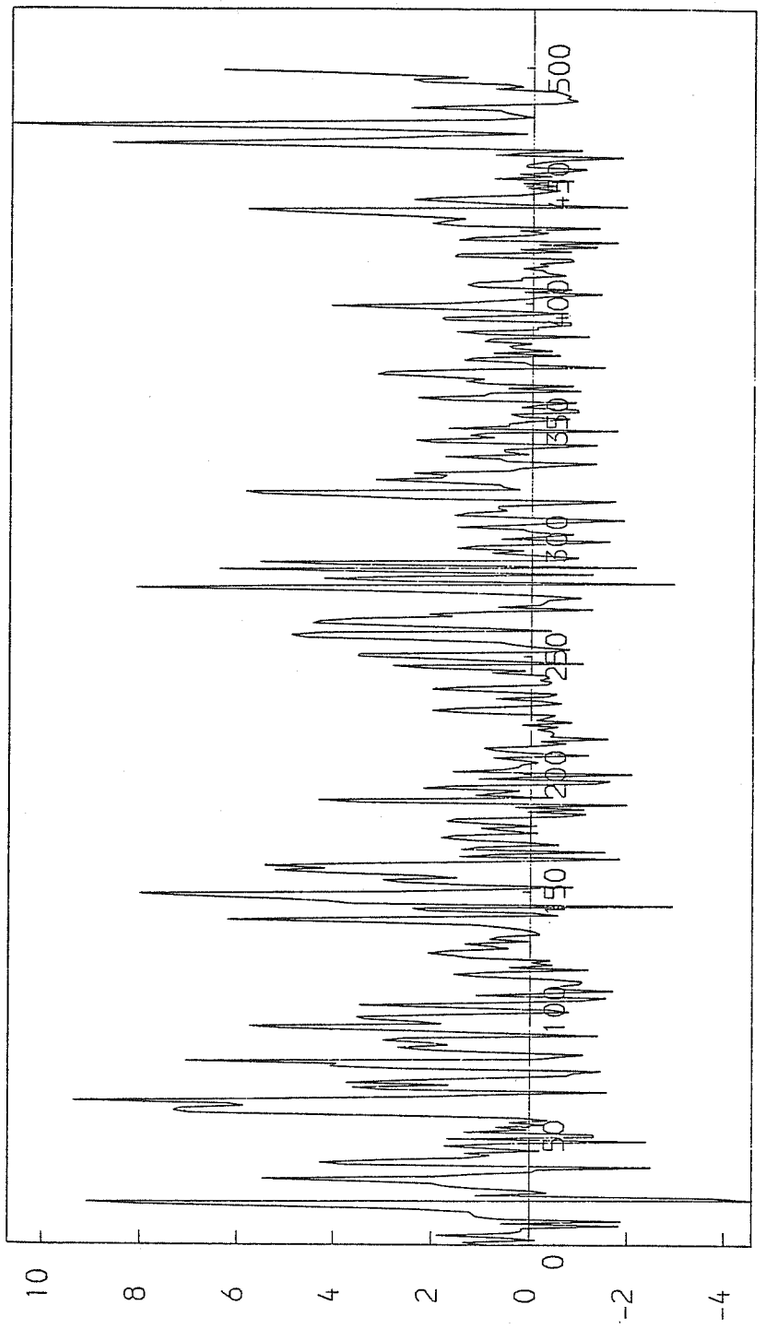


Figure 4.9: Series D, simulated from the BL(1,0,1,1), $X_t = .4\, X_{t-1} + .6\, X_{t-1}\, e_{t-1} + e_t$

Let us now consider the first stage of the test procedure for testing symmetry and linearity. For all these series the spectral density function is estimated using Daniell lag window, as explained in Chapter 2. The bispectral density function is also estimated, using the two-dimensional Daniell lag window and as described in Chapters 2 and 3. In estimating the spectral and bispectral density functions, we have used N = 500, M = 20 and truncation point = 30 for each series. The parameters K, L, d, r, P and n for constructing the T^2 and F_1 are as follows:

$$K = 6 \quad L = 4,\ P = 7,$$
$$r = 2 \quad n = 9,$$
$$d = 15.$$

For series A and C, the real and imaginary parts of the estimated bispectral density, at frequencies (ω_i, ω_j), as defined by (4.3.2), are given in Tables 4.2 and 4.4 respectively. Then the values of the bispectrum sample, at frequencies $(\omega_{i_p}, \omega_{j_q})$ as defined by (4.3.6), and after arranging them as in (4.3.7) and (4.3.8) are given in Tables 4.3 and 4.5 respectively. Then the sample mean $\hat{\underline{\eta}}$ and the Hermitian matrix A, definedby (4.3.9), are calculated and their values, together with the values of $A^{-1}\hat{\eta}$ are given in Tables 4.6 and 4.7 respectively. Finally, the value of $T^2 = n\, \underline{\eta}^*\, A^{-1}\hat{\underline{\eta}}$ is calculated and also given in the same tables. The values of T^2 and F_1 for series A, B, C and E are given in Table 4.8.

REAL PART				IMAGINARY PART				
.83	.000			.83	.001			
.67	-.024	.000	-.004	.67	.007	-.005	-.004	
.50	-.003	-.001		.50	-.047	.009		
.33	.012			.33	.065			
W2 W1	.167	.333	.500	W2 W1	.167	.333	.500	.667

TABLE 4.2: Bispectrum estimate $f(\omega_i, \omega_j)$ for the linear series A

SAMPLE BISPECTRUM DATA

REAL PART

.003	.031	.012	.002	.010	-.105	.022	-.015	.049
-.013	.038	-.003	-.031	.000	.024	.009	-.028	.010
.015	.009	-.024	.007	-.003	.006	.003	.001	-.012
.002	.022	.000	-.000	.004	.017	.016	-.021	-.001
.014	.004	-.001	-.006	-.007	.008	.006	-.012	-.013
-.000	.003	.000	-.001	.003	.001	.002	-.009	-.006
.002	-.000	-.004	-.001	-.003	.004	-.002	.003	.004

IMAGINARY PART

-.065	-.049	.065	.082	-.010	-.129	.005	-.017	-.009
.008	-.016	-.047	.016	-.008	-.022	-.040	.014	-.010
.057	.015	.007	.006	.006	-.013	.016	-.000	-.022
-.007	-.001	.001	-.008	-.008	.010	-.002	-.012	-.016
.005	.002	.009	.009	.001	.006	.003	.006	-.004
.010	-.003	-.005	.003	.002	-.003	-.000	.005	.001
.002	-.002	-.004	-.001	-.004	-.001	.001	.002	.002

TABLE 4.3: Bispectrum data matrix D, defined by (5.12), for the linear series A

REAL PART				IMAGINARY PART				
.83	-.021			.83	-.019			
.67	-.008	-.012	-.024	.67	-.028	-.011	-.002	
.50	-.030	-.010		.50	-.032	.019		
.33	.008			.33	.081			
W2 W1	.167	.333	.500	W2 W1	.167	.333	.500	.667

TABLE 4.4: Bispectrum estimate $f(\omega_i, \omega_j)$ for the bilinear series C

SAMPLE BISPECTRUM DATA

REAL PART

.180	.077	.008	.029	.071	.190	.102	.042	.057
.016	-.008	-.030	-.044	-.009	-.004	-.007	-.043	.004
-.063	-.033	-.008	.003	.002	.021	-.007	-.023	-.026
-.007	-.021	-.021	-.024	-.020	-.024	-.035	-.024	-.016
.026	.023	-.010	-.013	-.017	.001	.009	-.039	-.025
-.010	-.009	-.012	-.022	-.023	.007	.004	-.026	-.024
-.010	-.001	-.024	-.045	-.040	-.033	-.023	-.023	-.009

IMAGINARY PART

.015	.036	.081	.058	.029	-.001	.035	.063	.068
.030	.009	-.032	-.043	-.030	.039	-.022	-.008	-.009
.012	-.013	-.028	-.024	-.011	-.055	-.048	-.021	-.019
.019	-.006	-.019	-.012	.006	-.014	-.016	-.011	-.009
-.006	.026	.019	-.009	-.005	.004	.012	-.001	-.021
-.007	-.007	-.011	-.004	.008	-.008	-.003	-.009	-.002
.006	.000	-.002	-.010	-.018	.002	.002	-.007	.001

TABLE 4.5: Bispectrum data matrix D, defined by (5.12), for the bilinear series C.

THE HERMETIAN MATRIX A

1	.04828	0.00000	-.00071	-.00619	-.00171	-.00434	-.00198	-.00063	-.00055	-.00169	-.00028	-.00043	-.00078	-.00107
2	-.00071	.00619	.00811	0.00000	.00061	.00193	.00107	-.00196	.00060	-.00014	.00106	.00020	.00017	.00035
3	-.00171	.00434	.00061	-.00193	.00502	0.00000	.00035	.00016	.00062	.00105	.00055	.00002	.00014	-.00018
4	-.00198	.00063	.00107	.00196	.00035	-.00016	.00177	0.00000	.00073	.00045	.00019	.00041	-.00013	.00004
5	-.00055	.00169	.00060	.00014	.00062	-.00105	.00073	-.00045	.00084	0.00000	.00018	-.00000	-.00002	-.00003
6	-.00028	.00043	.00106	-.00020	.00055	-.00002	.00019	-.00041	.00018	.00000	.00032	0.00000	-.00001	.00010
7	-.00078	.00107	.00017	-.00035	.00014	.00018	-.00013	-.00004	-.00002	.00003	-.00001	-.00010	.00013	0.00000

$$A^{-1}\hat{\underline{n}}' = \begin{pmatrix} 1.77, & 1.20) \\ 3.88, & -7.03) \\ 18.96, & -.66) \\ 19.81, & -14.59) \\ -13.95, & 48.21) \\ -37.17, & 27.21) \\ 25.04, & -43.22) \end{pmatrix}, \quad \hat{\underline{n}} = \begin{pmatrix} .0010, & -.0142) \\ .0008, & -.0118) \\ .0003, & .0080) \\ .0042, & -.0049) \\ -.0008, & .0042) \\ -.0007, & .0011) \\ .0002, & -.0005) \end{pmatrix}$$

HOTELLINGS T SQUARE STATISTIC IS
T=(4.687,.000)

TABLE 4.6: The sum of squares matrix A, $A^{-1}\hat{\underline{n}}'$, $\hat{\underline{n}}$ and T^2 values for series A

THE HERMETIAN MATRIX A

1	.03805	0.00000	.00324	-.01496	-.00132	.00007	-.00028	-.00293	.00650	-.00237	.00387	-.00152	.00078	-.00182
2	.00324	.01496	.00981	0.00000	-.00143	-.00250	.00149	-.00045	.00231	.00273	.00039	.00121	.00235	.00085
3	-.00132	-.00007	-.00143	.00250	.00809	0.00000	.00060	.00241	-.00220	.00037	.00066	-.00134	-.00237	.00156
4	-.00028	.00293	.00149	.00045	.00060	-.00241	.00157	0.00000	-.00027	.00113	-.00006	-.00020	.00031	.00022
5	.00650	.00237	.00231	-.00273	-.00220	-.00037	-.00027	-.00113	.00547	0.00000	.00113	.00088	.00124	-.00032
6	.00387	.00152	.00039	-.00121	.00066	.00134	-.00006	.00020	.00113	-.00088	.00140	0.00000	-.00007	-.00066
7	.00078	.00182	.00235	-.00085	-.00237	-.00156	.00031	-.00022	.00124	.00032	-.00007	.00066	.00217	0.00000

$$A^{-1}\hat{\underline{n}}' = \begin{pmatrix} 24.42, & -1.13) \\ -.05, & -23.99) \\ -42.26, & -2.00) \\ 19.08, & -103.19) \\ 14.09, & 5.67) \\ -73.86, & -10.70) \\ -61.04, & 5.83) \end{pmatrix}, \quad \hat{\underline{n}} = \begin{pmatrix} .0839, & .0426) \\ -.0141, & -.0075) \\ -.0149, & -.0231) \\ -.0212, & -.0068) \\ -.0050, & .0021) \\ -.0129, & -.0048) \\ -.0232, & -.0028) \end{pmatrix}$$

HOTELLINGS T SQUARE STATISTIC IS
T=(49.460,.000)

TABLE 4.7: The sum of squares matrix A, $A^{-1}\hat{\underline{n}}'$, $\hat{\underline{n}}$ and T^2 values for series C

TABLE 4.8: Values of (T^2, F_1)

Series	T^2	F_1
A	4.687	1.4
B	5.713	1.6
C	49.460	14.1
D	96.248	27.5

Under the null hypothesis, the statistic F_1 has an F distribution with (14,4) degrees of freedom. As expected, the F_1 values for the series A and B are much less than the 5% point of F with (14,4) degrees of freedom (5.89). From this analysis, we can confidently conclude that series C and D are non-linear.

To confirm that the series A and B are linear and $\mu_3 = 0$, we go through the second stage. We test the hypothesis $f(\omega_1,\omega_2) = 0$ at $\omega_1 = \omega_2 = 0$ and on the line $\omega_1 = 0$ (for all ω_2). The values chosen are $\omega_1 = 0$ and $\omega_2 = 0.16\pi$, 0.48π and 0.80π. The values of the bispectrum estimate at the frequencies (note that the bispectrum on the boundary $\omega_1 = 0$ is real valued) are

ω_2	0.16π	0.48π	0.80π
$\hat{f}(0,\omega_2)$	.0676	-.0003	-.0018

A random sample of size 7 is chosen in the neighbourhood of the above frequencies (along the line $\omega_1 = 0$), giving P = 3 and n = 7. The values of the data matrix D, the matrix of sum of squares A, the sample mean vector $\hat{\underline{n}}$, $A^{-1}\hat{\underline{n}}'$ and T^2 values for series A are given in

BISPECTRUM DATA MATRIX D ON THE LINE $\omega_1 = 0$

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} -.804 & .562 & .199 & .068 & .112 & .054 & .041 \\ .022 & .018 & .006 & -.000 & .011 & .000 & -.005 \\ .001 & .002 & -.000 & -.002 & -.001 & -.002 & .001 \end{bmatrix}$$

THE COVARIANCE MATRIX A

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{bmatrix} 1.0155 & -.0073 & -.0005 \\ -.0073 & .0008 & .0001 \\ -.0005 & .0001 & .0000 \end{bmatrix}$$

$$A^{-1}\hat{\underline{n}}' = \begin{bmatrix} .154901 \\ 22.657772 \\ -90.598115 \end{bmatrix}, \quad \hat{\underline{n}} = \begin{bmatrix} .03322640 \\ .00738850 \\ -.00019906 \end{bmatrix}$$

HOTELLINGS T SQUARE STATISTIC IS

T= 1.33

TABLE 4.9: The data matrix D, the sum of squares matrix A, $A^{-1}\hat{\underline{n}}'$, $\hat{\underline{n}}$ and the statistic T^2 of the bispectrum estimate on the boundary line $\omega_1 = 0$ for series A.

Table 4.9. The same calculations for series B have been carried out and give the T^2 value as 0.88. The

$$F = \frac{n-P}{P} T^2$$

values for series A and B are respectively 1.77, 1.17. The 5% upper point of F with (3,4) degrees of freedom is 6.59. The F values are much smaller than the 5% points, confirming that in fact the series A and B are linear and $\mu_3 = 0$.

4.7 Applications to Real Time Series

The tests are also applied on Wolfer sunspot numbers, the Canadian lynx data and the transformed unemployment figures in W. Germany, which were considered in detail in Chapter 3. The spectral density and the bispectral density functions are estimated using one-dimensional and two-dimensional Daniell lag windows. The parameters M, d are given in Table 4.10. The values of K is chosen, as in Section 4.6, to be K = 6, which implies L = 4 and P = 7. Also, r = 2, which implies that n = 9 for all the above real series. The values of T^2, F_1 and 5% upper point of the F distribution with (14,4) degrees of freedom are given in Table 4.11).

Series	N	M	d
Sunspot Numbers	256	20	8
Canadian Lynx Data	114	16	3.5
Logarithm of Canadian Lynx Data	114	16	3.5
Transformed Unemployment Figures in W. Germany	389	20	12

TABLE 4.10: Values of (N, M, d)

TABLE 4.11: Values of (T^2, F_1)

	T^2	F_1	5% upper point of F(14.4)
Sunspot Numbers	7531.7	2151.9	5.89
Canadian Lynx Data	3044.6	869.9	5.89
Logarithm of Canadian Lynx Data	1073.7	306.8	5.89
Transformed Unemployment Figures in W. Germany	1224.5	349.9	5.89

The values of F_1 for the four series are much greater than the percentage point confirming the general belief that the series are non-linear. It is interesting to see that the logarithmic transformation of the Canadian lynx data has reduced F_1 value considerably, but still the series is not a linear Gaussian time series.

We now proceed to test the hypothesis that the series may be linear but $\mu_3 \neq 0$, i.e. X_{ij}'s are constant. To test the hypothesis we use the statistic F_2 which, under the null hypothesis, is distributed as F with (Q,n-Q) (Q = P-1) degrees of freedom. The same parameter values as given above are used, and the values of T^2 and F_2 are given in Table 4.12.

TABLE 4.12: Values of (T^2, F_2)

	T^2	F_2	5% upper point of F(6,3)
Sunspot Numbers	536.72	268.36	8.94
Canadian Lynx Data	1068.58	534.29	8.94
Logarithm of Canadian Lynx Data	9.72	4.86	8.94
Transformed Unemployment Figures of West Germany	443.05	221.53	8.94

From Table 4.12, it is clear that the sunspot numbers, Canadian lynx and the transformed unemployment figures in W. Germany are obviously non-linear, but, surprisingly, the logarithm of the Canadian lynx data is linear, though not Gaussian.

Finally, it must be pointed out that the above tests are not exhaustive in the sense that the bispectral density function can be zero for some non-linear processes. Still, the simulation studies show that in quite a number of situations tests based on bispectral density can provide a useful guide to the presence of non-linearity.

Recently, Hinich (1982) has modified the tests given above for tests for symmetry and linearity. He made use of the asymptotic expression for the variance and covariance matrix of the random vector $\underline{\xi}$ and the test reduces to a χ^2.

CHAPTER 5

BILINEAR TIME SERIES MODELS

5.1. NON-LINEAR REPRESENTATIONS IN TERMS OF INDEPENDENT RANDOM VARIABLES

In the theory of stationary random processes, Wold's theorem (Wold, 1938) plays a fundamental role. Briefly, the theorem can be stated as follows (for a proof of the statement, see e.g. Priestley, 1981). Let X_t be a zero mean second order stationary process. Then X_t can be expressed in the form

$$X_t = U_t + V_t \tag{5.1.1}$$

where

(i) U_t and V_t are uncorrelated processes

(ii) U_t is non-deterministic with a one-sided linear representation

$$U_t = \sum_{u=0}^{\infty} a_u \eta_{t-u}$$

with $a_0 = 1$, $\sum a_u^2 < \infty$, η_t is an uncorrelated process. The process η_t is uncorrelated with V_t,.i.e. $E(\eta_s V_t) = 0$, all s,t. The sequences $\{a_u\}$ and $\{\eta_t\}$ are uniquely determined.

(iii) V_t is deterministic, i.e. can be predicted from its own past with zero prediction variance.

For the present discussion, we assume that V_t is absent. It is important to note that the Wold's theorem only states that a second order stationary process has a one-sided infinite order moving average representation in terms of an uncorrelated process $\{\eta_t\}$, but it does not mean that the process necessarily conforms to a linear model.

The basic object in time series model building can be described as follows. Suppose $\{X_t\}$ denotes the given time series, then the object is to seek a function $h(\cdot)$ which is such that

$$h(X_t, X_{t-1}, \ldots) = e_t \tag{5.1.2}$$

where $\{e_t\}$ is a sequence of i.i.d. random variables. The class of linear models is given by restricting $h(\cdot)$ to be a linear function of $X_t, X_{t-1}, \ldots$, in which case (5.1.2) reduces to

$$\sum_{u=0}^{\infty} h_u X_{t-u} = e_t, \tag{5.1.3}$$

or

$$H(B)\, X_t = e_t,$$

where $H(z) = \sum_{u=0}^{\infty} h_u z^u$. If $H(z) \neq 0$, $|z| < 1$, (5.1.3) may be written as

$$X_t = H^{-1}(B)\, e_t = \Gamma(B)\, e_t = \sum_{u=0}^{\infty} g_u e_{t-u} \tag{5.1.4}$$

where $\Gamma(z) = \sum g_u z^u$.

The well known class of autoregressive moving average models of the order p, q (ARMA(p,q)) which were discussed earlier can be derived from (5.1.4) by choosing $\Gamma(z)$ to be,

$$\Gamma(z) = \frac{1 + b_1 z + \ldots + b_q z^q}{1 + a_1 z + \ldots + a_p z^p} \tag{5.1.5}$$

The resulting model for X_t now becomes

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q} \tag{5.1.6}$$

The autoregressive model of order p (AR(p)) is obtained from (5.1.6) by putting $b_1 = b_2 = \ldots = b_q = 0$.

Although the representations (5.1.1) and (5.1.4) look similar, the fundamental difference is that in the linear model (5.1.4) the $\{e_t\}$ are <u>independent</u> whereas the $\{\eta_t\}$ in the representation (5.1.1) are <u>merely uncorrelated</u>. As far as second order properties are concerned, e_t and η_t have identical properties, but they may differ substantially in many other respects. (See Granger and Andersen, 1978a; Priestley, 1978). If the process X_t is Gaussian, then of course the $\{\eta_t\}$ are independent and hence all Gaussian stationary processes conform to linear models. If $\{X_t\}$ is non-Gaussian, the representation (5.1.4) may no longer be valid. Motivated by the work of Volterra (1930) in series expansion of continuous functions, Wiener (1958) has made a systematic study of the non-linear representation of the continuous parameter stochastic process $\{X_t\}$. Wiener's representation, which is in terms of the Hermite polynomials, can be considered as an analogue of Wold's decomposition theorem (see Rosenblatt, 1979). Volterra (1930) has shown that under certain conditions the process $\{X_t\}$ can be written as

$$X_t = \sum_{i=1}^{\infty} \left[\sum_{u_1} \cdots \sum_{u_i} g_i(u_1, u_2, \ldots, u_i) \prod_{j=1}^{i} e_{t-u_j} \right] \tag{5.1.7}$$

The equation (5.1.7) is known as Volterra expansion and the kernels $\{g_1(u)\}$, $\{g_2(u_1,u_2)\}$ etc. are known as Volterra kernels.

It is well known that the transfer functions play a very useful role in describing linear models, but in the case of non-linear models there is no single transfer function which can completely characterise the model. Hence, we have to define an infinite sequence of generalised transfer functions (Brillinger, 1970; Priestley, 1978). The k-th order (k = 1,2,..,) generalised transfer functions can be defined as

$$\Gamma_k(\omega_1, \omega_2, \ldots, \omega_k) = \sum_{u_1} \sum_{u_2} \cdots \sum_{u_k} g_k(u_1, u_2, \ldots, u_k)\, e^{-i(u_1\omega_1 + \ldots + u_k\omega_k)} \qquad (k = 1,2,\ldots) \tag{5.1.8}$$

Although the Volterra and Wiener expansions are very general and mathematically elegant, in practice it is difficult to estimate the infinite set of parameters of Γ_k. In view of this, these expansions are not widely used in the context of time series analysis for representing non-linear time series. Hence we seek instead a finite parameter representation of sufficient generality to describe non-linear time series models. Recently control theorists have introduced the class of bilinear models (Mohler, (1973), Ruberti, et al (1972)) which have been found useful for describing many non-linear phenomena. Since the "bilinear models" are nearly linear, their structural properties are similar to those of linear models. In the following section we describe some types of bilinear time series models. The analysis of bilinear time series models has been considered by Granger and Andersen (1978a, 1978b), Subba Rao (1977, 1981a, 1981b).

Recently other non-linear models have been proposed by Haggan and Ozaki (1978), Priestley (1980), Tong and Lim (1980). The model proposed by Priestley (1980) known as "State dependent model" can be considered as quite a general non-linear model. Here we shall restrict our discussion to bilinear models only.

5.2 BILINEAR TIME SERIES MODELS

Let X_t be a discrete parameter time series satisfying the difference equation

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} = \sum_{j=0}^{r} c_j e_{t-j} + \sum_{\ell=1}^{k} \sum_{\ell'=1}^{k} b_{\ell\ell'} X_{t-\ell} e_{t-\ell'} \qquad (5.2.1)$$

where $\{e_t\}$ is a set of independent random variables and $c_0 = 1$.

We define the model (5.2.1) as a bilinear time series model BL(p,r,m,k) and the process $\{X_t\}$ as a bilinear process.

As shown earlier, it is convenient to study the properties of the models, such as conditions for stationarity, invertibility, etc., via equivalent state-space representations. Using the vector form of the bilinear model BL(p,0,p,1), Subba Rao (1981a) has derived the conditions for stationarity and the expressions for the covariances. The details are as follows.

Consider the bilinear model BL(p,0,p,1), i.e.

$$X_t + \sum_{j=1}^{p} a_j X_{t-j} = e_t + \sum_{\ell=1}^{p} b_{\ell 1} X_{t-\ell} e_{t-1} \qquad (5.2.2)$$

Let us define the matrices

$$\underline{A} = \begin{pmatrix} -a_1 & -a_2 & \cdots & -a_p \\ 1 & 0 & \cdots & 0 \\ & & & \\ 0 & 0 & & 0 \end{pmatrix} \qquad \underline{B} = \begin{pmatrix} b_{11} & b_{21} & b_{31} & \cdots & b_{p1} \\ 0 & 0 & 0 & \cdots & 0 \\ & & & & \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} \qquad (5.2.3)$$

and $\underline{C}' = (1,0,0,\ldots,0)$, $\underline{H}' = (1,0,\ldots,0)$, and let $\underline{x}_t' = (X_t, X_{t-1}, \ldots, X_{t-p+1})$. With this notation, we can write the model (5.2.2) in the form

$$\underline{x}_t = \underline{A}\,\underline{x}_{t-1} + \underline{B}\,\underline{x}_{t-1}\, e_{t-1} + \underline{C}\, e_t, \qquad (5.2.4)$$

$$X_t = \underline{H}'\,\underline{x}_t.$$

We define the model (5.2.4) as a vector form of the bilinear model BL(p,0,p,1) and denote it by VBL(p) (the initial letter emphasising the fact that (5.2.4) is written in the vector form).

It must be pointed out here that even though the random variables $\{e_t\}$ are an independent sequence, the representation $\{\underline{x}_t\}$ given by (5.2.4) is not a Markovian representation. However, as shown by Tuan Dinh Pham and Lanh Tat Tran (1981), by defining a new state vector $\underline{Z}_t = (\underline{A}+\underline{B}e_t)\underline{x}_t$, we can write (5.2.4) as

$$\underline{Z}_t = (A + \underline{B}e_t)\underline{Z}_{t-1} + (\underline{A} + \underline{B}e_t)e_t \qquad (5.2.4')$$

$$\underline{x}_t = \underline{Z}_{t-1} + \underline{C}e_t$$

The equivalent representation (5.2.4)' of $\underline{Z}_t$ is a Markovian representation. In fact this simple fact that $\underline{Z}_t$ is a Markovian representation can be used in obtaining all the moments of the process $\{\underline{Z}_t\}$ as shown by Tuan Dinh Pham (1983) in the case when $\underline{Z}_t$ is a scalar valued process.

Now the question arises whether all bilinear processes of the form (5.2.1) does admit a Markovian representation, and this aspect of the bilinear models will be considered in chapter 7, (for a detailed account see Tuan Dinh Pham (1983)).

Suppose we have the bilinear model BL(p,0,p,q). We define the matrix $\underline{A}$, and the vectors $\underline{C}$, $\underline{H}$, and $\underline{x}_t$ as before. Define the matrices

$$\underline{B}_j = \begin{bmatrix} b_{1j} & b_{2j} & \dots & b_{pj} \\ 0 & 0 & \dots & 0 \\ & & & \\ 0 & 0 & \dots & 0 \end{bmatrix}, \quad (j = 1,2,\dots,q) \qquad (5.2.5)$$

Then the vector form of the bilinear model BL(p,0,p,q) is (VBL(p,q)),

$$\underline{x}_t = \underline{A}\,\underline{x}_{t-1} + \sum_{j=1}^{q} \underline{B}_j\,\underline{x}_{t-1}\,e_{t-j} + \underline{C}\,e_t \qquad (5.2.6)$$

$$X_t = \underline{H}'\underline{x}_t$$

To study some features of bilinear time series, we have generated time series $\{X_t\}$ (see Subba Rao, 1979) from the models

(i) $X_t = 0.4\,X_{t-1} + 0.8\,X_{t-1}\,e_{t-1} + e_t \quad (t=1,2,\dots,1000)$

(ii) $X_t = 0.8\,X_{t-1} - 0.4\,X_{t-2} + 0.6\,X_{t-1}\,e_{t-1} + 0.7\,X_{t-2}\,e_{t-1} + e_t \quad (t=1,2,\dots,1000).$

The series (i) and (ii) are plotted in Fig. 5.1 and Fig. 5.2 respectively. An examination of the series 1, shows that at certain time points, there are high amplitude oscillations. In contrast to the series (i), the behaviour of the series(ii) is very remarkable. The sort of behaviour seen from this model is of common occurrence in seismological data, in particular explosion and earthquake data. For this type of data, the activity due to an event is of very short duration, and the rest of the record can be due to noise. For comparison purposes in Fig. 5.3, we give a seismic record obtained from an underground explosion that was carried out in the U.S.A. on 28th October 1966. The record is that of a P wave (pressure wave) and this event is nowadays commonly known as "Longshot". The analysis of this data has been reported in Subba Rao (1981) and Dargahi-Noubary, Laycock and Subba Rao (1978).

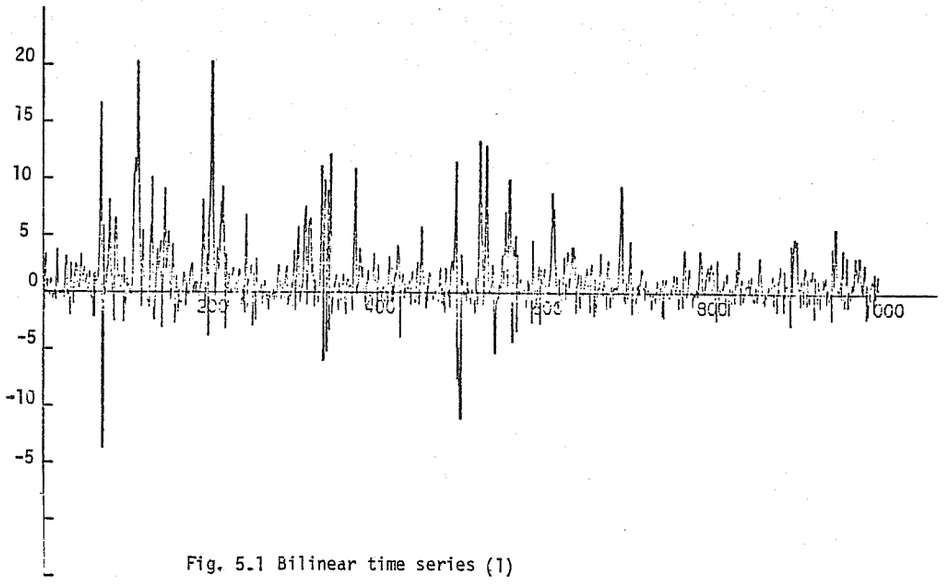


Fig. 5.1 Bilinear time series (1)

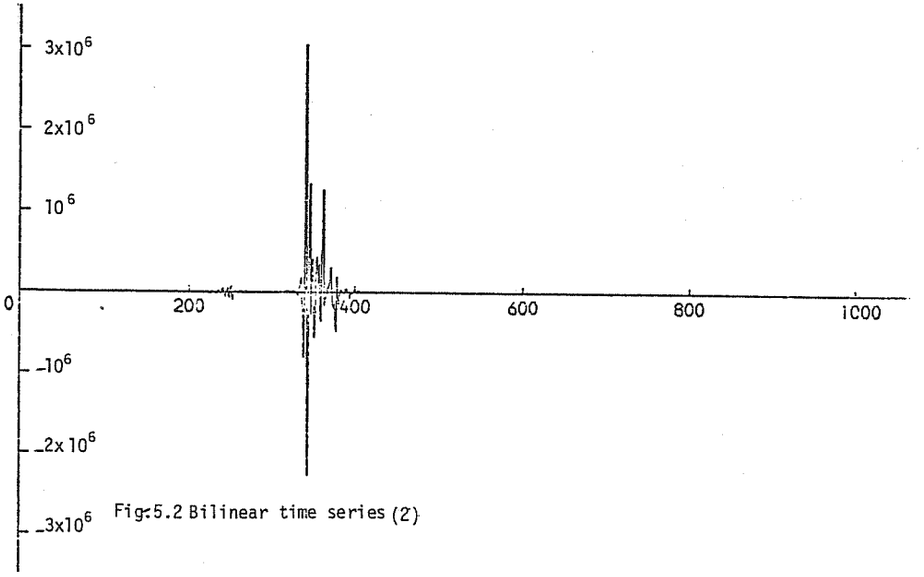


Fig.5.2 Bilinear time series (2)

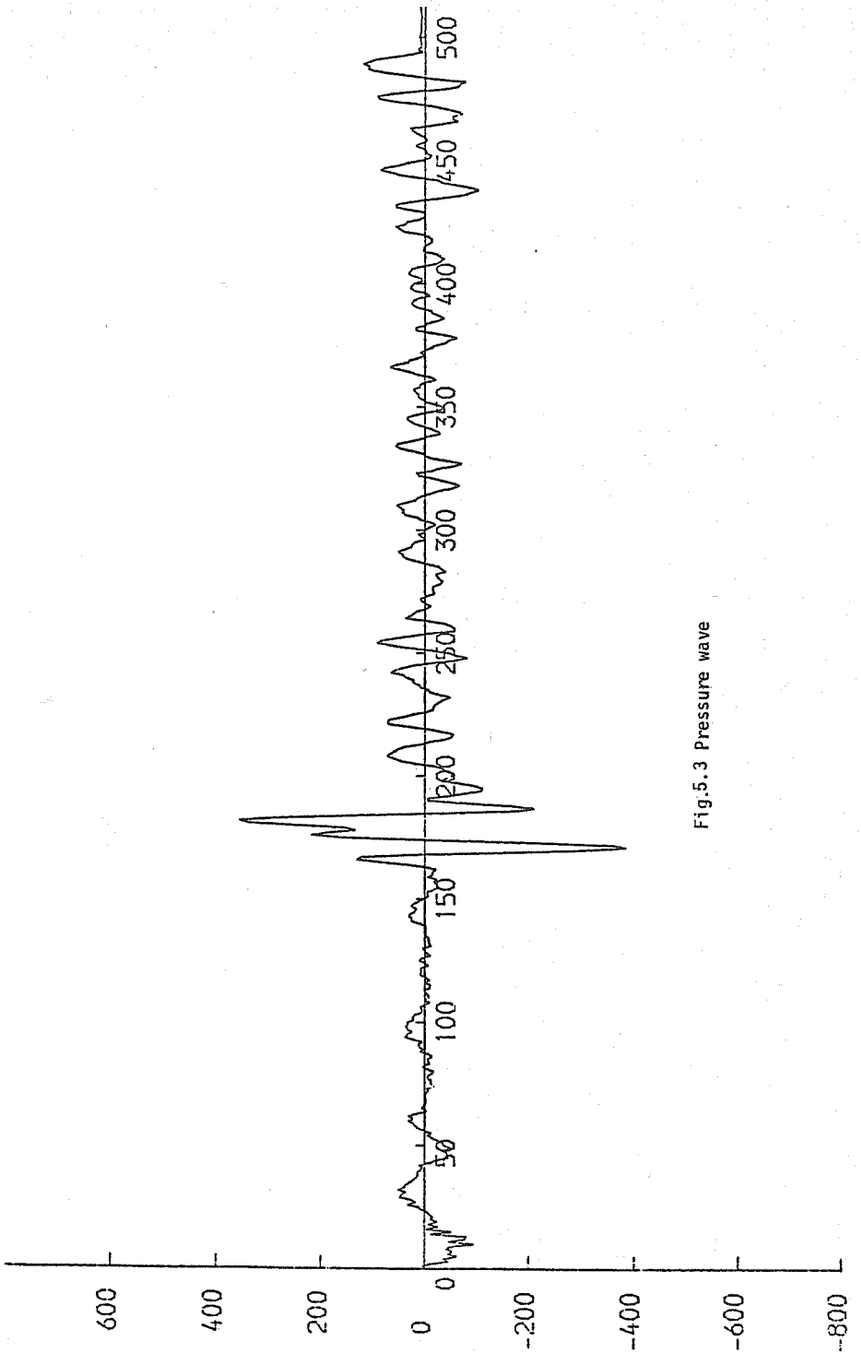


Fig.5.3 Pressure wave

5.3 VOLTERRA SERIES EXPANSION OF VBL(p) MODELS

In this section we obtain the Volterra expansion for VBL(p) model (see Subba Rao, 1981b), and thus derive the higher order transfer functions which characterise the model.

We now consider the VBL(p) model given by (5.2.4) and the solution of the equation (5.2.4) is sought in the form of a power series expansion. In this paper we use the "reversion method". (For further details see Cunningham, 1958, p.133). We consider the model

$$\underline{x}_t = \underline{A}\,\underline{x}_{t-1} + \lambda \underline{B}\,\underline{x}_{t-1}\,e_{t-1} + \lambda \underline{C}\,e_t \tag{5.3.1}$$

where λ is a numerical parameter introduced to facilitate the solution, but ultimately λ is allowed to become unity. A solution for (5.3.1) is sought in the form

$$\underline{x}_t = \sum_{j=1}^{\infty} \lambda^j \underline{x}_j(t). \tag{5.3.2}$$

Substituting (5.3.2) into (5.3.1) and then equating powers of λ on both sides, we get

$$\underline{x}_1(t) = \underline{A}\,\underline{x}_1(t-1) + \underline{C}\,e_t \tag{5.3.3}$$

$$\underline{x}_i(t) = \underline{A}\,\underline{x}_i(t-1) + \underline{B}\,\underline{x}_{i-1}(t-1)\,e_{t-1} \quad (i = 2,3,\ldots) \tag{5.3.4}$$

Assuming $\underline{x}_i(0) = \underline{0}$ $(i = 1,2,\ldots)$ and $e_t = 0$ $(t \le 0)$, we can show that

$$\underline{x}_1(t) = \sum_{j_1=0}^{t} \underline{A}^{j_1}\,\underline{C}\,e_{t-j_1}$$

$$\underline{x}_i(t) = \sum_{j_1=0}^{t-1} \underline{A}^{j_1}\underline{B}\,\underline{x}_{i-1}(t-1-j_1)\,e_{t-1-j_1}$$

$$= \sum_{u_1=0}^{t} \cdots \sum_{u_i=0}^{t} \underline{A}^{u_1-1}\,\underline{B}\,\underline{A}^{u_2-u_1-1}\,\underline{B} \cdots \underline{A}^{u_{i-1}-u_{i-2}-1}\,\underline{B}$$

$$\underline{A}^{u_i-u_{i-1}}\,\underline{C}\,\delta(u_1-1)\prod_{j=2}^{i-1}\delta(u_j-u_{j-1}-1)\,\delta(u_i-u_{i-1})\prod_{j=1}^{i} e_{t-u_j} \tag{5.3.5}$$

where the step function $\delta(u)$ is defined as

$$\delta(u) = \begin{cases} 1 & \text{if } u \ge 0 \\ 0 & \text{otherwise} \end{cases}$$

The Volterra series expansion of $\underline{x}_t$ is obtained by putting $\lambda = 1$ in (5.3.1) and (5.3.2) and the final solution of X_t (assuming we observe $X_t = \underline{H}'\underline{x}_t$) can be written in the form

$$X_t = \sum_{i=1}^{\infty} \underline{H}'\underline{x}_i(t) = \sum_{i=1}^{\infty} \left[\sum_{u_1=0}^{t} \cdots \sum_{u_i=0}^{t} V_i(u_1 \ldots u_i) \prod_{j=1}^{i} e_{t-u_j} \right] \tag{5.3.6}$$

where the Volterra kernels $V_i(u_1 \ldots u_i)$ are given by

$$V_1(u_1) = \underline{H}'\,\underline{A}^{u_1}\,\underline{C}$$

$$V_i(u_1,u_2,\ldots,u_i) = \underline{H}'\,\underline{A}^{u_1-1}\,\underline{B}\,\delta(u_1-1)\prod_{j=2}^{i-1}\left[\underline{A}^{u_j-u_{j-1}-1}\,\underline{B}\,\delta(u_j-u_{j-1}-1)\,\underline{A}^{u_i-u_{i-1}-1}\,\underline{C}\right]$$

$$(i = 2,3,..) \tag{5.3.7}$$

A comparison of the expansion (5.3.6) with (5.1.7) shows that for bilinear processes the coefficients $g_i(u_1,\ldots,u_i)$ of the Volterra expansion (5.1.7)

are zero if any of $u_1,u_2,\ldots,u_i$ is zero.

The kernels in the expansion (5.3.6) are not symmetric, but this expansion can be written in terms of symmetric kernels by defining

$$W_1(u_1) = V_1(u_1)$$
$$W_1(u_1,u_2,\ldots,u_i) = \frac{1}{i!} \sum_{per} V_i(u_1,u_2,\ldots,u_i) \qquad (5.3.8)$$

when the summation $\sum_{per}$ is taken over all possible permutations of the variables $u_1,u_2,\ldots,u_i$. The Volterra expansion can now be written as

$$X_t = \sum_{i=1}^{\infty} \left[\sum_{u_1=0}^{t} \cdots \sum_{u_i=0}^{t} W_i(u_1,u_2,\ldots,u_i) \prod_{j=1}^{i} e_{t-u_j} \right] \qquad (5.3.9)$$

where, for example, the first two kernels are given by

$$\underline{W}_1(u_1) = \underline{H}' \underline{A}^{u_1} \underline{C}$$

$$W_2(u_1,u_2) = \tfrac{1}{2} \left[\underline{H}' \underline{A}^{u_1-1} \underline{B} \underline{A}^{u_2-u_1} \underline{C}\, \delta(u_1-1)\, \delta(u_2-u_1) + \underline{H}' \underline{A}^{u_2-1} \underline{B} \underline{A}^{u_1-u_2} \underline{C}\, \delta(u_2-1)\, \delta(u_1-u_2) \right] . \qquad (5.3.10)$$

We can now define the k-th order transfer function as

$$\Gamma_k(\omega_1,\omega_2,\ldots,\omega_k) = \sum_{u_1=0}^{\infty} \cdots \sum_{u_k=0}^{\infty} W_k(u_1,u_2,\ldots,u_k)\, e^{-i(u_1\omega_1 + \ldots + u_k\omega_k)} \qquad (k = 1,2,\ldots) \qquad (5.3.11)$$

The transfer functions corresponding to the kernels (5.3.10) (assuming the spectral radius of $\underline{A}$ is less than 1) are

$$\Gamma_1(\omega) = \underline{H}' (\underline{I} - \underline{A}\, e^{-i\omega})^{-1} \underline{C} \qquad (5.3.12)$$

$$\Gamma_2(\omega_1,\omega_2) = \tfrac{1}{2} \left[\underline{H}' (\underline{I} - \underline{A}\, e^{-i(\omega_1+\omega_2)})^{-1} \underline{B} (\underline{I} - \underline{A}\, e^{-i\omega_1})^{-1} \underline{C} + \underline{H}' (\underline{I} - \underline{A}\, e^{-i(\omega_1+\omega_2)})^{-1} \underline{B} (\underline{I} - \underline{A}\, e^{-i\omega_2})^{-1} \underline{C} \right] e^{-i(\omega_1+\omega_2)} \qquad (5.3.13)$$

For the BL(1,0,1,1) model, the first two transfer functions are

$$\Gamma_1(\omega) = \frac{1}{1 + a\, e^{-i\omega}}$$

$$\Gamma_2(\omega_1,\omega_2) = \frac{b\, e^{i(\omega_1+\omega_2)}}{2(1 + a\, e^{-i(\omega_1+\omega_2)})} \left[\frac{1}{1 + a\, e^{-i\omega_1}} + \frac{1}{1 + a\, e^{-i\omega_2}} \right] \qquad (5.3.14)$$

These transfer functions agree with the corresponding expressions derived by Priestley (1978) for BL(1,0,1,1) model.

Although an infinite number of transfer functions have to be defined in the case of non-linear systems, one can see that in the case of bilinear models, when the norm of $\underline{B}$ is less than unity, the higher order transfer functions $\Gamma_k(\omega_1,\ldots,\omega_k)$ tend to zero as k tends to infinity. Hence, as such, a finite number of transfer functions would adequately characterise the bilinear model VBL(p).

5.4 EXPRESSIONS FOR COVARIANCES AND CONDITIONS FOR STATIONARITY

In this section we obtain the conditions for asymptotic stationarity of the time series X_t satisfying the model (5.2.4). We have

$$E(X_t) = \underline{H}'\, E(\underline{x}_t),$$
$$\mathrm{cov}(X_t, X_{t+s}) = \underline{H}' \left[E(\underline{x}_t - E(\underline{x}_t))(\underline{x}_{t+s} - E(\underline{x}_{t+s}))' \right] \underline{H}.$$

In the following derivation we assume that the random variables $\{e_t\}$ are independent and each e_t is distributed N(0,1), and obtain expressions for $\mathrm{cov}(\underline{x}_t\, \underline{x}_{t+s})$.

Let $\underline{\mu}_t = E(\underline{x}_t)$, $\underline{V}_t = E(\underline{x}_t \underline{x}'_t)$

$$\underline{S}_t = E\left[\underline{x}_t\, \underline{x}'_t\, e_t\right], \quad \underline{W}_t = E\left[\underline{x}_t\, \underline{x}'_t\, e_t^2\right] .$$

Taking expectations on both sides of(5.2.4)and noting $E(\underline{x}_t\ e_t) = \underline{C}$, we obtain

$$\underline{\mu}_{t+1} = \underline{A}\ \underline{\mu}_t + \underline{B}\ \underline{C} = \underline{A}^t\ \underline{\mu}_1 + \left(\sum_{j=0}^{t-1} \underline{A}^j\right) \underline{B}\ \underline{C} \qquad (5.4.1)$$

If $\underline{B} = \underline{0}$ and $\underline{\mu}_1 = 0$ then $\underline{\mu}_t = 0$ for all $t \geq 1$; and hence in this case no condition on the matrix $\underline{A}$ is necessary for the first order stationarity. Otherwise, we proceed as follows..

Let the spectral radius of a matrix $\underline{A}$, $\rho(\underline{A})$, be

$$\rho(\underline{A}) = \max_i \{|\lambda_i(\underline{A})|\} \qquad (5.4.2)$$

where $\lambda_i(\underline{A})$ is the i-th eigenvalue of $\underline{A}$ and it is known that $\rho(\underline{A}) \leq ||\underline{A}||$ where $||\underline{A}||$ is any norm (Wilkinson, 1965). A sufficient condition for $\lim_{t\to\infty} \left[\underline{A}^t\ \mu_1 + \left(\sum_{j=0}^{t-1} \underline{A}^j\right) \underline{B}\ \underline{C}\right]$ to be finite is that $\rho(\underline{A}) < 1$. Under this condition the mean value $\underline{\mu}$ is then given by

$$\underline{\mu} = (\underline{I} - \underline{A})^{-1}\ \underline{B}\ \underline{C} \qquad (5.4.3)$$

We now obtain the conditions for second order stationarity. From the model (5.2.4), we have $E(\underline{x}_t\ e_{t+1}) = \underline{0}$, $E(\underline{x}_t\ e_t\ e_{t+1}) = 0$. Also from (5.2.4) we obtain

$$\underline{V}_t = \underline{A}\ \underline{V}_{t-1}\ \underline{A}' + \underline{A}\ \underline{S}_{t-1}\ \underline{B}' + \underline{B}\ \underline{S}_{t-1}\ \underline{A}' + \underline{B}\ \underline{W}_{t-1}\ \underline{B}' + \underline{C}\ \underline{C}' \qquad (5.4.4)$$

where

$$\underline{S}_t = \underline{A}\ \underline{\mu}_{t-1}\ \underline{C}' + \underline{B}\ \underline{C}\ \underline{C}' + \underline{C}\ \underline{\mu}_{t-1}\ \underline{A}' + \underline{C}\ \underline{C}'\ \underline{B}', \qquad (5.4.5)$$

$$\underline{W}_t = \underline{A}\ \underline{V}_{t-1}\ \underline{A}' + \underline{A}\ \underline{S}_{t-1}\ \underline{B}' + \underline{B}\ \underline{S}_{t-1}\ \underline{A}' + \underline{B}\ \underline{W}_{t-1}\ \underline{B}' + 3\ \underline{C}\ \underline{C}'$$

$$= \underline{V}_t + 2\ \underline{C}\ \underline{C}'. \qquad (5.4.6)$$

In obtaining the expression for $\underline{W}_t$, we have made use of the fact that the random variables $\{e_t\}$ are Gaussian with $E(e_t) = 0$, $E(e_t^2) = 1$, so that

$E(e_t^4) = 3$. The above derivation is still valid even if e_t is not Gaussian, but in this case $E(e_t^4) = 3 + K_4$, where K_4 is the fourth order cumulant.

From (5.4.4) and (5.4.6) we have

$$\underline{V}_t = \underline{A}\ \underline{V}_{t-1}\ \underline{A}' + \underline{B}\ \underline{V}_{t-1}\ \underline{B}' + \underline{A}\ \underline{S}_{t-1}\ \underline{B}'$$

$$+ \underline{B}\ \underline{S}_{t-1}\ \underline{A}' + 2\ \underline{B}\ \underline{C}\ \underline{C}'\ \underline{B}' + \underline{C}\ \underline{C}' \qquad (5.4.7)$$

We now assume that the process $\{\underline{x}_t\}$ is first order stationary so that $\underline{\mu}_t = \underline{\mu}$ and this implies $\underline{S}_t = \underline{S}$, where

$$\underline{S} = \underline{A}\ \underline{\mu}\ \underline{C}' + \underline{B}\ \underline{C}\ \underline{C}' + \underline{C}\ \underline{\mu}'\underline{A}' + \underline{C}\ \underline{C}'\ \underline{B}'. \qquad (5.4.8)$$

The expression (5.4.7) can now be written as

$$\underline{V}_t = \underline{A}\ \underline{V}_{t-1}\ \underline{A}' + \underline{B}\ \underline{V}_{t-1}\ \underline{B}' + \Delta_1, \qquad (5.4.9)$$

where

$$\Delta_1 = \underline{A}\ \underline{S}\ \underline{B}' + \underline{B}\ \underline{S}\ \underline{A}' + 2\ \underline{B}\ \underline{C}\ \underline{C}'\ \underline{B}' + \underline{C}\ \underline{C}'$$

To find the conditions under which, as $t \to \infty$, $\underline{V}_t$ tends to, say, $\underline{V}$, where $\underline{V}$ does not depend on t, we proceed as follows.

Let $\underline{D}$, $\underline{E}$ and $\underline{F}$ be three square matrices, each of order p x p. Let $\underline{d}_{ij}$ be the element corresponding to the i-th row and j-th column of the matrix $\underline{D}$. Let $\underline{D}_{\cdot j}$ (j = 1,2,...,p) be the j-th column of $\underline{D}$. Define

$$\text{Vec}(\underline{D}) = \begin{pmatrix} \underline{D}_{\cdot 1} \\ \underline{D}_{\cdot 2} \\ \vdots \\ \underline{D}_{\cdot p} \end{pmatrix} \qquad (5.4.10)$$

and the Kronecker product $\underline{D} \otimes \underline{E}$, which is of order $p^2 \times p^2$, as $\underline{D} \otimes \underline{E} = (d_{ij}\,\underline{E})$. Then we have (Neudecker, 1969), $\mathrm{Vec}(\underline{D}\,\underline{E}\,\underline{F}) = (\underline{F}' \otimes \underline{D})\,\mathrm{Vec}(\underline{E})$, $\mathrm{Vec}(\underline{D}\,\underline{E}) = (\underline{I} \otimes \underline{D})\,\mathrm{Vec}(\underline{E})$. Using the above notation, we can write (5.4.9) as

$$\mathrm{Vec}(\underline{V}_t) = \left[\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B}\right] \mathrm{Vec}(\underline{V}_{t-1}) + \mathrm{Vec}\ \underline{\Delta}_1. \tag{5.4.11}$$

This is a first order difference equation in $\mathrm{Vec}(\underline{V}_t)$, and the solution of this equation can be written in power series of $(\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B})$.

For the solution of $\mathrm{Vec}(\underline{V}_t)$ to converge, it is sufficient that

$$\rho\left[\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B}\right] < 1 \tag{5.4.12}$$

This is the sufficient condition for the time series $\underline{x}_t$, generated from (5.2.4) to be asymptotically second order stationary. The condition (5.4.12) becomes

$$\rho\left[\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B}\ \sigma_e^2\right] < 1, \quad \text{if } \sigma_e^2 \neq 1.$$

Assuming the condition (5.4.12) is satisfied, we obtain the expression for the variance and covariance of $\{\underline{x}_t\}$. Let $\underline{V} = E(\underline{x}_t\,\underline{x}_t')$, then we have from (5.4.9),

$$\underline{V} = \underline{A}\,\underline{V}\,\underline{A}' + \underline{B}\,\underline{V}\,\underline{B}' + \underline{\Delta}_1, \tag{5.4.13}$$

which can be solved explicitly since the equation (5.4.13) is linear in $\underline{V}$. Here we do not need the explicit solution. From the model (5.2.4), we have

$$E(\underline{x}_{t+1}\,\underline{x}_t') = \underline{A}\,E(\underline{x}_t\,\underline{x}_t') + \underline{B}\,E(\underline{x}_t\,\underline{x}_t'\,e_t) \tag{5.4.14}$$

and for $s > 1$,

$$E(\underline{x}_{t+s}\,\underline{x}_t') = \underline{A}^{s-1}\,E(\underline{x}_{t+1}\,\underline{x}_t') + \left(\sum_{j=0}^{s-2} \underline{A}^j\,\underline{B}\,\underline{C}\right) \underline{\mu}' \tag{5.4.15}$$

Let $\underline{C}(s) = E(\underline{x}_{t+s} - \underline{\mu})\,(\underline{x}_t - \underline{\mu})'$, then we can show that

$$\underline{C}(0) = \underline{A}\,\underline{C}(0)\,\underline{A}' + \underline{B}\,\underline{C}(0)\,\underline{B}' + \underline{\Delta}_2, \tag{5.4.16}$$

$$\underline{C}(1) = \underline{A}\,\underline{C}(0) + \underline{\Delta}_3, \tag{5.4.17}$$

$$\underline{C}(s) = \underline{A}\,\underline{C}(s-1) = \underline{A}^{s-1}\,\underline{C}(1), \quad (s = 2,3,\ldots), \tag{5.4.18}$$

where

$$\underline{\Delta}_2 = \underline{B}\,\underline{\mu}\,\underline{\mu}'\underline{B}' + \underline{A}'\underline{\mu}\,\underline{\mu}'\underline{A}' + \underline{A}\,\underline{S}\,\underline{B}' + \underline{B}\,\underline{S}'\underline{A}' + 2\,\underline{B}\,\underline{C}\,\underline{C}'\underline{B}' + \underline{C}\,\underline{C}' - \underline{\mu}\,\underline{\mu}',$$

$$\underline{\Delta}_3 = \underline{A}\,\underline{\mu}\,\underline{\mu}' + \underline{B}\,\underline{S} - \underline{\mu}\,\underline{\mu}'.$$

If we now suppose $\underline{A}$ and $\underline{B}$ are of the form (5.2.3), we obtain from (5.4.18)

$$\gamma(s) + a_1\gamma(s-1) + \ldots + a_p\gamma(s-p) = 0, \quad s > 1, \tag{5.4.19}$$

where $\gamma(s) = \mathrm{cov}(X_{t+s}\,X_t)$. These equations are the same as the Yule-Walker equations for an ARMA(p,1) and thus show that the bilinear model BL(p,0,p,1) has the same covariance structure as an ARMA(p,1).

It is interesting to note that for a homogeneous bilinear system obtained from (5.2.4), by putting $\underline{C} = \underline{0}$, we have $\underline{\mu} = \underline{0}$, $\underline{S} = \underline{0}$ and $\underline{V}_t = \underline{A}\,\underline{V}_{t-1}\,\underline{A}' + \underline{B}\,\underline{V}_{t-1}\,\underline{B}'$. If $\underline{V}_1 = \underline{0}$ (and $\underline{C} = \underline{0}$, we have $\underline{V}_t = 0$, $t \geq 1$. Proceeding as above we can show that a homogeneous bilinear system degenerates into a deterministic system if $\rho(\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B}) < 1$ as $t \to \infty$, and if $\rho(\underline{A} \otimes \underline{A} + \underline{B} \otimes \underline{B}) > 1$, the system explodes.

If we now consider the model BL(1,0,1,1) given by

$$X_t + a_1 X_{t-1} = b_{11} X_{t-1}\,e_{t-1} + e_t,$$

the sufficient condition for the second order stationarity of the process $\underline{X}_t$ is that $a_1^2 + b_{11}^2 < 1$, and expressions for the covariances can be obtained from (5.4.18). The condition for stationarity and the expressions for variance and covariances agree with the results of Granger and Andersen (1978a).

In this section we have obtained the conditions for the existence of second order moments for the bilinear process satisfying the VBL(p) model. It must be noted that higher order moments need not always exist (Granger and Andersen, 1978a, p.40). The ergodicity of the process $\{X_t\}$ satisfying BL(1,0,1,1) model has been established by Akamanam (1983, p.35). Using the similar arguments the ergodicity of the process $\{X_t\}$ satisfying BL(p,0,p,1) can be established.

5.5 INVERTIBILITY OF THE VBL(p) MODEL

For a time series model to be useful for forecasting purposes, it is necessary that it should be invertible. The invertibility of linear time series models has been discussed by Box and Jenkins (1970). In this section, using the definition of Granger and Andersen (1978c), described in section (1.7), we obtain a sufficient condition for the invertibility of the VBL(p) model. The condition obtained by Granger and Andersen (1978c) for the BL(1,0,1,1) model is a special case of our result.

Consider the VBL(p) model given by (5.2.4) and let $X_t = \underline{H}'\underline{x}_t$. Let $\hat{e}_t$ be an estimate of e_t as defined in section 1.8 satisfying the difference equation

$$X_t = \underline{H}'\underline{A}\ \underline{x}_{t-1} + \underline{H}'\underline{B}\ \underline{x}_{t-1}\ \hat{e}_{t-1} + \underline{H}'\underline{C}\ \hat{e}_t, \tag{5.5.1}$$

From (5.2.4) and (5.5.1) we have

$$\underline{H}'\underline{C}\ \xi_1(t) = -\ \underline{H}'\underline{B}\ \underline{x}_{t-1}\ \xi_1(t-1), \tag{5.5.2}$$

where $\xi_1(t) = e(t) - \hat{e}(t)$. Assuming the process $\{\underline{x}_t\}$ to be ergodic and proceeding as in Granger and Andersen (1978c), we can show that

$$E(\xi_1^2(t)) \leq \left[E(\zeta^2(s))\right]^t, \tag{5.5.3}$$

where $\zeta(s) = (\underline{H}'\underline{B}\ \underline{x}_s)/\underline{H}'\underline{C}$. Taking limits on both sides of (5.5.3),

$$\lim_{t\to\infty} E(\xi_1^2(t)) \leq \lim_{t\to\infty} \left[E(\zeta^2(s))\right]^t. \tag{5.5.4}$$

The right-hand term of the inequality tends to zero as $t \to \infty$ if $E(\zeta^2(s)) < 1$, which implies that

$$\underline{H}'\underline{B}\ E(\underline{x}(s)\ \underline{x}'(s))\ \underline{B}'\underline{H} < (\underline{H}'\underline{C})^2. \tag{5.5.5}$$

For a given $\underline{A}$, $\underline{B}$, $\underline{H}$ and $\underline{C}$, one can evaluate explicitly $\underline{V} = E\{\underline{x}_s\ \underline{x}_s'\}$. The condition (5.5.5) is a sufficient condition for the invertibility of the VBL(p) model. The condition given by Granger and Andersen (1978a, p.74) for the bilinear model $X(t) = b_{11}X(t-1)\ e(t-1) + e(t)$ can be deduced from (5.5.5) by appropriate substitutions. Invertibility of the model $X_t = e_t + \beta e_{t-k}\ X_{t-\ell}$, $k,\ell > 0$ has also been discussed by Quinn (1982).

We now consider the conditions of stationarity of diagonal bilinear models.

5.6 CONDITIONS FOR STATIONARITY OF THE DIAGONAL BILINEAR MODEL, DBL(ℓ)

The diagonal bilinear model of order ℓ, DBL(ℓ), is defined as

$$X_t + \sum_{i=1}^{\ell} a_i\ X_{t-i} = \sum_{j=1}^{\ell} b_j\ X_{t-j}\ e_{t-j} + e_t \tag{5.6.1}$$

where $\{e_t\}$ is a sequence of i.i.d. N(0,1) random variables. Define the matrices

$$\underline{A} = \begin{bmatrix} -a_1 & -a_2 & \dots & -a_\ell \\ 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \quad \underline{B}_j = \begin{bmatrix} b_j & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

$$\underset{\sim}{x}_t = \begin{bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-\ell+1} \end{bmatrix}, \quad \underset{\sim}{C} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \underset{\sim}{H} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then, with these notations we can write the model (5.6.1) in the form

$$\left.\begin{aligned} \underset{\sim}{x}_t &= \underset{\sim}{A}\, \underset{\sim}{x}_{t-1} + \sum_{j=1}^{\ell} \underset{\sim}{B}_j\, \underset{\sim}{x}_{t-j}\, e_{t-j} + \underset{\sim}{C}\, e_t \\ X_t &= \underset{\sim}{H}'\, \underset{\sim}{x}_t \end{aligned}\right\} \qquad (5.6.2)$$

Following Subba Rao (1981a) the sufficient conditions for asymptotic stationarity of the diagonal bilinear model (5.6.1) can be obtained using the state space representation (5.6.2). We assume that $\{e_t\}$ are independent identically distributed N(0,1) random variables. Let

$$\underset{\sim}{\mu}_t = E\,(\underset{\sim}{x}_t)$$
$$V_t = E\,(\underset{\sim}{x}_t\, \underset{\sim}{x}_t'), \text{ and}$$
$$S_t(k) = E\,(\underset{\sim}{x}_t\, \underset{\sim}{x}_{t-k}'\, e_{t-k}).$$

Taking expectations on both sides of (5.6.2) and noting that

$$E(\underset{\sim}{x}_t\, e_t) = \underset{\sim}{C}, \text{ all } t,$$

we get

$$\underset{\sim}{\mu}_t = \underset{\sim}{A}\, \underset{\sim}{\mu}_{t-1} + \sum_{j=1}^{\ell} \underset{\sim}{B}_j\, \underset{\sim}{C}.$$

Define

$$\underset{\sim}{B} = \sum_{j=1}^{\ell} \underset{\sim}{B}_j = \begin{bmatrix} \sum_{j=1}^{\ell} b_j & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & & \\ 0 & 0 & \dots & 0 \end{bmatrix}$$

then

$$\underset{\sim}{\mu}_t = \underset{\sim}{A}\, \underset{\sim}{\mu}_{t-1} + \underset{\sim}{B}\, \underset{\sim}{C}. \qquad (5.6.3)$$

Hence a sufficient condition for $\{X_t\}$ to be asymptotically stationary of the first order is

$$\rho(\underset{\sim}{A}) = \max_{1 \le i \le \ell} \{\lambda_i(\underset{\sim}{A})\} < 1 \qquad (5.6.4)$$

where $\lambda_1(\underset{\sim}{A}), \dots, \lambda_\ell(\underset{\sim}{A})$ are the eigenvalues of the Matrix $\underset{\sim}{A}$.

If this condition is satisfied then

$$\underset{\sim}{\mu} = \lim_{t \to \infty} \underset{\sim}{\mu}_t = (I - \underset{\sim}{A})^{-1} B\, \underset{\sim}{C} \qquad (5.6.5)$$

Result

The asymptotic mean of X_t is

$$\mu = \lim_{t \to \infty} E(X_t) = \lim_{t \to \infty} \underset{\sim}{H}'\, \underset{\sim}{\mu}_t = \sum_{j=1}^{\ell} b_j / (1 + \sum_{j=1}^{\ell} a_j) \qquad (5.6.6)$$

Proof

Suppose that the inverse of (I-A) can be written as

$$(I - \underset{\sim}{A})^{-1} = \frac{1}{|I - \underset{\sim}{A}|} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1\ell} \\ & \dots & \dots & \\ \alpha_{\ell 1} & \alpha_{\ell 2} & \dots & \alpha_{\ell\ell} \end{bmatrix}$$

then $(|I - \underset{\sim}{A}| = \det(I - \underset{\sim}{A}))$

$$\underset{\sim}{\mu} = \underset{\sim}{H}'\, (I - \underset{\sim}{A})^{-1} B\, \underset{\sim}{C} = \frac{1}{|I - \underset{\sim}{A}|} (\alpha_{11}\ \alpha_{12}\ \dots\ \alpha_{1\ell}) \begin{bmatrix} \sum_{j=1}^{\ell} b_j \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \frac{\alpha}{|I - \underset{\sim}{A}|} \sum_{j=1}^{\ell} b_j$$

Adding the ℓ-th column of (I-A) to the $(\ell-1)$-th column and the resulting column to the $(\ell-2)$-th column, and so on, it is easy to show that

$$|(I - \underset{\sim}{A})| = \begin{vmatrix} 1+a_1 & a_2 & a_3 & \dots & a_\ell \\ -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & 0 & \dots & -1 & 1 \end{vmatrix} = \begin{vmatrix} 1 + \sum_{i=1}^{\ell} a_i & \sum_{i=1}^{\ell} a_i & \dots & a_\ell \\ 0 & 1 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & 0 & 1 \end{vmatrix}$$

$$= 1 + \sum_{i=1}^{\ell} a_i$$

and

$$\alpha_{11} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ -1 & 1 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots \\ 0 & \cdots & 0 & -1 & 1 \end{vmatrix} = 1$$

Hence the result (5.6.6).

We now obtain the conditions for second order stationarity. Because of the independence of e_t's, then from (5.6.2) we have

$$E(\underset{\sim}{x}_t\, e_{t+r}) = 0, \text{ for all } r \geq 1.$$

Also, from (5.6.2), we obtain

$$\begin{aligned} \underset{\sim}{V}_t &= E(\underset{\sim}{x}_t\, \underset{\sim}{x}_t') \\ &= \underset{\sim}{A}\, \underset{\sim}{V}_{t-1}\, \underset{\sim}{A}' + \Big[\underset{\sim}{A} \sum_{j=1}^{\ell} S_{t-1}(j-1)\, \underset{\sim}{B}_j'\Big]^{+1} \\ &\quad + \sum_{i=1}^{\ell}\sum_{j=1}^{\ell} B_j\, E(\underset{\sim}{x}_{t-j}\, \underset{\sim}{x}_{t-i}'\, e_{t-j}\, e_{t-i})\, \underset{\sim}{B}_i' + \underset{\sim}{C}\, \underset{\sim}{C}' \end{aligned} \tag{5.6.7}$$

where the notation $[D]^{+1}$ for any square matrix D, means

$$[D]^{+1} = D + D'.$$

Using (5.6.2) and (5.6.7), since

$$\begin{aligned} E(\underset{\sim}{x}_t\, \underset{\sim}{x}_t'\, e_t^2) &= E(\underset{\sim}{x}_t\, \underset{\sim}{x}_t') + 2\, \underset{\sim}{C}\, \underset{\sim}{C}' \\ &= \underset{\sim}{V}_t + 2\, \underset{\sim}{C}\, \underset{\sim}{C}', \end{aligned}$$

and since for $j < i$,

$$E(\underset{\sim}{x}_{t-j}\, \underset{\sim}{x}_{t-i}'\, e_{t-j}\, e_{t-i}) = \underset{\sim}{C}\, E(\underset{\sim}{x}_{t-i}'\, e_{t-i}) = \underset{\sim}{C}\, \underset{\sim}{C}'$$

(this expression also holds for $j > i$), we obtain

$$E(\underset{\sim}{x}_{t-j}\, \underset{\sim}{x}_{t-i}'\, e_{t-j}\, e_{t-i}) = \begin{cases} \underset{\sim}{V}_{t-j} + 2\, \underset{\sim}{C}\, \underset{\sim}{C}' & \text{if } i = j \\ \underset{\sim}{C}\, \underset{\sim}{C}' & \text{if } i \neq j \end{cases} \tag{5.6.8}$$

Therefore, (5.6.7) can be written in the form

$$\begin{aligned} \underset{\sim}{V}_t &= \underset{\sim}{A}\underset{\sim}{V}_{t-1}\, \underset{\sim}{A}' + \Big[\underset{\sim}{A} \sum_{j=1}^{\ell} S_{t-1}(j-1)\, \underset{\sim}{B}_j'\Big]^{+1} \\ &\quad + \Big[\sum_i \sum_j \underset{\sim}{B}_j\, \underset{\sim}{C}\, \underset{\sim}{C}'\, \underset{\sim}{B}_i'\Big] + \sum_{j=1}^{\ell} \underset{\sim}{B}_j\, (\underset{\sim}{V}_{t-j} + 2\, \underset{\sim}{C}\, \underset{\sim}{C}')\, \underset{\sim}{B}_j' + \underset{\sim}{C}\, \underset{\sim}{C}' \\ &= \underset{\sim}{A}\underset{\sim}{V}_{t-1}\, \underset{\sim}{A}' + \Big[\underset{\sim}{A} \sum_{j=1}^{\ell} S_{t-1}(j-1)\, \underset{\sim}{B}_j'\Big]^{+1} \\ &\quad + \sum_j \underset{\sim}{B}_j \underset{\sim}{V}_{t-j}\, \underset{\sim}{V}_j' + \underset{\sim}{\Delta} \end{aligned} \tag{5.6.9}$$

where

$$\underset{\sim}{\Delta} = \sum_{\substack{i}} \sum_{\substack{j \\ i \neq j}} \underset{\sim}{B}_j\, C\, C'\, \underset{\sim}{B}_i' + 2 \sum_{j=1}^{\ell} \underset{\sim}{B}_j\, \underset{\sim}{C}\, \underset{\sim}{C}'\, \underset{\sim}{B}_j' + \underset{\sim}{C}\, \underset{\sim}{C}'$$

where $\underset{\sim}{\Delta}$ is a constant matrix independent of t. Now,

$$\left.\begin{aligned} S_{t-1}(0) &= E(\underset{\sim}{x}_{t-1}\, \underset{\sim}{x}_{t-1}'\, e_{t-1}) = \Big[(A\, \underset{\sim}{\mu}_{t-2} + B\, \underset{\sim}{C})\, \underset{\sim}{C}'\Big]^{+1} \\ S_{t-1}(1) &= E(\underset{\sim}{x}_{t-1}\, \underset{\sim}{x}_{t-2}'\, e_{t-2}) = \underset{\sim}{A}\, S_{t-2}(0) + \underset{\sim}{B}_1 \underset{\sim}{V}_{t-2} + (B - \underset{\sim}{B}_1)\, \underset{\sim}{C}\, \underset{\sim}{C}' \\ S_{t-1}(2) &= E(\underset{\sim}{x}_{t-1}\, \underset{\sim}{x}_{t-3}'\, e_{t-3}) = \underset{\sim}{A}\, S_{t-2}(1) + \underset{\sim}{B}_2 \underset{\sim}{V}_{t-3} + (B - \underset{\sim}{B}_2)\, \underset{\sim}{C}\, \underset{\sim}{C}' \\ &\vdots \\ S_{t-1}(\ell-1) &= E(\underset{\sim}{x}_{t-1}\, \underset{\sim}{x}_{t-\ell}'\, e_{t-\ell}) = \underset{\sim}{A}\, S_{t-2}(\ell-2) + \underset{\sim}{B}_{\ell-1} \underset{\sim}{V}_{t-\ell} + (B - \underset{\sim}{B}_{\ell-1})\, \underset{\sim}{C}\, \underset{\sim}{C}' \end{aligned}\right\} \tag{5.6.10}$$

Postmultiplying both sides of the first, second, ..., ℓ-th equations of (5.6.10) respectively by $\underset{\sim}{B}_1'$, $\underset{\sim}{B}_2'$,...,$\underset{\sim}{B}_\ell'$ and summing, we get

$$\sum_{j=1}^{\ell} S_{t-1}(j-1)\, \underset{\sim}{B}_j' = \underset{\sim}{A}\Big(\sum_{j=2}^{\ell} S_{t-2}(j-2)\, \underset{\sim}{B}_j'\Big) + \Big(\sum_{j=2}^{\ell} \underset{\sim}{B}_{j-1} \underset{\sim}{V}_{t-j} \underset{\sim}{B}_j'\Big) + D \tag{5.6.11}$$

where D is again a constant matrix. By successive substitutions for the quantity

$$\sum_{j=1}^{\ell} S_{t-1}(j-1)\, \underset{\sim}{B}_j'$$

at t-2, t-3,...,t-ℓ from (5.6.11), we get

$$\sum_{j=1}^{\ell} S_{t-1}(j-1)\underline{B}_j' = \underline{A}^{\ell-1} S_{t-\ell}(0) + \underline{A}^{\ell-2}B_1\underline{V}_{t-\ell}\underline{B}_\ell'$$

$$= \underline{A}^{\ell-3}\Big(\sum_{j=\ell-1}^{\ell} \underline{B}_{j-\ell+2}\underline{V}_{t-j}\underline{B}_j'\Big) + \ldots$$

$$+ \underline{A}\sum_{j=3}^{\ell} B_{j-2}\underline{V}_{t-j}\underline{B}_j' + \sum_{j=2}^{\ell} B_{j-1}\underline{V}_{t-j}\,B_j'$$

$$+ (\underline{A}^{\ell-1} + \underline{A}^{\ell-2} + \ldots + \underline{I})\, D.$$

This quantity can be rewritten as

$$\sum_{j=1}^{\ell} S_{t-1}(j-1)B_j' = B_1\underline{V}_{t-2}B_2' + B_2\underline{V}_{t-3}B_3' + B_3\underline{V}_{t-4}B_4' + .. + B_{\ell-1}\underline{V}_{t-\ell}B_\ell'$$

$$+ \underline{A}B_1\underline{V}_{t-3}B_3' + \underline{A}B_2\underline{V}_{t-4}B_4' + .. + \underline{A}B_{\ell-2}\underline{V}_{t-\ell}B_\ell'$$

$$+ \underline{A}^2B_1\underline{V}_{t-4}B_4' + \ldots + \underline{A}^2B_{\ell-3}\underline{V}_{t-\ell}B_\ell'$$

$$\vdots \qquad \ddots \qquad \vdots$$

$$+ \underline{A}^{\ell-2}B_1\underline{V}_{t-\ell}B_\ell'$$

$$+ \underline{A}^{\ell-1} S_{t-\ell}(0) + (\underline{A}^{\ell-1} + \underline{A}^{\ell-2} + \ldots + \underline{I})\, D$$

$$= B_1\underline{V}_{t-2}B' + (\underline{A}B_1 + B_2)V_{t-3}B_3' + (\underline{A}^2B_1 + \underline{A}B_2 + \underline{B}_3)V_{t-4}B_4'$$

$$+ \ldots + (\underline{A}^{\ell-2}B_1 + \underline{A}^{\ell-3}B_2 + \ldots + \underline{A}B_{\ell-2} + B_{\ell-1})V_{t-\ell}B_\ell'$$

$$+ \underline{A}^{\ell-1}S_{t-\ell}(0) + (\underline{A}^{\ell-1} + \underline{A}^{\ell-2} + \ldots + \underline{I})\, D.$$

Resubstituting again in (5.6.12), we get

$$\underline{V}_t = \underline{A}V_{t-1}\,A' + \sum_{j=1}^{\ell} \underline{B}_j\underline{V}_{t-j}\underline{B}_j'$$

$$+ \Big[\underline{A}B_1\underline{V}_{t-2}\underline{B}_2' + \underline{A}(\underline{A}B_1+B_2)\underline{V}_{t-3}\underline{B}_3' + \ldots + A(\underline{A}^{\ell-2}B_1 + \ldots + B_{\ell-1})$$

$$\underline{V}_{t-\ell}B_\ell'\Big]^{+1}$$

$$+ |\underline{A}^{\ell}S_{t-\ell}(0)|^{+1} + \text{a constant matrix.}$$

Hence,

$$\text{vec}(\underline{V}_t) = \{\underline{A}\otimes\underline{A} + B_1\otimes B_1\}\text{vec}(\underline{V}_{t-1}) + \{\underline{B}_2\otimes(\underline{A}B_1+B_2) + \underline{A}B_1\otimes B_2\}\text{vec}(\underline{V}_{t-2})$$

$$+ \{\underline{B}_3\otimes(\underline{A}^2B_1+\underline{A}B_2+\underline{B}_3) + (\underline{A}^2B_1+\underline{A}B_2)\otimes B_1\}\text{vec}(\underline{V}_{t-3}) + \ldots$$

$$+ \{\underline{B}_\ell\otimes(\underline{A}^{\ell-1}B_1+\underline{A}^{\ell-2}B_2+ .. + \underline{A}B_{\ell-1}+B_\ell)$$

$$+ (\underline{A}^{\ell-1}B_1+\underline{A}^{\ell-2}B_2 + .. + \underline{A}B_{\ell-1})\otimes B_\ell\} + \text{constant.}$$

Therefore a sufficient conditicn for $\{X_t\}$ to be second order asymptotically stationary is all the roots (in modulus) of the equation

$$|\lambda^{\ell} I - \lambda^{\ell-1}\Gamma_1 - \ldots - \lambda\Gamma_{\ell-1} - \Gamma_\ell| = 0 \qquad (5.6.12)$$

lie inside the unit circle, where

$$\Gamma_1 = \underline{A}\otimes\underline{A} + B_1\otimes B_1,$$

$$\Gamma_j = B_j\otimes(\underline{A}^{j-1}B_1+\underline{A}^{j-2}B_2 + .. + \underline{A}B_{j-1}+B_j)$$

$$+ (\underline{A}^{j-1}B_1+A^{j-2}B_2+ .. + \underline{A}B_{j-1})\otimes B_j,$$

$$j = 2,3,\ldots, \ell.$$

This condition is equivalent to the condition

$$\rho(\Gamma) < 1 \qquad (5.6.13)$$

where the block matrix $\underline{\Gamma}$ is given by

$$\Gamma = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \ldots & \Gamma_\ell \\ I & 0 & \ldots & 0 \\ 0 & I & \ldots & 0 \\ 0 & \ldots & \ldots & 0 \end{bmatrix}$$

5.7 CONDITIONS FOR STATIONARITY OF THE LOWER TRIANGULAR BILINEAR MODEL, LTBL (ℓ,ℓ)

We now consider the properties of lower triangular bilinear models.

A time series $\{X_t\}$ is said to be a lower triangular bilinear process if X_t satisfies the difference equation

$$X_t + \sum_{i=1}^{\ell} a_i X_{t-i} = e_t + \sum_{\substack{i=1 \\ i \geq j}}^{\ell} \sum_{j=1}^{\ell} \beta_{ij} X_{t-i} e_{t-j} \qquad (5.7.1)$$

where $\{e_t\}$ are defined as above. We denote the model (5.7.1) by LTBL(ℓ,ℓ). The lower triangular bilinear model is defined as

$$\begin{aligned} X_t + \sum_{i=1}^{\ell} a_i X_{t-i} &= e_t + \beta_{11} X_{t-1} e_{t-1} \\ &\quad + \beta_{21} X_{t-2} e_{t-1} + \beta_{22} X_{t-2} e_{t-2} \\ &\quad \vdots \\ &\quad + \beta_{\ell 1} X_{t-\ell} e_{t-1} + \beta_{\ell 2} X_{t-\ell} e_{t-2} + \cdots + \beta_{\ell\ell} X_{t-\ell} e_{t-\ell} \end{aligned} \qquad (5.7.2)$$

This model can be rewritten in state space form as follows. Define the matrices

$$\underset{\sim}{B}_j = \begin{bmatrix} \overbrace{\beta_{jj} \quad \beta_{j+1,j} \quad \cdots \quad \beta_{\ell j}}^{(\ell-j+1)} & \overbrace{0 \quad \cdots \quad 0}^{(\ell-1) \text{ times}} \\ 0 \quad 0 \quad \cdots & \cdots \quad 0 \\ \\ 0 \quad \cdots \quad \cdots & \cdots \quad \cdots \quad 0 \end{bmatrix} \qquad j = 1,2,\ldots,\ell$$

and define the vectors

$$\underset{\sim}{x}_t = \begin{Bmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-\ell+1} \end{Bmatrix} \qquad \underset{\sim}{C} = \begin{Bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{Bmatrix}$$

With these definitions the model (5.7.1) can be written in the form (5.6.2). Therefore the sufficient conditions for the stationarity of the process (5.7.1) are the same as the conditions given by (5.6.4) and (5.6.12) (5.6.12) respectively. The stationarity conditions for the upper triangular bilinear model, or the general bilinear model (5.2.1) are quite complicated to study.

5.8 ESTIMATION OF THE PARAMETERS OF BILINEAR MODELS

We now consider the problem of fitting a BL(p,0;m,K) model

$$X_t + \sum_{i=1}^{p} a_i X_{t-i} = \alpha + \sum_{i=1}^{m} \sum_{j=1}^{K} b_{ij} X_{t-i} e_{t-j} + e_t \qquad (5.8.1)$$

to observed data (note that we added the constant α to the model). Subba Rao (1981a) has considered the estimation of the BL(p,0,p,q) model. The procedure we describe here for estimating BL(p,0,m,K) model is similar to the technique developed by Subba Rao (1981a). The process of fitting this model to the data consists of two separate stages, namely,

(i) the estimation of the parameters of the model;

(ii) the determination of the order of the model.

Here we consider the first stage, and the second stage will be considered in the next section.

Given a realization $(x_1,x_2,\ldots,x_N)$ on a stationary time series $\{X_t\}$ the problem is to estimate the unknown parameters $\{a_1,\ldots,a_p; b_{11},\ldots,b_{mK}\}$ (assuming that p, m and K are known a priori). The method we use is the Newton-Raphson method which has been used by many authors (see for example Astrom and Bohlin (1966), Box and Jenkins (1970), and Subba Rao (1981a).

For convenience, let us denote

$$\theta_i = a_i,\ i = 1,2,\ldots,p,$$

$$\theta_{p+1} = b_{11}, \theta_{p+2} = b_{12}, \ldots, \theta_{p+mK} = b_{mK};\ \theta_{p+mK+1} = \alpha,$$

so that the set of parameters may be written as

$$\underset{\sim}{\theta}' = (\theta_1, \theta_2, \ldots \theta_n)$$

$$= (a_1, a_2, \ldots, b_{11}. \ldots b_{mK}; \alpha),$$

where $n = p + mK + 1$.

If we assume that $\{e_t\}$ are independent identically distributed $N(0,\sigma_e^2)$, then the joint density function of $\{e_{\gamma+1}, e_{\gamma+2}, \ldots, e_N\}$, where $\gamma = \max(p,m,K)$, is given by

$$(2\pi\ \sigma_e^2)^{-(N-\gamma)/2} \exp\{-\frac{1}{2\sigma_e^2} \sum_{t=\gamma+1}^{N} e_t^2\}.$$

Therefore, the conditional maximum likelihood estimate $\hat{\underset{\sim}{\theta}}$ is then obtained by minimizing

$$Q(\underset{\sim}{\theta}) = \sum_{t=\gamma+1}^{N} e_t^2 \tag{5.8.2}$$

(note that the terms $e_1,\ldots,e_\gamma$ cannot be computed from the observed x_t's since x_0, x_{-1}, x_{-2}, ... are not observed). The Newton-Raphson iterative equation for minimizing $Q(\underset{\sim}{\theta})$ is given by

$$\underset{\sim}{\theta}^{(i+1)} = \underset{\sim}{\theta}^{(i)} - H^{-1}(\underset{\sim}{\theta}^{(i)})\ \underset{\sim}{G}(\underset{\sim}{\theta}^{(i)}) \tag{5.8.3}$$

where $\underset{\sim}{\theta}^{(i)}$ is the set of estimates obtained at the i-th stage of iteration, and

$$\left.\begin{aligned} \underset{\sim}{G}'(\underset{\sim}{\theta}) &= \left[\frac{\partial Q(\underset{\sim}{\theta})}{\partial\theta_1}, \frac{\partial Q(\underset{\sim}{\theta})}{\partial\theta_2}, \ldots, \frac{\partial Q(\underset{\sim}{\theta})}{\partial\theta_n}\right], \\ H(\underset{\sim}{\theta}) &= \left[\frac{\partial^2 Q(\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j}\right]_{nxn} \end{aligned}\right\} \tag{5.8.4}$$

The partial derivatives of $Q(\underset{\sim}{\theta})$ with respect to $\underset{\sim}{\theta}$ are given by

$$\left.\begin{aligned} \frac{\partial Q(\underset{\sim}{\theta})}{\partial\theta_i} &= 2 \sum_{t=\gamma+1}^{N} e_t \frac{\partial e_t}{\partial\theta_i},\quad i = 1,2,\ldots,n \\ \frac{\partial^2 Q(\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} &= 2 \sum_{t=\gamma+1}^{N} \frac{\partial e_t}{\partial\theta_i}\frac{\partial e_t}{\partial\theta_j} + 2 \sum_{t=\gamma+1}^{N} e_t \frac{\partial^2 e_t}{\partial\theta_i \partial\theta_j} \\ & \qquad\qquad i,j = 1,2,\ldots,n, \end{aligned}\right\} \tag{5.8.5}$$

where these partial derivatives of e_t satisfy the recursive equations

$$\left.\begin{aligned} &\frac{\partial e_t}{\partial a_i} + \phi(a_i) = X_{t-i},\ i = 1,2,\ldots,p \\ &\frac{\partial e_t}{\partial b_{ij}} + \phi(b_{ij}) = -X_{t-i}\ e_{t-j},\ i = 1,\ldots,m,\ j = 1,\ldots,K \\ &\frac{\partial e_t}{\partial \alpha} + \phi(\alpha) = -1, \end{aligned}\right\} \tag{5.8.6}$$

where $\phi(\theta_\ell) = \sum_{i=1}^{m} \sum_{j=1}^{K} b_{ij}\ X_{t-i} \frac{\partial e_{t-j}}{\partial\theta_\ell}$.

We assume

$$e_t = \frac{\partial e_t}{\partial\theta_i} = 0,\quad t = 1,\ldots,\gamma,\ i = 1,\ldots,n,$$

so the second order partial derivatives satisfy the recursive equations

$$\left.\begin{aligned} &\frac{\partial e_t}{\partial a_i \partial a_i} = 0,\quad \frac{\partial^2 e_t}{\partial a_i \partial\alpha} = 0;\ i, i' = 1,\ldots,p \\ &\frac{\partial^2 e_t}{\partial a_i \partial b_{j\ell}} + \psi(a_i, b_{j\ell}) = -X_{t-j} \frac{\partial e_{t-\ell}}{\partial a_i} \\ &\frac{\partial^2 e_t}{\partial b_{ij} \partial b_{i'j'}} + \psi(b_{ij}, b_{i'j'}) = -X_{t-i}\frac{\partial e_{t-j}}{\partial b_{i'j'}} - X_{t-i'} \frac{\partial e_{t-j'}}{\partial b_{ij}} \\ &\frac{\partial^2 e_t}{\partial b_{ij} \partial\alpha} + \psi(b_{ij}, \alpha) = -X_{t-i} \frac{\partial e_{t-j}}{\partial\alpha} \\ &\frac{\partial^2 e_t}{\partial\alpha^2} = 0, \end{aligned}\right\} \tag{5.8.7}$$

where $\psi(\theta_r, \theta_\ell) = \sum_{i=1}^{m} \sum_{j=1}^{K} b_{ij}\ X_{t-i} \frac{\partial^2 e_{t-j}}{\partial\theta_r \partial\theta_\ell}$

Hence, for a given set of values of $\{a_i\}$, $\{b_{ij}\}$ and α one can evaluate the first and second order partial derivatives using the recursive equations (5.8.5), (5.8.6) and (5.8.7), then evaluate the gradient vector $\underset{\sim}{G}$ and the Hessian matrix $\underset{\sim}{H}$ defined by (5.8.4) and use the iterative equation (5.8.3). The iteration stops when the required accuracy is attained (the iterative equation usually converges if we start with good initial estimates). The residual variance is estimated as

$$\hat{\sigma}_e^2 = \frac{1}{N-\gamma} Q(\underset{\sim}{\hat{\theta}}) = \frac{1}{N-\gamma} \sum_{t=\gamma+1}^{N} \hat{e}_t^2 \qquad (5.8.8)$$

where $\underset{\sim}{\hat{\theta}}$ is the convergent estimate, and $\{e_t,\ t = \gamma+1,\ldots,N\}$ are calculated from (5.8.1) when the parameters $\underset{\sim}{\theta}$ are replaced by the convergent estimate $\underset{\sim}{\hat{\theta}}$.

Note: In the numerical illustration considered in section (5.10) the second order partial derivatives, defined by (5.8.5) are approximated, as is done in the Marquardt algorithm (see Marquardt, 1963), by

$$\frac{\partial^2 Q(\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j} = 2\sum \frac{\partial e_t}{\partial\theta_i} \frac{\partial e_t}{\partial\theta_j}$$

Initial Values

In order that the iterative equation (5.8.3) converge to a good set of estimates of the parameters, it is necessary to start the iteration with a good set of initial values. The procedure for obtaining the initial estimates is as follows:-

(a) Fit an AR(p) model (with constant α present) and calculate the corresponding residual variance, denoted by $\hat{\sigma}_e^2(AR)$. (See Appendix C).

(b) Calculate the residual estimates $\{\hat{e}_t,\ t = \gamma+1,\ldots,N\}$ from (a), by using the "repeated residual method" proposed by Subba Rao (1977).

This method can be described as follows. Given $\{\hat{e}_t,\ t = \gamma+1,\ldots,N\}$ from step (a), and the observed data $\{x_t,\ t = 1,\ldots,N\}$ we find the estimates of $\underset{\sim}{\theta}$ for which

$$Q(\underset{\sim}{\theta}) = \sum_{t=\gamma+1}^{N} \{x_t + \sum_{i=1}^{p} a_i x_{t-i} - \sum_{i=1}^{m}\sum_{j=1}^{K} b_{ij} x_{t-i}\, \hat{e}_{t-j} - \alpha\}^2$$

is minimum using the standard least squares approach. The least squares estimates, in this case, could be found using the Householder transformation. Replace $\underset{\sim}{\theta}$ in (5.8.1) by these least squares estimates and evaluate $\{\hat{e}_t,\ t = \gamma+1,\ldots,N\}$ and find, again, a new set of least squares estimates. Since this method may not always lead to convergent values, the iteration will be continued for a few steps to get a reasonable decrease in the residual variance $\hat{\sigma}_e^2$ from $\hat{\sigma}_e^2(AR)$. So if the residual variance using the repeated residual method is smaller than $\hat{\sigma}_e^2(AR)$, take the obtained least squares estimates as initial values for the Newton-Raphson iteration. If not, take the coefficients of the fitted AR(p), in step (a), as the initial values for the autoregressive part in the bilinear model BL(p,0;m,K), and set $b_{ij} = 0$, $i = 1,\ldots,m$; $j = 1,\ldots,K$. An alternative method for obtaining the initial values is considered in the following section.

5.9 DETERMINATION OF THE ORDER OF BILINEAR MODELS

We now consider the choice of the order p, m, K of the bilinear model BL(p,0;m,K). The procedure is to estimate the parameters of the model for different values of p,m,K, and in each case calculate the residual variance $\hat{\sigma}_e^2$ and the information criterion due to Akaike (1977).

The information criterion (AIC) is defined as

$$\text{AIC} = -2\ (\text{max log likelihood}) + 2\ (\text{number of independent parameters})$$
$$= (N-\gamma)\ \log\ \hat{\sigma}_e^2 + 2\ (\text{number of independent parameters}) \qquad (5.9.1)$$

and the normalized AIC(NAIC) is defined as

$$\text{NAIC} = \text{AIC}/(N-\gamma)$$

where $\hat{\sigma}_e^2$ is defined by (5.8.8), $(N-\gamma)$ is the number of effective observations (i.e. the number of observations used for calculating the likelihood function). The chosen order of the model is the one for which AIC value is minimum. The algorithm for choosing the order of the bilinear model BL(p,0;m,K) is described as follows.

(i) Choose a fixed integer γ (γ should be greater than or equal to the order of the best AR model for the data).

(ii) Fit the linear AR(p) model and let the corresponding residual variance be $\hat{\sigma}_e^2(\text{AR})$.

(iii) Take the coefficients obtained from (i) as initial estimates of the autoregressive part of the BL(p,0;1,1) model and set $b_{11} = 0$. Use the Newton-Raphson iteration method described in the previous section (5.8). Calculate the corresponding $\hat{\sigma}_e^2$ and AIC values for the fitted model.

(iv) From (iii) fit the BL(p,0;1,2) and BL(p,0;2,1) models using the coefficients obtained from (iii) as initial values of the parameters for the Newton-Raphson iteration and set the remaining bilinear parameter equal to zero. Calculate the corresponding $\hat{\sigma}_e^2$ and AIC values for both fitted models.

(v) Take the coefficients obtained from BL(p,0;1,2) or BL(p,0;2,1), whichever has the smaller residual variance, as the initial values for fitting the BL(p,0;2,2) model. The procedure is continued, as shown in Fig. 5.1 for all possible combinations (m,K), such that $m,K < \gamma$.

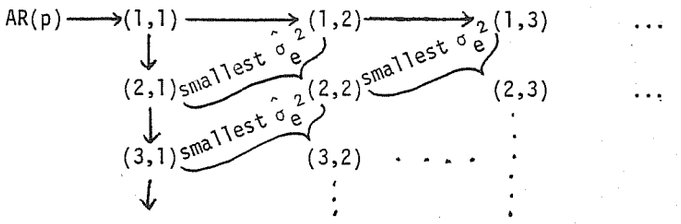


Fig. 5.1

(vi) Repeat all the steps (ii) to (v) for $p = 1,2,\ldots,\gamma$ and for each value of P the procedure stops if the residual variance $\hat{\sigma}_e^2$ increases as m and K increase.

Finally we chose that model which had the minimum AIC value.

5.10 NUMERICAL ILLUSTRATIONS

For our illustration we consider the annual sunspot numbers, which have been considered in detail in Chapter 3, for the years 1700-1955, giving 256 observations.

The best AR model and BL model are fitted to the first 221 observations (i.e. N = 221) and then predictions are calculated for the next 35 observations. The maximum order, γ, is chosen to be equal to 10; this is because the order of the best AR model is 9. The best

AR model is 9. The best AR model, selected by the AIC, fitted to the mean deleted observations $\{x_t\}$, $(x_t = X_t - \bar{X})$ is the AR(9) model,

$$x_t - 1.2163x_{t-1} + .467x_{t-2} + .1416x_{t-3} - 0.1691x_{t-4} + .1473x_{t-5} - 0.543x_{t-6} + .0534x_{t-7} - .0667x_{t-8} - .1129x_{t-9} = e_t, \qquad (5.10.1)$$

The residual variance and the AIC values for this model are

$$\hat{\sigma}_e^2(AR) = 199.27, \quad AIC = 1137.17, \quad NAIC = 5.389.$$

We applied the algorithm, described in section (5.9), on this series with the same values of N, γ as before. The AIC selects the bilinear model BL(8,0;5,4). The estimated values of the parameters of this model are as follows.

$$\left.\begin{array}{lll} \hat{a}_1 = -1.2923, & \hat{a}_2 = .3167, & \hat{a}_3 = -.0159, \\ \hat{a}_2 = 0.6077, & \hat{a}_5 = -.4812, & \hat{a}_6 = .1099, \\ \hat{a}_7 = 0.0038, & \hat{a}_8 = -.1530, & \hat{\alpha} = 4.8113 \end{array}\right\} \qquad (5.1012)$$

and the values of $\hat{b}_{ij}$ $(i = 1,\ldots,5;\ j = 1,\ldots,4)$ are given below in matrix form.

$$[\hat{b}_{ij}] = \begin{bmatrix} -.00042 & .00053 & -.00641 & .01305 \\ -.00842 & .00277 & -.01452 & -.02209 \\ .01213 & -.00636 & .01111 & .02377 \\ -.03434 & -.00673 & -.01126 & -.00446 \\ .02830 & .00584 & -.01347 & .001192 \end{bmatrix} \qquad (5.10.3)$$

The values of the residual variance and AIC are

$$\hat{\sigma}_e^2 = 110.62, \quad AIC = 1051.00, \quad NAIC = 4.981$$

It is clear that the mean sum of squares of residual and the AIC values for the BL(8,0,5,4) are much smaller than the corresponding values

obtained from the best AR model fitted to the data. It should also be noted that this model has smaller AIC value and mean sum of squares of residuals than the BL(3,0,3,4) model fitted by Subba Rao (1981a) to the last 238 observations of the same data (see also Birkenfeld, 1981).

Prediction

Suppose we have a semi-infinite realisation $\{X_s,\ s \le t\}$ from a zero mean, second order stationary time series $\{X_t\}$ and we wish to predict X_{t+m} $(m > 0)$. Let the predictor be $\tilde{X}_t(m)$. Then it is well known that the mean square prediction error, $M(m) = E[X_{t+m} - \tilde{X}_t(m)]^2$ is minimum if and only if $\tilde{X}_t(m) = E(X_{t+m}/X_s,\ s \le t)$. Now suppose the time series $\{X_t\}$ satisfies the BL(p,0;m,K) model (5.8.1) and we have a sample $(X_1, X_2, \ldots, X_N)$ and wish to find $\tilde{X}_N(1)$. To obtain this, we write (5.8.1) as

$$X_{N+1} = -\sum_{i=1}^{p} a_i X_{N+1-i} + \alpha + \sum_{i=1}^{m}\sum_{j=1}^{K} b_{ij} X_{N+1-i} e_{N+1-j} + e_{N+1} \qquad (5.10.4)$$

and then take conditional expectations both sides. We obtain the predictor

$$\tilde{X}_N(1) = -\sum_{i=1}^{p} a_i X_{N+1-i} + \alpha + \sum_{i=1}^{m}\sum_{j=1}^{K} b_{ij} X_{N+1-i} e_{N+1-j} \qquad (5.10.5)$$

and its error as $X_{N+1} - \tilde{X}_N(1) = e_{N+1}$

Using (5.10.4), we can write an expression for a one-step-ahead predictor from any time origin, say, $N+\ell-1$ $(\ell = 1,2,\ldots,M)$ as,

$$X_{N+\ell-1}(1) = -\sum_{i=1}^{p} a_i X_{N+\ell-i} + \alpha + \sum_{i=1}^{m}\sum_{j=1}^{K} b_{ij} X_{N+\ell-i} e_{N+\ell-j} \qquad (5.10.6)$$

$$(\ell = 1,2,\ldots,M)$$

and the prediction error is

$$e_{N+\ell} = X_{N+\ell} - \tilde{X}_{N+\ell-1}(1) \quad (\ell = 1,2,\ldots,M)$$

To calculate the predictors (5.10.6), we need to know the parameters. Typically we substitute the least squares estimates for these parameters, and let us denote these predictors as $\hat{X}_{N+\ell-1}(1)$ and the errors as $e_{N+\ell} = X_{N+\ell} - \hat{X}_{N+\ell-1}(1)$ ($\ell = 1,2,\ldots,M$). Define the mean square prediction error (M.S.P.E.) by

$$\hat{\sigma}_e^2(1) = \frac{1}{M} \sum_{\ell=1}^{M} e_{N+\ell}^2 \qquad (5.10.7)$$

These values can be calculated for various models for a possible comparison. In the following chapter we evaluate the predictors from subset bilinear models and then obtain the mean square prediction errors.

5.11 SAMPLING PROPERTIES OF PARAMETER ESTIMATES FOR THE BL(1,0,1,1) MODEL

The consistency of the estimates of the parameters of the BL(1,0,1,1) model has been discussed by Tuan Dinh Pham and Lanh Tat Tran (1981), but the distributional properties of these estimates are not investigated. Our object here is to report some results that have been obtained by Sesay (1982) about these aspects via simulations. We briefly summarise these results.

Let X_t satisfy the BL(1,0,1,1) model, namely,

$$X_t + a X_{t-1} = b X_{t-1} e_{t-1} + e_t \,, \qquad (5.11.1)$$

where $\{e_t\}$ is a sequence of i.i.d variables each distributed as N(0,1). The realisations that have been generated are from (5.11.1) for various values of a and b. In order to study the asymptotic distribution of these estimates, we need to define the measures of skewness, Kurtosis for single and several random variables.

Let X be a random variable with mean μ. Define the rth central moment

$$\mu_r = E(X - \mu)^r \,, \qquad (r=1,2,\ldots)$$

and
$$\beta_{r-2} = \frac{\mu_r}{(\mu^2)^{r/2}} \qquad (r > 2) \qquad (5.11.2)$$

When $r=3$, $\beta_1 = \frac{\mu_3}{\mu_2^{3/2}}$, and $r=4$, $\beta_2 = \frac{\mu_4}{\mu_2^2}$. β_1 is a measure of skewness of the distribution and β_2 is a measure of Kurtosis.

If the random variable X is normal, it is well known that $\beta_1 = 0$ and $\beta_2 = 3$.

Suppose a sample of size N is available on the random variable X. Then suppose $\hat{\mu}_r$ is the estimate of μ_r, we can define the sample measures of skewness and Kurtosis as

$$b_1 = \frac{\hat{\mu}_3}{(\hat{\mu}_2)^{3/2}} \,, \quad b_2 = \frac{\hat{\mu}_4}{(\hat{\mu}_2)^2} \,.$$

We say the distribution is skewed to the left or right according as $b_1 < 0$ or $b_1 > 0$. The distribution is said to be more peaked around the centre or flatter around the centre, when compared to normal distribution, according as b_2 is greater than 3 or less than 3.

We need to extend these measures to higher order dimensions if we want to study the joint distributions. These measures of skewness and Kurtosis have been difined by Mardia (1970).

Multivariate Measures of Skewness and Kurtosis

Let $\underline{X}' = (X_1, X_2, \ldots, X_p)$ be a p dimensional random vector with mean $\underline{\mu}' = (\mu_1, \mu_2, \ldots, \mu_p)$ and variance covariance matrix $\underline{\Sigma}$. To measure the skewness, Mardia (1970) has proposed the quantity

$$\beta_{1p} = \sum_{r,s,t} \sum_{r',s',t'} \sigma^{rr'} \sigma^{ss'} \sigma^{tt'} \mu_{111}^{rst} \mu_{11}^{r's't'} \qquad (5.11.3)$$

where $\mu_{111}^{rst} = E(X_r - \mu_r)(X_s - \mu_s)(X_t - \mu_t)$, σ^{ij} is the i,j th element of $\underline{\Sigma}^{-1}$.

When p=1, $\beta_{11} = \beta_1$ and when p=2,

$$\beta_{12} = (1-\rho^2)^{-3}\Big[\gamma_{30}^2 + \gamma_{03}^2 + 3(1+2\rho^2)(\gamma_{12}^2 + \gamma_{21}^2) - 2\rho^3\gamma_{30}\gamma_{03} + 6\rho\{(\gamma_{30}\rho\gamma_{12} - \gamma_{21}) + \gamma_{03}(\rho\gamma_{21} - \gamma_{12}) - (2+\rho^2)\gamma_{12}\gamma_{21}\}\Big] \qquad (5.11.4)$$

where $\sigma_i^2 = \text{var}(X_i)$, $i=1,2,$,

$$\rho = \frac{\text{cov}(X_1X_2)}{\sqrt{\sigma_1^2\sigma_2^2}} , \quad \gamma_{rs} = \frac{\mu_{rs}}{\sigma_1^r\sigma_2^s} , \quad \mu_{rs} = E(X_1 - \mu_2)^r(X_2 - \mu_2)^s .$$

If the random vector $\underline{X}' = (X_1, X_2)$ has a bivariate normal distribution, then $\beta_{12}=0$. Given a sample of size N on $\underline{X}$, we can estimate μ_{rs} by m_{rs} where

$$m_{rs} = \frac{1}{N}\sum_{i=1}^{N}(X_{1i} - \bar{X}_1)^r(X_{2i} - \bar{X}_2)^s .$$

and then substitute the sample quantities for the corresponding population parameters in the equation (5.11.3). Let us denote the estimate of β_{1p} by b_{1p} . Mardia has shown that , under the null hypothesis that $\beta_{1p} = 0$, the statistic $A = \frac{Nb_{1p}}{6}$ is approximately distributed as a χ^2 with degrees of freeedom equal to $\frac{p(p+1)(p+2)}{6}$.

Kurtosis Measure

Define $$\beta_{2p} = E\Big[(\underline{X} - \underline{\mu})'\underline{\Sigma}^{-1}(\underline{X} - \underline{\mu})\Big]^2 \qquad (5.11.5)$$

$$= \Sigma(\sigma^{ii})^2\,\mu_4^{(i)} + 4\Sigma\,\sigma^{ii}\sigma^{ij}\mu_{31}^{(ij)} + \Sigma\{2(\sigma^{ij})^2 + \sigma^{ii}\sigma^{jj}\}\mu_2^{(ij)} + \Sigma\{4\sigma^{ij}\sigma^{jk} + 2\sigma^{ii}\sigma^{jk}\}\mu_{211}^{ijk} + \Sigma\sigma^{ij}\sigma^{k\ell}\,\mu_{1111}^{(ijk\ell)} \qquad (5.11.6)$$

where $$\mu_{i_1 i_2 \cdots i_s}^{(j_1 \cdots j_s)} = E\{\prod_{r=1}^{s}(X_{j_r} - \mu_{j_r})^{i_r}\}.$$

and σ^{ij} is the (i,j)-th element of $\underline{\Sigma}^{-1}$ and the summation Σ is defined over all possible values of the subscripts.

When p=2,

$$\beta_{22} = \{\gamma_{40} + \gamma_{04} + 4\rho(\rho\gamma_{22} - \gamma_{13}\gamma_{31})\}/(1-\rho^2)^2 \qquad (5.11.7)$$

Given a random sample $(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_N)$ on the random vector $\underline{X}$, we can estimate β_{2p} by

$$b_{2p} = \frac{1}{N}\sum_{i=1}^{N}\{(\underline{X}_i - \bar{\underline{X}})'\underline{S}^{-1}(\underline{X}_i - \bar{\underline{X}})\}^2,$$

where $\bar{\underline{X}} = \frac{1}{N}\sum_{i=1}^{N}\underline{X}_i$, $\underline{S} = \frac{1}{N}\sum_{i=1}^{N}(\underline{X}_i - \bar{\underline{X}})(\underline{X}_i - \bar{\underline{X}})'$.

Mardia has shown that for normal populations,

$$E(b_{2p}) \simeq p(p+2)$$
$$\text{var}(b_{2p}) \simeq 8p(p+2) .$$

For large samples under the null hypothesis of bivariate normality the statistic

$$B = \left[\frac{b_{2p} - \frac{p(p+2)(n-1)}{n+1}}{\left(\frac{8p(p+2)}{N}\right)^{\frac{1}{2}}}\right]$$

is approximately distributed as normal with mean zero and variance one.

The univariate measures of skewness and Kurtosis and their multivariate analogues are calculated for the estimates obtained for the bilinear models and these are summarised in the tables 5.1 and 5.2.

Series 1 100 realisations were generated from the model

$$X_t - 0.2\,X_{t-1} = 0.2\,X_{t-1}\,e_{t-1} + e_t \quad (t=1,2,\ldots,N)$$

when $\{e_t\}$ is a sequence of independent random variables each distributed

normally, with mean zero and variance one. To the above realisation, the following model is fitted using the method described earlier.

$$X(t) + a_1X(t-1) + a_0 = b_{11}X(t-1)e(t-1) + e(t).$$

The means, variances, measures of skewness, Kurtosis, joint skewness, joint Kurtosis of these estimates, based on 100 realisations have been summarised in table 5.1. Here $\rho(a_1,a_0)$ is the sample correlation coefficient between the estimates a_1 and a_0. $A(a_1,a_0)$ is the skewness statistic calculated between the estimates a_1 and a_0, $B(a_1,a_0)$ is the Kurtosis statistic calculated between the estimates a_1 and a_0 etc.

From table 5.1, it is clear that these estimates are approximately unbiased and consistent. From the univariate measures, it is clear that their marginal distributions are symmetric, but there is some evidence of departure from normality. From the joint measures of skewness and Kurtosis, it is clear that their joint distributions are also symmetric. Their Kurtosis statistics $B(a_1,a_0)$, $B(a_1,b_{11})$ etc. are significant at 5% level of significance, confirming that there is evidence of departure from bivariate normality.

Series 2 100 realisations of time series $\{X_t\}$ are generated from the model

$$X(t) - 0.2\ X(t-1) = 0.4\ X(t-1)\ e(t-1) + e(t) \quad (t=1,2,\ldots,N)$$

where the sequence $\{e(t)\}$ is generated as before. The results from this model also confirm that the marginal distributions, and their joint distributions, are symmetrical, and there is a strong departure from bivariate normality of the estimates of the coefficients (a_1,b_{11}) when N=200, 400 and 600.

The simulations carried out for other models with different values a_1 and b_{11} have resulted in similar conclusions (for details, see Sesay, 1982).

TABLE 5.1SAMPLE STATISTICS FOR SERIES 1

$$X(t) - 0.2x(t-1) = 0.2x(t-1)e(t-1)+e(t)$$

SAMPLE SIZE

N	100	200	400	600
mean(a_1)	-0.1988	-0.2000	-0.1942	-0.2009
var(a_1)	0.0088	0.0052	0.0024	0.0016
mean(a_0)	0.0033	-0.0023	-0.0003	-0.0036
var(a_0)	0.0146	0.0063	0.0030	0.0021
mean(b_{11})	0.2002	0.1998	0.2012	0.1994
var(b_{11})	0.0048	0.0023	0.0013	0.0007
skew(a_1)	-0.0032	0.0120	-0.0122	-0.0006
kurt(a_1)	2.4332	2.2637	3.1247	2.6397
skew(a_0)	0.0455	0.0112	0.0153	0.0031
kurt(a_0)	3.0549	2.6063	2.9046	2.2239
skew(b_{11})	-0.0100	0.0043	-0.0041	-0.0020
kurt(b_{11})	3.2007	2.9140	3.1032	2.6914
$\rho(a_1,a_0)$	-0.0716	-0.0694	0.0053	-0.1051
$\rho(a_1,b_{11})$	0.1096	0.1396	0.0417	0.1650
$\rho(a_0,b_{11})$	-0.4639	-0.5309	-0.5371	-0.6851
$B12(a_1,a_0)$	0.0022	0.0003	0.0004	0.0000
$B12(a_1,b_{11})$	0.0001	0.0002	0.0002	0.0000
$BT2(a_0,b_{11})$	0.0046	0.0005	0.0007	0.0000
$B22(a_1,a_0)$	5.5451	4.9174	6.0297	4.9729
$B22(a_1,b_{11})$	5.7717	5.3855	6.2496	5.6337
$B22(a_0,b_{11})$	10.1592	10.7040	11.8656	17.4518
$A(a_1,a_0)$	0.0359	0.0046	0.0064	0.0002
$A(a_1,b_{11})$	0.0023	0.0028	0.0028	0.0001
$A(a_0,b_{11})$	0.0772	0.0075	0.0110	0.0011
$B(a_1,a_0)$	-3.0687	-3.8533	-2.4629	-3.7839
$B(a_1,b_{11})$	-2.7854	-3.2681	-2.1881	-2.9578
$B(a_0,b_{11})$	2.6990	3.3800	4.8320	11.8148

TABLE5.2 SAMPLE STATISTICS FOR SERIES 2

$X(t)-0.2x(t-1)=0.4x(t-1)e(t-1)+e(t)$

SAMPLE SIZE

N	100	200	400	600
mean(a_1)	-0.1996	-0.2028	-0.1987	-0.1964
var(a_1)	0.0092	0.0038	0.0015	0.0016
mean(a_o)	0.0221	0.0130	0.0064	0.0054
var(a_o)	0.0146	0.0070	0.0029	0.0024
mean(b_{11})	0.3739	0.3817	0.3914	0.3884
var(b_{11})	0.0060	0.0044	0.0014	0.0021
skew(a_1)	0.0119	-0.0292	-0.0051	-0.0231
kurt(a_1)	3.4521	3.2380	3.5023	5.3889
skew(a_o)	0.0640	0.0143	0.0389	0.0614
kurt(a_o)	3.9835	3.0978	4.1583	6.6525
skew(b_{11})	-0.1208	-0.1480	-0.1617	-0.1770
kurt(b_{11})	5.7703	9.4037	31.2439	19.5823
$\rho(a_1, a_)$	-0.2743	-0.2909	-0.1974	-0.3534
$\rho(a_1,b_{11})$	0.1168	0.2616	0.1546	0.4393
$\rho(a_ ,b_{11})$	-0.5163	-0.5679	-0.4914	-0.6883
B12(a_1,a_o)	0.0055	0.0013	0.0017	0.0062
B12(a_1,b_{11})	0.0153	0.0273	0.0281	0.0577
B12(a_o,b_{11})	0.0420	0.0661	0.0588	0.1910
B22(a_1,a_o)	8.6956	7.5611	8.2944	15.7239
B22(a_1,b_{11})	9.4796	14.5673	36.4697	38.3401
B22(a_o,b_{11})	18.1343	27.2387	61.5303	94.7407
A(a_1,a_o)	0.0913	0.0223	0.0288	0.1031
A(a_1,b_{11})	0.2553	0.4550	0.4680	0.9621
A(a_o,b_{11})	0.6997	1.1008	0.9807	3.1841
B(a_1,a_o)	0.8695	-0.5486	0.3680	9.6549
B(a_1,b_{11})	1.8495	8.2091	35.5871	37.9251
B(a_o,b_{11})	12.6679	24.0484	66.9129	108.4258

So far no theoretical studies of the sampling distribution of these estimates are available. One of the problems one may come across in the study of asymptotic distributions is the existence of higher order moments. It has been stated earlier that all higher order moments need not exist, and the existence of the moments does strongly depend on the nonlinear coefficients, for example, b_{11} .

Some theoretical studies of t he asymptotic distribution of the estimates of other types of nonlinear models have been reported recently, see for example, Robinson (1977).

CHAPTER SIX

ESTIMATION AND PREDICTION FOR SUBSET BILINEAR TIME SERIES MODELS WITH APPLICATIONS

6.1 INTRODUCTION

It has been pointed out earlier that some of the coefficients for the full bilinear models of the form (5.8.1) when fitted to a realisation may be "small" when compared to other coefficients. Therefore, it is useful to see whether it is possible to fit a subset bilinear model to the data which leads to a parsimonious representation. In this chapter, we consider the estimation of a subset bilinear model and we give an algorithm for the estimation of its parameters (see also Gabr and Subba Rao, 1981). The method is illustrated with real data. A comparison is then made between the forecasts obtained from the subset bilinear models and other time series models. Some comments about the transformation of the series are included. (See Subba Rao and Gabr, 1981).

6.2 AN ALGORITHM FOR FITTING SUBSET BILINEAR MODELS

In general, bilinear models are models which are linear in the 'state' and linear in the errors, but not jointly. In the description of the model (5.8.1), the 'linear part' of the model as described by the autoregressive model, explains the 'linear variation' in the series, and the rest of the variation is explained by the non-linear terms $\{b_{ij}\ X_{t-i}\ e_{t-j}\}$. Of course, this interpretation is only heuristic, since the linear terms and the non-

linear terms of the model (5.8.1) are not orthogonal in the usual sense of analysis of variance.

In view of this heuristic interpretation, it seems reasonable to consider first fitting a best AR model to the realisation and then seeing whether a further reduction in the mean sum of squares of residuals (or AIC) can be achieved by introducing extra non-linear terms. This is what motivated the development of the present algorithm.

Let $\{X_1, X_2, \ldots, X_N\}$ be a realisation from the time series $\{X_t\}$, and let $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$ be the sample mean. Suppose γ is the maximum order of the AR models we wish to fit to the data. The maximum lag of the non-linear terms $\{X_{t-r}\ e_{t-s}\}$ we include in the subset bilinear model is always less than or equal to γ. This choice of the maximum order restricts the effective number of observations to $n = N-\gamma$. When the order of the time series model is chosen, it is necessary to use the same number of observations over which we wish to fit and compare models of various orders.

The decision regarding the order of the model is made on the basis of the information criterion of Akaike (AIC), defined earlier.

We fit full autoregressive models of various orders for the realisation $(x_{m_1}, x_{m_1+1}, \ldots, x_N)$, where $x_t = X_t - \bar{X}$, $m_1 = \gamma+1$; and choose that model for which AIC is minimum. Let the order of this full linear AR model be p and let the model be

$$x_t + \sum_{i=1}^{p} a_i\ x_{t-i} = e_t \qquad (6.2.1)$$

Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2(1)}$ and its AIC be equal to AIC(1). The estimation of the full AR models have been discussed by

Box and Jenkins (1970), Chatfield (1980), Hannan (1962,1970) and Priestley (1981), and hence the details are omitted here. (For a brief account, see Appendix C).

Having fitted the full AR model (6.2.1) we can now fit the best subset AR model using the algorithm described by Haggan and Oyetunji (1980) (see also Oyetunji, 1979).

Let the best subset AR model be

$$x_t + a_{k_1} x_{t-k_1} + a_{k_2} x_{t-k_2} + \ldots + a_{k_\ell} x_{t-k_\ell} = e_t \qquad (6.2.2)$$

where $k_1, k_2, \ldots, k_\ell$ are subsets of the integers $(1, 2, \ldots, p)$ and let $1 \leq k_1 < k_2 < \ldots < k_\ell \leq p$. Let the mean sum of squares of the residuals be $\hat{\sigma}_e^{2\,(2)}$ and the AIC value be AIC (2), AIC (2) $\leq$ AIC (1).

The model (6.2.2) can be rewritten in the form

$$X_t + a_{k_1} X_{t-k_1} + a_{k_2} X_{t-k_2} + \ldots + a_{k_\ell} X_{t-k_\ell} - \alpha = e_t \qquad (6.2.3)$$

where

$$\alpha = \bar{X}\left[1 + \sum_{i=1}^{\ell} a_{k_i}\right].$$

We now define the bilinear model

$$X_t + a_{k_1} X_{t-k_1} + \ldots + a_{k_\ell} X_{t-k_\ell} = \alpha + e_t + b_{ij} X_{t-i} e_{t-j} \qquad (6.2.4)$$

where $k_1, k_2, \ldots, k_\ell$ are defined as above, and the term $b_{ij} X_{t-i} e_{t-j}$ is defined for all i and j such that $1 \leq i, j \leq \xi$, ξ being a suitable integer satisfying $p \leq \xi \leq \gamma$. Define the set $T_1 = \{(i,j);\ i = 1, 2, \ldots, \xi;\ j = 1, 2, \ldots, \xi\}$. Using the estimation method to be described in the following section, we fit models of the form (6.2.4) for all $(i,j) \in T_1$. We choose that combination, say (r_1,s_1) for which AIC (or the mean sum of

squares of residuals) is minimum. This subset bilinear model is defined as SBL(k_ℓ;1) and is given by

$$X_t + a_{k_1} X_{t-k_1} + \ldots + a_{k_\ell} X_{t-k_\ell} = \alpha + e_t + b_{r_1 s_1} X_{t-r_1} e_{t-s_1} \qquad (6.2.5)$$

Let the mean sum of squares of residuals be $\hat{\sigma}_e^2(\mathrm{SBL}(k_\ell,1))$ and its AIC be AIC(SBL(k_ℓ,1)).

IN the next stage of our estimation, we define the bilinear model SBL(k_ℓ,2) by

$$X_t + a_{k_1} X_{t-k_1} + \ldots + a_{k_\ell} X_{t-k_\ell} = \alpha + b_{r_1 s_1} X_{t-r_1} e_{t-s_1} + b_{r,s} X_{t-r} e_{t-s} + e_t \qquad (6.2.6)$$

where $(r,s) \in T_2$, $T_2 = T_1 - (r_1,s_1)$. In other words, (r,s) may be any of the elements of the set T_1 except (r_1,s_1). Let us suppose for the combination $(r_2,s_2) \in T_2$, the AIC is minimum. In this case the model SBL(k_ℓ,2) is

$$X_t + a_{k_1} X_{t-k_1} + \ldots + a_{k_\ell} X_{t-k_\ell} = \alpha + b_{r_1 s_1} X_{t-r_1} e_{t-s_1} + b_{r_2 s_2} X_{t-r_2} e_{t-s_2} + e_t \qquad (6.2.7)$$

We repeat the above procedure by adding extra bilinear terms until a subset model with minimum AIC is found. Let us suppose the model SBL(k_ℓ,m) given by

$$X_t + \sum_{i=1}^{\ell} a_{k_i} X_{t-k_i} = \alpha + \sum_{j=1}^{m} b_{r_j s_j} X_{t-r_j} e_{t-s_j} + e_t \qquad (6.2.8)$$

has attained the minimum AIC and this is our subset bilinear model. Here the pairs of integers $(r_1,s_1) \in T_1$, $(r_2,s_2) \in T_2 = T_1 - (r_1,s_1)$, ..., $(r_m,s_m) \in T_m = T_{m-1} - (r_{m-1},s_{m-1})$.

In the following section we describe the estimation of the model (6.2.8)

6.3 Estimation of the Parameters of SBL(k_ℓ,m)

In the estimation procedure to be discussed in this section we assume that the sets of integers $\{k_1, k_2, \ldots, k_\ell\}$ and $\{(r_1,s_1), (r_2,s_2), \ldots, (r_m,s_m)\}$ are fixed and known. Proceeding as in Subba Rao (1981a), and as in section 5.8 we can show that maximising the likelihood function of $(X_{\gamma+1}, X_{\gamma+2}, \ldots, X_N)$ is the same as minimising the function $Q(\theta) = \sum_{t=\gamma+1}^{N} e_t^2$ with respect to the parameters

$$(a_{k_1}, a_{k_2}, \ldots, a_{k_\ell}, b_{r_1s_1}, \ldots, b_{r_ms_m}, \alpha).$$

Let $\underline{\theta}' = (a_{k_1}, a_{k_2}, \ldots, a_{k_\ell}; b_{r_1s_1}, \ldots, b_{r_ms_m}, \alpha)$. For convenience we shall write

$$\theta_1 = a_{k_1}, \theta_2 = a_{k_2}, \ldots, \theta_\ell = a_{k_\ell}, \theta_{\ell+1} = b_{r_1s_1}, \ldots, \theta_{\ell+m} = b_{r_ms_m}, \theta_n = \alpha$$

where $n = \ell + m + 1$. We now apply the Newton-Raphson iteration method, discussed in section 5.8. The second order partial derivatives are approximated as is done in Marquardt algorithm (see Marquardt, 1963) by

$$\frac{\partial^2 Q(\theta)}{\partial\theta_i \partial\theta_j} \simeq 2 \sum_{t=\gamma+1}^{N} \left(\frac{\partial e_t}{\partial\theta_i}\right)\left(\frac{\partial e_t}{\partial\theta_j}\right)$$

and the first order partial derivatives satisfy the recursive equations

$$\left.\begin{aligned}
\frac{\partial e_t}{\partial a_{k_r}} &= X_{t-k_r} - \sum_{j=1}^{m} b_{r_js_j} X_{t-r_j} \frac{\partial e_{t-s_j}}{\partial a_{k_r}}, \\
\frac{\partial e_t}{\partial b_{r_qs_q}} &= -X_{t-r_q} e_{t-s_q} - \sum_{j=1}^{m} b_{r_js_j} X_{t-r_j} \frac{\partial e_{t-s_j}}{\partial b_{r_qs_q}}, \\
\frac{\partial e_t}{\partial \alpha} &= -1 - \sum_{j=1}^{m} b_{r_js_j} X_{t-r_j} \frac{\partial e_{t-s_j}}{\partial \alpha}.
\end{aligned}\right\} \qquad (6.3.1)$$

To start the iteration, we need to have good sets of initial values of the parameters. These can be obtained as follows.

Suppose we wish to fit the subset model SBL(k_ℓ,1). We choose the coefficients of the AR part of this model equal to the corresponding best subset AR model and set $b_{r_1s_1} = 0$. When we wish to fit the subset model SBL(k_ℓ,2), we choose the coefficients of this model to the corresponding coefficients of SBL(k_ℓ,1) and set $b_{r_2s_2} = 0$. Similarly, we calculate the initial estimates for other models.

6.4 Residuals

An important assumption we have made in fitting the linear and non-linear models to the data is that the errors $\{e_t\}$ of the model are mutually independent and Gaussian. Once a model is fitted to the data, it is therefore necessary to see whether the assumptions are satisfied. In the following we consider tests for dealing with the above problems.

Let $(y_1, y_2, \ldots, y_N)$ be a realisation from a discrete parameter, fourth order stationary time series $\{y_t\}$ with $E(y_t) = 0$, $\text{var}(y_t) = \sigma_y^2$, $R_y(s) = E(y_t\, y_{t+s})$, $\rho_y(s) = R_y(s)/R_y(0)$, $R_y(s_1,s_2) = E(y_t\, y_{t+s_1}\, y_{t+s_2})$, $R_y(s_1,s_2,s_3) = E(y_t\, y_{t+s_1}\, y_{t+s_2}\, y_{t+s_3})$.

Let
$$\bar{y} = \frac{1}{N}\sum_{t=1}^{N} y_t,$$

$$\hat{R}_y(s) = \frac{1}{N}\sum_{t=1}^{N-|s|} (y_t-\bar{y})(y_{t+|s|}-\bar{y}),\ s = 0, \pm1, \pm2, \ldots, \pm(N-1),$$

$$\hat{\rho}_y(s) = \frac{\hat{R}_y(s)}{\hat{R}_y(0)}.$$

The asymptotic properties of the sample covariances $\hat{R}_y(s)$ and sample correlations $\hat{\rho}_y(s)$ are well known (for example, see Anderson, 1971).

We can easily show that

$$\text{var}(\hat{\rho}_y(s)) \simeq \frac{\sigma_y^4}{N} \sum_{r=-\infty}^{\infty} \left[R_y(s,r,r+s) - 2\rho_y(s)\, R_y(0,r,r+s) + \rho_y^2(s)\, R_y(0,r,r) \right], \tag{6.4.1}$$

and in particular when s = 1,

$$\text{var}(\hat{\rho}_y(1)) \simeq \frac{\sigma_y^4}{N} \Big[R_y(0,1,1) - 2\rho_y(1)\, R_y(0,0,1) + \rho_y^2(1)\, R_y(0,0,0) + 2 \sum_{r=1}^{\infty} \left(R_y(1,r,r+1) - 2\rho_y(1)\, R_y(0,r,r+1) + \rho_y^2(1) R_y(0,r,r) \right) \Big] \tag{6.4.2}$$

If the y_t are mutually independent then $\rho_y(1) = 0$ and $R_y(0,r,r+1) = 0$, $R_y(1,r,r+1) = 0$, $r \neq 0$ and hence the asymptotic variance of $\hat{\rho}_y(1)$, under the null hypothesis of independence, is

$$\text{var}(\hat{\rho}_y(1)) \simeq \frac{\sigma_y^4}{N} \quad R_y(0,1,1) = \frac{1}{N}.$$

Under suitable conditions (see Parzen, 1957), one can show that if y_t are independent, $\hat{\rho}_y(1)$ is asymptotically normal with mean zero and variance $\frac{1}{N}$.

An immediate application of the above result is as follows.

Suppose we have fitted a time series model to the data, and obtained the residuals $(e_1, e_2, \ldots, e_N)$ from the fitted model. If the fit is adequate, the residuals must be mutually independent. To test for the independence, Granger and Andersen (1978b) have suggested second order covariance analysis on the squares of the residuals and Subba Rao (1978) has suggested the third order moments,

$$\frac{1}{N} \sum_{t=1}^{N-t_1} (e_t - \bar{e})(e_{t+t_1} - \bar{e})^2, \quad t_1 > 0.$$

In this chapter we consider the covariance analysis of the squares of the residuals.

Let $e_t^2 = y_t$ $(t = 1,2,\ldots,N)$. If e_t are independent (which implies the independence of y_t), $\hat{R}_y^{-2}(0)\, \hat{R}_y(0,1,1)$, where

$$\hat{R}_y(0,1,1) = \frac{1}{N} \sum_{t=1}^{N-1} (y_t - \bar{y})^2 (y_{t+1} - \bar{y})^2$$

must be near unity. Any substantial departure from unity must indicate that the y_t's are not independent. In fact we have calculated the statistic

$$W = \frac{\hat{\rho}(1)\, \hat{R}_y(0) \sqrt{N}}{\sqrt{\hat{R}_y(0,1,1)}} \tag{6.4.3}$$

for the residuals obtained from the fitted model. This statistic is asymptotically distributed as normal with mean zero and variance unity.

We now consider testing for the normality of $\{e_t\}$, and use the test proposed by Lomnicki (1961).

Let

$$m_k = \frac{1}{N} \sum_{t=1}^{N} (e_t - \bar{e})^k, \quad k = 2,3,4.$$

$$G_1 = \frac{m_3}{\sqrt{m_2^3}}, \quad G_2 = \frac{m_4}{m_2^2} - 3.$$

Lomnicki (1961) has shown that, for large N, if e_t are Gaussian and stationary, G_1 and G_2 are asymptotically normal with $E(G_1) = 0$, $E(G_2) = 0$, $\text{var}(G_1) = 6 \sum_{s=-\infty}^{\infty} \rho_e^2(s)$, $\text{var}(G_2) = 24 \sum_{s=-\infty}^{\infty} \rho_e^4(s)$. Since, in practice, $\rho_e(s)$ are unknown the variances of G_1 and G_2 are estimated by replacing $\rho_e(s)$ by $\hat{\rho}_e(s)$ and by replacing the infinite sum by a finite sum. Lomnicki (1961) has shown that the statistics

$$u_1 = \frac{G_1}{\sqrt{\text{var}(G_1)}}, \quad u_2 = \frac{G_2}{\sqrt{\text{var}(G_2)}} \tag{6.4.4}$$

are asymptotically distributed as standard normal variables. In the following section these tests have been performed on the residuals obtained from the fitted models. If $\{e_t\}$ are independent, $\rho_e(s) = 0$, $s \neq 0$.

Remarks

In the above analysis we should strictly use the estimates $\hat{e}_t$ for e_t ($t = 1,2,\ldots,N$) because the models are fitted and the residuals $\{e_t\}$ are estimated from the models. So the expressions given above are only approximately valid. We assume the difference $\hat{e}_t - e_t$ is negligible asymptotically.

6.5 FITTING SUBSET BILINEAR MODELS TO TIME SERIES DATA

(i) Wolfer Sunspot Numbers

For our first illustration we consider the annual sunspot numbers for the years 1700-1955, giving 256 observations. These series are given in Appendix D.

The linear and bilinear models are fitted to the first 221 observations, and then predictors are calculated for the next 35 observations. The maximum order, γ, is chosen to be equal to 10; this is because the order of the best AR model is 9.

The fitted models are as follows:-

(a) Full AR Model

The fitted model to the mean deleted observations x_t, $(x_t = X_t - \bar{X})$ is

$$x_t - 1.2163x_{t-1} + 0.4670x_{t-2} + 0.1416x_{t-3} - 0.1691x_{t-4} + 0.1473x_{t-5} - 0.0543x_{t-6} + 0.0534x_{t-7} - 0.0667x_{t-8} - 0.1129x_{t-9} = e_t \qquad (6.5.1)$$

(b) Best subset AR model

The best subset AR model is

$$x_t - 1.2496x_{t-1} + 0.5510x_{t-2} - 0.1450x_{t-9} = e_t \qquad (6.5.2)$$

(c) Best subset bilinear model

The best subset bilinear model is

$$\begin{aligned} X_t + \hat{a}_1X_{t-1} + \hat{a}_2X_{t-2} + \hat{a}_9X_{t-9} = \hat{\alpha} &+ \hat{b}_{2,1}X_{t-2}e_{t-1} + \hat{b}_{8,1}X_{t-8}e_{t-1} \\ &+ \hat{b}_{1,3}X_{t-1}e_{t-3} + \hat{b}_{4,3}X_{t-4}e_{t-3} + \hat{b}_{1,6}X_{t-1}e_{t-6} \\ &+ \hat{b}_{2,4}X_{t-2}e_{t-4} + \hat{b}_{3,2}X_{t-3}e_{t-2} + e_t \end{aligned} \qquad (6.5.3)$$

where

$\hat{a}_1 = -1.5012$, $\hat{a}_2 = .7670$, $\hat{a}_9 = -.1152$, $\hat{\alpha} = 6.886$

$\hat{b}_{2,1} = -0.1458$, $\hat{b}_{8,1} = .006312$, $\hat{b}_{1,3} = -.007152$, $\hat{b}_{4,3} = -006047$

$\hat{b}_{1,6} = .003619$, $\hat{b}_{2,4} = .004334$, $\hat{b}_{3,2} = .001782$.

(d) Threshold Autoregressive Model

Tong and Lim (1980, p.261, eqn.(9.2)) have fitted the following self-exciting threshold autoregressive model, SETAR(2;4,12) to the above data,

$$X_t = \begin{cases} 10.544 + 1.692X_{t-1} - 1.1592X_{t-2} + .2367X_{t-3} + .1503X_{t-4} \\ \qquad + \varepsilon_t(1) \quad \text{if } X_{t-3} < 36.6 \\ \\ 7.8041 + .7432X_{t-1} - .0409X_{t-2} - .202X_{t-3} + .173X_{t-4} \\ \qquad - .2266X_{t-5} + .0189X_{t-6} + .1612X_{t-7} - .256X_{t-8} \\ \qquad + .319X_{t-9} - .3891X_{t-10} + .4306X_{t-11} - 0.397X_{t-12} \\ \\ \qquad + \varepsilon_t(2) \quad \text{if } X_{t-3} > 36.6, \end{cases}$$

where the maximum order chosen for this model is $\gamma = 20$. The one-step-ahead predictors of this model are given in their paper.

The mean sum of squares of residuals, and the normalized AIC values of all the four models (a), (b), (c) and (d) are given in Table 6.1. The model (6.5.3) has the smallest mean sum of squares of residuals and NAIC when compared to other models.

Predictive Performance of the models In order to compare the performance of the above models, it is necessary that we should obtain the forecasts. The forecasts and their errors are obtained as follows.

Let $\tilde{X}_{N+\ell-1}(m) = E(X_{N+\ell+m-1}/X_{N+\ell-1}, X_{N+\ell-2}, \dots)$, and it is known that $\tilde{X}_{N+\ell-1}(m)$ is the optimal m step ahead predictor of $X_{N+\ell+m-1}$. Let the error of this predictor be $e_{N+\ell}(m) = X_{N+\ell+m-1} - \tilde{X}_{N+\ell-1}(m)$ ($\ell = 1,2,\dots,M$; $m = 1,2,\dots$). Let the predictors obtained by replacing the unknown parameters in $\tilde{X}_{N+\ell-1}(m)$ by their least squares estimates be denoted by $\hat{X}_{N+\ell-1}(m)$ and the errors by $\hat{e}_{N+\ell}(m)$. Define the mean sum of squares of errors by

$$\hat{\sigma}_e^2(m) = \frac{1}{M} \sum_{\ell=1}^{M} \hat{e}_{N+\ell}^2(m) \quad (m = 1,2,\dots) \tag{6.5.5}$$

For linear models, it is quite easy to calculate $\hat{X}_{N+\ell-1}(m)$ for all values of m. For the subset bilinear model (6.2.8), the optimum ℓ-step-ahead predictors $\tilde{X}_{N+\ell-1}(m)$ can be obtained by making use of the following properties:

(a) Under the assumption of invertibility, knowledge of $\{X_1,\dots,X_N\}$ implies the knowledge of $\{e_1,\dots,e_N\}$, and $\{e_{N+1},e_{N+2},\dots\}$ are independent identically distributed $N(0,\sigma_e^2)$ random variables;

(b) The variables X_{t+r} and e_{t+s} are independent for $s > r$;

(c)
$$E(X_{t+\ell-i}\, e_{t+\ell-j}/X_t, X_{t-1}, \dots) = \begin{cases} X_{t+\ell-i}\, e_{t+\ell-j} & \text{if } i, j \geq \ell \\ \tilde{X}_t(\ell-i)\, e_{t+\ell-j} & \text{if } i < \ell, j < \ell \\ 0 & \text{if } j < i, j < \ell \\ - & \text{if } i < j < \ell \end{cases} \tag{6.5.5}$$

where, in the last case, when $i < j < \ell$, we have to resubstitute for $X_{t+\ell-i}$ from (6.2.8) to decrease the lag $(\ell-i)$ and then take the conditional expectations. For example, the ℓ-step-ahead predictors $\hat{X}_t(\ell)$ ($\ell = 1, \dots, 5$) for the sunspot numbers using the fitted subset bilinear model (6.5.3) are given as follows.

$$\hat{X}_t(1) = -\hat{a}_1 X_t - \hat{a}_2 X_{t-1} - \hat{a}_9 X_{t-8} + \hat{\alpha} + \hat{b}_{2,1} X_{t-1}\, e_t + \hat{b}_{1,3} X_t\, e_{t-2} + \hat{b}_{4,3} X_{t-3}\, e_{t-2} + \hat{b}_{1,6} X_t\, e_{t-5} + \hat{b}_{2,4} X_{t-1}\, e_{t-3} + \hat{b}_{3,2} X_{t-2}\, e_{t-1}$$

$$\hat{X}_t(2) = -\hat{a}_1 \hat{X}_t(1) - \hat{a}_2 X_t - \hat{a}_9 X_{t-7} + \hat{\alpha} + \hat{b}_{1,3} \hat{X}_t(1)\, e_{t-1} + \hat{b}_{4,3} X_{t-2}\, e_{t-1} + \hat{b}_{1,6} \hat{X}_t(1)\, e_{t-4} + \hat{b}_{2,4} X_t\, e_{t-2} + b_{3,2} X_{t-1}\, e_t,$$

$$\hat{X}_t(3) = -\hat{a}_1 \hat{X}_t(2) - \hat{a}_2 \hat{X}_t(1) - \hat{a}_9 X_{t-6} + \hat{\alpha} + \hat{b}_{1,3} \hat{X}_t(2)\, e_t + \hat{b}_{4,3} X_{t-1}\, e_t + \hat{b}_{1,6} \hat{X}_t(2)\, e_{t-3} + \hat{b}_{2,4} \hat{X}_t(1)\, e_{t-1},$$

$$\hat{X}_t(4) = -\hat{a}_1 \hat{X}_t(3) - \hat{a}_2 \hat{X}_t(2) - \hat{a}_9 X_{t-5} + \hat{\alpha} + \hat{b}_{1,3}\, \hat{XE}_t(3,1) + \hat{b}_{1,6} \hat{X}_t(3)\, e_{t-2} + \hat{b}_{2,4} \hat{X}_t(2)\, e_t, \quad \text{and}$$

$$\hat{X}_t(5) = -\hat{a}_1 \hat{X}_t(4) - \hat{a}_2 X_t(3) - \hat{a}_9 X_{t-4} + \hat{\alpha} + \hat{b}_{1,3}\, \hat{XE}_t(4,2) + \hat{b}_{1,6} \hat{X}_t(3)\, e_{t-1} + b_{2,4} \hat{XE}_t(3,1),$$

where

$$XE_t(r,s) = E(X_{t+r}\, e_{t+s}/X_t, X_{t-1}, \dots),$$

and can be obtained, recursively, as follows. Since for all $r > 0$

$$E(X_{t+r}\, e_{t+r}/X_t, X_{t-1}, \dots) = E(e_{t+r}^2) = \sigma_e^2$$

$XE_t(r,r)$ can be estimated by

$$\hat{XE}_t(r,r) = \hat{\sigma}_e^2, \quad (r = 1,2,\ldots,)$$

where $\hat{\sigma}_e^2(1)$ is the residual variance (mean sum of squares of residuals). Now replace t by (t+2) in (6.2.8) and then multiply both sides by e_{t+1}. Taking the conditional expectations both sides, we obtain, after replacing the parameters by their estimates,

$$\hat{XE}_t(2,1) = -\hat{a}_1\hat{\sigma}_e^2 + \hat{b}_{2,1}x_t\hat{\sigma}_e^2 + \hat{b}_{8,1}x_{t-6}\hat{\sigma}_e^2 + \hat{b}_{1,3}e_{t-1}\hat{\sigma}_e^2 + \hat{b}_{1,6}e_{t-4}\hat{\sigma}_e^2.$$

Similarly,

$$\hat{XE}_t(3,1) = -\hat{a}_1 XE_t(2.1) - \hat{a}_2\hat{\sigma}_e^2 + \hat{b}_{1,3}\hat{XE}_t(2,1) + \hat{b}_{1,6} + \hat{b}_{1,6}\hat{XE}_t(2,1)e_{t-3} + \hat{b}_{2,4}\hat{\sigma}_e^2 e_{t-1} + \hat{b}_{3,2}x_t\hat{\sigma}_e^2;$$

$$\hat{XE}_t(3,2) = -\hat{a}_1\hat{\sigma}_e^2 + \hat{b}_{2,1}\hat{X}_t(1)\hat{\sigma}_e^2 + \hat{b}_{8,1}x_{t-5}\hat{\sigma}_e^2 + \hat{b}_{1,3}\hat{\sigma}_e^2 e_t + \hat{b}_{1,6}\hat{\sigma}_e^2 e_{t-3}$$

$$\hat{XE}_t(4,2) = -\hat{a}_1\hat{XE}_t(3,2) - \hat{a}_2\hat{\sigma}_e^2 + \hat{b}_{1,3}\hat{b}_{2,1}\hat{\sigma}_e^2\,\hat{\sigma}_e^2 + \hat{b}_{1,6}\hat{XE}_t(3,2)e_{t-2} + \hat{b}_{2,4}\hat{\sigma}_e^2 e_t + \hat{b}_{3,2}\hat{X}_t(1)\,\hat{\sigma}_e^2.$$

Finally, for the threshold model (6.5.4), the calculation of the optimal forecasts for more than three steps does not seem to be possible. However, as an approximation, we have replaced the threshold values, namely $X_{t+\ell-3} > 36.6$ ($\ell = 4,5,\ldots$) and $X_{t+\ell-3} < 36.6$ by the corresponding estimated values $\hat{X}_t(\ell-3) > 36.6$ ($\ell = 4,5,\ldots$) and $\hat{X}_t(\ell-3) < 36.6$, etc. To circumvent the problems just mentioned, Tong and Wu (1981) suggest shifting the 'trigger' and obtaining several steps ahead forecasts.

Model	Full AR(9)	Subset AR	SETAR	Bilinear Subset
$\hat{\sigma}_e^2$	199.27	203.21	153.71	124.33
NAIC	5.39	5.352	5.00	4.927
No. of parameters	10	4	19	11
$\hat{\sigma}_e^2(1)$	190.89	214.1	148.205	123.77
$\hat{\sigma}_e^2(2)$	414.83	421.4	383.9	337.54
$\hat{\sigma}_e^2(3)$	652.21	660.38	675.59	569.79
$\hat{\sigma}_e^2(4)$	725.85	716.08	773.51	659.047
$\hat{\sigma}_e^2(5)$	771.04	756.39	784.27	718.866

TABLE 6.1: Values of $\hat{\sigma}_e^2$, NAIC and $\hat{\sigma}_e^2(h)$ (Sunspot data)

The one-step-ahead predictor values, i.e. $\hat{X}_{N+\ell-1}(1)$ for various values of ℓ for the models (6.5.1), (6.5.2), and (6.5.3) for the period 1922-1935 are given in Table 6.2; and for the same period Tong and Lim (1980) have also given the $\hat{X}_{N+\ell-1}(1)$ values for the threshold model (6.5.2).

The mean sum of squares of the errors of the predictors obtained by all the four models for the period 1922-1935 and for $\ell = 1,2,\ldots,5$ are given in Table 6.1.

From Table 6.1, it is clear that the bilinear model (6.5.3) not only has smaller mean sum of squares of residuals when compared to other models, but also the performance of this model seems to be better when it is used for forecasting several steps ahead.

TABLE 6.2: 1-STEP-AHEAD PREDICTION OF SUNSPOT NUMBERS

I	X(I)	SBL (6.5.3)	FULL AR (6.5.1)	SUBSET AR (6.5.2)
222	26.1	24.55	24.60	19.07
223	14.2	11.43	13.13	18.69
224	5.8	10.08	13.74	11.39
225	16.7	9.90	9.96	13.12
226	44.3	37.52	35.25	32.82
227	63.9	75.67	67.83	68.32
228	69.0	77.86	74.57	74.11
229	77.8	73.93	67.11	67.14
230	64.9	75.04	72.20	71.42
231	35.7	57.03	49.87	48.73
232	21.2	29.34	16.12	17.56
233	11.1	9.37	14.05	14.27
234	5.7	.68	11.49	11.28
235	8.7	10.94	13.28	14.23
236	36.1	21.84	24.12	23.90
237	79.7	63.85	57.50	57.25
238	114.4	105.82	98.80	97.95
239	109.6	106.99	114.89	115.36
240	88.8	85.10	86.85	85.86
241	67.8	58.95	60.36	60.34
242	47.5	38.76	44.48	44.04
243	30.6	24.79	27.11	29.44
244	16.3	20.28	17.92	19.95
245	9.6	16.20	14.69	15.51
246	33.2	16.58	21.63	21.55
247	92.6	73.04	59.19	59.94
248	151.6	140.92	119.53	120.44
249	136.3	148.47	157.61	158.32
250	134.7	121.87	104.42	103.54
251	83.9	87.77	104.82	106.93
252	69.4	50.79	45.94	41.80
253	31.5	30.73	40.88	49.52
254	13.9	7.01	11.57	9.15
255	4.4	14.83	4.20	11.58
256	38.0	7.65	22.63	18.31
$\hat{\sigma}_e^2(1)$		123.77	190.89	214.07

Testing of Residuals Using the tests described in section 6.4, we have tested for the independence and Gaussianity of the residuals. The statistics W, defined by (6.4.3), and u_1,u_2 defined by (6.4.4) are calculated for the residuals obtained from the linear models (6.5.1), (6.5.2) and the bilinear models (6.5.3). The results are shown in Table 6.3. From these results, we can conclude that the residuals obtained from the SBL model (6.5.3) are independent and Gaussian.

Model	W	u_1	u_2
Full AR(9) (6.5.1)	2.459	0.333	0.360
Subset AR (6.5.2)	2.520	0.306	0.331
SBL (6.5.3)	1.561	0.302	0.200

TABLE 6.3: Testing of independence and Gaussianity of the residuals of models fitted to the sunspot data

(ii) Canadian Lynx Data

For our second illustration we consider the logarithm (to the base 10) of the annual record of the numbers of Canadian lynx data for the years 1821-1934 (inclusive) giving a total of 114 observations (see Appendix D).

The linear and bilinear models are fitted to the observations over the period 1821-1920, giving N = 100, and the predictions are obtained for the next 14 observations. Since the best AR model is found to be AR(12), the maximum order is chosen to be $\gamma = 12$.

(a) Full AR model

The full AR model fitted to the data is

$$x_t - 1.0541x_{t-1} + 0.4539x_{t-2} - 0.32597x_{t-3} + 0.37912x_{t-4} - 0.23452x_{t-5} + 0.17570x_{t-6} - 0.09598x_{t-7} + 0.12843x_{t-8} - 0.27435x_{t-9} - 0.11427x_{t-10} + 0.18534x_{t-11} + 0.17128x_{t-12} = e_t \quad (6.5.6)$$

(b) The best subset AR model

The fitted subset AR model is

$$x_t - 1.01705x_{t-1} + 0.39997x_{t-2} - 0.25851x_{t-3} + 0.22037x_{t-4} - 0.21099x_{t-9} + 0.25343x_{t-12} = e_t \quad (6.5.7)$$

(c) The best subset bilinear model

The fitted best SBL model is

$$X_t + \hat{a}_1X_{t-1} + \hat{a}_2X_{t-2} + \hat{a}_3X_{t-3} + \hat{a}_4X_{t-4} + \hat{a}_9X_{t-9} + \hat{a}_{12}X_{t-12} = \hat{\alpha} + \hat{b}_{3,9}X_{t-3}\,e_{t-9} + \hat{b}_{9,9}\,X_{t-9}\,e_{t-9} + \hat{b}_{6,2}X_{t-6}\,e_{t-2} + \hat{b}_{1,1}X_{t-1}\,e_{t-1} + \hat{b}_{2,7}\,X_{t-2}\,e_{t-7} - \hat{b}_{4,2}X_{t-4}\,e_{t-2} + e_t \quad (6.5.8)$$

where

$\hat{a}_1 = -.77277$, $\hat{a}_2 = .09157$, $\hat{a}_3 = -.08307$,
$\hat{a}_4 = .26149$, $\hat{a}_9 = -.22559$, $\hat{a}_{12} = .24584$
$\hat{\alpha} = -1.48629$, $\hat{b}_{3,9} = -.7893$, $\hat{b}_{9,9} = .4798$,
$\hat{b}_{6,2} = .3902$ $\hat{b}_{1,1} = .1326$, $\hat{b}_{2,7} = .07944$,
$\hat{b}_{4,2} = -.3212$.

(d) Tong (1980) has fitted a threshold AR model SETAR (2;6,3) to the above data, and the model is

$$X_t = \begin{cases} 0.8023 + 1.0676\,X_{t-1} - 0.2069X_{t-2} + 0.1712X_{t-3} - 0.4528X_{t-4} \\ + 0.2237X_{t-5} - 0.0331X_{t-6} + \varepsilon_t^{(1)} \quad \text{if } 0 < X_{t-2} < 3.05 \\ 2.2964 + 1.4246X_{t-1} - 1.0795X_{t-2} - 0.0907X_{t-3} + \varepsilon_t^{(2)} \quad \text{if } 10 > X_{t-2} > 3.05 \\ 0 \quad \text{otherwise} \end{cases}$$

The mean sum of squares of residuals and the normalized AIC values of the above four models are given in Table 6.4. It is useful to note here that the mean sum of squares of residuals for the model SETAR (2;6,3) is greater than the full linear AR model and the best subset AR model. The mean sum of squares of residuals of the subset bilinear model (6.5.8) has the smallest value when compared to all the four models.

One-step-ahead forecasts from the linear models (6.5.6), (6.5.7) and the bilinear model (6.5.8) are calculated and shown in Table 6.5, while one-step-ahead forecasts from model (6.5.9) are given in Tong and Lim (1980), p.286.

We have calculated the forecasts $\hat{X}_{N+\ell-1}(m)$ ($m = 1,2,\ldots,6$; $N = 100$) from all the above four models as discussed in the previous example (i). In the case of the threshold AR models, as pointed out in example (i), it is not possible to obtain optimal forecasts for steps 3 and beyond, and therefore we have calculated approximate forecasts as before (see Tong and Wu, 1981). The mean sum of squares of the prediction errors $\hat{\sigma}_e^2(h)$ ($h = 1,2,\ldots,6$) are given in Table 6.4.

The mean sum of squares of the one-step-ahead prediction errors obtained from the bilinear model is smaller than all the models considered here. However, it is interesting to note that for more than one step, the performance of the threshold model is better than all other models, with the exception of six-step-ahead, despite the fact that it has the largest sum of squares of residuals.

In our experience with this series, we have found it is possible to find subset bilinear models which have slightly higher mean sum of squares of residuals, but giving smaller sum of squares of prediction errors. This seems to be so because the model in this case is very sensitive for

any slight change in the parameters and the number of observations over which the model is fitted.

Model	Full AR	AR Subset	SETAR (2,6,3)	Bilinear Subset
$\hat{\sigma}_e^2$	0.0358	0.0378	0.0415	0.0223
NAIC	-3.033	-3.116	-2.985	-3.508
No. of parameters	13	7	12	13
$\hat{\sigma}_e^2(1)$	0.02549	0.02233	0.01448	0.01331
$\hat{\sigma}_e^2(2)$	0.07377	0.07240	0.0259	0.04432
$\hat{\sigma}_e^2(3)$	0.11612	0.11992	0.0329	0.06282
$\hat{\sigma}_e^2(4)$	0.16121	0.16873	0.03744	0.07657
$\hat{\sigma}_e^2(5)$	0.18488	0.20211	0.0481	0.08596
$\hat{\sigma}_e^2(6)$	0.18560	0.20690	0.12268	0.07562

TABLE 6.4: Values of $\hat{\sigma}_e^2$, NAIC and $\hat{\sigma}_e^2(h)$ (Canadian Lynx Data)

The residuals obtained from models (6.5.6), (6.5.7) and (6.5.8) are tested for independence and Gaussianity. The results are shown in Table 6.6. From these results we can conclude that the residuals obtained from the SBL model (6.5.8) are independent and Gaussian.

It may be pointed out that Nicholls and Quinn (1982, p.46) have also fitted autoregressive models with random coefficients to the same data, and the one step ahead forecasts obtained from their models are quite comparable with the above models.

TABLE 6.5: 1-STEP-AHEAD PREDICTION OF LOG CANADIAN LYNX DATA

I	X(I)	SBL (6.5.8)	FULL AR (6.5.6)	SUBSET AR (6.5.7)
101	2.360	2.410	2.389	2.362
102	2.601	2.745	2.812	2.792
103	3.054	2.911	2.788	2.863
104	3.386	3.211	3.197	3.206
105	3.553	3.341	3.354	3.338
106	3.468	3.438	3.431	3.303
107	3.187	3.152	2.860	2.946
108	2.723	2.569	2.624	2.636
109	2.686	2.796	2.485	2.435
110	2.821	2.825	2.853	2.832
111	3.000	3.056	2.973	2.978
112	3.201	3.175	3.255	3.262
113	3.424	3.291	3.397	3.425
114	3.531	3.444	3.563	3.562
$\hat{\sigma}_e^2(1)$		.0133	.0255	.0223

Model	W	u_1	u_2
Full AR	.0871	.0098	-.0542
Subset AR	-1.313	.1222	-.0162
SBL	-1.049	.1046	-.0270

TABLE 6.6: Testing of independence and normality of the residuals of models fitted to log Canadian lynx data

Original Canadian Lynx Data

The models fitted above are to the logarithmically transformed Canadian lynx data. The forecasts obtained for the period 1921-1934 (both years inclusive) are once again for the transformed data. Of course, in practice, we would take the antilogarithms of these forecasts and compare them with the original data to see how good the forecasts are. It is well known that these forecasts are biased and lead to higher mean square error (Granger and Newbold, 1976) and we substantiate this in this section. It is, therefore, interesting to see whether we can fit a subset bilinear model to the same period 1821-1920 (using 100 observations) and obtain one-step-ahead forecasts for the rest of the period, i.e. 1921-1934. We compare these forecasts with the forecasts obtained from the model (6.5.8). The subset bilinear model fitted for the orginal Canadian lynx data is

$$X(t) = 1.2598X(t-1) + 0.6548X(t-2) + 0.1564X(t-7) - 0.3392X(t-8) - 336.2$$

$$= -(0.4151 \times 10^{-3})\, X(t-8)\, e(t-10) + (0.8036 \times 10^{-4})\, X(t-9)\, e(t-10) - (0.1946 \times 10^{-3})\, X(t-4)\, e(t-1) + (0.2380 \times 10^{-3})\, X(t-3)\, e(t-5) - (0.5296 \times 10^{-4})\, X(t-7)\, e(t-6) - (0.2801 \times 10^{-3})\, X(t-3)\, e(t-7) - (0.6984 \times 10^{-4})\, X(t-1)\, e(t-1) + e(t) \qquad (6.5.9)$$

with mean sum of squares $\hat{\sigma}_e^2 = 477964.68$. The one-step-ahead forecasts obtained from the model (6.5.9) for the period 1921-1934 are given in Table 6.7, together with the mean of the errors and mean sum of squares.

A comparison of the results from Table 6.7 (column 6) shows that the errors of the forecasts obtained from the inverse transforms are very positively biased, and the mean sum of squares of the errors is nearly three and a half times that of the mean sum of squares of the errors obtained from the model fitted to the original data. This analysis shows that if one is

dealing with non-Gaussian time series (and non-linear as well), it may be better to fit a model to the original data (if it is possible) than to try to transform the data, and fit a linear or non-linear model to the transformed data.

Nicholls and Quinn (1982, p.146) have fitted an autoregressive model with random coefficients for the original lynx data, and the root mean square prediction error (i.e. $\hat{\sigma}_e(1)$) for this model is 307.00 which is substantially larger than for the subset bilinear model for which $\hat{\sigma}_e(1)$ is 264.25.

Year	True Values X(t)	Transformed Data: Y(t) = log X(t)	Transformed Data: Predictors for Y(t)	Transformed Data: $\hat{X}(t)=10^{\hat{Y}(t)}$	Transformed Data: error $\hat{=}$ X(t)-$\hat{X}$(t)	Original Data: Predictors from (6.5.9)	Original Data: Errors
1921	229	2.360	2.41	257.0	- 28.0	470.4	-241.4
1922	399	2.601	2.745	555.9	-156.9	679.4	-280.4
1923	1132	3.054	2.911	814.7	317.3	1045.7	86.3
1924	2432	3.386	3.211	1625.5	806.5	2586.2	-154.2
1925	3574	3.553	3.341	2192.8	1381.2	2842.6	731.4
1926	2935	3.468	3.438	2741.6	193.4	2855.5	79.5
1927	1537	3.187	3.152	1419.1	117.9	1609.8	-72.8
1928	529	2.723	2.569	370.7	158.3	689.8	-160.8
1929	485	2.686	2.796	625.2	-140.2	216.1	268.9
1930	662	2.821	2.825	668.3	- 6.3	650.1	11.9
1931	1000	3.000	3.056	1137.6	- 137.6	896.4	103.6
1932	1590	3.201	3.175	1496.2	93.8	1545.2	44.8
1933	2657	3.424	3.291	1954.3	702.7	2288.6	368.4
1934	3396	3.531	3.444	2779.7	616.3	3522.8	-126.8
		$\hat{\sigma}_e^2(1)$=.01331		Mean of the errors = 279.9	$\hat{\sigma}_e^2(1)$=262974.3	Mean of the errors = 8.62	$\hat{\sigma}_e^2(1)$=69825.6

TABLE 6.7. Comparison of the Forecasts

(iii) Unemployment Data of West Germany

For our third illustration we consider the number of people registered as unemployed in West Germany for the period January 1948 - May 1980 (inclusive) The data is monthly, and the total number of observations is 389. The linear and subset bilinear models are fitted to the differenced data $\{X_t\}$, as suggested earlier, with difference operator $(1-B)(1-B^{12})$. The first 365 observations are used for fitting the models, and the next 24 observations are used for prediction purposes. The maximum order is chosen to be equal to $\gamma = 20$.

(a) Full AR Model

The fitted model to the mean deleted observations, $x_t = X_t - \bar{X}$, is

$$x_t + 0.093167x_{t-1} + 0.142049x_{t-2} + 0.011562x_{t-3} + 0.050863x_{t-4} - 0.006261x_{t-5} + 0.022672x_{t-6} - 0.028367x_{t-7} - 0.005289x_{t-8} - 0.070334x_{t-9} + 0.047378x_{t-10} - 0.295109x_{t-11} + 0.0378162x_{t-2} = e_t \quad (6.5.10)$$

(b) Subset AR Model

$$x_t + 0.097011x_{t-1} + 0.1365x_{t-2} - 0.0768x_{t-9} - 0.3032x_{t-11} + 0.3741x_{t-12} = e_t \quad (6.5.11)$$

(c) The best subset bilinear model is of the form

$$X_t + \hat{a}_1X_{t-1} + \hat{a}_2X_{t-2} + \hat{a}_9X_{t-9} + \hat{a}_{11}X_{t-11} + \hat{a}_{12}X_{t-12} = \hat{\alpha} + \hat{b}_{1,10}X_{t-1}e_{t-10} + \hat{b}_{2,5}X_{t-2}e_{t-5} + \hat{b}_{5,4}X_{t-5}e_{t-4} + \hat{b}_{11,7}X_{t-11}e_{t-7} + \hat{b}_{12,4}X_{t-12}e_{t-4} + \hat{b}_{12,2}X_{t-12}e_{t-2} + \hat{b}_{4,10}X_{t-4}e_{t-10} + \hat{b}_{10,8}X_{t-10}e_{t-8} + \hat{b}_{1,9}X_{t-1}e_{t-9} + e_t \quad (6.5.12)$$

where $\hat{a}_1 = -0.087412$, $\hat{a}_2 = 0.12610$, $\hat{a}_9 = -0.04263$, $\hat{a}_{11} = -0.255587$, $\hat{a}_{12} = 0.506737$, $\hat{\alpha} = -4598.325$;

$\hat{b}_{1,10} = -0.1315 \times 10^{-4}$, $\hat{b}_{11,7} = 0.1902 \times 10^{-5}$, $\hat{b}_{4,10} = -0.9507 \times 10^{-6}$

$b_{2,5} = -0.1279 \times 10^{-5}$, $b_{12,4} = 0.1513 \times 10^{-5}$, $b_{10,8} = -0.1948 \times 10^{-5}$

$b_{5,4} = -0.3790 \times 10^{-6}$, $b_{12,2} = -0.2267 \times 10^{-5}$, $b_{1,9} = 0.2715 \times 10^{-5}$

The mean sum of squares of residuals and AIC values are given in Table 6.8, from which we see that the AIC value and the mean sum of squares obtained from the bilinear model are smaller than other linear models.

The one-step-ahead forecasts $\hat{X}_{N+\ell-1}(1)$ ($\ell = 1,2,\ldots$) for the differenced series $\{X_t\}$ are calculated for the above model. The one-step-ahead forecasts for the original data $\{Y_t\}$, where

$$X_t = (1-B)(1-B^{12})Y_t$$
$$= Y_t - Y_{t-1} - Y_{t-12} + Y_{t-13}$$

are calculated by

$$\hat{Y}_t(1) = \hat{X}_t(1) + Y_t + Y_{t-11} - Y_{t-12}$$

Similarly,

$$\hat{Y}_t(2) = \hat{X}_t(2) + \hat{Y}_t(1) + Y_{t-10} - Y_{t-11}$$

and so on.

In Table 6.10.we have given the one-step-ahead forecasts from the above three models, over the period June 1978 - May 1980 (both months included). The mean sum of squares of the errors $\hat{\sigma}_e^2(\ell)$ ($\ell = 1, 2, \ldots, 5$) obtained from the models (6.5.10), (6.5.11), and (6.5.12) are given in Table 6.8.

Model	Full AR	Subset AR	Bilinear Subset
σ_e^2	0.81048*	0.81831*	0.36665*
AIC	7600.82	7590.011	7341.469
$\hat{\sigma}_e^2(1)$	0.51109**	0.50388**	0.36897**
$\hat{\sigma}_e^2(2)$	0.11254*	0.10981*	0.75978**
$\hat{\sigma}_e^2(3)$	0.16619*	0.162685*	0.12446*
$\hat{\sigma}_e^2(4)$	0.19934*	0.19677*	0.15105*
$\hat{\sigma}_e^2(5)$	0.24927*	0.24865*	0.19703*

TABLE 6.8 Values of $\hat{\sigma}_e^2$, AIC and $\hat{\sigma}_e^2(h)$ (Unemployment data)

* All these numbers must be multiplied by 10^{10}

** All these numbers must be multiplied by 10^9

The residuals obtained from models (6.5.10), (6.5.11), (6.5.12) are tested and the results are shown in Table 6.9. The residual obtained from (6.5.12) are found to be independent and Gaussian.

Model	W	u_1	u_2
Full AR(12)	3.05	-.394	2.228
Subset AR	2.95	-.369	2.270
SBL	1.53	-.267	0.600

TABLE 6.9 Testing of independence and normality of the residuals obtained from the fitted models to the unemployment data

TABLE 6.10 1-STEP-AHEAD PREDICTION OF UNEMPLOYMENT FIGURES IN W.GERMANY

I	X(I)	SBL (6.5.12)	FULL AR (6.5.10)	SUBSET AR (6.5.11)
366	877319.	887551.	883577.	887083.
367	922230.	892999.	896832.	895746.
368	923963.	942116.	936876.	936266.
369	864274.	864625.	868762.	868920.
370	901636.	912798.	914751.	911619.
371	927043.	938933.	943428.	945840.
372	1006724.	994642.	1011585.	1011201.
373	1171353.	1154829.	1153556.	1155095.
374	1134060.	1150142.	1166575.	1161692.
375	957711.	964916.	987000.	986540.
376	875452.	884646.	894399.	894233.
377	775117.	788336.	777893.	780898.
378	763173.	754035.	762048.	758275.
379	803653.	772547.	778214.	776907.
380	798867.	813662.	810957.	809399.
381	736809.	738518.	740761.	739147.
382	761724.	767786.	768253.	771427.
383	798973.	787855.	789539.	790325.
384	866783.	884275..	890446.	888688.
385	1036519.	989472.	1003350.	1000164.
386	992520.	1001631.	1000566.	1000098.
387	875909.	846978.	839789.	840173.
388	825374.	797121.	776477.	776632.
389	766768.	741394.	722383.	723537.
$\hat{\sigma}_e^2(1)$		.369E+09	.511E+09	.504E+09

Conclusions

We have described an algorithm for estimating subset bilinear time series models, and the estimation procedure is illustrated with three real time series which are believed to be non-linear. For these non-linear models, it is possible to calculate optimal forecasts for several steps ahead. On the basis of the forecasting performance of these, it seems that these subset bilinear models can form a useful class of non-linear models. It is possible that still better forecasts are obtainable by the combination of several forecasts (see Granger and Newbold, 1977) for the three series considered here, but this is not dealt with here.

CHAPTER 7

MARKOVIAN REPRESENTATION AND EXISTENCE THEOREMS FOR BILINEAR TIME SERIES MODELS

7.1 MARKOVIAN REPRESENTATIONS

The Markovian representation (or state space form) plays an important role in the theory of linear time series models. Akaike (1974) has shown that the autoregressive - moving average model admits a linear Markovian representation and vice versa. A natural extension is to generalise the above results for the linear models to the bilinear models. It has been shown in section 5.2 that the bilinear model BL(p,0,p,1) given by (5.2.2) can be written in the vector form (5.2.4) and this in turn can be rewritten in terms of a Markovian representation. The general construction of Markovian representation for the bilinear models has been considered by Tuan Pham Dinh (1983), and here we briefly review his contribution.

Let $\{X_t, t=\ldots -2, -1, 0, 1, 2, \ldots\}$ be a zero mean, strictly stationary process with finite second moments and let $\mathcal{H}_t$ be the space of all square integrable random variables measurable with respect to the σ-field $\mathcal{F}_t$ generated by $(X_t, X_{t-1} \ldots)$. Define the best (nonlinear) predictor of a random variable Y_s based on $(X_t, X_{t-1}, \ldots)$ as the element of $\mathcal{H}_t$ which is closest to Y_s in mean square sense. Let this predictor be $\hat{Y}(s/\mathcal{H}_t) = E(Y_s/\mathcal{F}_t)$. Let $\hat{Y}_i(s/\mathcal{H}_t)$, $(i=1,2, \ldots)$ be the components of the random vector $\hat{Y}(s/\mathcal{H}_t)$. Define the (nonlinear) predictor space $\mathcal{P}_t$ to be the closed subspace spanned by $\hat{X}(t+k/\mathcal{F}_t) = E(X_{t+k}/\mathcal{F}_t)$, $k=1,2, \ldots$.

We now make the following assumptions about $\mathcal{P}_t$.

Assumption 1. The predictor space $\mathcal{P}_t$ has finite dimension.

Let $Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t}$ form the basis of $\mathcal{P}_t$. Then we have $\hat{X}(t/\mathcal{H}_{t-1}) = \underline{H}'\underline{Z}_{t-1}$ and $\hat{\underline{Z}}(t/\mathcal{H}_{t-1}) = \underline{A}\,\underline{Z}_{t-1}$, where $\underline{Z}'_t = (Z_{1,t}, Z_{2,t}, \ldots, Z_{n,t})$,

$\underline{H}$ is a vector, $\underline{A}$ is a square matrix. Define the innovation $e_t = X_t - \hat{X}(t/\mathcal{H}_{t-1})$. Then the elements of the column vector $\underline{Z}(t) - \hat{\underline{Z}}(t/\mathcal{H}_{t-1})$ belong to the closed subspace of $\mathcal{F}_t$ of $\mathcal{H}_t$, the subspace $\mathcal{F}_t$ is spanned by $\hat{X}(t+k/\mathcal{H}_t) - \hat{X}(t+k/\mathcal{H}_{t-1})$, $k=1, 2, \ldots$. This space $\mathcal{F}_t$ can be defined as the innovation space, and it is orthogonal to the random variables defined in $\mathcal{H}_{t-1}$. We make the following assumption about $\mathcal{F}_t$.

Assumption 2 The elements of $\mathcal{F}_t$ can be expressed as

$$e_t Y_t + C e_t + d(e^2_t - \sigma^2) \qquad (7.1.1)$$

where Y_{t-1} spans subspace $\mathcal{F}^o_{t-1}$ of $\mathcal{F}_{t-1}$, C, d are constants.

Assumption 3 The space $\mathcal{F}^o_t$ is contained in $\mathcal{P}_t$. Tuan Pham Dinh (1983) has shown that the assumptions 1, 2 and 3 imply the existence of the representation

$$\underline{Z}_t = \underline{A}\,\underline{Z}_{t-1} + \underline{B}\,\underline{Z}_{t-1}\, e_t + Ce_t + D(e^2_t - \sigma^2) \qquad (7.1.2)$$

$$X_t = \underline{H}\,\underline{Z}_{t-1} + e_t.$$

The assumption 3 may not be satisfied by all bilinear models, and in view of this, this assumption can be replaced by assumption 3' which is slightly weaker than the assumtion 3.

Assumption 3' The space $\mathcal{F}^o_t$ is contained in the space $P_t(m)$ for some $m \geq 0$ where $P_t(m)$ denotes the space spanned by $\mathcal{P}_t$ and $X_t, X_{t-1}, \ldots X_{t-m+1}$.

If the sequence of random variables $\{e_t\}$ are independent, and e_t is independent of $\underline{Z}_{t-1}, \underline{Z}_{t-2}, \ldots$, the process $\underline{Z}_t$ defined by (7.1.2) is Markovian. Hence, Tuan Pham Dinh (1983) has defined the representation (7.1.2) as a bilinear Markovian representation.

Since quite a number of bilinear processes have non zero mean, an obvious way to define a non-zero mean Markovian representation is to define it as

$$\underline{Z}_t = \underline{A}\,\underline{Z}_{t-1} + \underline{B}\,\underline{Z}_{t-1}\,e_t + \underline{C}\,e_t + \underline{D}\,e^2{}_t + \underline{E},$$
$$X(t) = \underline{H}'\underline{Z}_{t-1} + e_t \qquad (7.1.3)$$

The next thing is to find out what types of bilinear models admit a Markovian representation of the form (7.1.2). Let the scalar valued time series X_t satisfy the representation

$$X_t = \sum_{j=1}^{p} a_j\,X_{t-j} + e_t + \sum_{j=1}^{q} b_j\,e_{t-j} + \sum_{k=1}^{Q}\sum_{j=0}^{P} b_{kj}\,X_{t-j-k}\,e_{t-k} \qquad (7.1.4)$$

where $\{e_t\}$ is a sequence of independent random variables with zero mean and variance σ^2. The model (7.1.4) is known as a subdiagonal bilinear model. Tuan Pham Dinh (1983) has shown that the process X_t satisfying (7.1.4) satisfies all the assumptions 1, 2, 3' and can be rewritten in the form

$$\underline{Z}_t = \underline{A}\,\underline{Z}_{t-1} + \underline{B}\,\underline{Z}_{t-1}\,e_t + \underline{C}\,e_t + \underline{D}\,e^2{}_t$$
$$X_t = \underline{H}\,\underline{Z}_{t-1} + e_t\,. \qquad (7.1.5)$$

where the matrices A, B, C, D are functions of the coefficients of the model (7.1.4). Let the dimension of the state vector $\underline{Z}_t$ be n. Pham Dinh Tuan has shown that n = max(p, q+P, Q+p), and he explicitly constructed the elements of the vector $\underline{Z}_t$ in terms of $(X_t, X_{t-1} \ldots)$. The existence of a stationary solution of (7.1.5) has been discussed by Tuan Pham Dinh (1983). (See Hannan, 1982.)

Covariance properties of the Bilinear Markovian Representation

Consider the representation

$$\underline{Z}_t = \underline{A}\,\underline{Z}_{t-1} + \underline{B}\,\underline{Z}_{t-1}\,e_t + \underline{n}_t\,, \qquad (7.1.6)$$

$$X_t = \underline{H}'\,\underline{Z}_{t-1} + e_t\,, \qquad (7.1.7)$$

where $n_t = \underline{C}e_t + \underline{D}(e^2{}_t - \sigma^2)$.

Let $\underline{Q} = E(\underline{Z}_{t-1} - E\,\underline{Z}_{t-1})(Z_{t-1} - E\,\underline{Z}_{t-1})'$. Then the autocovariance function of lag zero of the process X_t is given by

$$R_X(0) = \underline{H}'\,\underline{Q}\,\underline{H} + \sigma^2.$$

Now we obtain the covariance of lag $k \geq 1$, $R_X(k) = E(X_{t+k}X_t)$. We observe $R_X(k) = E[E(X_{t+k}\,X_t/\mathcal{H}_t)] = E[E(X_{t+k}/\mathcal{H}_t)X_t] = E(\hat{X}(t+k/\mathcal{H}_t)X_t) = E(\underline{H}'\,\hat{\underline{Z}}(t+k-1/\mathcal{H}_t)\,X_t)$ in view of the relation (7.1.7). From the representation of (7.1.6), we can show that

$$E(Z_{t+k}/H_t) = \hat{\underline{Z}}(t+k/H_t) = \underline{A}\,\underline{Z}(t+k-1/H_t) = \ldots = \underline{A}^k\,\hat{\underline{Z}}(t/H_t) = \underline{A}^k\,\underline{Z}_t\,.$$

Hence $R_X(k) = E(\underline{H}'\underline{A}^{k-1}\,\hat{\underline{Z}}(t/H_t)X_t) = E(\underline{H}'\underline{A}^{k-1}\underline{Z}_tX_t) = \underline{H}'\underline{A}^{k-1}\,E(\underline{Z}_tX_t)$.

But

$$E(\underline{Z}_tX_t) = \underline{A}\,\underline{Q}\,\underline{H} + \underline{C}\sigma^2 + \underline{D}\,E(e^3{}_t).$$

Hence

$$R_X(k) = \underline{H}'\,\underline{A}^k\underline{Q}\,\underline{H} + \underline{H}'\,\underline{A}^{k-1}\,\underline{C}\,\sigma^2 + \underline{H}'\underline{A}^{k-1}\underline{D}\,E(e^3{}_t) \quad k \geq 1 \qquad (7.1.8)$$

The result (7.1.8) shows that the autocovariance function of the bilinear Markovian representation has the same form as that of an autoregressive moving average process, and this conculsion is similar to the conclusion we arrived at already in section 5.4. Specifically Tuan Pham Dinh (1983) has shown that the autocovariance function of the process X_t satisfying subdiagonal bilinear model (7.1.4) has the same form as that of an ARMA process of order p, max (q,Q) with autoregressive coefficients $a_1, a_2, \ldots, a_p$ and moving average coefficients being functions of $\{a_j, b_j, b_{kj}\}$.

7.2 EXISTENCE OF THE BILINEAR MODEL BL(p,0,p,1)

In chapter 5, we have discussed extensively the covariance properties of the BL(p,0,p,1) model when put in a vector form (5.2.4). In this section we consider the existence of the process $\underset{\sim}{X}_t$ satisfying (5.2.4). Our approach is on the same lines as given in Bhaskara Rao, Subba Rao and Walker (1983). The existence of the BL(1,0,1,1) model has been proved by Tuan Dinh Pham and Lanh Tat Tran (1981) using Law of large numbers. Our proofs are based on the calculation of moments, and are much simpler and straightforward.

We make use of some properties of Kronecker products of matrices, and these properties are summarised in appendix A. Some probabilistic definitions we make use of in the proofs are as follows:

A statement depending upon ω in Ω is said to hold a.e. [P] within the purview of a probability space $(\Omega,\mathcal{B}, P)$ if $P\{\omega \in \Omega$; the statement fails to hold at $\omega\} = 0$.

Let $\{Y_n, n \geq 1\}$ be a sequence of random vectors each of the same order pxl defined on some probability space $(\Omega,\mathcal{B},P)$. We say that $\sum_{n\geq 1} Y_n$ converges absolutely almost surely [P] if

$$\sum_{n\geq 1} |(Y_n)_i| < \infty \quad \text{a.e.[P]}$$

for every i = 1,2, ..., p.

We say that $\sum_{n\geq 1} \underset{\sim}{Y}_n$ converges in the mean (quadratic mean) if there exists a random vector $\underset{pxl}{Y}$ such that

$$\lim_{m\to\infty} E\Big|\sum_{n=1}^{m} (\underset{\sim}{Y}_n)_i - (\underset{\sim}{Y})_i\Big| = 0 \qquad \Big(\lim_{m\to\infty} E\Big|\sum_{n=1}^{m} (\underset{\sim}{Y}_n)_i - (\underset{\sim}{Y})_i\Big|^2 = 0\Big)$$

for every i = 1,2, ... , p.

In this section, we give a sufficient condition for the existence of a strictly stationary process conforming to the bilinear model (5.2.4). The

following theorem is the main result of this section.

THEOREM Let $\{e_t, t \in Z\}$ be a sequence of independent identically distributed random variables defined on a probability space $(\Omega,\mathcal{B}, P)$ such that $Ee_t = 0$ and $Ee_t^2 = \sigma^2 < \infty$. Let $\underline{A}$ and $\underline{B}$ be two matrices of order pxp such that $\rho(\underline{A}\otimes\underline{A} + \sigma^2\,\underline{B}\otimes\underline{B}) = \lambda < 1$. Let C be any column vector with components $c_1, c_2, \ldots, c_p$. Then the series of random vectors

$$\sum_{r\geq 1} \prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\,\underline{C}e_{t-r}$$

converges absolutely almost surely [P] as well as in the mean for every fixed t in Z. Further, if

$$\underset{\sim}{X}_t = \underline{C}e_t + \sum_{r\geq 1} \prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\,\underline{C}e_{t-r} \quad , \quad t \in Z,$$

then $\{\underset{\sim}{X}_t, t \in Z\}$ is a strictly stationary process conforming to the bilinear model

$$\underset{\sim}{X}_t = \underline{C}e_t + \underline{A}\,\underset{\sim}{X}_{t-1} + \underline{B}\,\underset{\sim}{X}_{t-1}\,e_{t-1}$$

for every t in Z.

Conversely, if $\{\underset{\sim}{X}_t, t \in Z\}$ is a strictly stationary process satisfying

$$\underset{\sim}{X}_t = \underline{C}e_t + \underline{A}\,\underset{\sim}{X}_{t-1} + \underline{B}\,\underset{\sim}{X}_{t-1}\,e_{t-1} \qquad \text{a.e. [P]} \qquad (7.2.1)$$

for every t in Z for some sequence $\{e_t, t \in Z\}$ of independent identically distributed random variables with $Ee_t = 0$ and $Ee_t^2 = \sigma^2 < \infty$ and for some matrices $\underline{A}$, $\underline{B}$, $\underline{C}$ of orders pxp, pxp, pxl respectively with $\rho(\underline{A}\otimes\underline{A} + \sigma^2\,\underline{B}\otimes\underline{B}) = \lambda < 1$, then

$$\underset{\sim}{X}_t = \underline{C}e_t + \sum_{r\geq 1} \prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\,\underline{C}e_{t-r} \qquad \text{a.e. [P]}$$

for every t in Z.

Proof The proof is broken down into the following steps.

1. For almost sure convergence, we show that

$$\sum_{r\geq 1} E\Big|\prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\ \underline{C}e_{t-r})_i\Big| < \infty \qquad (7.2.2)$$

for every i = 1, 2, ... ,p. This then implies that the series

$$\sum_{r\geq 1} \prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\ \underline{C}e_{t-r}$$

is absolutely convergent almost surely [P] as well as in the mean. See Chung [1974, (xi), p.42].

2. We establish (7.2.2) for i = 1. The general case is clear. First, we note that for every t in Z, r≥1 and s = 1, 2, ... ,p,

$$\begin{aligned} & E|((\underline{A} + \underline{B}e_{t-r})\ \underline{C}e_{t-r})_s| \\ &= E\Big|\sum_{j=1}^{p} (A)_{sj}\, c_j\, e_{t-r} + \sum_{j=1}^{p} (B)_{sj}\, c_j\, e_{t-r}^2\Big| \\ &\leq K_0, \end{aligned}$$

where K_0 is a constant which depends only on $\underline{A}$, $\underline{B}$, $\underline{C}$ and σ^2.

3. If r≥2, we show that

$$E\Big|\Big(\prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\ \underline{C}e_{t-r}\Big)_1\Big| \leq K_1\, p\, \lambda^{\frac{r-1}{2}} \qquad (7.2.3)$$

for some constant $K_1 > 0$.

Observe that

$$\begin{aligned} & E\Big|\Big(\prod_{j=1}^{r} (\underline{A} + \underline{B}e_{t-j})\ \underline{C}e_{t-r}\Big)_1\Big| \\ &= E\Big|\Big(\Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)((\underline{A} + \underline{B}e_{t-r})\ \underline{C}e_{t-r})\Big)_1\Big| \end{aligned}$$

$$\begin{aligned} &= E\Big|\sum_{s=1}^{p} \Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)_{1s}((\underline{A} + \underline{B}e_{t-r})\ \underline{C}e_{t-r})_s\Big| \\ &\leq \sum_{s=1}^{p} \Big(E\Big|\Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)_{1s}\Big|\Big)\ \big(E|(\underline{A} + \underline{B}e_{t-r})\ \underline{C}e_{t-r})_s|\big) \end{aligned}$$

(In the above derivation, we have used the fact that $\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})$ and $(\underline{A} + \underline{B}e_{t-r})\ \underline{C}e_{t-r}$ are independently distributed.)

The last expression is not greater than

$$K_0 \sum_{s=1}^{p} \big(E\big(\big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\big)_{1s}\big)^2\big)^{\frac{1}{2}},$$

by 2, and the Cauchy-Schwartz inequality. Now, for any s = 1, 2, ... ,p,

$$\begin{aligned} & \Big(\Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)_{1s}\Big)^2 \\ &= \Big(\Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big) \otimes \Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)\Big)_{1s;1s} \\ &= \Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j}) \otimes (\underline{A} + \underline{B}e_{t-j})\Big)_{1s;1s}, \text{ by (A.3)} \end{aligned}$$

Consequently,

$$\begin{aligned} & E\Big(\Big(\prod_{j=1}^{r-1} (\underline{A} + \underline{B}e_{t-j})\Big)_{1s}\Big)^2 \\ &= \prod_{j=1}^{r-1} (E(\underline{A} + \underline{B}e_{t-j}) \otimes (\underline{A} + \underline{B}e_{t-j}))_{1s;1s} \\ &= ((E[(\underline{A} + \underline{B}e_t) \otimes (\underline{A} + \underline{B}e_t)])^{r-1})_{1s;1s} \\ &= ((E(\underline{A} \otimes \underline{A} + e_t\, \underline{A} \otimes \underline{B} + e_t\, \underline{B} \otimes \underline{A} + e_t^2\, \underline{B} \otimes \underline{B}))^{r-1})_{1s;1s} \\ &= ((\underline{A} \otimes \underline{A} + \sigma^2 \underline{B} \otimes \underline{B})^{r-1})_{1s;1s} \\ &\leq K\, \lambda^{r-1}, \text{ by (A.4).} \end{aligned}$$

Hence

$$E\,|(\prod_{j=1}^{r}(\underline{A}+\underline{B}e_{t-j})\,\underline{C}e_{t-r})_1| \le K_1\,P\,\lambda^{\frac{r-1}{2}}$$

for a suitable choice of K_1.

4. Since $\lambda < 1$, we have

$$\sum_{r\ge 1} E|(\prod_{j=1}^{r}(\underline{A}+\underline{B}e_{t-j})\,\underline{C}e_{t-r})_1| < \infty.$$

Thus (7.2.1) is established.

5. It is obvious that the vector-valued stochastic process $\{\underset{\sim}{X}_t,\ t \in Z\}$ defined by

$$\underset{\sim}{X}_t = \underline{C}e_t + \sum_{r\ge 1}\prod_{j=1}^{r}(\underline{A}+\underline{B}e_{t-j})\,\underline{C}e_{t-r}\ ,\ t \in Z$$

is strictly stationary. Further, we have

$$\underset{\sim}{X}_t = \underline{C}e_t + (\underline{A}+\underline{B}e_{t-1})(\underline{C}e_{t-1} + \sum_{r\ge 1}\prod_{j=1}^{r}(\underline{A}+\underline{B}e_{t-1-j})\,\underline{C}e_{t-1-r})$$

$$= \underline{C}e_t + \underline{A}\,\underset{\sim}{X}_{t-1} + \underline{B}\,\underset{\sim}{X}_{t-1}\,e_{t-1}\ ,\ t \in Z.$$

6. Conversely, if $\{\underset{\sim}{X}_t,\ t \in Z\}$ conforms to the bilinear model (7. 2.1) above, by repeated application of (7.2.1) we observe that for any $n\ge 2$

$$\underset{\sim}{X}_t = \underline{C}e_t + \sum_{j=1}^{n-1}\prod_{k=1}^{j}(\underline{A}+\underline{B}e_{t-k})\,\underline{C}e_{t-j} + \prod_{k=1}^{n}(\underline{A}+\underline{B}e_{t-k})\,\underset{\sim}{X}_{t-n} \quad \text{a.e. [P]} \qquad (7.2.4)$$

for every t in Z.

As in 3., we note that for any $1\le r,\ s\le p$

$$E\,|(\prod_{k=1}^{n}(\underline{A}+\underline{B}e_{t-k}))_{rs}|$$

$$\le (E((\prod_{k=1}^{n}(\underline{A}+\underline{B}e_{t-k}))_{rs})^2)^{\frac{1}{2}}$$

$$\le K'\,\lambda^{n/2}\ ,\ \text{for some positive constant } K'.$$

Since $\lambda < 1$,

$$\lim_{n\to\infty} E\prod_{k=1}^{n}(\underline{A}+\underline{B}e_{t-k}) = 0.$$

Since $\{\underset{\sim}{X}_t,\ t \in Z\}$ is a strictly stationary process, $\{\underset{\sim}{X}_{t-n},\ n\ge 1\}$ converges to $\underset{\sim}{X}_1$ in distribution. Consequently,

$$\prod_{k=1}^{n}(\underline{A}+\underline{B}e_{t-k})\,\underset{\sim}{X}_{t-n}\ ,\quad n\ge 2 \quad \text{converges to } 0$$

in distribution and hence in probability. (See Chung, 1974, Theorem 4.4.6, p.92.) We can find a subsequence of this sequence which converges to 0 a.e. [P]. Taking limits along this subsequence in (7.2.4) we obtain

$$\underset{\sim}{X}_t = \underline{C}e_t + \sum_{j\ge 1}\prod_{k=1}^{j}(\underline{A}+\underline{B}e_{t-k})\,\underline{C}e_{t-j} \quad \text{a.e. [P]}$$

for every t in Z. (The almost sure convergence of the above series follows from the first part of the theorem.)

This completes the proof.

Remarks

(1) If we are looking for a real valued process $\{X_t,\ t \in Z\}$ conforming to the bilinear model

$$X_t = e_t + a\,X_{t-1} + b\,X_{t-1}\,e_{t-1} \quad \text{a.e. [P]}$$

for every t in Z under the usual assumptions on the e_t's, a sufficient condition for its existence is given by $a^2 + b^2 \sigma^2 < 1$. This result follows at once by taking p = 1 in the preceding theorem. Also a separate proof is easily constructed using the same approach as that presented above, the details being much simpler, and, moreover, this is substantially simpler than the proof based on the strong law of large numbers given by Tuan Dinh Pham and Lanh Tat Tran (1981, Theorem 2.1, p.618) in the univariate case. Further their Theorem 2.1 is strengthened in the above theorem by the inclusion of mean convergence of the series.

(2) Let $\{e_t, t \in Z\}$ be a sequence of independent identically distributed real random variables with common mean zero and variance $\sigma^2 < \infty$. Suppose $a^2 + b^2 \sigma^2 = 1$ and $|a| < 1$. If e_1 is not two-valued, then there exists a strictly stationary real valued process $\{X_t, t \in Z\}$ such that

$$X_t = e_t + a X_{t-1} + b X_{t-1} e_{t-1} \qquad \text{a.e. [P]}$$

for every t in Z. To show the existence of the process, we first show that

$$\sum_{r\geq 1} E \left| \prod_{j=1}^{r} (a + be_{t-j}) e_{t-r} \right| < \infty$$

from which it follows that the series

$$\sum_{r\geq 1} \prod_{j=1}^{r} (a + be_{t-j}) e_{t-r}$$

converges absolutely almost surely [P] as well as in the mean. Note that for any r≥2

$$E \left| \prod_{j=1}^{r} (a + be_{t-j}) e_{t-r} \right|$$

$$= E \left| \prod_{j=1}^{r-1} (a + be_{t-j})(a + be_{t-r}) e_{t-r} \right|$$

$$= \left(\prod_{j=1}^{r-1} E|a + be_{t-j}| \right) E|a + be_{t-r}||e_{t-r}|$$

$$= K d^{r-1},$$

where $K = E|a + be_1| \, |e_1|$ and $d = E|a + be_1|$. Now, we claim that $d < 1$. Two cases arise. If e_1 is degenerate, then $e_1 = 0$ a.e. [P]. Consequently, $d = E|a + be_1| = E|a| = |a| < 1$. If e is not degenerate and not two-valued, then $|a + be_1|$ is not degenerate. So, $d = E|a + be_1| < (E(a + be_1)^2)^{\frac{1}{2}} = (a^2 + b^2 \sigma^2)^{\frac{1}{2}} = 1$. This settles the claim. Consequently,

$$\sum_{r\geq 1} E \left| \prod_{j=1}^{r} (a + be_{t-j}) e_{t-r} \right| < \infty.$$

If we set $X_t = e_t + \sum_{r\geq 1} \prod_{j=1}^{r} (a + be_{t-j}) e_{t-r}$, $t \in Z$, then $\{X_t \; t \in Z\}$ is the desired process.

(3) If $Ee_1^4 < \infty$, then we can show that the series

$$\sum_{r\geq 1} \prod_{j=1}^{r} (A + Be_{t-j}) Ce_{t-r}$$

converges in the quadratic mean. The proof given above for the theorem can easily be adapted to establish this.

(4) Guegan (1981) has studied the real valued stochastic process $\{X_t, t \in Z\}$ satisfying

$$X_t = bX_{t-2} e_{t-1} + e_t \qquad \text{a.e. [P]}$$

for every t in Z. It can easily be seen that this model can be written in the form

$$\underset{\sim}{X}_t = \underline{C} e_t + \underline{A} \underset{\sim}{X}_{t-1} + \underline{B} \underset{\sim}{X}_{t-1} e_{t-1} \qquad \text{a.e. [P]}$$

for every t in Z where

$$A = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & b & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \underset{\sim}{C}^T = (1,0,0).$$

It can easily be checked that $\rho(\underset{\sim}{A} \otimes \underset{\sim}{A} + \sigma^2 \underset{\sim}{B} \otimes \underset{\sim}{B}) = \sigma^2 b^2$.

5. If the stochastic process to be modelled for a given time series data started only a finite number of steps ago, the same condition stipulated in the above theorem guarantees that the process involved is asymptotically stationary. To be more specific, suppose the p-variate process starts at time t = 0 with the initial random vector being X_0 and satisfies

$$\underset{\sim}{X}_t = \underset{\sim}{C} e_t + \underset{\sim}{A}\, \underset{\sim}{X}_{t-1} + \underset{\sim}{B}\, \underset{\sim}{X}_{t-1}\, e_{t-1} \quad \text{a.e. [P]} \qquad (7.2.5)$$

for t = 1, 2, 3, ... for some sequence $\{e_0, e_1, e_2, \ldots\}$ of independent identically distributed random variables with common mean $Ee_0 = 0$ and common variance $Ee_0^2 = \sigma^2 < \infty$ and for some constant matrices $\underset{\sim}{A}$, $\underset{\sim}{B}$ and $\underset{\sim}{C}$ of orders pxp, pxp and pxl respectively. Repeated use of (7.2.5) gives

$$\underset{\sim}{X}_t = \underset{\sim}{C} e_t + \sum_{r=1}^{t-1} \prod_{j=1}^{r} (\underset{\sim}{A} + \underset{\sim}{B} e_{t-j}) \underset{\sim}{C} e_{t-r} + \prod_{j=1}^{t} (\underset{\sim}{A} + \underset{\sim}{B} e_{t-j}) \underset{\sim}{X}_0$$

for every t = 2, 3, 4, ...

The process $\{\underset{\sim}{Y}_t, t = 2, 3, 4, \ldots\}$ defined by

$$\underset{\sim}{Y}_t = \underset{\sim}{C} e_0 + \sum_{r=1}^{t-1} \prod_{j=1}^{r} (\underset{\sim}{A} + \underset{\sim}{B} e_t) \underset{\sim}{C} e_r + \prod_{j=1}^{t} (\underset{\sim}{A} + \underset{\sim}{B} e_j) \underset{\sim}{X}_0$$

has the property that $\underset{\sim}{X}_t$ and $\underset{\sim}{Y}_t$ have the same distribution for every t = 2, 3, 4, ... This follows from the fact that $\{e_0, e_1, e_2, \ldots\}$ are independently identically distributed. Under the condition $\rho(\underset{\sim}{A} \otimes \underset{\sim}{A} + \sigma^2 \underset{\sim}{B} \otimes \underset{\sim}{B}) < 1$,

$$\sum_{r \geq 1} \prod_{j=1}^{r} (\underset{\sim}{A} + \underset{\sim}{B} e_j)\, \underset{\sim}{C} e_r$$

converges absolutely a.e. [P] and

$$\{\prod_{j=1}^{t} (\underset{\sim}{A} + \underset{\sim}{B} e_j) \underset{\sim}{X}_0 \,, \; t \geq 2\}$$

converges to 0 in probability. Consequently, the process $\{\underset{\sim}{Y}_t, t = 2, 3, 4, \ldots\}$, and hence the process $\{\underset{\sim}{X}_t, t = 2, 3, 4, \ldots\}$ converges to the random vector

$$\underset{\sim}{C} e_0 + \sum_{r \geq 1} \prod_{j=1}^{r} (\underset{\sim}{A} + \underset{\sim}{B} e_j)\, \underset{\sim}{C} e_r$$

in distribution. (See Chung, 1974, Theorem 4.4.6, p.92.) One cannot fail to notice that the distribution of the limiting random vector above is the same as that of X_t whose representation is given in the above theorem.

For further extension of the above results to the other models, see Bhaskara Rao, Subba Rao and Walker (1983), Akamanam (1983).

APPENDIX A

ON THE KRONECKER MATRIX PRODUCT

For any two matrices $D = (d_{ij})$ or $P=(p_{ij})$ of orders mxn and rxs respectively, we denote the Kronecker product of D and P by $D \otimes P$, defined to be the following matrix of order mr x ns:

$$D \otimes P = \begin{bmatrix} d_{11}P & d_{12}P & \dots & d_{1n}P \\ d_{21}P & d_{22}P & \dots & d_{2n}P \\ & \dots\dots\dots\dots & & \\ d_{m1}P & d_{m2}P & & d_{mn}P \end{bmatrix}$$

$(D \otimes P)_{ij;uv}$ denotes the element $d_{ij}P_{uv}$ which is the $((i-1)r+u, (j-1)s+v)$ -th element of the matrix $D \otimes P$ for i = 1,2, ... , m; j = 1,2, ...,n; u = 1,2, ... ,r and v = 1,2, ...,s. For any matrix F, we denote the (i,j) -th element of F by $(F)_{ij}$ or $(F)_{i,j}$, if the elements of F are not indicated specifically otherwise. If C is a column vector, the i-th component of C is denoted by $(C)_i$ or $(C)_{i1}$ when the elements of C are not explicitly indicated otherwise. In terms of this notation, we have

$$(D \otimes P)_{ij;uv} = d_{ij}P_{uv} = (D \otimes P)_{(i-1)r+u,(j-1)s+v} .$$

The following are some basic properties of Kronecker products. The reader may refer to Neudecker (1969) and Rogers (1980) for details.

(A.1) For any three matrices, A, B and C

$$(A \otimes B) \otimes C = A \otimes (B \otimes C).$$

(A.2) For any four matrices A, B, C and D, where A and B are of the same order and C and D are of the same order,

$$(A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$$

(A.3) $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$ provided the matrices involved are conformable for multiplication.

(A.4) Let $\lambda_1,\lambda_2,\dots,\lambda_n$ be the eigenvalues of the square matrix $\underline{A}$. Let $\rho(A) = \max_{1\le i\le n} |\lambda_i|$. Then there exists a positive constant K such that for any postive integer m,

$$|(A^m)_{ij}| \le K(\rho(A))^m \quad \text{for all i and j.}$$

(A.5) $(\underline{A} \otimes \underline{B})^{-1} = \underline{A}^{-1} \otimes \underline{B}^{-1}$, $(\underline{A} \otimes \underline{B})^{1} = \underline{A}^{1} \otimes \underline{B}^{1}$,

If $\underline{A}$ and $\underline{B}$ are square matrices of order n, s respectively, then

$|\underline{A} \otimes \underline{B}| = |\underline{A}|^s \ |\underline{B}|^n$,

$\mathrm{Tr}(\underline{A} \otimes \underline{B}) = \mathrm{Tr}(\underline{A})\ \mathrm{Tr}(\underline{B})$.

The eigenvalues of the Kronecker product $\underline{A} \otimes \underline{B}$ are ns numbers $\lambda_i(\underline{A})\ \lambda_j(\underline{B})$, (i=1,2,...,n; j=1,2,...,s), where $\lambda_i(\underline{A})$, (i=1,2,...,n) are the eigenvalues of A, and similarly $\lambda_j(B)$.

Let

$$A = [a_{ij}]_{mxn} = [A_1 \quad A_2 \quad \dots \quad A_n]$$

where A_j (j = 1, 2, ..., n) denotes the j-th colum of A, then the (m n) column vector vec A is defined as

$$\mathrm{vec}\ A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix}$$

(In particular, if $\underline{y}$ is a column vector then vec $\underline{y} = \underline{y}$). We state the following important results (for proofs, see Neudecker, 1969).

(a) $\mathrm{vec}\ (A\ B\ C) = (C' \otimes A)\ \mathrm{vec}\ (B)$

(b) $\mathrm{vec}\ (A\ B') = \Sigma\ B_i' \otimes A_i$

(c) $\mathrm{tr}\ (A\ B\ C) = [\mathrm{vec}\ (A)]'\ (I \otimes B)\ \mathrm{vec}\ C$

(d) $\mathrm{tr}\ (A\ B) = [\mathrm{vec}\ (A')]'\ \mathrm{vec}\ B$

APPENDIX B

LINEAR LEAST SQUARES SOLUTIONS BY HOUSEHOLDER TRANSFORMATIONS

Let A be a given n x m real matrix with $m \geq n$ and of rank n, $\underset{\sim}{y}_{nx1}$ a given vector and $\underset{\sim}{\theta}_{nx1}$ is a vector of unknown parameters such that

$$\underset{\sim}{y} = A \underset{\sim}{\theta} + \underset{\sim}{e} \qquad (B.1)$$

We wish to determine the least square estimates $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$, i.e. to find $\hat{\underset{\sim}{\theta}}$ such that

$$S(\underset{\sim}{\theta}) = \underset{\sim}{e}^T \underset{\sim}{e} = ||\underset{\sim}{e}||^2 = ||\underset{\sim}{y} - A \underset{\sim}{\theta}||^2$$

is minimum. ($||\underset{\sim}{x}||^2$ denotes the norm; $||\underset{\sim}{x}||^2 = \underset{\sim}{x}^T \underset{\sim}{x}$). The Householder transformation enables us to reduce A to an upper (or lower) triangular (or tridiagonal) form. Let $\underset{\sim}{u}$ be a real non-zero column vector and let

$$b = \tfrac{1}{2} \underset{\sim}{u}^T \underset{\sim}{u} = \tfrac{1}{2} ||\underset{\sim}{u}||^2 .$$

The Householder matrix is defined as

$$P = I - \frac{\underset{\sim}{u}\,\underset{\sim}{u}^T}{b} \quad ; \text{ I is the identity matrix.}$$

For practical computation the above notation is convenient, but for purposes of theory it is simpler to define

$$\underset{\sim}{v} = \frac{\underset{\sim}{u}}{||\underset{\sim}{u}||^2} ,$$

so that $\underset{\sim}{v}$ is a unit vector and the Householder matrix is

$$P = I - 2\underset{\sim}{v}\,\underset{\sim}{v}^T \qquad (B.2)$$

It follows immediately that

$$P^T = I^T - 2(\underset{\sim}{v}\,\underset{\sim}{v}^T)^T = I - 2\underset{\sim}{v}\,\underset{\sim}{v}^T = P$$

so that P is symmetric. Furthermore,

$$P^2 = (I - 2\underset{\sim}{v}\,\underset{\sim}{v}^T)(I - 2\underset{\sim}{v}\,\underset{\sim}{v}^T)$$
$$= I - 4\,\underset{\sim}{v}\,\underset{\sim}{v}^T + 4\underset{\sim}{v}\,\underset{\sim}{v}^T . \underset{\sim}{v}\,\underset{\sim}{v}^T .$$

But $\underset{\sim}{v}^T \underset{\sim}{v} = 1$, so that the final terms cancel and

$$P^2 = I \text{ or } P^{-1} = P = P^T.$$

Therefore the Householder matrix P is <u>symmetric</u> and <u>orthogonal</u>. We now examine the effect of multiplying a matrix A by a Householder matrix P. Consider

$$P\,A = (I - 2\underset{\sim}{v}\,\underset{\sim}{v}^T)\,A = A - 2\underset{\sim}{v}\,\underset{\sim}{v}^T A \quad .$$

If the components of $\underset{\sim}{v}$ are $v_1, v_2, \ldots, v_n$ and the raw data of A are denoted by $R_1, R_2, \ldots, R_n$, then $\underset{\sim}{v}^T$ A is the row vector

$$v_1R_1 + v_2R_2 + \ldots + v_nR_n .$$

Then $2\underset{\sim}{v}\,\underset{\sim}{v}^T A$ is the matrix whose K-th row is just

$$2v_K (v_1R_1 + v_2R_2 + \ldots + v_nR_n),$$

and $PA = A - 2\underset{\sim}{v}\,\underset{\sim}{v}^T A$ has the K-th row equal to

$$R_K - 2v_K (v_1R_1 + v_2R_2 + \ldots + v_nR_n).$$

In other words, PA is obtained from A by subtracting from each row of A a linear combination of rows of A. In particular, if $v_K = 0$, then the K-th row of PA is the same as the K-th row of A.

Let $\underline{A} = \underline{A}^{(1)}$, and let $\underline{A}^{(2)}, \underline{A}^{(3)}, \ldots, \underline{A}^{(m+1)}$ be defined as follows:-

$$A^{(K+1)} = P^{(K)}\, A^{(K)}, \quad K = 1, 2, \ldots, m.$$

$P^{(K)}$ is a Householder matrix of the form (B.2) where the elements of $P^{(K)}$ are derived so that

$$a_{i,K}^{(K+1)} = 0, \quad \text{for } i = K+1, \ldots, n$$

where $A^{(r)} = [a_{i,j}^{(r)}]$. It can be shown (see Wilkinson, 1965) that we can find a sequence of Householder matrices $P^{(1)}, P^{(2)}, \ldots, P^{(m)}$ whose product Q, which is an orthogonal matrix, is such that

$$QA = R = \begin{bmatrix} \hat{R} \\ \hline 0 \end{bmatrix} \begin{matrix} \} \, m \\ \} \, n-m \end{matrix}$$

where R is an upper triangular matrix. Then

$$S(\underset{\sim}{\theta}) = ||\underset{\sim}{y}-A\underset{\sim}{\theta}||^2 = ||Q(A\underset{\sim}{\theta}-\underset{\sim}{y})||^2 = ||QA\underset{\sim}{\theta}-Q\underset{\sim}{y}||^2 = ||R\underset{\sim}{\theta}-\underset{\sim}{Z}||^2$$

where

$$\underset{\sim}{Z} = Q\,\underset{\sim}{y} = \begin{bmatrix} \hat{Z} \\ \hline Z \end{bmatrix} \begin{matrix} \} \, m \\ \} \, n-m \end{matrix}$$

since the orthogonal matrix Q does not change the norm. Thus we can equivalently minimize the quantity $||R\underset{\sim}{\theta}-\underset{\sim}{Z}||^2$. Since

$$S(\underset{\sim}{\theta}) = ||R\underset{\sim}{\theta}-\underset{\sim}{Z}||^2 = ||\hat{R}\underset{\sim}{\theta}-\hat{\underset{\sim}{Z}}||^2 + ||\tilde{\underset{\sim}{Z}}||^2$$

and $\hat{\underset{\sim}{\theta}}$ is now independent of $||\tilde{\underset{\sim}{Z}}||^2$, then to minimize $S(\underset{\sim}{\theta})$ we choose $\underset{\sim}{\theta}$ to minimize $||\hat{R}\underset{\sim}{\theta}-\underset{\sim}{Z}||^2$. Since R is of rank m (by assumption), we can make the latter norm equal to zero by simply solving the triangular system

$$\hat{R}\,\underset{\sim}{\theta} = \hat{Z}$$

exactly. The residual sum of squares is given by

$$S(\hat{\underset{\sim}{\theta}}) = ||\tilde{\underset{\sim}{Z}}||^2.$$

APPENDIX C

FITTING THE BEST AR MODEL

The problem of fitting a full AR model of order P,

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} + \ldots + a_p X_{t-p} = e_t \qquad (C.1)$$

has been considered in many time series books, e.g. Box and Jenkins (1970) and Priestley (1981). Let $\{X_1, X_2, \ldots, X_N\}$ be a realisation from $\{X_t\}$ with autocovariance and autocorrelation functions R(s) and $\rho(s)$ respectively. The sample estimates of R(s) and $\rho(s)$ are given by

$$\hat{R}(s) = \frac{1}{N} \sum_{t=1}^{N-s} (X_t - \bar{X})(X_{t+s} - \bar{X}), \qquad S > 0$$

and

$$\hat{\rho}(s) = \hat{R}(s)/\hat{R}(0),$$

where

$$\bar{X} = \frac{1}{N} \sum_{t=1}^{N} X_t.$$

Under Gaussian assumption of $\{e_t\}$, the maximum likelihood estimates of the coefficients $a_1, \ldots, a_p$, in the AR(p) model (C.1), may be estimated by solving the Yule-Walker equations (1.5.3) which in matrix form are given by

$$\hat{\underset{\sim}{R}}_p = - \hat{\underset{\sim}{a}}\, \hat{\Gamma}_p$$

where

$$\hat{R}_p = \begin{bmatrix} \hat{R}(1) \\ \hat{R}(2) \\ \vdots \\ \hat{R}(p) \end{bmatrix}, \quad \hat{\Gamma}_p = \begin{bmatrix} \hat{R}(0) & \hat{R}(1) & \ldots & \hat{R}(P-1) \\ \hat{R}(1) & \hat{R}(0) & \ldots & \hat{R}(P-2) \\ \ldots & \ldots & \ldots & \ldots \\ \hat{R}(P-1) & \hat{R}(P-2) & \ldots & \hat{R}(0) \end{bmatrix}, \quad \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix}$$

and the solution is, since $\hat{\Gamma}_p$ is symmetric positive definite (non-singular),

$$\hat{\underset{\sim}{a}} = -\hat{\underset{\sim}{r}}_p^{-1}\hat{\underset{\sim}{R}}_p \qquad (C.2)$$

The maximum likelihood estimate of the residual variance σ_e^2 is given by

$$\hat{\sigma}_e^2 = (1-\hat{\underset{\sim}{R}}_p' \underset{\sim}{r}_p^{-1}\hat{\underset{\sim}{R}}_p)\ \hat{R}(0),$$

and the corresponding AIC value is

$$AIC(p) = (N-\gamma)\ \text{ln}\ \hat{\sigma}_e^2 + 2(P+1)$$

where γ is the maximum order (fixed at the beginning). The above procedure is repeated with $p = 1, 2, \ldots, \gamma$ and in each case the corresponding values of σ_e^2 and AIC are evaluated. The best AR model is that model which attained the minimum AIC value.

Note: The least squares estimates $\hat{\underset{\sim}{a}}$, given by (C.2), may be found also using the Householder transformation, as explained in Appendix B. In this case, equation (C.1) can be written as (B.1), with

$$\underset{\sim}{y} = \begin{bmatrix} x_{\gamma+1} \\ x_{\gamma+2} \\ \vdots \\ x_N \end{bmatrix}, \quad A = -\begin{bmatrix} x_\gamma & x_{\gamma-1} & \cdots & x_{\gamma-P+1} \\ x_{\gamma+1} & x_\gamma & \cdots & x_{\gamma-P} \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-P} \end{bmatrix},$$

$$\underset{\sim}{\theta} = \underset{\sim}{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix}, \quad \underset{\sim}{e} = \begin{bmatrix} e_{\gamma+1} \\ e_{\gamma+2} \\ \vdots \\ e_N \end{bmatrix}$$

where $x_t = X_t - \bar{X}$, $t = 1, 2, \ldots, N$.

APPENDIX D

TIME SERIES DATA

SUN-SPOT NUMBERS

Year	X_t	Year	X_t	Year	X_t	Year	X_t	Year	X_t
1700	5.0	1726	78.0	1752	47.8	1778	154.4	1804	47.5
1701	11.0	1727	122.0	1753	30.7	1779	125.9	1805	42.2
1702	16.0	1728	103.0	1754	12.2	1780	84.8	1806	28.1
1703	23.0	1729	73.0	1755	9.6	1781	68.1	1807	10.1
1704	36.0	1730	47.0	1756	10.2	1782	38.5	1808	8.1
1705	58.0	1731	35.0	1757	32.4	1783	22.8	1809	2.5
1706	29.0	1732	11.0	1758	47.6	1784	10.2	1810	0.0
1707	20.0	1733	5.0	1759	54.0	1785	24.1	1811	1.4
1708	10.0	1734	16.0	1760	62.9	1786	82.9	1812	5.0
1709	8.0	1735	34.0	1761	85.9	1787	132.0	1813	12.2
1710	3.0	1736	70.0	1762	61.2	1788	130.9	1814	13.9
1711	0.0	1737	81.0	1763	45.1	1789	118.1	1815	35.4
1712	0.0	1738	111.0	1764	36.4	1790	89.9	1816	45.8
1713	2.0	1739	101.0	1765	20.9	1791	66.6	1817	41.1
1714	11.0	1740	73.0	1766	11.4	1792	60.0	1818	30.1
1715	27.0	1741	40.0	1767	37.8	1793	46.9	1819	23.9
1716	47.0	1742	20.0	1768	69.8	1794	41.0	1820	15.6
1717	63.0	1743	16.0	1769	106.1	1795	21.3	1821	6.6
1718	60.0	1744	5.0	1770	100.8	1796	16.0	1822	4.0
1719	39.0	1745	11.0	1771	81.6	1797	6.4	1823	1.8
1720	28.0	1746	22.0	1772	66.5	1798	4.1	1824	8.5
1721	26.0	1747	40.0	1773	34.8	1799	6.8	1825	16.6
1722	22.0	1748	60.0	1774	30.6	1800	14.5	1826	36.3
1723	11.0	1749	80.9	1775	7.0	1801	34.0	1827	49.6
1724	21.0	1750	83.4	1776	19.8	1802	45.0	1828	64.2
1725	40.0	1751	47.7	1777	92.5	1803	43.1	1829	67.0

SUN-SPOT NUMBERS (contd.)

Year	X_t	Year	X_t	Year	X_t	Year	X_t	Year	X_t
1830	70.9	1856	4.3	1882	59.7	1908	48.5	1934	8.7
1831	47.8	1857	22.7	1883	63.7	1909	43.9	1935	36.1
1832	27.5	1858	54.8	1884	63.5	1910	18.6	1936	79.7
1833	8.5	1859	93.8	1885	52.2	1911	5.7	1937	114.4
1834	13.2	1860	95.8	1886	25.4	1912	3.6	1938	109.6
1835	56.9	1861	77.2	1887	13.1	1913	1.4	1939	88.8
1836	121.5	1862	59.1	1888	6.8	1914	9.6	1940	67.8
1837	138.3	1863	44.0	1889	6.3	1915	47.4	1941	47.5
1838	103.2	1864	47.0	1890	7.1	1916	57.1	1942	30.6
1839	85.7	1865	30.5	1891	35.6	1917	103.9	1943	16.3
1840	64.6	1866	16.3	1892	73.0	1918	80.6	1944	9.6
1841	36.7	1867	7.3	1893	85.1	1919	63.6	1945	33.2
1842	24.2	1868	37.6	1894	78.0	1920	37.6	1946	92.6
1843	10.7	1869	74.0	1895	64.0	1921	26.1	1947	151.6
1844	15.0	1870	139.0	1896	41.8	1922	14.2	1948	136.3
1845	40.1	1871	111.2	1897	26.2	1923	5.8	1949	134.7
1846	61.5	1872	101.6	1898	26.7	1924	16.7	1950	83.9
1847	98.5	1873	66.2	1899	12.1	1925	44.3	1951	69.4
1848	124.7	1874	44.7	1900	9.5	1926	63.9	1952	31.5
1849	96.3	1875	17.0	1901	2.7	1927	69.0	1953	13.9
1850	66.6	1876	11.3	1902	5.0	1928	77.8	1954	4.4
1851	64.5	1877	12.4	1903	24.4	1929	64.9	1955	38.0
1852	54.1	1878	3.4	1904	42.0	1930	35.7		
1853	39.0	1879	6.0	1905	63.5	1931	21.2		
1854	20.6	1880	32.3	1906	53.8	1932	11.1		
1855	6.7	1881	54.3	1907	62.0	1933	5.7		

* 256 observations

CANADIAN LYNX DATA

YEAR	X_t	YEAR	X_t	YEAR	X_t	YEAR	X_t	YEAR	X_t
1821	269	1844	213	1867	4245	1890	49	1913	3800
1822	321	1845	546	1868	687	1891	59	1914	3091
1823	585	1846	1033	1869	255	1892	188	1915	2985
1824	871	1847	2129	1870	473	1893	377	1916	3790
1825	1475	1848	2536	1871	353	1894	1292	1917	674
1826	2821	1849	957	1872	784	1895	4031	1918	81
1827	3928	1850	361	1873	1594	1896	3495	1919	80
1828	5943	1851	377	1874	1676	1897	587	1920	108
1829	4950	1852	225	1875	2251	1898	105	1921	229
1830	2577	1853	360	1876	1426	1899	153	1922	399
1831	523	1854	731	1877	756	1900	387	1923	1132
1832	98	1855	1638	1878	299	1901	758	1924	2432
1833	184	1856	2725	1879	201	1902	1307	1925	3574
1834	279	1857	2871	1880	229	1903	3465	1926	2935
1835	409	1858	2119	1881	469	1904	6991	1927	1537
1836	2285	1859	684	1882	736	1905	6313	1928	529
1837	2685	1860	299	1883	2042	1906	3794	1929	485
1838	3409	1861	236	1884	2811	1907	1836	1930	662
1839	1824	1862	245	1885	4431	1908	345	1931	1000
1840	409	1863	552	1886	2511	1909	382	1932	1590
1841	151	1864	1623	1887	389	1910	808	1933	2657
1842	45	1865	3311	1888	73	1911	1388	1934	3396
1843	68	1866	6721	1889	39	1912	2713		

* 114 observations

UNEMPLOYMENT FIGURES IN W. GERMANY

YEAR	JAN	FEB	MAR	APR	MAY	JUN
1948	481971	476353	471803	469382	446943	451091
1949	962866	1068885	1168127	1232381	1256889	1383302
1950	2200486	2288368	2155962	2074220	1942134	1808534
1951	2113553	1948422	1850960	1736166	1673661	1611908
1952	2106836	2172973	1848101	1728250	1602178	1534867
1953	2081227	2060651	1631613	1479538	1400709	1312201
1954	2217243	2275347	1629817	1473474	1299533	1198475
1955	1975159	2000102	1578827	1047886	876414	790579
1956	1390335	1982469	1158631	754433	653274	586675
1957	1601562	1222789	804115	690039	587952	544987
1958	1533557	1418192	1201913	678161	556518	481151
1959	1445508	1203311	667087	466799	386004	314389
1960	684297	581440	297704	225051	184564	162558
1961	422744	321910	187840	153046	130387	115126
1962	286398	273789	205467	135430	109403	97466
1963	410047	416889	216323	143659	124037	112083
1964	337497	304690	227188	146634	126654	112166
1965	286334	291236	200978	126862	106541	95419
1966	268848	235816	141428	121288	107743	100697
1967	621156	673572	576047	501303	458461	400773
1968	672617	589707	459853	330851	264674	226552
1969	368585	374124	243212	155181	122967	110744

UNEMPLOYMENT FIGURES IN W.GERMANY(CONTD)

YEAR	JUL	AUG	SEP	OCT	NOV	DEC
1948	665016	784232	784126	739423	715128	759623
1949	1302857	1308091	1313691	1316572	1383832	1558469
1950	1739507	1635604	1566588	1508348	1595491	1976461
1951	1584067	1543866	1502799	1476741	1570796	1931002
1952	1431499	1372614	1309563	1276009	1496764	1955635
1953	1237767	1186715	1148914	1169558	1331378	1747757
1954	1109003	1042697	982184	977389	1117323	1464489
1955	692089	630088	610855	627627	728116	1185850
1956	529246	503216	501687	516321	744959	1202533
1957	467085	435506	436205	435751	557143	1320321
1958	422220	393066	387484	421322	497959	1022634
1959	258048	235253	223206	235428	273606	512409
1960	140701	132931	130861	142001	158424	302812
1961	107704	111304	107883	112935	127792	239975
1962	93939	91344	91383	101984	131356	232653
1963	106111	104243	104507	114335	133024	252329
1964	105399	102835	100266	111462	126844	202086
1965	89018	85677	84974	92231	118962	177908
1966	101476	105743	112726	145804	216382	371623
1967	377235	359473	341078	360846	395004	526218
1968	202689	187778	174467	180223	196056	266372
1969	108018	103753	100477	107770	118849	192174

UNEMPLOYMENT FIGURES IN W. GERMANY(CONTD)

YEAR	JAN	FEB	MAR	APR	MAY	JUN
1970	286266	264080	197784	120550	103407	94767
1971	286171	254753	206472	160356	142890	135157
1972	375564	368952	268461	231219	208289	190224
1973	356352	347053	286576	240734	211276	200950
1974	620494	620154	561762	517365	456965	450684
1975	1154295	1183501	1114048	1087078	1017716	1002135
1976	1350990	1346723	1190159	1093693	953538	921037
1977	1248918	1213741	1084229	1039228	946498	930974
1978	1213498	1224309	1098969	1000429	912997	877319
1979	1171353	1134060	957711	875452	775117	763173
1980	1036519	992520	875909	825374	766768	781396

UNEMPLOYMENT FIGURES IN W.GERMANY(CONTD)

YEAR	JUL	AUG	SEP	OCT	NOV	DEC
1970	98562	99460	97338	110849	129476	175058
1971	141975	145835	146740	170111	207990	269810
1972	196774	198266	194660	214880	235379	279237
1973	216616	221905	219271	266969	331839	485631
1974	490894	527051	556981	672312	799337	945916
1975	1035235	1031122	1005495	1061128	1114190	1223396
1976	944609	939528	898701	943685	984699	1089935
1977	927624	963468	911239	954376	1004325	1090708
1978	922230	923963	864274	901636	927043	1006724
1979	803653	798867	736809	761724	798973	866783
1980	853077	864519	822565	888100	967533	1118500

* 396 observations

CHAPTER 8

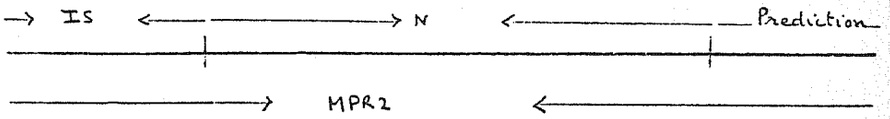
Listing of Programs

Program 1

(1) This program fits the full bilinear model (5.8.1) described in section 5.8. It is assumed that the data is on a file, and in the present case it is called ZZICAND, (see line 4). Change this to suit the required name.

(2) The total number of observations is denoted by MPR2. (For the Canadian Lynx data, MPR2 = 114.; line 32.) Decide on the Maximum order for AR part of the model = IS. In the Canadian Lynx data it is chosen as 10. The model is fitted over N observations.

The one step ahead predictors are calculated over the next MPR2 - (N+IS) observation



(3) line 42 of the program. Format of the input.

(4) lines 57 and 59. These DO loops correspond to the coefficients b_{ij} (i=1,2, ... ,5: j=1,2, ... ,10) of the model (5.8.1). Change to the orders wanted.

(5) the output will be (i) the best AR model, (ii) different bilinear models with estimated parameters (if these are convergent) together with mean sum of squares of the errors, AIC and minimum AIC.


```
 0 DISPOSE(OUTPUT,*PR,ST=SoACEN)
 1 FTN.
 2 ATTACH(MYLIB,LIBNAGFTNSCM,ID=LIBAPPL)
 3 LIBRARY(MYLIB)
 4 ATTACH(DATA5,Z21CAND,ST=SoA,FO=CONF)
 5 REWIND(DATA1)
 6 REWIND(DATA5)
 7 COPYP(DATA5,DATA1)
 8 REWIND(DATA1)
 9 LDSET(PRESET=NGINF,MAP=B/ZZZZMP)
10 LGO(PL=100000)
11 \\\\S
12       PROGRAM FITBIL(INPUT,OUTPUT,DATA1,TAPE1=INPUT,TAPE2=OUTPUT,
13      1TAPE3=DATA1)
14       COMMON/A2/A(602,30)
15       COMMON/A3/B(602),ALPHA(30),IPIV(30),ZZ(30)
16       COMMON/A4/X(602),E(602)
17       COMMON/A5/EE(20,602)
18       COMMON/A8/EAR(602)
19       COMMON/A9/Y(30),G(30)
20       COMMON/NR1/H(30,30),V(602),AJT(30,30)
21       COMMON/NR2/Z(30,30),DELT(30),WK1(30),WK2(30)
22       COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1
23      *,JPQ,JPQ1,IAVR,EROR,HGH,MPR1,MPR2
24       DIMENSION YC(30),VR(25)
25       N=90
26       IS=10
27       IS1=IS+1
28       NN=N+IS
29       IC=1
30       IAVR=1
31       MPR1=NN+1
32       MPR2=114
33       HGH=10.**50
34       VARMM=HGH
35       AICM=HGH
36       EROR=1./(10.**5)
37       SCAL=10.**2
38       REWIND3
39       READ(3,152)(X(I),I=1,NN)
40       DO 151 I=1,NN
41   151 X(I)=X(I)/SCAL
42   152 FORMAT(6X,12F5.0)
43     6 FORMAT(1H1,10X,2HN=,I4,5X,3HIP=,I2,5X,3HIQ=,I2,5X,3HJP=,
44      *I2,5X,3HJQ=,I2,5X,3HIC=,I2,5X,5HIAVR=,I3,5X,/,5X,60(2H<>))
45       DO 5500 IIP=2,10
46       IP=IIP
47       CALL FITARM(N,IP,VARAR,AICAR)
48       VARM=VARAR
49       VARMM=AMIN1(VARMM,VARAR)
50       AICM=AMIN1(AICM,AICAR)
51       IF(IAVR.EQ.0)GO TO 3
52       REWIND 3
53       READ(3,152)(X(I),I=1,MPR2)
54       DO 4 I=1,MPR2
55     4 X(I)=X(I)/SCAL
56     3 IQ=0
57       DO 5500 JJQ=1,5
58       JQ=JJQ
59       DO 5500 JJP=1,10
60       JP=JJP
```

```
      IP1=IP+1
      IPQ=IP+IQ
      IPQ1=IPQ+1
      JPQ=JP*JQ
      JPQ1=JPQ+1
      NP=IPQ+JPQ
      NP1=NP+IC
      IF(NP1.GT.29)GO TO 5500
      WRITE(2,6)N,IP,IQ,JP,JQ,IC
      IF(JQ.GT.1)GO TO 33
      IF(JP.GT.1)GO TO 78
      DO 31 K=1,NN
   31 E(K)=EAR(K)
      GO TO 78
   33 IF(JP.GT.1)GO TO 36
      DO 35 K=1,NN
   35 E(K)=EE(1,K)
      VARM=VR(1)
      GO TO 78
   36 VAR1=VR(JP-1)
      VAR2=VR(JP)
      IF(VAR1-VAR2)71,75
   71 VARM=VAR1
      DO 73 K=1,NN
   73 E(K)=EE(JP-1,K)
      GO TO 78
   75 VARM=VAR2
      DO 77 K=1,NN
   77 E(K)=EE(JP,K)
   78 CALL LSTSQ(N,NP1,VARLS,AICLS,IFAIL)
      VARMM=AMIN1(VARMM,VARLS)
      AICM=AMIN1(AICM,AICLS)
      IF(IFAIL.EQ.1)GO TO 5500
      CALL FTNTRN(N,NP1,VARNR,AICNR,JFAIL)
      IF(JFAIL.EQ.1)GO TO 5500
      VARMM=AMIN1(VARMM,VARNR)
      AICM=AMIN1(AICM,AICNR)
      VR(JP)=VARNR
      IF(VARNR.GT.VARM)GO TO 5500
      DO 81 K=1,NN
   81 EE(JP,K)=E(K)
      WRITE(2,5110)VARMM,AICM
 5110 FORMAT(1HO,20X,'MIN(VAR)=',E16.10,10X,'MIN(AIC)=',E16.10)
 5500 CONTINUE
      STOP
      END
C-----------------------------------------------------------------------
C
      SUBROUTINE LSTSQ(N,NP1,VAR,CAIC,IFAIL)
      COMMON/A2/A(602,30)
      COMMON/A3/B(602),ALPHA(30),IPIV(30),ZZ(30)
      COMMON/A4/X(602),E(602)
      COMMON/A9/Y(30),G(30)
      COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1,JPQ,JPQ1,IAVR
     *,EROR,HGH
      NP=NP1-IC
      WRITE(2,8)
    8 FORMAT(1HO,50X,27H(3)REPEATED RESIDUAL METHOD,//,
     *50X,14(2H--),//)
      NN=N+IS
```



```
      IT=0
      VARF=0.0
 5000 IT=IT+1
      DO 38 I=1,N
      B(I)=X(I+IS)
   38 CONTINUE
      DO 51 I=1,N
      II=I+IS
      IF(IP.EQ.0)GO TO 210
      DO 39 J=1,IP
      A(I,J)=-X(II-J)
   39 CONTINUE
  210 IF(IQ.EQ.0)GO TO 220
      DO 40 J=IP1,IPQ
      A(I,J)=E(II-J+IP+1)
   40 CONTINUE
  220 IF(JP.EQ.0.OR.JQ.EQ.0)GO TO 50
      J=IPQ
      DO 41 J1=1,JP
      DO 41 J2=1,JQ
      J=J+1
   41 A(I,J)=E(II-J2)*X(II-J1)
   50 IF(IC.EQ.1)A(I,NP1)=1.0
   51 CONTINUE
      IFAIL=1
      CALL F01AXF(N,NP1,A,602,ALPHA,IPIV,Y,ZZ,IFAIL)
      IF(IFAIL.EQ.0)GO TO 1000
      WRITE(2,1001)IFAIL
      GO TO 7700
 1000 CALL F04ANF(N,NP1,A,602,ALPHA,IPIV,B,Y,ZZ)
      CALL RESID(NN,NP1,Y,VAR,IFL)
      IF(IFL.EQ.1)GO TO 7700
      WRITE(2,4400)IT,VAR
      DIFR=ABS(VAR-VARF)
      IF(DIFR.LE.EROR)GO TO 5400
      IF(IT.GE.5)GO TO 5400
      VARF=VAR
      GO TO 5000
 5400 CAIC=FLOAT(N)*ALOG(VAR)+FLOAT(2*NP1)
      WRITE(2,28)IT
      IF(IP.EQ.0)GO TO 17
      WRITE(2,20)(Y(I),I=1,IP)
   17 IF(IQ.EQ.0)GO TO 18
      WRITE(2,21)(Y(I),I=IP1,IPQ)
   18 IF(JP.GT.0.AND.JQ.GT.0)WRITE(2,23)(Y(I),I=IPQ1,NP)
      IF(IC.EQ.0)GO TO 60
      WRITE(2,25)Y(NP1)
   60 WRITE(2,24)VAR,CAIC
   20 FORMAT(1HO,5X,'PARAMETERS <A> OF X(I) :',7(2X,F10.6))
   21 FORMAT(1HO,5X,'PARAMETERS <C> OF E(I) :',7(2X,F10.6))
   23 FORMAT(1HO,5X,'PARAMETERS <B> OF X(I)*E(I) :',8(2X,E10.4),
     */,10X,10(2X,E10.4))
   24 FORMAT(1HO,5X,'VARIANCE=',E16.10,15X,'AIC=',E16.10)
   25 FORMAT(1HO,5X,'CONSTANT=',E16.10)
   28 FORMAT(1HO,5X,'NO OF ITERATION=',I3)
 1001 FORMAT(1HO,5X,6HIFAIL=,I3)
 4400 FORMAT(1H ,5X,I3,5X,E16.10)
 7700 RETURN
      END
```

```
C-------------------------------------------------------------------
C---NEWTON RAPHSON ITERATION BY MARQUARDT(NAGLIB)-------------------
C-------------------------------------------------------------------
      SUBROUTINE FTNTRN(N,NP1,VARNR,AICNR,JFAIL)
      COMMON/A4/X(602),E(602)
      COMMON/NR2/Z(30,30),DELT(30),WK1(30),WK2(30)
      COMMON/A9/Y(30),G(30)
      COMMON/NR1/H(30,30),V(602),AJT(30,30)
      COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1,JPQ,JPQ1,IAVR,
     *EROR,HGH,MPR1,MPR2
      NP=NP1-IC
      NN=N+IS
      AN=FLOAT(N)
      WRITE(2,1002)
      VARF=0.0
      IT=0
 6543 IT=IT+1
      CALL LSQ(NN,NP1)
      JFAIL=1
      CALL F04ATF(H,30,G,NP1,DELT,Z,30,WK1,WK2,JFAIL)
      IF(JFAIL.EQ.0)GO TO 450
      WRITE(2,1001)JFAIL
      GO TO 7700
  450 DO 455 J=1,NP1
  455 Y(J)=Y(J)-DELT(J)
      CALL RESID(NN,NP1,Y,VARNR,JFL)
      IF(JFL.EQ.1)GO TO 7700
      WRITE(2,4400)IT,VARNR
 4400 FORMAT(1H ,2X,I3,2X,E16.10)
      DIFR=ABS(VARNR-VARF)
      VARF=VARNR
      IF(DIFR.LT.EROR)GO TO 470
      IF(IT.GE.20)GO TO 470
      GO TO 6543
  470 AICNR=AN*ALOG(VARNR)+2.*FLOAT(NP1)
      WRITE(2,28)IT
      IF(IP.GT.0)WRITE(2,20)(Y(I),I=1,IP)
      IF(IQ.GT.0)WRITE(2,21)(Y(I),I=IP1,IPQ)
      IF(JP.GT.0.AND.JQ.GT.0)WRITE(2,23)(Y(I),I=IPQ1,NP)
      IF(IC.EQ.1)WRITE(2,25)Y(NP1)
      WRITE(2,24)VARNR,AICNR
C     ----------PREDICTION-----------------
      SRE=0.0
      DO 490 I=MPR1,MPR2
      RE1=X(I)-Y(NP1)
      DO 475 J=1,IP
  475 RE1=RE1+Y(J)*X(I-J)
      J=IPQ
      DO 480 J1=1,JP
      DO 480 J2=1,JQ
      J=J+1
      RE1=RE1-Y(J)*E(I-J2)*X(I-J1)
  480 CONTINUE
      E(I)=RE1
      SRE=SRE+RE1**2
  490 CONTINUE
      SRE=SRE/FLOAT(MPR2-MPR1+1)
      WRITE(2,2003)SRE
   20 FORMAT(1HO,5X,25HPARAMETERS  'A' OF X(I) :,7(2X,F10.6))
   21 FORMAT(1HO,5X,25HPARAMETERS  'C' OF U(I) :,5(2X,F8.4))
   22 FORMAT(1HO,5X,25HPARAMETERS  'B' OF E(I) :,7(2X,E10.4))
   23 FORMAT(1HO,5X,'PARAMETERS "U" OF X(I)*E(I) :',8(2X,E10.4),//,
     *10X,10(2X,E10.4))
```



```
   24 FORMAT(1HO,10X,9HVARIANCE=,E16.10,/,
     *10X,9HAIC     =,E16.10)
   25 FORMAT(1HO,10X,10HCONSTANT =,E16.10)
   28 FORMAT(1HO,10X,16HNO OF ITERATION=,I3)
 1001 FORMAT(1HO,10X,6HIFAIL=,I3)
 1002 FORMAT(1H ,/,30X,14HNEWTON RAPHSON ,//)
 1500 FORMAT(1H1,30X,20HTHE HESSIEN MATRIX H=,/,30X,10(2H--),//)
 1501 FORMAT(1HO,5X,I2,5X,10(E10.4,2X))
 1502 FORMAT(1HO,//,60(2H--),//,30X,22HTHE COVARIANCE MATRIX  ,//)
 2001 FORMAT(1H1,30X,19HPREDICTION OF X(I) ,//,10X,1HI,5X,
     *4HX(I),6X,4HPRD1,9X,4HEROR,10X,4HPRD2,10X,4HEROR,//)
 2002 FORMAT(1HO,9X,I3,4X,F6.1,3X,F7.1,6X,F7.2,7X,F7.1,7X,F7.2)
 2003 FORMAT(1HO,//,10X,10HVAR(PRD1)=,E16.10,20X,10HVAR(PRD2)=,F12.3)
 7700 RETURN
      END

C  -------------------------------------------------------------------

      SUBROUTINE FITARM(N,NP,VAR,AMN)
      COMMON/A2/A(602,30)
      COMMON/A3/B(602),ALPHA(30),IPIV(30),ZZ(30)
      COMMON/A4/X(602),E(602)
      COMMON/A8/EAR(602)
      COMMON/A9/Y(30),G(30)
      COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1
     *,JPQ,JPQ1,IAVR,EROR,HGH
      WRITE(2,6)
    6 FORMAT(1HO,50X,28H(1)FITTING THE BEST AR MODEL,/,50X,14(2H--))
      NN=N+IS
      SX=0
      DO 50 I=1,NN
   50 SX=SX+X(I)
      AVR=SX/FLOAT(NN)
      DO 51 I=1,NN
   51 X(I)=X(I)-AVR
      DO 8 I=1,N
      B(I)=X(I+IS)
    8 CONTINUE
      DO 9 I=1,IS
      E(I)=0.0
      EAR(I)=0.0
    9 CONTINUE
      DO 20 I=1,N
      DO 20 J=1,NP
      A(I,J)=X(I+IS-J)
   20 CONTINUE
      IFAIL=1
      CALL F01AXF(N,NP,A,602,ALPHA,IPIV,Y,ZZ,IFAIL)
      IF(IFAIL.EQ.0)GO TO 1000
      WRITE(2,1001)IFAIL
      STOP
 1000 CALL F04ANF(N,NP,A,602,ALPHA,IPIV,B,Y,ZZ)
      RSS=0.0
      DO 35 I=IS1,NN
      E(I)=X(I)
      DO 34 K=1,NP
      E(I)=E(I)-Y(K)*X(I-K)
      EAR(I)=E(I)
   34 CONTINUE
```

```
      RSS=RSS+E(I)**2
   35 CONTINUE
      VAR=RSS/FLOAT(N)
       AMN=FLOAT(N)*ALOG(VAR)+2.*FLOAT(NP+1)
      WRITE(2,1003)NP,VAR,AMN
      WRITE(2,1002)(Y(I),I=1,NP)
 1002 FORMAT( 1H0,5X,'THE PARAMETERS ARE :-',///,10X,12(F8.5,2X))
 1003 FORMAT(1H0,10X,3HNP=,I3,5X,4HVAR=,E16.10,5X,4HAIC=,E16.10)
 1004 FORMAT(1H0,//,10X,4HVAR=,E16.10)
 1001 FORMAT(1H0,10X,6HIFAIL=,I3)
      RETURN
      END

C   ------------------------------------------------------------------

      SUBROUTINE RESID(NN,NP1,YC,VAR,IFL)
      COMMON/A4/X(602),E(602)
      COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1
     *,JPQ,JPQ1,IAVR,EROR,HGH
      DIMENSION YC(30)
      HGH=10.**50
      RSS=0.0
      NP=NP1-IC
      DO 210 I=1,IS
  210 E(I)=0.0
      DO 220 I=IS1,NN
      SUM=X(I)
      IF(IC.EQ.1)SUM=SUM-YC(NP1)
      IF(IP.EQ.0)GO TO 300
      DO 212 J=1,IP
  212 SUM=SUM+YC(J)*X(I-J)
  300 IF(IQ.EQ.0)GO TO 310
      DO 215 J=IP1,IPQ
  215 SUM=SUM-YC(J)*E(I-J+IP)
  310 IF(JP.EQ.0.OR.JQ.EQ.0)GO TO 218
      J=IPQ
      DO 315 J1=1,JP
      DO 315 J2=1,JQ
      J=J+1
      SUM=SUM-YC(J)*E(I-J2)*X(I-J1)
  315 CONTINUE
  218 E(I)=SUM
      RSS=RSS+E(I)**2
      IF(RSS.GT.HGH)GO TO 230
  220 CONTINUE
      VAR=RSS/FLOAT(NN-IS)
      IFL=0
      GO TO 240
  230 IFL=1
  240 RETURN
      END


      SUBROUTINE LSQ(NN,NP1)
      COMMON/A2/A(602,30)
      COMMON/A4/X(602),E(602)
      COMMON/NR1/H(30,30),V(602),AJT(30,30)
      COMMON/A9/Y(30),G(30)
      COMMON/GEN/IS,IS1,IC,JP,JQ,IPQ,IPQ1,IQ,IP,IP1
     *,JPQ,JPQ1,IAVR,EROR,HGH
      NP=NP1-IC
      DO 310 I=1,IS
      DO 310 J=1,NP1
      A(I,J)=0.0
  310 CONTINUE
```



```
C
C  -- CALCULATING THE FIRST DERIVATIVES D(J,L)
C   ----------------------------------------
      DO 250 J=IS1,NN
      DO 240 L=1,NP1
      IF(L.LE.IP)A(J,L)=X(J-L)
      IF(L.GT.IP.AND.L.LE.IPQ)A(J,L)=-E(J-L+IP)
      IF(L.GT.IPQ.AND.L.LE.NP)GO TO 240
      IF(IC.EQ.1.AND.L.EQ.NP1)A(J,NP1)=-1.0
  240 CONTINUE
      IF(JP.EQ.0.OR.JQ.EQ.0)GO TO 250
      L=IPQ
      DO 243 J1=1,JP
      DO 243 J2=1,JQ
      L=L+1
  243 A(J,L)=-E(J-J2)*X(J-J1)
  250 CONTINUE
      DO 460 L=1,NP1
      SUM=0
      DO 430 J=IS1,NN
      SMB=0
      IF(JP.EQ.0.OR.JQ.EQ.0)GO TO 412
      K=IPQ
      DO 410 J1=1,JP
      DO 410 J2=1,JQ
      K=K+1
      SMB=SMB+Y(K)*X(J-J1)*A(J-J2,L)
  410 CONTINUE
  412 IF(IQ.EQ.0)GO TO 420
      DO 415 J3=1,IQ
      SMB=SMB+Y(J3+IP)*A(J-J3,L)
  415 CONTINUE
  420 A(J,L)=A(J,L)-SMB
      SUM=SUM+A(J,L)*E(J)
  430 CONTINUE
      G(L)=SUM
      LL=L
      DO 450 K=1,LL
      SUM=0.0
      DO 440 J=IS1,NN
  440 SUM=SUM+A(J,K)*A(J,L)
      AJT(K,L)=SUM
  450 AJT(L,K)=SUM
  460 CONTINUE
      DO 400 J=1,IS
  400 V(J)=0.0
C
C  --  CALCULATING THE SECOND DERIVATIVES AND THE HASSIEN MATRIX H(I1,I2)
C   ------------------------------------------------------------------
      DO 700 I1=1,NP1
      DO 690 I2=1,I1
      H(I1,I2)=2.*AJT(I1,I2)
      H(I2,I1)=H(I1,I2)
  690 CONTINUE
  700 CONTINUE
      RETURN
      END
\\\\S
```

Program 2

Estimation of the Bispectral density function using optimum Window.

In this Program the bispectral estimate, $f(w_i, w_j)$, is obtained for sunspot numbers using optimum Window.

The data (input) (X(t), t=1, 2, ... ,N) should be in a file, called DATA1.

The parameters N,M, MW, NP should be given . For the sunspot numbers, the parameters chosen are

N = 256

M = 30, the number of third order covariances used

MW = 20, Window parameter (note M = MW if the Window is truncated). In general MW ≤ M.

NP = Number of frequencies (w_i, w_j).


```
C   <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>
C FILE Z210BISPSN : ESTIMATION OF BISPECTRAM USING OPTIMUM WINDOW
C   <><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><><>

      PROGRAM NBISPC2(INPUT,OUTPUT,DATA1,TAPE1=INPUT,TAPE2=OUTPUT,
     *TAPE3=DATA1)
      COMMON/A1/W(75),SP(75),RU(75),BP(75)
      COMMON/A2/C(75,75)
      COMMON/A3/V(75,75)
      LEVEL2,V,C
      COMMON/A4/FN(22,22),FMOD(22,22),FARG(22,22)
      COMMON/A6/X(1002)
      N=256
      M=30
      MW=20
      NP=20
      B=0.4
      REWIND 3
      READ(3,152)(X(I),I=1,N)
  152 FORMAT(20F5.1)
      WRITE(2,190)N,M,B
  190 FORMAT(1H1,10X,21HNR OF OBSERVATIONS=N=,I4,//,10X,
     124HNR OF COVARIANCE LAGS=M=,I2,//,10X,7HB=BETA=,F4.2,//,65(2H==))
      SUMX=0
      DO 90 I=1,N
   90 SUMX=SUMX+X(I)
      XMEAN=SUMX/FLOAT(N)
      DO 140 I=1,N
  140 X(I)=X(I)-XMEAN
      M1=2*M-1
      M2=M-1
      M4=M+1
      DO 15 J=1,M4
      SUMV=0.0
      NJ=N-J+1
      DO 12 I=1,NJ
      SUMV=SUMV+X(I)*X(I+J-1)
   12 CONTINUE
      RU(J)=SUMV/FLOAT(N)
   15 CONTINUE
      CALL WINDOW(M,MW)
      DO 10 J2=1,M4
      DO 10 J1=J2,M4
      SUM=0
      NL=N-J1+1
      DO 20 K=1,NL
      K1=J1-1+K
      K2=J2-1+K
   20 SUM=SUM+X(K1)*X(K2)*X(K)
      C(J1,J2)=(SUM/FLOAT(N))*V(J1,J2)
   10 CONTINUE
      W(1)=0.
      PI=4.*ATAN(1.)
      NP1=NP+1
      W(NP1)=1.
      NP3=2*NP+1
      DO 40 I=2,NP3
      W(I)=FLOAT(I-1)/FLOAT(NP)
   40 CONTINUE
```

```
      G=(1./(2.*PI))**2
      CALL SPEC(M,NP)
      DO 60 I1=1,NP1
      DO 60 I2=I1,NP1
      ARF=0.
      AIF=0.
      DO 50 J2=2,M
      JJ=J2+1
      DO 50 J1=JJ,M4
      A1=FLOAT(J1-1)*PI
      A2=FLOAT(J2-1)*PI
      A12=A1-A2
      R=C(J1,J2)
      ARF=ARF+R*(COS(A1*W(I1)+A2*W(I2))+COS(A1*W(I2)+A2*W(I1))+
     1COS(A1*W(I1)+A12*W(I2))+COS(A1*W(I2)+A12*W(I1))+COS(A12*W(I1)-
     1A2*W(I2))+COS(A12*W(I2)-A2*W(I1)))
      AIF=AIF-R*(SIN(A1*W(I1)+A2*W(I2))+SIN(A1*W(I2)+A2*W(I1))-SIN(A1*
     1W(I1)+A12*W(I2))-SIN(A1*W(I2)+A12*W(I1))+SIN(A12*W(I1)-A2*W(I2))+
     1SIN (A12*W(I2)-A2*W(I1)))
   50 CONTINUE
      DO 51 I=2,M4
      AI=FLOAT(I-1)*PI
      ARF=ARF+(C(I,1)+C(I,I))*(COS(AI*W(I1))+COS(AI*W(I2))+COS(AI*(W(I1)
     1+W(I2))))
      AIF=AIF+(C(I,1)-C(I,I))*(SIN(AI*(W(I1)+W(I2)))-SIN(AI*W(I1))-
     1SIN(AI*W(I2)))
   51 CONTINUE
      ARF=ARF+C(1,1)
      FMOD(I1,I2)=G*SQRT(ARF**2+AIF**2)
      FMOD(I2,I1)=FMOD(I1,I2)
      FN(I1,I2)=FMOD(I1,I2)/SQRT(SP(I1)*SP(I2)*SP(I1+I2-1))
      FN(I2,I1)=FN(I1,I2)
      FARG(I1,I2)=ATAN(AIF/ARF)
   60 CONTINUE
      WRITE(2,220)
      WRITE(2,221)
  220 FORMAT(1H1,////,35X,9(2H--),'ESTIMATED BISPECTRAL MODULUS',9(2H--))
  221 FORMAT(1H ,57X,'USING OPTIMUM WINDOW')
  230 FORMAT(1HO,5X,F4.2,2X,14F8.1)
  231 FORMAT(1HO,5X,F4.2,2X,14F7.3)
  211 FORMAT(1H ,7X,2HW2,/,10X,2HW1,F6.2,13F7.2)
      LP=2*NP1/3
      LP1=LP+1
      NPL=NP1-LP
      DO 300 L2=1,NPL
      I2=NP1-L2+1
      I21=2*L2-1
      WRITE(2,230)W(I2),(FMOD(I1,I2),I1=1,I21)
  300 CONTINUE
      DO 400 L2=1,LP
      I2=LP-L2+1
      WRITE(2,230)W(I2),(FMOD(I1,I2),I1=1,I2)
  400 CONTINUE
      WRITE(2,210)(W(J),J=1,LP)
  210 FORMAT(1H ,7X,2HW2,/,10X,2HW1,F6.2,13F8.2,/,15X,
     123HW1(AS A FRACTION OF PI))
      WRITE(2,240)
  240 FORMAT(1H1,35X,10(2H--),29HESTIMATED BISPECTRAL ARGUMENT,10(2H--))
      DO 310 L2=1,NPL
```



```
      I2=NP-L2
      I21=2*L2-1
      WRITE(2,231)W(I2),(FARG(I1,I2),I1=1,I21)
  310 CONTINUE
      DO 410 L2=1,LP
      I2=LP-L2+1
      WRITE(2,231)W(I2),(FARG(I1,I2),I1=1,I2)
  410 CONTINUE
      WRITE(2,211)(W(J),J=1,LP)
      WRITE(2,221)
      WRITE(2,220)
      DO 320 L2=1,NPL
      I2=NP1-L2+1
      I21=2*L2-1
      WRITE(2,230)W(I2),(FN(I1,I2),I1=1,I21)
  320 CONTINUE
      DO 415  L2=1,LP
      I2=LP-L2+1
      WRITE(2,230)W(I2),(FN(I1,I2),I1=1,I2)
  415 CONTINUE
      WRITE(2,210)(W(J),J=1,LP)
      CALL PAPER(1)
      CALL CTRSET(2)
      CALL CTRMAG(7)
      ZMA=0.0
      ZMN=0.0
      DO 420 I=1,NP1
      DO 420 J=1,I
      ZMA=AMAX1(ZMA,FMOD(I,J))
      ZMN=AMAX1(ZMN,FN(I,J))
  420 CONTINUE
      CALL PSPACE(0.2,0.65,0.5,0.9)
      CALL ISOPLT(FMOD,1,21,22,1,21,22,0.0,ZMA,0,1,0)
      CALL PSPACE(0.2,0.65,0.1,0.5)
      CALL ISOPLT(FN,1,21,22,1,21,22,0.0,ZMN,0,1,0)
      CALL FRAME
      CALL GREND
      STOP
      END
      SUBROUTINE WINDOW(M,MW)
      COMMON/A3/V(75,75)
      LEVEL2,V
      M2=2*M-1
      M4=M+1
      PI=4.*ATAN(1.)
      A=PI/FLOAT(MW)
      V(1,1)=1
      DO 10 I=2,M4
      AI=A*FLOAT(I-1)
      CI=COS(AI)
      SI=SIN(AI)
      V(I,1)=4.*(30.-(30.+9.*AI*AI)*CI+(-6.*AI+AI**3)*SI)/(21.*AI**4)
      V(I,I)=V(I,1)
   10 CONTINUE
      DO 11 J=2,M
      JJ=J+1
      DO 11 I=JJ,M4
```

```
      SI=FLOAT(I-1)*A
      SJ=FLOAT(J-1)*A
      SI=-SI
      SJ=-SJ
      SM=SJ-SI
      C1=COS(SI)
      C2=COS(SJ)
      C12=COS(SM)
      S1=SIN(SI)
      S2=SIN(SJ)
      S12=SIN(SM)
      D1=(2.*SI*SI+5.*SJ*SJ-5.*SI*SJ)*C1*SI**3
      D2=(5.*SI*SI+2.*SJ*SJ-5.*SI*SJ)*C2*SJ**3
      D12=(2.*SI*SI+2.*SJ*SJ+SI*SJ)*C12*SM**3
      E1=SI**3*S1
      E2=SJ**3*S2
      E12=SM**3*S12
      HM=SI*SJ*SM
      V(I,J)=8.*((D12+D1-D2)/HM**3-(E1+E2+E12)/HM**2)/7.
   11 CONTINUE
      RETURN
      END
      SUBROUTINE SPEC(M,NP)
      COMMON/A1/W(75),SP(75),RU(75),BP(75)
       PI=4.*ATAN(1.)
      NP1=NP+1
      NP2=2*NP+1
      M4=M+1
      BP(1)=1.0
      DO 10 I=2,M4
      AI=FLOAT(I-1)/FLOAT(M)
      CI=AI*PI
      BP(I)=SIN(CI)/CI
   10 CONTINUE
      DO 20 I=1,NP2
      SM=RU(1)
      DO 15 J=1,M
      AA=2.*RU(J+1)*COS(FLOAT(J)*W(I)*PI)
      SM=SM+BP(J+1)*AA
   15 CONTINUE
      SP(I)=SM/(2.*PI)
   20 CONTINUE
      WRITE(2,100)
      DO 30 I=1,NP1
   30 WRITE(2,200)W(I),SP(I)
  100 FORMAT(1H1,10X,'NON-NORMALIZED SPECTRUM',///,10X,'W',10X,
     *'DANIEL ESTIMATE',/)
  200 FORMAT(1H0,8X,F4.2,10X,E12.6)
      RETURN
      END
```


Program 3

This program estimates the bispectral density function and tests the hypothesis H_0: $f(w_i, w_j) = 0$ for all w_i and w_j, except on the boundary.

(1) Attach the data (card 3).

(2) Declare the following parameters (card 21).

N = Number of observations.

M = Number of third order covariances used in the estimation of the bispectral density function.

MW = Window Parameter. (MW ≤ M).
(For truncated Windows MW = M).

NP = K
K5 = K
} as described in section 4.5.

IW =
- 1 for Daniell Window
- 2 for Tukey Window
- 3 for Parzen Window

Card 152 Data Format.

```
0 FTN.
1 LIBRARY(PROCLIB)
2 NAG.
3 ATTACH(DATA5,HERDATA,ST=SoA,FO=CONF)
4 REWIND(DATA1)
5 REWIND(DATA5)
6 COPYP(DATA5,DATA1)
7 REWIND(DATA1)
8 LDSET(PRESET=NGINF,MAP=B/ZZZZMP)
9 LGO.
10 \\\\S
11       PROGRAM SBISPC(INPUT,OUTPUT,DATA1,TAPE1=INPUT,TAPE2=OUTPUT,
12      1TAPE3=DATA1)
13        COMMON/A1/W(10,12)
14        COMMON/A2/C(72,72)
15        COMMON/A3/V(72)
16        COMMON/A4/FR(10,10,22),FI(10,10,22)
17        COMMON/A5/X(1202)
18        COMMON/A6/WKSCE(20)
19        COMPLEX F(20,42),FAVR(20),FTC(42,20),Z(20,20),A(20,1),Q,T
20        LEVEL2,W,C,FR,FI
21        READ(1,5)N,M,NP,IW,K5
22      5 FORMAT(I4,4I2)
23        KS=K5
24        WRITE(2,190)N,M,NP,KS
25    190 FORMAT(1H1,10X,21HNR OF OBSERVATIONS=N=,I4,//,10X,
26       124HNR OF COVARIANCE LAGS=M=,I2,//,10X,3HNP=,I2,//,10X,3HKS=,I2,//,
27       165(2H--))
28        REWIND3
29        READ(3,*)(X(I),I=1,N)
30
31             DO 777 I=1,N
32
33
34    777     X(I)=SQRT(X(I)+1)
35
36     152 FORMAT(6X,12F5.0)
37         SUMX=0.
38         DO 90 I=1,N
39      90 SUMX=SUMX+X(I)
40         XMEAN=SUMX/FLOAT(N)
41         DO 140 I=1,N
42     140 X(I)=X(I)-XMEAN
43         M4=M+1
44         DO 10 J2=1,M4
45         DO 10 J1=J2,M4
46         SUM=0
47         NL=N-J1+1
48         DO 20 K=1,NL
49         K1=J1-1+K
50         K2=J2-1+K
51      20 SUM=SUM+X(K1)*X(K2)*X(K)
52         C(J1,J2)=SUM/FLOAT(N)
53      10 CONTINUE
54         CALL WINDOW(M,IW)
55         PI=4.*ATAN(1.)
56         KF=2*KS+1
57         NP1=NP+1
58         NP2=NP-1
59         SN=FLOAT(N)/FLOAT(NP)
60         PK=6.
```



```
61         DO 40 I=1,NP2
62         DO 40 KK=1,KF
63         W(I,KK)=(FLOAT(I)*SN+FLOAT(KK-KS-1)*PK)/FLOAT(N)
64      40 CONTINUE
65         G=(1./(2.*PI))**2
66         LP=2*NP/3
67         DO 60 I1=1,LP
68         V1=W(I1,KS+1)
69         I12=NP-I1/2-1
70         I11=I1+1
71         IF(I11.GT.I12)GO TO 61
72         DO 60 I2=I11,I12
73         U2=W(I2,KS+1)
74         DO 60 I3=1,KF
75         U1=W(I1,I3)
76         V2=W(I2,I3)
77         ARF2=C(1,1)
78         ARF1=C(1,1)
79         AIF1=0
80         AIF2=0
81         DO 50 J2=2,M
82         JJ=J2+1
83         DO 50 J1=JJ,M4
84         R=V(J1)*V(J2)*V(J1-J2)*C(J1,J2)
85         A1=FLOAT(J1-1)*PI
86         A2=FLOAT(J2-1)*PI
87         A12=A1-A2
88         ARF1=ARF1+R*(COS(A1*U1+A2*U2)+COS(A1*U2+A2*U1)+COS(A1*U1+A12*U2)
89        *+COS(A1*U2+A12*U1)+COS(A12*U1-A2*U2)+COS(A12*U2-A2*U1))
90         ARF2=ARF2+R*(COS(A1*V1+A2*V2)+COS(A1*V2+A2*V1)+COS(A1*V1+A12*V2)
91        *+COS(A1*V2+A12*V1)+COS(A12*V1-A2*V2)+COS(A12*V2-A2*V1))
92         AIF1=AIF1-R*(SIN(A1*U1+A2*U2)+SIN(A1*U2+A2*U1)-SIN(A1*U1+A12*U2)
93        *-SIN(A1*U2+A12*U1)+SIN(A12*U1-A2*U2)+SIN(A12*U2-A2*U1))
94         AIF2=AIF2-R*(SIN(A1*V1+A2*V2)+SIN(A1*V2+A2*V1)-SIN(A1*V1+A12*V2)
95        *-SIN(A1*V2+A12*V1)+SIN(A12*V1-A2*V2)+SIN(A12*V2-A2*V1))
96      50 CONTINUE
97         DO 51 I=2,M4
98         AI=FLOAT(I-1)*PI
99         R=(C(I,1)+C(I,I))*(V(I)**2)
100        H=(C(I,1)-C(I,I))*(V(I)**2)
101        ARF1=ARF1+R*(COS(AI*U1)+COS(AI*U2)+COS(AI*(U1+U2)))
102        ARF2=ARF2+R*(COS(AI*V1)+COS(AI*V2)+COS(AI*(V1+V2)))
103        AIF1=AIF1+H*(SIN(AI*(U1+U2))-SIN(AI*U1)-SIN(AI*U2))
104        AIF2=AIF2+H*(SIN(AI*(V1+V2))-SIN(AI*V1)-SIN(AI*V2))
105     51 CONTINUE
106        FR(I1,I2,I3)=G*ARF1
107        FI(I1,I2,I3)=G*AIF1
108        IF(I3-K5-1)57,60,52
109     57 II3=I3+KF
110        GO TO 53
111     52 II3=I3+KF-1
112     53 FR(I1,I2,II3)=G*ARF2
113        FI(I1,I2,II3)=G*AIF2
114     60 CONTINUE
115     61 WRITE(2,220)
116    220 FORMAT(1H1,35X,8(2H--),'REAL PART OF THE ESTIMATED BISPECTRUM',
117       18(2H--))
118    221 FORMAT(1HO,10X,21HUSING BARTLETT WINDOW)
119    222 FORMAT(1HO,10X,18HUSING TUKEY WINDOW)
120    223 FORMAT(1HO,10X,19HUSING PARZEN WINDOW)
121        IF(IW-0)501,502,503
122    501 WRITE(2,221)
```

```
      GO TO 504
  502 WRITE(2,222)
      GO TO 504
  503 WRITE(2,223)
      NPL=NP-LP-1
  504 DO 300 L2=1,NPL
      I2=NP-L2
      I21=2*L2-1
      WRITE(2,230)W(I2,K5+1),(FR(I1,I2,K5+1),I1=1,I21)
  300 CONTINUE
  230 FORMAT(1HO,2X,F4.2,2X,8(E10.4,2X))
      NL1=LP-1
      DO 400 L2=1,NL1
      I2=LP-L2+1
      I21=I2-1
      WRITE(2,230)W(I2,K5+1),(FR(I1,I2,K5+1),I1=1,I21)
  400 CONTINUE
      WRITE(2,210)(W(J,K5+1),J=1,LP)
  210 FORMAT(1HO,4X,2HW2,/,6X,2HW1,8(F7.3,5X)//,10X,
     123HW1(AS AFRACTION OF PI ))
      WRITE(2,240)
  240 FORMAT(1H1,35X,8(2H--),38HIMAGINARY PART OF ESTIMATED BISPECTRUM
     1,8(2H--))
      DO 310 L2=1,NPL
      I2=NP-L2
      I21=2*L2-1
      WRITE(2,230)W(I2,K5+1),(FI(I1,I2,K5+1),I1=1,I21)
  310 CONTINUE
      DO 410 L2=1,NL1
      I2=LP-L2+1
      I21=I2-1
      WRITE(2,230)W(I2,K5+1),(FI(I1,I2,K5+1),I1=1,I21)
  410 CONTINUE
      WRITE(2,210)(W(J,K5+1),J=1,LP)
      KFF=2*KF-1
      WRITE(2,509)
      DO 560 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 561
      DO 560 I2=I11,I12
      WRITE(2,510)I1,I2,(FR(I1,I2,I3),I3=1,KFF)
  560 CONTINUE
  509 FORMAT(1H1,30X,10(2H--),22HSAMPLE BISPECTRUM DATA,10(2H--),///,
     154X,9HREAL PART,//)
  510 FORMAT(1HO,1X,I2,1X,I2,10E11.3,//)
  511 FORMAT(1HO,//,65(2H--),///,55X,14HIMAGINARY PART,//)
  561 WRITE(2,511)
      DO 570 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 571
      DO 570 I2=I11,I12
      WRITE(2,510)I1,I2,(FI(I1,I2,I3),I3=1,KFF)
  570 CONTINUE
  571 L=0
      DO 660 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 662
      DO 660 I2=I11,I12
      L=L+1
```



```
      DO 660 K=1,KFF
      D1=FR(I1,I2,K)
      D2=FI(I1,I2,K)
      F(L,K)=CMPLX(D1,D2)
  660 CONTINUE
  662 N=7
      K=KFF
      DO 8 I=1,N
      FAVR(I)=(.0,.0)
      DO 7 J=1,K
      FAVR(I)=FAVR(I)+F(I,J)
    7 CONTINUE
      FAVR(I)=FAVR(I)/K
    8 CONTINUE
      DO 9 I=1,N
      DO 9 J=1,K
      F(I,J)=F(I,J)-FAVR(I)
      FTC(J,I)=CONJG(F(I,J))
    9 CONTINUE
      DO 19 I1=1,N
      DO 19 I2=1,N
      Z(I1,I2)=(.0,.0)
      DO 19 J=1,K
      Z(I1,I2)=Z(I1,I2)+F(I1,J)*FTC(J,I2)
   19 CONTINUE
      WRITE(2,21)
   21 FORMAT(1H1,65(2H==),///,28X,12(2H--),2X,20HTHE HERMETIAN MATRIX,
     12X,12(2H--))
      DO 880 I=1,N
      WRITE(2,22)I,(Z(I,J),J=1,N)
  880 CONTINUE
   22 FORMAT(1HO,2X,I2,1X,14E9.2)
      M=1
      IFALL=1
      IZ=20
      IC=20
      IFAVR=20
      CALL F04ADF(Z,IZ,FAVR,IFAVR,N,M,A,IC,WKSPCE,IFALL)
      IF(IFALL.EQ.0)GO TO 100
      WRITE(2,24)IFALL
      STOP
  100 WRITE(2,27)(A(I,1),FAVR(I),I=1,N)
   27 FORMAT(1HO,//,10X,1H(,E12.6,1H,,E12.6,1H),5X,1H(,E12.6,1H,,E12.6
     1,1H))
   24 FORMAT(1HO,65(2H==),///,6HIFALL=,I2)
      Q=(0.0,0.0)
      DO 18 I=1,N
      Q=Q+CONJG(FAVR(I))*A(I,1)
   18 CONTINUE
      T=K*Q
      WRITE(2,25)T
   25 FORMAT(1HO,30X,20(2H==),//35X,32HHOTELLINGS T SQUARE STATISTIC IS
     1,///,48X,3HT=(,F18.3,1H,,F8.3,1H),///,30X,20(2H==))
      STOP
      END
      SUBROUTINE WINDOW(M,IW)
      COMMON/A3/V(72)
      IF(IW-2)1,2,3
    1 CALL BTLETT(M)
      GO TO 4
```

```
245      2 CALL TUKEY(M)
246        GO TO 4
247      3 CALL PARZEN(M)
248      4 RETURN
249        END
250        SUBROUTINE BTLETT(M)
251        COMMON/A3/V(72)
252        M4=M+1
253        DO 12 I=1,M4
254        V(I)=1.-FLOAT(I-1)/FLOAT(M)
255     12 CONTINUE
256        RETURN
257        END
258        SUBROUTINE TUKEY(M)
259        COMMON/A3/V(72)
260        M4=M+1
261        PI=4.*ATAN(1.)
262        DO 12 I=1,M4
263        V(I)=0.54+0.46*COS(PI*FLOAT(I-1)/FLOAT(M))
264     12 CONTINUE
265        RETURN
266        END
267        SUBROUTINE PARZEN(M)
268        COMMON/A3/V(122)
269        M4=M+1
270        DO 12 I=1,M4
271        K=I-1
272        H1=FLOAT(K)/FLOAT(M)
273        IF(K-M/2)13,13,14
274     13 V(I)=1.-6.*(H1**2-H1**3)
275        GO TO 12
276     14 V(I)=2.*((1.-H1)**3)
277     12 CONTINUE
278        RETURN
279        END
280 \\\\S
281  19360 6 3 2
282
283 ****
```


Program 4

In this program, as in Program 3, the bispectral density function $f(\omega_i,\omega_j)$ and the spectral density function $f(\omega_i)$ are estimated. The null hypothesis H_o:

$$\frac{|f(\omega_i,\omega_j)|^2}{f(\omega_i)f(\omega_j)f(\omega_i+\omega_j)} = \text{constant, is tested.}$$

The program 4 works as in Program 3, except for the differences.

(1) All the parameters, N, M, NP, IW, K5 must be given as input in one card.

(2)
$$IW = \begin{cases} 1 & \text{for Bartlett Window} \\ 2 & \text{for Tukey Window} \\ 3 & \text{for Parzen Window} \end{cases}$$

```
0 FTN.
1 LIBRARY(PROCLIB)
2 NAG.
3 ATTACH(DATA5,HERDATA,ST=SOA,FO=CONF)
4 REWIND(DATA1)
5 REWIND(DATA5)
6 COPYP(DATA5,DATA1)
7 REWIND(DATA1)
8 LDSET(PRESET=NGINF,MAP=B/ZZZZMP)
9 LGO.
10 \\\\S
11       PROGRAM SBISPC(INPUT,OUTPUT,DATA1,TAPE1=INPUT,TAPE2=OUTPUT,
12      1TAPE3=DATA1)
13        COMMON/A1/W(10,12)
14        COMMON/A2/C(72,72)
15        COMMON/A3/V(72)
16
17          COMMON/A4/FM(8,8,17),FN(8,8,17)
18              COMMON/A5/X(1202),IR(1202),RU(1202),E(1202)
19             COMMON/A6/WK1(15),WK2(15)
20               COMMON/A7/F(15,17),FAVR(15),Z(15,15),Y(15),FT(17,15)
21                COMMON/A8/R(72)
22        LEVEL2,W,C,FM,FN
23        READ(1,5)N,M,NP,IW,K5
24     5  FORMAT(I4,4I2)
25        KS=K5
26        WRITE(2,190)N,M,NP,KS
27   190  FORMAT(1H1,10X,21HNR OF OBSERVATIONS=N=,I4,///,10X,
28       124HNR OF COVARIANCE LAGS=M=,I2,///,10X,3HNP=,I2,///,10X,3HKS=,I2,///,
29       165(2H--))
30        REWIND3
31        READ(3,*)(X(I),I=1,N)
32
33        SUMX=0.
34        DO 90 I=1,N
35     90 SUMX=SUMX+X(I)
36        XMEAN=SUMX/FLOAT(N)
37        DO 140 I=1,N
38    140 X(I)=X(I)-XMEAN
39        M4=M+1
40        DO 10 J2=1,M4
41        DO 10 J1=J2,M4
42        SUM=0
43        NL=N-J1+1
44        DO 20 K=1,NL
45        K1=J1-1+K
46        K2=J2-1+K
47     20 SUM=SUM+X(K1)*X(K2)*X(K)
48        C(J1,J2)=SUM/FLOAT(N)
49     10 CONTINUE
50
51
52          CALL WINDOW(M,IW)
53          DO 15 I=1,M4
54          SUM=0.
55          NI=N-I+1
56          DO 16 J=1,NI
57    16    SUM=SUM+X(J)*X(I+J-1)
58    15       R(I)=V(I)*SUM/FLOAT(N-I)
59         PI=4.*ATAN(1.)
60         KF=2*KS+1
61         NP1=NP+1
```



```
62         NP2=NP-1
63         SN=FLOAT(N)/FLOAT(NP)
64         PK=6.
65         DO 40 I=1,NP2
66         DO 40 KK=1,KF
67         W(I,KK)=(FLOAT(I)*SN+FLOAT(KK-KS-1)*PK)/FLOAT(N)
68      40 CONTINUE
69         G=(1./(2.*PI))**2
70         LP=2*NP/3
71         DO 60 I1=1,LP
72         V1=W(I1,KS+1)
73
74           D3=B(V1)
75         I12=NP-I1/2-1
76         I11=I1+1
77         IF(I11.GT.I12)GO TO 61
78         DO 60 I2=I11,I12
79
80
81           U2=W(I2,KS+1)
82           D2=B(U2)
83         DO 60 I3=1,KF
84         U1=W(I1,I3)
85         V2=W(I2,I3)
86         ARF2=C(1,1)
87         ARF1=C(1,1)
88         AIF1=0
89         AIF2=0
90         DO 50 J2=2,M
91         JJ=J2+1
92         DO 50 J1=JJ,M4
93         S=V(J1)*V(J2)*V(J1-J2)*C(J1,J2)
94         A1=FLOAT(J1-1)*PI
95         A2=FLOAT(J2-1)*PI
96         A12=A1-A2
97         ARF1=ARF1+S*(COS(A1*U1+A2*U2)+COS(A1*U2+A2*U1)+COS(A1*U1+A12*U2)
98        **COS(A1*U2+A12*U1)+COS(A12*U1-A2*U2)+COS(A12*U2-A2*U1))
99         ARF2=ARF2+S*(COS(A1*V1+A2*V2)+COS(A1*V2+A2*V1)+COS(A1*V1+A12*V2)
100       **COS(A1*V2+A12*V1)+COS(A12*V1-A2*V2)+COS(A12*V2-A2*V1))
101        AIF1=AIF1-S*(SIN(A1*U1+A2*U2)+SIN(A1*U2+A2*U1)-SIN(A1*U1+A12*U2)
102       *-SIN(A1*U2+A12*U1)+SIN(A12*U1-A2*U2)+SIN(A12*U2-A2*U1))
103        AIF2=AIF2-S*(SIN(A1*V1+A2*V2)+SIN(A1*V2+A2*V2)-SIN(A1*V1+A12*V2)
104       *-SIN(A1*V2+A12*V1)+SIN(A12*V1-A2*V2)+SIN(A12*V2-A2*V1))
105    50 CONTINUE
106        DO 51 I=2,M4
107        AI=FLOAT(I-1)*PI
108        S=(C(I,1)+C(I,I))*(V(I)**2)
109        H=(C(I,1)-C(I,I))*(V(I)**2)
110        ARF1=ARF1+S*(COS(AI*U1)+COS(AI*U2)+COS(AI*(U1+U2)))
111        ARF2=ARF2+S*(COS(AI*V1)+COS(AI*V2)+COS(AI*(V1+V2)))
112        AIF1=AIF1+H*(SIN(AI*(U1+U2))-SIN(AI*U1)-SIN(AI*U2))
113        AIF2=AIF2+H*(SIN(AI*(V1+V2))-SIN(AI*V1)-SIN(AI*V2))
114    51 CONTINUE
115         FM(I1,I2,I3)=G**2*(ARF1**2+AIF1**2)
116        FN(I1,I2,I3)=FM(I1,I2,I3)/(D2*B(U1)*B(U1+U2))
117        IF(I3-KS-1)57,60,52
118    57 J3=I3+KF
119       GO TO 53
120    52 J3=I3+KF-1
```

```
      
53          FM(I1,I2,J3)=G**2*(ARF2**2+AIF2**2)
            FN(I1,I2,J3)=FM(I1,I2,J3)/(D3*B(V2)*B(V1+V2))
   60 CONTINUE
   61 WRITE(2,220)

220     FORMAT(1H1,35X,8(2H--),'ESTIMATED BISPEC MODULUS SQUARE'
     18(2H--))
  221 FORMAT(1HO,10X,21HUSING BARTLETT WINDOW)
  222 FORMAT(1HO,10X,18HUSING TUKEY WINDOW)
  223 FORMAT(1HO,10X,19HUSING PARZEN WINDOW)
      IF(IW-2)501,502,503
  501 WRITE(2,221)
      GO TO 504
  502 WRITE(2,222)
      GO TO 504
  503 WRITE(2,223)
      NPL=NP-LP-1
  504 DO 300 L2=1,NPL
      I2=NP-L2
      I21=2*L2-1
      WRITE(2,230)W(I2,K5+1),(FM(I1,I2,K5+1),I1=1,I21)
  300 CONTINUE
230       FORMAT(1HO,2X,F4.2,2X,9F13.6)
      NL1=LP-1
      DO 400 L2=1,NL1
      I2=LP-L2+1
      I21=I2-1
      WRITE(2,230)W(I2,K5+1),(FM(I1,I2,K5+1),I1=1,I21)
  400 CONTINUE
      WRITE(2,210)(W(J,K5+1),J=1,LP)
  210 FORMAT(1HO,4X,2HW2,/,6X,2HW1,9(8X,F5.3)//,10X,
     123HW1(AS AFRACTION OF PI ))
      WRITE(2,240)

240     FORMAT(1H1,35X,8(2H--),'NORMALIZE BISPEC MODOLUS SQUARE'
     1,8(2H--))
      DO 310 L2=1,NPL
      I2=NP-L2
      I21=2*L2-1
      WRITE(2,230)W(I2,K5+1),(FN(I1,I2,K5+1),I1=1,I21)
  310 CONTINUE
      DO 410 L2=1,NL1
      I2=LP-L2+1
      I21=I2-1
      WRITE(2,230)W(I2,K5+1),(FN(I1,I2,K5+1),I1=1,I21)
  410 CONTINUE
      WRITE(2,210)(W(J,K5+1),J=1,LP)
      KFF=2*KF-1
      WRITE(2,509)
      DO 560 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 561
      DO 560 I2=I11,I12
      WRITE(2,510)I1,I2,(FM(I1,I2,I3),I3=1,KFF)
  560 CONTINUE
```



```
  509     FORMAT(1H1,30X,10(2H--),'MODOULUS SQUARED DATA',10(2H--),//)
  510 FORMAT(1HO,1X,I2,1X,I2,9F10.6,//)
  511 FORMAT(1HO,//,65(2H--),//,55X,'NORMALIZED DATA',//)
  561 WRITE(2,511)
      DO 570 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 571
      DO 570 I2=I11,I12
      WRITE(2,510)I1,I2,(FN(I1,I2,I3),I3=1,KFF)
  570 CONTINUE
  571 L=0
      DO 660 I1=1,LP
      I12=NP-I1/2-1
      I11=I1+1
      IF(I11.GT.I12)GO TO 662
      DO 660 I2=I11,I12
      L=L+1
      DO 660 K=1,KFF
      F(L,K)=FN(I1,I2,K)

  660 CONTINUE
  662 N=7
      K=KFF

        N1=N-1
        DO 678 I=1,N1
       DO 678 J=1,K
        F(I,J)=F(I,J)-F(I+1,J)
678      CONTINUE
          N=N-1
      DO 8 I=1,N
      FAVR(I)=0.0
      DO 7 J=1,K
      FAVR(I)=FAVR(I)+F(I,J)
    7 CONTINUE
      FAVR(I)=FAVR(I)/K
    8 CONTINUE
      DO 9 I=1,N
      DO 9 J=1,K
      F(I,J)=F(I,J)-FAVR(I)

        FT(J,I)=F(I,J)
    9 CONTINUE
      DO 19 I1=1,N
      DO 19 I2=I1,N
      Z(I1,I2)=.0
      DO 19 J=1,K
      Z(I1,I2)=Z(I1,I2)+F(I1,J)*FT(J,I2)


   19   CONTINUE
        IFALL=1
        CALL F04ASF(Z,15,FAVR,N,Y,WK1,WK2,IFALL)
        DO 881 I=2,N
           I1=I-1
```

```
234              DO 881 J=1,I1
235             Z(I,J)=Z(J,I)
236  881      CONTINUE
237         WRITE(2,21)
238       21 FORMAT(1H1,65(2H=),///,28X,12(2H--),2X,'THE COVARIANCE MATRIX',
239          12X,12(2H--))
240         DO 880 I=1,N
241         WRITE(2,22)I,(Z(I,J),J=1,N)
242       880 CONTINUE
243       22 FORMAT(1HO,2X,I2,1X,10F11.7)
244         IF(IFALL.EQ.0)GO TO 100
245         WRITE(2,24)IFALL
246         STOP
247      100 WRITE(2,27)(Y(I),FAVR(I),I=1,N)
248       27 FORMAT(1HO,///,10X,1H(,E12.6,1H,,E12.6,1H),5X,1H(,E12.6,1H,,E12.6
249          1,1H))
250       24 FORMAT(1HO,65(2H=),///,6HIFALL=,I2)
251            Q=0.
252         DO 18 I=1,N
253         Q=Q+(FAVR(I))*Y(I)
254       18 CONTINUE
255         T=K*Q
256         WRITE(2,25)T
257       25 FORMAT(1HO,30X,20(2H=),///35X,32HHOTELLINGS T SQUARE STATISTIC IS
258          1,///,48X,3HT=(,F18.3,1H,,F8.3,1H),///,30X,20(2H=))
259         STOP
260         END
261         SUBROUTINE WINDOW(M,IW)
262         COMMON/A3/V(72)
263         IF(IW-2)1,2,3
264          1 CALL BTLETT(M)
265            GO TO 4
266          2 CALL TUKEY(M)
267            GO TO 4
268          3 CALL PARZEN(M)
269          4 RETURN
270            END
271         SUBROUTINE BTLETT(M)
272         COMMON/A3/V(72)
273         M4=M+1
274         DO 12 I=1,M4
275         V(I)=1.-FLOAT(I-1)/FLOAT(M)
276        12 CONTINUE
277         RETURN
278         END
279         SUBROUTINE TUKEY(M)
280         COMMON/A3/V(72)
281         M4=M+1
282         PI=4.*ATAN(1.)
283         DO 12 I=1,M4
284         V(I)=0.54+0.46*COS(PI*FLOAT(I-1)/FLOAT(M))
285        12 CONTINUE
286         RETURN
287         END
```



```
288         SUBROUTINE PARZEN(M)
289         COMMON/A3/V(122)
290         M4=M+1
291         DO 12 I=1,M4
292         K=I-1
293         H1=FLOAT(K)/FLOAT(M)
294         IF(K-M/2)13,13,14
295        13 V(I)=1.-6.*(H1**2-H1**3)
296         GO TO 12
297        14 V(I)=2.*((1.-H1)**3)
298        12 CONTINUE
299         RETURN
300         END
301
302          FUNCTION B(U)
303          COMMON/A8/R(72)
304          M1=29
305          PI=4.*ATAN(1.)
306          SUM=0.0
307          DO 10 I=1,M1
308        10    SUM=SUM+R(I+1)*COS(PI*U*FLOAT(I))
309          B=(R(1)+2.*SUM)/(2.*PI)
310          RETURN
311          END
312 \\\\S
313  19330 6 3 2
314
315 ****
```

REFERENCES

AKAIKE, H. (1974) Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes, Ann. Inst. Statist. Math., 26, 363-387.

AKAIKE, H. (1977) On entropy maximisation principle, Proc. Symp. on Applications of Statistics, Dayton, Ohio, Ed. P. R. Krishnaiah.

AKAMANAM, J. S. (1983) Some contributions to the study of time series models. Unpublished Ph.D. thesis submitted to the University of Sheffield, England.

ANDERSON, T. W. (1958) "An Introduction to Multivariate Statistical Analysis," John Wiley.

ANDERSON, T. W. (1971) "The Statistical Analysis of Time Series". Wiley, New York.

ASTROM, K. J. and Bohlin, T. (1966) Numerical identification of linear dynamic systems from normal operating records. Proc. IFAC Symp. on Thoery of Self Adapting Control Systems, Sept. 1965, Plenum, New York.

BARTLETT, M. S. (1948) Smoothing periodogram from time series with continuous spectra, Nature, 161, 686-687.

BARTLETT, M. S. (1950) Periodogram analysis and continuous spectra, Biometrika, 37, 1-16.

BARTLETT, M. S. (1966) "An Introduction to Stochastic Processes", 2nd edition, Cambridge University Press.

BHASKARA RAO, M., SUBBA RAO, T. and WALKER, A. M. (1983) On the existence of some bilinear time series models. Jour. Time. Series Anal. 4, 2, 95-110.

BIRKENFELD, W. (1981) On bilinear models for time series analysis forecasting, Computational Statistics, De Gruyter Verlag, Berlin, Ed. H. Buning and P. Naeve.

BOX, G. E. P. and JENKINS, G. M. (1970) "Time Series, Forecasting and Control", Holden-Day, San Francisco.

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

BRILLINGER, D. R. (1980) The comparison of least squares and third order periodogram procedures in the estimation of bifrequency, J. Time Series Anal., 1, 2, 95-101.

BRILLINGER, D. R. and ROSENBLATT, M. (1967a) Asymptotic theory of estimates of k-th order spectra. In Spectral Analysis of Time Series (ed. by B. Harris) 153-188, Wiley, New York.

BRILLINGER, D. R. and ROSENBLATT, M. (1967b) Computation and interpretation of k-th order spectra. In "Spectral Analysis of Time Series", (ed. by B. Harris)189-232, Wiley, New York.

CAMPBELL, H. J. and WALKER, A. M. (1977) A survey of statistcal work on the MacKenzie River series of annual Canadian lynx trappings for the years 1871-1934 and a new analysis, J. Roy. Statist. Soc. (A) 140, 411-431.

CHATFIELD, C. (1980) "The Analysis of Time Series - An Introduction", Chapman and Hall.

CHUNG, K. L. (1974) A course in probability theory, 2nd ed., Academic Press, London.

CRANDALL, S. H. (1963) Zero crossings, peaks and other statistical measures of random responses. J. Acous. Soc. Amer. 35, 11, 1693-1699.

CUNNINGHAM, W. J. (1958) "Introduction to Nonlinear Analysis", McGraw Hill.

DARGAHI-NOUBARY, G. R., LAYCOCK, P. J. and SUBBA RAO, T. (1978) Nonlinear Stochastic Models for seismic events with applications in event identification. Geophys. J. Roy. Astro. Soc., 55, 655-558.

DOOB, J. L. (1953) "Stochastic Processes", Wiley, New York.

DUFFING, G. (1918) Erwingen Schwingungen bei Veranderlichen Eigen Frequenz F. Vieweg u. Sohn: Braunschweig.

GABR, M. M. and SUBBA RAO, T. (1981) The estimation and prediction of subset bilinear time series models with applications, J. Time Series Anal., Vol. 2, 153-171.

GIRI, N. (1965) "On the complex analogues of T^2 and R^2 tests, Ann. Math. Statist. 36, 664-670.

GODFREY, M. D. (1965) An exploratory study of the bispectrum of economic time series, J. Roy. Statist. Soc. (C), 14, 48-69.

GRANGER, C. W. J. and ANDERSEN, A. P. (1978a) "An Introduction to Bilinear Time Series Models", Vandenhoek and Ruprecht, Gottingen.

GRANGER, C. W. J. and ANDERSEN, A. P. (1978b) Nonlinear time series modelling in Applied Time Series Analysis", ed. D. F. Findley, Academic Press.

GRANGER, C. W. J. and ANDERSEN, A. P. (1978c) On the invertibility of time series models. Stochastic Processes and their Applications, 8, 83-92.

GRANGER, C. W. J. and NEWBOLD, P. (1976) Forecasting transformed series, J. Roy. Statist. Soc. B, 38, 189-203.

GRANGER, C. W. J. and NEWBOLD, P. (1977) "Forecasting Economic Time Series", Academic Press.

GUEGAN, D. (1981) Etude d'un Modele nonlineaire, le modele superdiagonal d'ordre C.R. Acad. Sci. Paris, 293, Series 1, 95-98.

HAGGAN, V. and OZAKI, T. (1978) Amplitude-dependent AR model fitting for nonlinear random vibrations. Dept. of Maths. Tech. Report 103, UMIST.

HAGGAN, V. and OYETUNJI, O. (1980) On the selection of subset autoregressive time series models, UMIST Tech. Report 124, Dept. of Maths., UMIST.

HANNAN, E. J. (1962) "Time Series Analysis", Methuen Monographs.

HANNAN, E. J. (1970) "Multiple Time Series",John Wiley, New York.

HANNAN, E. J. (1982) A note on bilinear time series models. Stochastic Processes and their applications, 12, 221-224.

HASSELMAN, K., MUNK, W. and MacDONALD, G. (1963) Bispectra of ocean waves, in "Proc. Symp. Time Series Analysis", ed. M. Rosenblatt, 135-139, John Wiley.

HELLAND, K. N., LII, K. S. and ROSENBLATT, M. (1979) Bispectra and energy transfer in grid-generated turbulence, in "Developments in Statistics", (Ed. P. R. Krishnaiah), Vol. 2, 125-155, Academic Press, New York.

HINICH, M. (1982) Testing for Gaussianity and Linearity of a stationary time series. Jour. Time. Ser. Analysis 3, No. 3, 169-176.

HUBER, P. J., KLEINER, B. and GASSER, T. (1971) "Statistical Methods for investigating phase relations in stationary stochastic processes", I.E.E.E. Trans. on Audio and Electroacoustics, Vol. AU-19-20.

HUNG, G., BRILLINGER, D. R. and STARK, L. (1979) Interpretation of kernels II - Kernels of the Human Pupillary System. Mathematical Biosciences, 46, 159-187.

KHATRI, C. G. (1965) Classical Statistical Analysis based on a certain multivariate complex Gaussian distribution. Ann. Math. Statist. 36, 98-107.

KSHIRSAGAR, A. M. (1972) "Multivariate Analysis", New York, Marcel Dekker.

LEONOV, V. P. and SHIRYAEV, A. N. (1959) On a method of calculation of semi-invariants. Theor. Prob. Appl. 4, 319-329.

LII, K. S. and ROSENBLATT, M. (1982) Deconvolution and estimation of transfer function phase and coefficients for non-Gaussian linear processes. Ann. Statist. 10, 1195-1208.

LII, K. S. and ROSENBLATT, M. (1983) A fourth order deconvolution technique for non-Gaussian linear processes. (Presented at 6th International Symposium on Multivariate Analysis, University of Pittsburgh, U.S.A.)

LII, K. S., ROSENBLATT, M. and VAN ATTA, C. (1976) Bispectral measurements in turbulence, J. Fluid. Mech. 77, 45.

LOMNICKI, Z. A. (1961) Tests for departure from normality in the case of linear stochastic processes, Metrika, 4, 37-62.

LOMNICKI, A. A. and ZAREMBA, S. K. (1957) On estimating the spectral density function of a stochastic process, J. Roy. Statist. Soc., B, 19, 13-37.

LYON, R. H. (1963) Empirical evidence for non-linearity and direction for future work, J. Acous. Soc. Amer., 35, 11, 1712-1721.

MARDIA, K. V. (1970) Measures of multivariate skewness and Kurtosis with applications. Biometrika, 57, 519-530.

MARQUARDT, D. W. (1963) An algorithm for least squares estimation of nonlinear parameters, J. Soc. Industrial and Applied Mathematics, Vol. 11, 431-441.

MEHRA, R. K. (1977) Kalman filters and their applications to forecasting, Management Sciences (special issue of "Forecasting"), reprinted in "TIMS Studies in the Management Sciences", (ed. Makridethi & Wheelwright),Vol. 12, North-Holland, Amsterdam.

MILLER, K. S. (1974) "Complex Stochastic Processes", Addison-Wesley.

MOHLER, R. R. (1973) "Bilinear control processes", Academic Press.

NEAVE, H. R. (1972) A comparison of lag window generators, J. Amer. Statist. Assoc. 67, 152-158.

NEUDECKER, H. (1969) Some theorems on matrix differentiation with special references to Kronecker matrix products, J. Amer. Statist. Assoc., 69, 153-159.

NICHOLLS, D. F. and QUINN, B. G. (1982) Random coefficient autoregressive models: An introduction. Springer-Verlag, New York. Lecture Notes in Statistics 11.

OYETUNJI, O. B. (1980) On selection of subset time series models, unpublished Ph.D. thesis submitted to Dept. of Maths., UMIST.

PARZEN, E. (1957) On consistent estimates of the spectrum of a stationary time series. Ann. Math. Statist. 28, 329-348.

PARZEN, E. (1958) On asymptotically efficient consistent estimates of the spectral density function of a stationary time series, J. Roy. Statist. Soc., B, 20, 303-302.

PERROCHAUD, J. B. (1982) Bispectral analysis of nonlinear systems. Unpublished M.Sc. dissertation, University of Southampton.

PRIESTLEY, M. B. (1962) Basic considerations in the estimation of spectra, Technometrics, 4, 551-564.

PRIESTLEY, M. B. (1978) Nonlinear models in time series analysis, the Statistician, 27, 159-176.

PRIESTLEY, M. B. (1980) State dependent models: A general approach to nonlinear time series analysis, Jour. Time Ser. Analysis, 1, no. 1, 47-71.

PRIESTLEY, M. B. (1981) "Spectral Analysis and Time Series", Vol. 1, Academic Press.

PRIESTLEY, M. B. (1981) "Spectral Analysis and Time Series", Vol. 2, Academic Press.

PRIESTLEY, M. B. and SUBBA RAO, T. (1969) A test for stationarity of time series, J. Roy. Statist. Soc, B, 31, 140-149.

QUINN, B. G. (1982) Stationarity and invertibility of bilinear models. Stochastic Processes and their Applications, 12, 225-230.

ROBINSON, P. M. (1977) The estimation of a nonlinear moving average model. Stochastic Processes and their Applications, 5, 81-90.

ROGERS, G. S. (1980) Matrix Derivatives, Lecture Notes in Statistics, Vol. 2, Marcel Dekker Inc., New York.

ROSENBLATT, M. (1966) Remarks on higher order spectra, "Multivariate Analysis", ed. P. R. Krishnaiah, 383-389, Academic Press, New York.

ROSENBLATT, M. (1979) Linearity and nonlinearity in time series. Paper presented at the ISI Meeting, Manila, Phillipines.

ROSENBLATT, M. (1980) Linear Processes and Bispectra, Jour. Appl. Prob. 17, 265-270.

ROSENBLATT, M. (1981) Cumulants and cumulant spectra, in Handbook of Statistics, Vol. 3, (ed. P. R. Krishnaiah), North-Holland, Amsterdam.

ROSENBLATT, M. and VAN NESS, J. W. (1965) Estimation of the bispectrum, Ann. Math. Statist., 36, 1120-1135.

RUBERTI, A., ISIDORI, A. and d'ALLESSANDRO, P. (1972) "Theory of Bilinear Dynamical Systems", Springer Verlag.

SESAY, S. A. O. (1982) Sampling Properties of the estimates of the parameters of the Bilinear model BL(1,0,1,1). Unpublished M.Sc. dissertation, UMIST, Manchester.

STOKER, J. J. (1950) "Nonlinear Vibrations", New York: Interscience.

SUBBA RAO, T. (1977) On the estimation of bilinear time series models, Bull. Inst. Internat. Statist. 41 (paper presented at the 41st session of ISI, New Delhi, India).

SUBBA RAO, T. (1978) A note on higher order moments of stationary time series and their use in examination of the residuals, Technical Report No. 95, Dept. of Maths., UMIST.

SUBBA RAO, T. (1981a) On the theory of bilinear time series models, J. R. Statist. Soc., B. 43, 2, 244-255.

SUBBA RAO, T. (1981b) The bispectral analysis of nonlinear stationary time series with reference to bilinear time series models, to appear in Handbook of Statistics, Vol. 3 (ed. P. R. Krishnaiah), North-Holland.

SUBBA RAO, T. and GABR, M. M. (1980) A test for linearity of stationary time series, J. Time Series Analysis, 1, 2, 145-158.

SUBBA RAO, T. and GABR, M. M. (1981) The properties of bilinear time series models and their usefulness in forecasting - with examples, paper presented at the Golden Jubilee celebration of the Indian Statistical Institute, Calcutta.

TONG, H. (1980) A view of nonlinear time series model building, Proc. Int. Conf. held at Nottingham University, North Holland, (ed. O. D. Anderson).

TONG, H. and LIM, K. S. (1980) Threshold autoregression, limit cycles and cyclical data, J. R. Statist. Soc. B, 42, 245-292.

TONG, H. and WU, Z. M. (1981) Multi-step-ahead forecasting of cyclical data by threshold autoregression, Technical Report 144, Dept. of Mathematics, UMIST.

TUAN, Pham Dinh, (1983) Bilinear Markovian Representation and Bilinear models, Technical Report no. 161, Dept. of Mathematics, UMIST.

TUAN, Pham Dinh and LANH, Tat Tran (1981) On the first order bilinear time series model. J. Appl. Prob. 18, 617-627.

TUKEY, J. W. (1969) The estimation of power spectra and related quantities, in "On Numerical Approximation", p. 389-411, Madison, Univ. of Wisconsin Press.

VAN NESS, J. W. (1966) Asymptotic normality of bispectral estimates, Ann. Math. Statist. 37, 1257-1275.

VOLTERRA, (1930) "Theory of Functionals and Integral and Integro-differential Equations", London, Dover.

WALDMEIRER, M. (1961) "The sunspot activity in the years 1610-1960", Schultheses, Zurich.

WHITTLE, P. (1954) Curve and periodogram smoothing, J. R. Statist. Soc. B, Vol. 19, 38-47.

WHITTLE, P. (1954) A statistical investigation of sunspot observations with reference to H. Alven's sunspot model, Astrophys. J. 120, 251-260.

WIENER, N. (1958) "Nonlinear Problems in Random Theory", John Wiley & Sons.

WILKINSON, J. H. (1965) "The Algebraic Eigenvalue Problem", Oxford: Clarendon.

WOLD, H. (1938) A study in the analysis of stationary time series", Almquist & Wiskell, Uppsala.

YULE, G. U. (1927) On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, Phils. Trans. Roy. Soc., London, A, 22, 267-298.

AUTHOR INDEX

SUBJECT INDEX