ASPECTS OF NON-LINEAR TIME SERIES ANALYSIS

by

RAHIM MOEANADDIN, B.Sc., M.Sc.

A Thesis Submitted to the University of Kent

at Canterbury for the Degree of Doctor of Philosophy in Statistics

Statistics Department

Mathematical institute

University of Kent

At Canterbury

January 1989

ACKNOWLEDGEMENTS

I am extremely grateful to my supervisor Professor Howell Tong for his efficient guidance and constructive criticisms throughout the course of this research. The constant discussion with him has been a stimulating experience for me. I am also thankful to the other staff members of the Mathematical Institute for their helpful discussion and encouragement from time to time.

I am extremely happy to find an occasion to express my gratitude to my parents and my family, though whatever I express may not fully capture my feelings of respect for them. Their ideas of hard work and devotion to duty are always with me and I dedicate this work to them.

ABSTRACT

Some difficulties of non-linear time series modelling are discussed. The importance of likelihood plots and information matrix calculation is demonstrated. A constructive use of reparametrization is illustrated. Possible non-invertibility of two bilinear models fitted to real data is discussed. The relation between the variance of error term and the variance of the estimate of the bilinear term is also studied. We have examined numerical solutions for a recursive relation for computing the m-step-ahead conditional density of a non-linear autoregressive model by using the Chapman-Kolmogorov formula. The stationary marginal probability density function of the model is approximated by the m-step-ahead conditional density, for sufficiently large m. The advantage of incorporating the matrix squaring procedure is also studied. The conditional mean (regression function) and conditional variance of some non-linear models are studied briefly .The importance of non-parametric estimation of regression functions in non-linear time series analysis is demonstrated. A comparison of likelihood ratio tests using the newly available asymptotic results with the non-likelihood ratio approach such as the modified Petruccelli-Davies's test and Tsay's test is studied . The 5% empirical critical values for the some cases of likelihood ratio test is also obtained. The effect of outliers on the tests is briefly studied. The importance of the profile likelihood plots in locating the threshold and estimating the delay parameter is demonstrated. The blowfly data (raw and transformed) are analysed. The importance of influential data is also discussed. The non-monotonicity of the conditional variance of the error of a m-step non-linear least squares predictor is discussed . We have also studied methods of evaluating the conditional variance for non-linear autoregressive models and illustrated these with both real and simulated data. Bias correction is included. Moreover, the possibility of combinations of forecasts is explored. The performances of linear, bilinear and SETAR models in predicting the sunspot numbers are compared.

Table of Contents

Chapter One: Introduction 1

Chapter Two: Linear Stationary Processes 4

2.1: Stationary stochastic processes and Gaussian process 4

2.2: Some linear processes 7

2.2.1: General linear processes 7

2.2.2: Autoregressive and moving average processes 8

Chapter There: Non-linear Models 9

3.1: Threshold autoregressive models 9

3.1.1: The stability of a special class of threshold autoregressive models 12

3.1.1.1: Notations and definitions 12

3.1.1.2: Stability in time series analysis 13

3.2: Bilinear models 22

3.3: Exponential autoregressive models 25

Chapter Four: Some Structural Properties of Linear and Non-linear Models 26

4.1: Moving average 28

4.1.1: Exact maximum likelihood estimation of MA(1) 28

4.1.2: Conditional maximum likelihood estimation 30

4.2: Exponential autoregressive models 31

4.2.1: Reparametrization 38

4.3: Bilinear models 40

4.3.1: Invertibility 47

4.4: SETAR models 49

4.5: Real data 49

Chapter Five: Stationary Probability Density Function of Non-linear Autoregressive Process 52

5.1: Numerical solution 54

5.1.1: Series expansion 55

5.1.2: Iterative numerical quadrature method 55

5.2: Conditional density approach 55

5.3: Application 57

5.4: Acceleration by matrix squaring 70

Chapter Six: Conditional Mean and Conditional Variance In Time Series Analysis 73

6.1: Gaussian processes 74

6.2: Bilinear models 75

6.2.1: Superdiagonal bilinear models 75

6.2.2: Diagonal models 76

6.2.3: Subdiagonal models 77

6.3: Non-linear autoregressive models 78

6.4: Real data 80

Chapter Seven: A Comparison of Likelihood Ratio Test and CUSUM and Tsay's Tests For Threshold Autoregression 85

7.1: The likelihood ratio test for threshold autoregression 85

7.2: Tests based on predictive residuals 88

7.3: Empirical critical value 90

7.4: A comparative study of likelihood ratio test and the reverse CUSUM test 99

7.5: Power plots 106

7.5.1: Piecewise constant models 107

7.5.2: Piecewise linear models 108

Chapter Eight: SETAR Modelling of Blowfly Data 110

8.1: Some background of A.J. Nicholson data 110

8.2: A test for multi-modality based on kernel density estimation 110

8.3: Influential data 115

8.4: Profile likelihood 116

8.5: A case study 118

8.6: Transformation 125

8.7: Model diagnostic checking 125

8.7.1: Test of whiteness 126

8.7.2: Test of normality 127

Chapter Nine: Multi-step Non-linear Least Squares Prediction 131

9.1: Introduction 131

9.2: Non-monotonicity 134

9.3: Non-linear models 135

9.3.1: Bilinear models 135

9.3.1.1: Superdiagonal models 136

9.3.1.2: Diagonal models 137

9.3.1.3: Subdiagonal models 138

9.3.2: Non-linear autoregressive models 139

9.3.3: Two experiments 142

9.3.4: Prediction from SETAR models 145

9.4: Bias correction of forecasting transformed series 149

9.5: A comparative study of linear, bilinear and SETAR models in predicting sunspot numbers 150

9.6: Combination of forecasts 154

9.7: Risky prediction of Wolf's sunspot numbers 157

9.8: Linear non-Gaussian autoregressive models 159

References 162

Three computer programs for plotting the contours, approximating the stationary probability density function and calculating the m-step-ahead forecasting of SETAR models. 169

CHAPTER ONE

INTRODUCTION

It is probably fair to say that linear time series modelling was born at the beginning of this century, with the introduction of such linear models as the autoregressive models and moving average models, and has reached its height in recent years. For example, the book by Box and Jenkins (1970) and the structural clarification by Akaike (1974) suffice as demonstrations of maturity of the subject of linear time series modelling. In fact the impact of Box and Jenkins was such that the field in the 1970's was pre-dominated by linear and Gaussian. However, these two assumptions have finally been abandoned and time series analysis has moved in a new direction. This new direction is, of course, the study of non-linear models. The mathematical ideas involved are much more complex than those of linear models, and the statistical problems of model identification and parameter estimation are similarly more intricate. (See e.g. Tong, 1987a and Priestley, 1988)

Although the era of non-linear time series models started with Wiener (1958) and was discussed by Nelson and Van Ness (1973) it is in the last ten years that most progress has been made in the field of non-linear time series analysis. Granger and Andersen (1978) and Subba Rao and Gabr (1984) studied bilinear time series models in some detail. These models were originally developed by control engineers to describe input-output relationships for a deterministic non-linear system. Tong (1987,1983) has discussed another important class of non-linear time series models commonly known as threshold autoregressive models. These models are general enough to capture the notion of limit cycles, which play a key role in the modelling of cyclical data and in physical and biological sciences. Other non-linear time series models are exponential autoregressive models introduced by Ozaki and Oda (1978), non-linear threshold models, exponential moving average and other related models introduced by Lawrance and Lewis (1977) and some other non-Gaussian time series models introduced by Mckenzie (1984) and Raftery (1982). All the models, bilinear, threshold autoregressive and exponential autoregressive as well as linear ARMA models can be

considered as special cases of the State-dependent models, introduced by Priestley (1980).

In this thesis, we have discussed some aspects of non-linear time series analysis.

Chapter Two contains some basic idea of time series analysis and stochastic processes.

Some of the non-linear models are introduced in Chapter Three. The Lagrange stability of a special class of threshold autoregressive model has been discussed.

In Chapter Four, some difficulties of non-linear time series modelling are discussed. The importance of likelihood plots and information matrix calculation is demonstrated. A constructive use of reparametrization is illustrated. Possible non-invertibility of two bilinear models fitted to real data is discussed. The relation between the variance of error term and the variance of the estimate of the bilinear term is also studied.

In Chapter Five we have examined numerical solutions for a recursive relation for computing the m-step-ahead conditional density of a non-linear autoregressive model by using the Chapman-Kolmogorov formula. We have approximated the stationary marginal probability density function of the model by the m-step-ahead conditional density, for sufficiently large m. We have also studied the advantage of incorporating the matrix squaring procedure.

Chapter Six contains a brief study of the conditional mean (regression function) and conditional variance of some non-linear models. The importance of non-parametric estimation of regression functions in non-linear time series analysis is demonstrated.

A comparison of likelihood ratio tests using the newly available asymptotic results with the non-likelihood ratio approach such as the modified Petruccelli-Davies's test and Tsay's test is studied in Chapter Seven. In the highlight of Monte Carlo method the 5% empirical critical values for the some cases of likelihood ratio test is also discussed. The effect of outliers on the tests is briefly studied.

In Chapter Eight the kernel density estimate method is employed to construct a test for multimodality of the density of a time series data set. The importance of the profile likelihood plots in locating the threshold and estimating the delay parameter is also demonstrated. The blowfly data (raw and transformed) are analysed. The importance of influential data is also discussed.

The non-monotonicity of the conditional variance of the error of a m-step non-linear least squares

predictor is discussed in Chapter Nine. We have also studied methods of evaluating the conditional variance for non-linear autoregressive models and illustrated these with both real and simulated data. Bias correction is included. Moreover, the possibility of combinations of forecasts is explored. The performances of linear, bilinear and SETAR models in predicting the sunspot numbers are compared.

CHAPTER TWO

LINEAR STATIONARY PROCESSES

In this chapter we introduce some basic idea of time series analysis and stochastic processes. A time series is a set of observations x_t ,each one being recorded at a specified time t. A discrete-time series is one in which the set T_o of time at which observations are made is a discrete set. Continuos-time series are obtained when observations are recorded continuously over some time interval.

The first step in the analysis of a time series is the selection of a suitable mathematical model (or class of models) for the data. To allow for the possibly unpredictable nature of future observations it is natural to suppose that each observation x_t is a realized value of a certain random variable X_t.

The time series $\{x_t , t \in T_o\}$ is then a realization of the family of random variables $\{X_t , t \in T_o\}$. These considerations suggest modeling the data as a realization (or part of a realization) of a stochastic process $\{X_t , t \in T\}$ where very often $T = \{0,\pm1,\pm2,...\} \supseteq T_o$.

2.1. Stationary Stochastic Processes and Gaussian Processes

Definition 2.1.1. A time sequence, $\{X_t , t\in T\}$, is said to be *deterministic* if there exist a function of past and present values $g_t = g(X_{t-j} , j=0,1,...)$ such that

$$E[(X_{t+1}-g_t)^2]=0 .$$

Definition 2.1.2. A *stochastic process* is a family of random variables $\{X_t , t\in T\}$ defined on a probability space (Ω,F,P).

We note that for each fixed $t \in T$, X_t is a function of $X_t(.)$ on Ω. On the other hand for each fixed $\omega \in \Omega$, $X_.(\omega)$ is a function on T.

Definition 2.1.3. The functions $\{X_.(\omega) , \omega \in \Omega\}$ on T are known as the realizations or sample-paths of the process $\{X_t , t \in T\}$.

Definition 2.1.4. (The Distribution Function of a Stochastic Process $\{X_t, t\in T\subset \mathbf{R}\}$). Let **T** be the set of all vectors $\{\mathbf{t}=(t_1,\ldots,t_n)\in T^n: t_1<t_2<\cdots<t_n,\ n=1,2,3,\ldots\}$. Then the (finite-dimensional) distribution functions of $\{X_t,\ t\in T\}$ are the functions $\{F_{\mathbf{t}}(.),\ \mathbf{t}\in \mathbf{T}\}$ defined for $\mathbf{t}=(t_1,\ldots,t_n)$ by

$$F_{\mathbf{t}}(\mathbf{X})=P(X_{t_1}\le x_1,\ldots,X_{t_n}\le x_n),\ \mathbf{x}=(x_1,\ldots,x_n)'\in \mathbf{R}^n .$$

Theorem 2.1.1. (Kolmogorov's Theorem). The probability distribution functions $\{F_{\mathbf{t}}(.),\ \mathbf{t}\in \mathbf{T}\}$ are the distribution functions of some stochastic process if and only if for any $n\in\{1,2,3,\ldots\}$, $\mathbf{t}=(t_1,\ldots,t_n)\in \mathbf{T}$ and $1\le i\le n$,

$$\lim_{x_i\to\infty} F_{\mathbf{t}}(\mathbf{x})=F_{\mathbf{t}(i)}(\mathbf{x}(i)), \tag{2.1.2}$$

where $\mathbf{t}(i)$ and $\mathbf{X}(i)$ are the (n-1)-component vectors obtained by deleting the i^{th} components of **t** and **X** respectively. Hence forth, unless the contrary explicitly stated, we will assume that, for each t, X_t is a continuous random variable with probability density function $f_t(x)$ defined for all x, so that, the mean and the variance of X_t will be given by

$$\mu_t=E(X_t)=\int_{-\infty}^{\infty} x f_t(x)dx ,$$

$$\sigma_t^2=V(X_t)=\int_{-\infty}^{\infty}(x-\mu_t)^2 f_t(x)dx ,$$

and the covariance between X_t and X_s will be

$$\gamma_{t,s}=cov(X_t,X_s)=E[(X_t-\mu_t)(X_s-\mu_s)],$$

if they exist.

In time series analysis, it is often impossible to make more than one observation at a given time so that we only have an observation on the random variable at time t. With a single realization, it is not possible to estimate with any precision μ_t, for every t if the μ_t sequence is allowed to take any set of values. To overcome such problem, the time series analyst is forced to assume the process to be stationary and ergodic.

Definition 2.1.5 The time series $\{X_t,\ t\in \mathbf{Z}\}$, with index set $\mathbf{Z}=\{0,\pm1,\pm2,\ldots\}$, is said to be stationary if

(*i*) $E(X_t^2)<\infty$ for *all* $t\in \mathbf{Z}$,

(*ii*) $E(X_t)=m$ for *all* $t\in \mathbf{Z}$,

and

(*iii*) $\gamma_{r,s}=\gamma_{r+t,s+t}$ for *all* $r,s,t\in \mathbf{Z}$.

Stationarity as just defined is frequently referred to in the literature as weak stationarity, covariance stationarity, stationarity in wide sense or second order stationarity.

Definition 2.1.6 The time series $\{X_t, t\in Z\}$ is said to be strictly stationary if the joint distributions of $(X_{t_1},\ldots,X_{t_k})'$ and $(X_{t_1+h},\ldots,X_{t_k+h})'$ are the same for all positive integer k and for all $t_1,\ldots,t_k$, $k\in Z$.

A strictly stationary process with finite second moments is stationary. The converse of the previous statement is not true.

The methods which have been used to estimate the mean and autocovariance function using observations from a single realization are all based on the basic strategy of replacing *ensemble*[1] *average* by their corresponding *time average*. If the resulting estimates are mean square consistent it follows that the time average over a period of n points converges in mean squares, as $n\to\infty$, to the corresponding ensemble average. Processes which possess this type of property are called ergodic.

Consider a stationary time series $\{X_t\}$, it follows from definition (2.1.5) that $cov(X_{t_1},X_{t_2})$ is simply a function of $|t_1-t_2|$. This function is called *autocovariance function* of $\{X_t\}$ at lag (t_2-t_1), which will be denoted by $\gamma_{t_2-t_1}$. It has the following properties:

i) $\gamma_o = Var(X_t)$,.

ii) $|\gamma_\tau|\le\gamma_o$, $\forall\tau\in Z$,

iii) $\gamma_{-\tau}=\gamma_\tau$, $\forall\tau\in Z$,

iv) $\forall t_1,t_2,\ldots,t_n\in Z$, for *all positive integer* n, and
for all real number $z_1,\ldots,z_n\in R^n$,

$$\sum_{r=1}^{n}\sum_{s=1}^{n}\gamma_{t_r-t_s}\,z_r\,z_s\ge 0\,.$$

The ratio $\rho_\tau=\dfrac{\gamma_\tau}{\gamma_o}$, $\tau\in Z$, is called *autocorrelation function* of $\{X_t\}$ of lag τ. Properties (*ii*),(*iii*) and (*iv*) still hold if the γ's are replaced by the ρ's with corresponding suffixes. It is well known that ρ_τ can be interpreted as a measure of *linear* association between X_t and $X_{t\pm\tau}$.

Definition 2.1.7. The process $\{X_t\}$ is a *Gaussian* time series if and only if the distribution functions of $\{X_t\}$ are all multivariate normal.

† The collection of all possible records of random variable X_t.

If $\{X_t, t\in \mathbf{Z}\}$ is a stationary Gaussian process then $\{X_t\}$ is strictly stationary, since for all $n \in \{1,2,3,...\}$ and for all $h, t_1,t_2,... \in \mathbf{Z}$, the random vectors $(X_{t_1},\ldots,X_{t_n})'$ and $(X_{t_1+h},\ldots,X_{t_n+h})'$ have the same mean and covariance matrix, and hence the same distribution.

2.2. Some Linear Processes

In many respects the simplest kind of time series $\{X_t\}$is one in which the random variables X_t, t=0,±1,±2, $\cdots$ are independently and identically distributed with zero mean and variance σ^2.

Definition 2.2.1. The process $\{e_t\}$ is said to be white noise with mean zero and variance σ^2 if and only if $\{e_t\}$ has zero mean and covariance function

$$\gamma_h = \begin{cases} \sigma^2 & \text{if } h=0 \\ 0 & \text{if } h\neq 0 . \end{cases}$$

It will then be called *strict white noise* if $\{e_t\}$ is a sequence of independent random variables.

A very wide class of stationary processes can be generated by using white noise as the forcing terms in a set of linear difference equations.

2.2.1. General Linear Processes

A general linear process $\{X_t\}$ is one that can be represented as a weighted linear combination of the present and past terms of a white noise process $\{e_t\}$:

$$X_t = e_t+\psi_1 e_{t-1}+\psi_2 e_{t-2}+\cdots , \tag{2.2.1}$$

where $\{\psi_i\}$ is a sequence of constant real numbers satisfying $\sum_{i=1}^{\infty} \psi_i^2 < \infty$.

By using the backward shift operator ,B, ($B^k X_t = X_{t-k}$), equation (2.2.1) is equivalent to

$$X_t = \Psi(B)e_t , \tag{2.2.2}$$

where $\Psi(B) = 1+\psi_1 B+\psi_2 B^2+\cdots$. In fact $\Psi(B)$ is a linear operator which transforms $\{e_t\}$ into $\{X_t\}$. We can write (2.2.2) symbolically in the equivalent form

$$e_t = \frac{1}{\Psi(B)} X_t = \Phi(B)X_t , \tag{2.2.3}$$

where $\Phi(B) = 1+\phi_1 B+\phi_2 B^2+\cdots$, if $\Psi(B)$ is invertible. It implies that equation (2.2.3) can be written as

$$\Phi(B)X_t = e_t ,$$

or

$$X_t+\phi_1 X_{t-1}+\phi_2 X_{t-2}+\cdots = e_t \ . \tag{2.2.4}$$

The condition for stationarity of the general linear model (2.2.2) is that $\Psi(B)$, which is the generating function of ψ weights, must converge for $|B| \leq 1$, and the invertibility condition for the model (2.2.4) is that $\Phi(B)$ must converge for $|B| \leq 1$. That is, on or within the unit circle.

2.2.2. Autoregressive and Moving Average Processes

It may be said that the era of linear time series modeling began with such linear model as Yule's *autoregressive* models (1927), first introduced in the study of sunspot numbers.

Definition 2.2.2. The process (2.2.4) with all except the first p coefficients equal to zero can be written as

$$X_t+\phi_1 X_{t-1}+...+\phi_p X_{t-p} = e_t \ , \tag{2.2.5}$$

and $\{X_t\}$ is called an *autoregressive* process of order p, $AR(p)$. The equation (2.2.5) equivalently can be written in the form

$$\Phi(B)X_t = e_t , \tag{2.2.6}$$

where $\Phi(B) = 1+\phi_1 B+\cdots+\phi_p B^p$. The $AR(p)$ model is stationary if all the roots of $\Phi(B)$ lie outside a unit circle.

Definition 2.2.5. In general linear model (2.2.1) if $\psi_{q+i}=0$, $i=1,2,...$ then

$$X_t = e_t+\psi_1 e_{t-1}+\cdots+\psi_p e_{t-q} , \tag{2.2.7}$$

or

$$X_t = \Psi(B)e_t ,$$

where $\Psi(B) = 1+\psi_1 B+...+\psi_q B^q$, and the process will be called *moving average processes* ; $MA(q)$. The $MA(q)$ process is invertible if all the roots of $\Psi(B)$ lie outside the unit circle.

A more general class of linear models is obtained by replacing e_t by a weighted average of $e_t, e_{t-1}, \cdots, e_{t-q}$; that is

$$X_t+\phi_1 X_{t-1}+\cdots+\phi_p X_{t-p} = e_t+\psi_1 e_{t-1}+\cdots+\psi_q e_{t-q} , \tag{2.2.8}$$

or $\Phi(B)X_t = \Psi(B)e_t$. This is known as the mixed *autoregressive moving average process* of order p and q, denoted by $ARMA(p,q)$. This process is stationary if all the roots of $\Phi(B)$ lie outside the unit circle and is invertible if all the roots of $\Psi(B)$ lie outside the unit circle.

CHAPTER THREE

NON-LINEAR MODELS

Many processes occurring in nature and in a variety of physical and engineering fields display some form of non-linear behaviour. The most well-known non-linear phenomenon is the *limit cycle*, which has proved extremely useful in such widely differing areas as ecology, hydrology, medical engineering, etc. Such behaviour is quite unknown to any form of linear representational model, and hence many sets of time series data can not be adequately represented by a linear model. This and other facts have led to the development, in recent years, of many interesting non-linear time series models. Some of these models have been shown to be capable of modelling a wide variety of non-linear behaviour and have also been shown to provide extremely good fits to real data. We now describe some of these models below.

3.1. Threshold Autoregressive Models

The idea of using piecewise linear models in a systematic way for the modeling of discrete time series data was first mentioned in Tong (1977a). A comprehensive account is available in Tong (1983).

Let $\{\mathbf{X}_t\}$ be a k-dimensional time series and, for each t, let J_t be an observable (indicator) random variable, taking integer values $\{1,2,3,...,l\}$. A canonical form of a *threshold* model is given by (for $t=0,\pm1,\pm2,...$)

$$\mathbf{X}_t = \mathbf{B}^{(J_t)}\mathbf{X}_t+\mathbf{A}^{(J_t)}\mathbf{X}_{t-1}+\mathbf{H}^{(J_t)}\mathbf{e}_t+\mathbf{C}^{(J_t)}, \tag{3.1.1}$$

where, for $J_t=j$, $\mathbf{A}^{(j)}$, $\mathbf{B}^{(j)}$ and $\mathbf{H}^{(j)}$ are $k\times k$ (nonrandom) matrix coefficients, $\mathbf{C}^{(j)}$ is $k\times1$ vector of constants, and $\{\mathbf{e}_t\}$ is a sequence of independent and identically distributed k-dimensional random vector with zero mean and a covariance matrix and $\mathbf{e}_t$ independent of $\mathbf{X}_s$, $s<t$.

For each t, the value taken by J_t indicates the *regime* the system (X_t) is currently in and the switching from one regime to the next may be related to the crossing of a threshold by an exogenous or endogeneous variable.

Let $\{r_o, r_1, \cdots, r_l\}$ denote a linearly ordered subset of the real numbers, such that $r_o < r_1 < \cdots < r_l$, where r_o and r_l are taken to be $-\infty$ and $+\infty$ respectively. They define a partition of the real line $\mathbf{R}$, i.e.

$$\mathbf{R} = R_1 \cup R_2 \cup \cdots \cup R_l ,$$

where

$$R_i = (r_{i-1}, r_i] .$$

Writing $\mathbf{X}_t = (X_t, X_{t-1}, \cdots, X_{t-k+1})'$,

$$\mathbf{A}_{(j)} = \left[\begin{array}{ccc|c} a_1^{(j)} & a_2^{(j)} & \cdots & a_{k-1}^{(j)} & a_k^{(j)} \\ \hline & & \mathbf{I}_{k-1} & & \mathbf{0} \end{array}\right],$$

$$\mathbf{B}^{(j)} = \mathbf{0} , \ \mathbf{H}^{(j)} = \begin{bmatrix} h_1^{(j)} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix},$$

$$\mathbf{e}_t = (e_t, e_{t-1}, \cdots, e_{t-k+1})' , \ \mathbf{C}^{(j)} = (a_o^{(j)}, 0, ..., 0)' .$$

and $R_{j,d}^{(k)} = \mathbf{R}\times\mathbf{R}\times\cdots\times\mathbf{R}\times R_j\times\mathbf{R}\times\cdots\times\mathbf{R}$, the cylinder set in the Cartesian product of k real lines, on the interval R_j with d-th coordinate space, $d \in \{1,2,...,k\}$. Setting $J_t = j$ if $\mathbf{X}_{t-1} \in R_{j,d}^{(k)}$, we have

$$X_t = a_o^{(J)} + \sum_{i=1}^{k} a_i^{(J)} X_{t-i} + h_1^{(J)} e_t , \tag{3.1.2}$$

conditional on $X_{t-d} \in R_j$, $j=1,2,...,l$. Since $\{J_t\}$ is now a function of $\{X_t\}$ itself, we call the univariate time series $\{X_t\}$ given by (3.1.2) a *self-exciting threshold autoregressive model of order* $(l;k,...,k)$ or SETARMA$(l;k,...,k)$, where k is repeated l times.

If for $j = 1,2,...,l$

$$a_i^{(j)} = 0 \text{ for } i = k_j+1, k_j+2, \cdots, k ,$$

then $\{X_t\}$ is called SETAR$(l;k_1,k_2,\cdots,k_l)$. The parameters $r_1, \ldots, r_{l-1}$ are *threshold parameters* and d is called *delay* parameter.

If the first row of $\mathbf{H}^{(j)}$ is of the form

$$(h_1^{(j)}, h_2^{(j)}, \cdots, h_k^{(j)}) , \ h_k(j) \neq 0 , \ (j = 1,2,...,l) ,$$

then the generalization of SETAR to a *self-exciting threshold autoregressive moving average* model of order $(l;k,...,k;k-1,...,k-1)$ or simply SETAR$(l;k,...,k;k-1,...,k-1)$, takes the form

$$X_t = a_o^{(j)} + \sum_{i=1}^{k} a_i^{(j)} X_{t-i} + \sum_{i=0}^{k-1} h_i^{(j)} e_{t-i} \,, \tag{3.1.3}$$

conditional on $X_{t-d} \in R_j$, $j = 1,2,...,l$.

Jones (1976,1978) has given a comprehensive study of the probabilistic structure of a first order non-linear autoregressive time series model . Some other authors have studied ergodicity of specific non-linear time series models mostly SETAR models of order one, Tong(1983), Petruccelli and Woolford (1984), Chan et al (1985). The case of non-linear models of higher orders is insufficiently studied although the approach of Chan and Tong (1985) seems promising.

Definition 3.1.1: The Markov process $\{X_t\}_{t=0}^{\infty}$ defined on (X,F) with transition probabilities

$$p^n(x,A) = p_r(X_n \in A \mid X_o = x) \,, \; x \in \mathbf{X} \,, \; A \in \mathbf{F} \,,$$

is ergodic if there exists a unique, finite invariant measure π on F such that

$$\pi(A) = \int \pi(dy) p^1(y,A) \,.$$

The link between stability and ergodicity of a stochastic difference equations have been studied by Chan and Tong (1985). Let

$$h(x_1, x_2, \cdots, x_m) = c_i + \sum_{j=1}^{m} a_{ij} x_j \;\text{ if } r_{i-1} \le x_d \le r_i \,,$$

where $\{-\infty = r_o < r_1 < \cdots < r_l = \infty\}$ is an ordered partition of R, $d \le m$, c_i and a_{ij} are constants. The function h is the autoregressive function for the SETAR model. If we define $T: \mathbf{R}^m \to \mathbf{R}^m$ by

$$T(\mathbf{x}) = \begin{bmatrix} h(\mathbf{x}) \\ x_1 \\ \vdots \\ x_{m-1} \end{bmatrix} \text{ where } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \in \mathbf{R}^m \,,$$

then this T with $\mathbf{e}_n = \begin{bmatrix} e_n' \\ 0 \\ \dots \\ 0 \end{bmatrix}$ with e_n' i.i.d., each having an absolutely continuous distribution and the probability density function $f(.)$ is positive every where in R, constitutes the Markovian state-space equation for the SETAR model. (See Chan and Tong 1985 p. 670).

Lemma 3.1.1: If $\max_i \sum_j |a_{ij}| < 1$ and e_n' possesses first absolute moment, then $\mathbf{X}_{n+1} = T(\mathbf{X}_n) + \mathbf{e}_n$ is ergodic. (See Chan and Tong 1985).

Suppressing the random forcing term $\mathbf{e}_{n+1}$ in $\mathbf{X}_{n+1}=T(\mathbf{X}_n)+\mathbf{e}_{n+1}$, we obtain $\mathbf{X}_{n+1}=T(\mathbf{X}_n)$ which will be called the associated deterministic difference equation or the deterministic part of the stochastic difference equation $\mathbf{X}_{n+1}=T(\mathbf{X}_n)+\mathbf{e}_{n+1}$. The study of the deterministic difference equation $\mathbf{X}_{n+1}=T(\mathbf{X}_n)$ may be the skeleton of the study of the stochastic difference equation $\mathbf{X}_{n+1}=T(\mathbf{X}_n)+\mathbf{e}_{n+1}$. If the range of T is bounded, it is clear that $\mathbf{X}_{n+1}=T(\mathbf{X}_n)+\mathbf{e}_n$ is ergodic. Boundedness is one form of stability of the dynamics $\mathbf{X}_{n+1}=T(\mathbf{X}_n)$.

3.1.1. The Stability of a Special Class of Threshold Autoregressive Models

The term *stability* is so expressive that it almost tells its own story. A device of some sort operates under certain conditions. These conditions are slightly changed. Does the change have little or considerable effect upon the device? In the first case it is thought as stable, and in the second as unstable. Nowadays there is more and more reason for studying difference equations systematically. First let us consider the difference equation

$$x_{n+1}=T(x_n),n=0,\pm1,\pm2,\cdots.$$

The solution to the initial value problem

$$x_{n+1}=T(x_n),\ x_o=x^o, \tag{3.1.1.1}$$

is $x_n=T^n(x^o)$, where T^n is the nth iterate of T: $T^{n+1}=T(T^n)$ and $T^o=I$, the identity mapping. The product of functions is composite.

There are many reasons for being interested in what happens to $T^n(x)$ for large values of n. This concern with the asymptotic behaviour of $T^n(x)$ is what stability theory is all about. More than that, stability has to do with dynamic behaviour. Basic to methods of successive approximation of solutions of $x=T(x)$ is the fact that, if $T^n(x^o)$ converges, its limit is a solution (a *fixed* or *invariant point*).

3.1.1.1. Notations and Definitions

For $x\in R^m$ and S any set in R^m

$$\rho(x,S)=inf\{|y-x|;y\in S\},$$

the distance of x from S;

$T^n(x) \to S$ as $n \to \infty$ means $\rho(T^n(x), S) \to 0$ as $n \to \infty$;

$\bar{S} = \{x; \rho(x,S) = 0\}$ is the *closure* of S.

A set S is *closed* if $\bar{S} = S$ and open if its complement is closed.

$T(S) = \{T(x); x \in S\}$

Definition 3.1.1.1.1 (Birkhoff). A point y is a *limit point* of $T^n(x)$ if there is a sequence of integers n_i such that $T^{n_i}(x) \to y$ and $n_i \to \infty$ as $i \to \infty$. The *limit set* $\Omega(x)$ of the *motion* (the motion or trajectory $T^n(x)$ from x refers to the sequence of states $x, T(x), \cdots, T^n(x), \cdots$) $T^n(x)$ from x is the set of all limit points of $T^n(x)$.

Definition 3.1.1.1.2. A motion $T^n(x)$ is said to be *periodic* (or *cyclic*) if for some $k > 0$, $T^k(x) = x$. The least such integer k is called the *period* of the motion or the order of cycle. If $k = 1$, x is a fixed point of T and is called an *equilibrium state* of (3.1.1.1).

Definition 3.1.1.1.3. A motion $T^n(x)$ which is bounded for all $n > 0$, is said to be stable in the sense of *Lagrange.*

3.1.1.2. Stability in Time Series Analysis

The idea of *innovations* is now well-developed and may be formalized by the expression $X_t - E[X_t | F_{t-1}]$ where F_S denotes the sigma field generated by $X_s, X_{s-1}, \cdots$. (See e.g. Tong, 1987b).

Let the innovation sequence be denoted by $\{\varepsilon_t\}$, then we may decompose a time series $\{X_t\}$ by

$$X_t = \hat{X}_t + \varepsilon_t , \qquad (3.1.1.2.1)$$

where $\hat{X}_t = E[X_t | F_{t-1}]$, $\varepsilon_t = X_t - \hat{X}_t$. We refer to $X_t = \hat{X}_t$ as the *skeleton* of the time series specified by (3.1.1.2.1).

Consider the class of non-linear autoregressive model, i.e. models of the form

$$X_t = \lambda(X_{t-1}) + e_t , \qquad (3.1.1.2.2)$$

and its higher order generalization. In this case

$$\hat{X}_t = \lambda(X_{t-1})$$

and $\varepsilon_t = e_t$. If the skeleton (difference equation $X_t = \lambda(X_{t-1})$) is asymptotically stable at the origin in the

sense that the recursion $X_t = \lambda(X_{t-1})$ always tends to zero as $t \to \infty$ regardless of initial X_o, then motion of $\lambda^n(x)$ is Lagrange stable and the time series defined by (3.1.1.2.2) is ergodic under quite general conditions on e_t and λ. (Chan and Tong, 1985).

Now, except for the linear case, there is no systematic way of checking stability. Chan and Tong (1984) and Tong (1987b) have considered an application of symbolic dynamics in studying the Lagrange stability of difference equation

$$X_n = \begin{cases} \phi_1 X_{n-1} + \phi_2 X_{n-2} & \text{if } X_{n-2} > 0 \\ \phi_1' X_{n-1} & \textit{otherwise}, \end{cases} \qquad (3.1.1.2.3)$$

where $\phi_1^2 + 4\phi_2 < 0$. His method will be used in studying the stability of the following models.

Consider

$$X_n = \begin{cases} \phi_1 X_{n-1} & \text{if } X_{n-1} < r' \\ \phi_1' X_{n-1} + \phi_2' X_{n-2} & \text{if } r' \le X_{n-1} \le r \\ \phi_1'' X_{n-1} & \text{if } X_{n-1} > r, \end{cases} \qquad (3.1.1.2.4)$$

where ϕ_1, ϕ_1', ϕ_2', ϕ_1'', r, r', are real constant. Without loss of generality let $r > 0$ and $r' < 0$. To understand the stability of (3.1.1.2.4), it is more convenient to use the state space representation.

Let

$$Z_n = \begin{bmatrix} z_n^1 \\ z_n^2 \end{bmatrix} = \begin{bmatrix} X_n \\ X_{n-1} \end{bmatrix}, A = \begin{bmatrix} \phi_1 & 0 \\ 1 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} \phi_1' & \phi_2' \\ 1 & 0 \end{bmatrix} \text{ and } C = \begin{bmatrix} \phi_1'' & 0 \\ 1 & 0 \end{bmatrix}.$$

Note that A, B and C are companion matrices, A and C being degenerate. Then (3.1.1.2.4) is equivalent to

$$Z_n = \begin{cases} CZ_{n-1} & \text{if } z_{n-1}^1 < r' \\ BZ_{n-1} & \text{if } r' \le z_{n-1}^1 \le r \\ AZ_{n-1} & \text{if } z_{n-1}^1 > r, \end{cases} \qquad (3.1.1.2.5)$$

$X_n = (1 \;\; 0) Z_n$. Obviously, X_n is Lagrange stable if and only if Z_n is so.

Define

$$R=\{(x,y):x>r\}\ ,\ M=\{(x,y):r'\le x\le r\},$$
$$L=\{(x,y):x<r'\}\ ,\ Q_1=\{(x,y)\in R:y<0\},$$
$$Q_2=\{(x,y)\in R:y\ge 0\}\ ,\ Q_3=\{(x,y)\in M\},$$
$$Q_4=\{(x,y)\in L:y\ge 0\}\ ,\ Q_5=\{(x,y)\in L:y<0\}.$$

For each $Z_n\in R$, Z_{n+1} is obtained by the action of A on Z_n and similarly for any $Z_n\in M$ and $Z_n\in L$, Z_{n+1} is obtained by the action of B and C on Z_n respectively.

Case (i) Let $\phi_1\ge 0$, $\phi_1''\ge 0$, then we may have the pictorial representation of Fig. 3.1.1.2.1a and 3.1.1.2.1b which displayed the action of A on R.

$$A\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\phi_1 & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=x\begin{bmatrix}\phi_1\\1\end{bmatrix},\ x\ \&\ r>0.$$

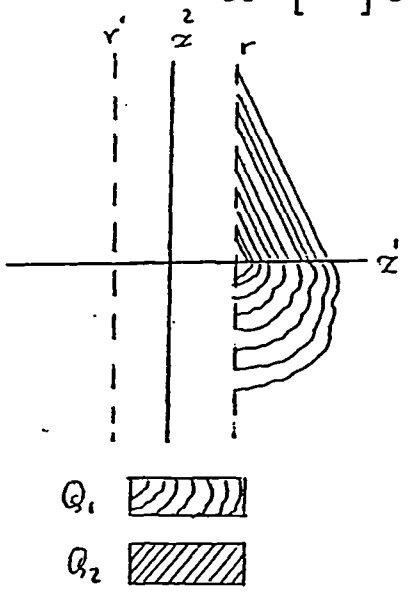


Fig. 3.1.1.2.1a

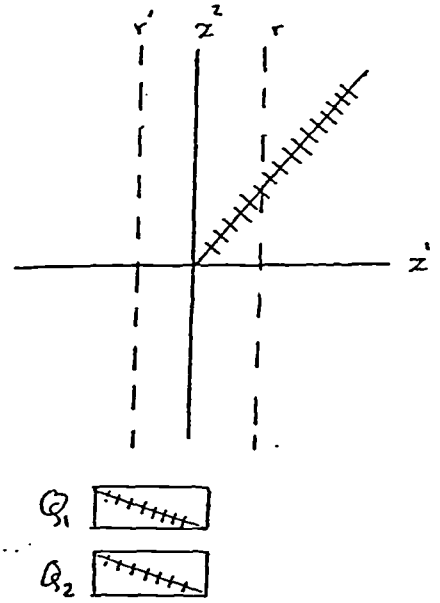


Fig. 3.1.1.2.1b

The action of C on L (Figs. 3.1.1.2.2a and 3.1.1.2.2b) can be represented as

$$C\begin{bmatrix}x\\y\end{bmatrix}=x\begin{bmatrix}\phi_1''\\1\end{bmatrix},\ x\ \&\ r'<0.$$

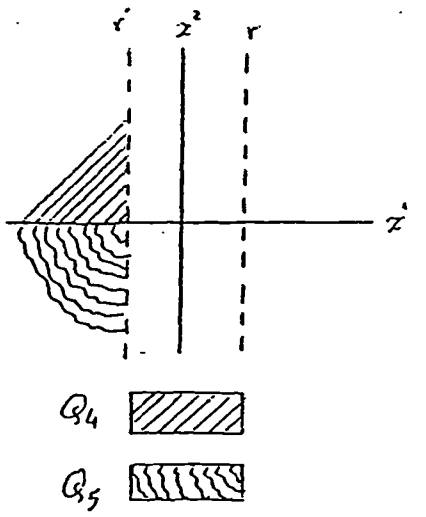


Fig. 3.1.1.2.2a

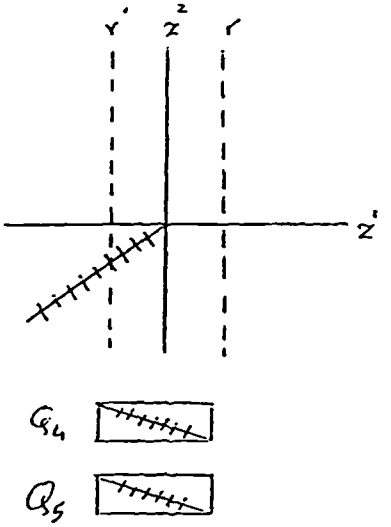


Fig. 3.1.1.2.2b

For each point $Z_n \in R$, $Z_{n+1} = AZ_n \in Q_2$ or Q_3. In the first case, Z_{n+1} will be under action A again. In the second case, Z_{n+1} under action B may move to any other regime or stay in Q_3. Note that in this case, we will show that Z_n can not stay in Q_3 and move along the z^2 axis to $\pm\infty$. i.e After finite actions under B it will move to R or L or may absorb to the origin. For each $Z_n \in L$, $Z_{n+1} = CZ_n \in Q_5$ or M. In the first case Z_{n+1} will be under action C and in the second case under a finite number of actions of B, may move to R or L or stay at the origin. Therefore for $\phi_1 \geq 0$, $\phi_1'' \geq 0$, the tail of trajectories, which are important, are of the form A^n or B^n. In this case the system is Lagrange stable if $\phi_1 \leq 1$ and $\phi_1'' \leq 1$.

Similar to the case (i), when $\phi_1 < 0$, $\phi_1'' < 0$, we can show that the system is Lagrange stable if $\phi_1 \phi_1'' \leq 1$.

Case (ii) Let $\phi_1 < 0$ and $\phi_1'' \geq 0$. the action of A on R and C on L are given by Figs. 3.1.1.2.3a , 3.1.1.2.3b and 3.1.1.2.4a , 3.1.1.2.4b respectively.

$$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \phi_1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} = x\begin{bmatrix} \phi_1 \\ 1 \end{bmatrix}, x \& r > 0 , \phi_1 < 0 .$$

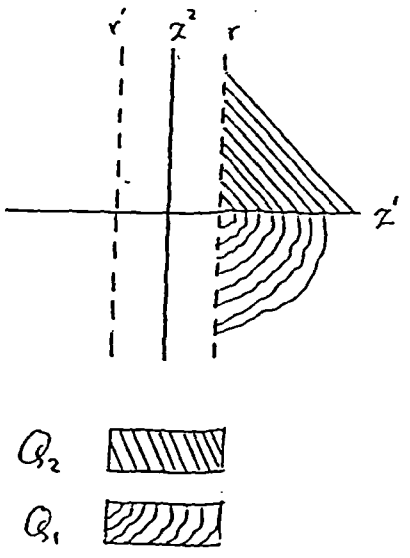


Fig. 3.1.1.2.3a

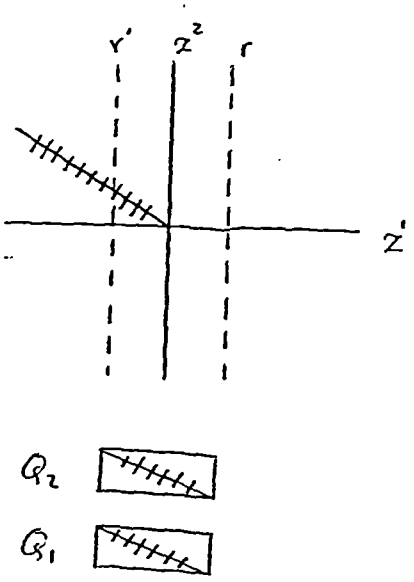


Fig. 3.1.1.2.3b

And

$$C\begin{bmatrix}x\\y\end{bmatrix}=\begin{bmatrix}\phi_1^{\sim} & 0\\ 1 & 0\end{bmatrix}\begin{bmatrix}x\\y\end{bmatrix}=x\begin{bmatrix}\phi^{\sim}\\1\end{bmatrix},\ x\,\&\,r'<0,\ \phi^{\sim}>0.$$

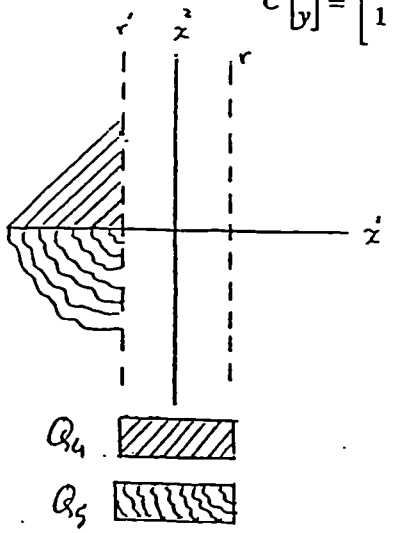


Fig. 3.1.1.2.4a

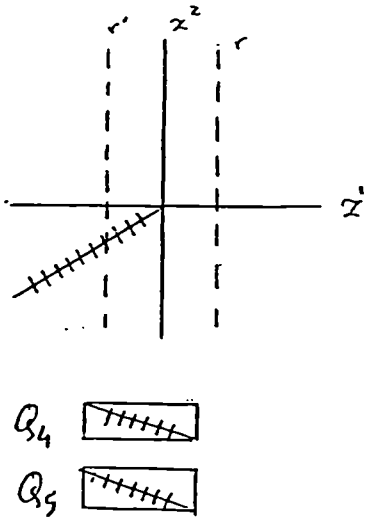


Fig. 3.1.1.2.4b

It is easy to show that the important tail of trajectories takes the form $(A;C^n)$. Thus the system is Lagrange stable if $\phi_1^{\sim}\le 1$ and $\phi_1\le 0$. Similarly for the case $\phi_1^{\sim}<0$, $\phi_1\ge 0$, the system is Lagrange stable if $\phi_1\le 1$ and $\phi_1^{\sim}<0$.

Lemma 3.1.1.2.1: Let Z_n satisfy (3.1.1.2.4), then $\{Z_n\}$ is lagrange stable iff one of the following conditions hold:

(i) $0<\phi_1\le 1$ and $0<\phi_1^{\sim}\le 1$,

(ii) $\phi_1<0$, $\phi_1^{\sim}<0$ and $\phi_1\phi_1^{\sim}\le 1$,

(iii) $\phi_1^{\sim}\le 0$, $0\le\phi_1\le 1$,

(vi) $\phi_1\le 0$, $0\le\phi_1^{\sim}\le 1$.

Example 3.1.1.2.1.: The system

$$X_n=\begin{cases}-5.0X_{n-1} & \text{if } X_{n-1}<-1.0\\ -3.0X_{n-1}+8.0X_{n-2} & \text{if } -1.0\le X_{n-1}\le 2\\ 0.9X_{n-1} & X_{n-1}>2.0,\end{cases}$$

has two unstable sub systems but the whole system is stable. (ϕ_1=–5.0<0 and $0<\phi_1^{\sim}=0.9<1$.) This system has a limit cycle with a long period.

Now we will show that the system (3.1.1.2.4) can not move parallel to a single axis. Suppose after n steps transition $z_{n-1}^1 \in [r', r]$ for some integer n. Then Z_{n+1} can be obtained by the action of B on Z_n. Thus

$$Z_{n+1} = BZ_n = \begin{bmatrix} \phi_1' & \phi_2' \\ 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \phi_1' x + \phi_2' y \\ x \end{bmatrix}$$

where $Z_n = \begin{bmatrix} x \\ y \end{bmatrix}$. It is clear that

$$Z_{n+2} = \begin{cases} AZ_{n+1} & \text{if } \phi_1' x + \phi_2' y < r' \\ BZ_{n+1} & \text{if } r' \le \phi_1' x + \phi_2' y \le r \\ CZ_{n+1} & \text{if } \phi_1' x + \phi_2' y > r . \end{cases}$$

Case (i) $Z_{n+2} = AZ_{n+1} = \begin{bmatrix} \phi_1 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1' x + \phi_2' y \\ x \end{bmatrix} = \begin{bmatrix} \phi_1(\phi_1' x + \phi_2' y) \\ \phi_1' x + \phi_2' y \end{bmatrix}$.

Case (ii) $Z_{n+2} = BZ_{n+1} = \begin{bmatrix} \phi_1' & \phi_2' \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1' x + \phi_2' y \\ x \end{bmatrix} = \begin{bmatrix} \phi_1'(\phi_1' x + \phi_2' y) + \phi_2' x \\ \phi_1' x + \phi_2' y \end{bmatrix}$,

and

Case (iii) $Z_{n+2} = CZ_{n+1} = \begin{bmatrix} \phi_1'' & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \phi_1' + \phi_2' y \\ x \end{bmatrix} = \begin{bmatrix} \phi_1''(\phi_1' x + \phi_2' y) \\ \phi_1' x + \phi_2' y \end{bmatrix}$.

In all cases the second component of Z_{n+2} is equal to the first component of Z_{n+1}. Therefore as $n \to \infty$, Z_n can not move parallel to one of the axis.

It may be useful to investigate the affect of changing the delay parameter on the stability conditions of system (3.1.1.2.4).

Consider

$$X_n = \begin{cases} \phi_1 X_{n-1} & \text{if } X_{n-2} < r' \\ \phi_1' X_{n-1} + \phi_2' X_{n-2} & \text{if } r' \le X_{n-2} \le r \\ \phi_1'' X_{n-1} & \text{if } X_{n-2} > r , \end{cases} \qquad (3.1.1.2.6)$$

where ϕ_1, ϕ_1', ϕ_2', ϕ_1'', r and r' are real constants. Using the state space representation we may have

$$Z_n = \begin{bmatrix} z_n^1 \\ z_n^2 \end{bmatrix} = \begin{bmatrix} X_n \\ X_{n-1} \end{bmatrix}, A = \begin{bmatrix} \phi_1 & 0 \\ 1 & 0 \end{bmatrix},$$

$$B = \begin{bmatrix} \phi_1 & \phi_2' \\ 1 & 0 \end{bmatrix}, C = \begin{bmatrix} \phi_1'' & 0 \\ 1 & 0 \end{bmatrix}.$$

Then the system (3.1.1.2.6) is equivalent to

$$Z_n = \begin{cases} AZ_{n-1} & \text{if } z_{n-1}^2 < r' \\ BZ_{n-1} & \text{if } r' \le z_{n-1}^2 \le r \\ CZ_{n-1} & \text{if } z_{n-1}^2 > r, \end{cases} \qquad (3.1.1.2.7)$$

$X_n = (1 \quad 0)Z_n$.

Define

$$U = \{(x,y) : y > r\}, \; M = \{(x,y) : r' \le y \le r\},$$
$$L = \{(x,y) : y < r'\}, \; Q_1 = \{(x,y) \in U : x > 0\},$$
$$Q_2 = \{(x,y) \in U : x < 0\}, \; Q_3 = \{(x,y) \in U : x = 0\},$$
$$Q_4 = \{(x,y) \in M\}, \; Q_5 = \{(x,y) \in L : x > 0\},$$
$$Q_6 = \{(x,y) \in L : x = 0\}, \; Q_7 = \{(x,y) \in L : x < 0\}.$$

For every point Z_n in Q_3 or Q_6, Z_{n+1} is obtained by the action of C or A on Z_n. Let $Z_n = \begin{bmatrix} 0 \\ y \end{bmatrix} \in M$, then $AZ_n = CZ_n = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. For each $Z_n \in U$, Z_{n+1} is obtained by the action of C on Z_n and similarly for every point $Z_n \in M$ and $Z_n \in L$, Z_{n+1} is obtained by the action of B and A on Z_n respectively.

Case 1 : Let $0 \le \phi_1$, $0 \le \phi_1''$, the action of C on U (Figs. 3.1.1.2.5a and 3.1.1.2.5.b) is given by

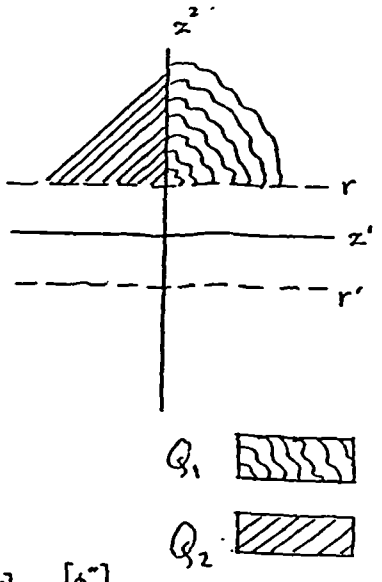


Fig. 3.1.1.2.5a

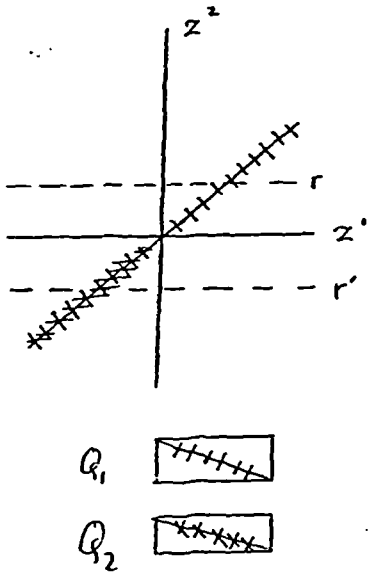


Fig. 3.1.1.2.5b

$$C\begin{bmatrix} x \\ y \end{bmatrix} = x\begin{bmatrix} \phi_1'' \\ 1 \end{bmatrix}$$

The action of A on L (Fig. 3.1.1.2.6a and Fig. 3.1.1.2.6b) can be represented as

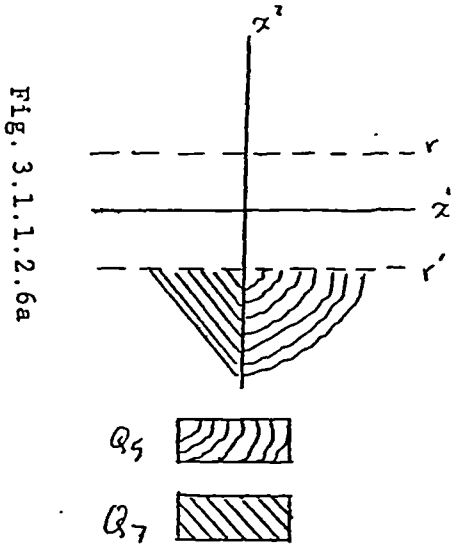


Fig. 3.1.1.2.6a

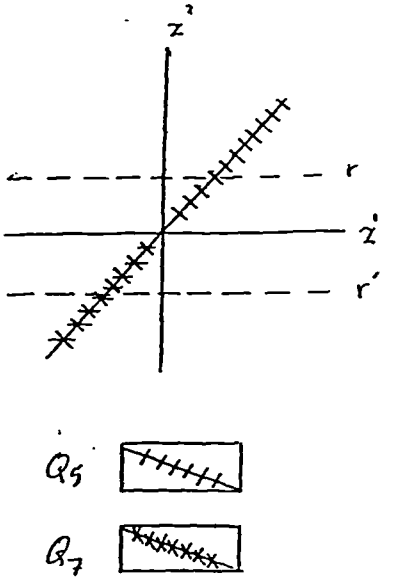


Fig. 3.1.1.2.6b

Let $Z_n \in Q_1$, then $Z_{n+1} = CZ_n \in Q_1$, therefore the system is stable if $|\phi_1''| \leq 1$. If $Z_n \in Q_2$, then $Z_{n+1} = CZ_n \in M$ or $CZ_n \in Q_7$. If $Z_{n+1} \in Q_7$, then $Z_{n+2} = AZ_{n+1} \in M$ or $\in Q_7$, thus the condition $|\phi_1''| \leq 1$ ensures the stability of the system. If $Z_n \in Q_5$, then $Z_{n+1} \in M$ or Q_1 . More generally, by starting from any point, Z_n under (finite number) action of A or C or both along the line (ϕ_1 , 1) or (ϕ_1'' , 1) will lie in M, and then under action B may move to the other region or to stay in M. Thus, in this case the system is stable if $0 \leq \phi_1 \leq 1$, $0 \leq \phi_1'' \leq 1$.

Case 2 : $0 \leq \phi_1$, $\phi_1'' < 0$, the actions of A on U and C on L can be represented in the same way as the Case 1.

By starting from Q_1, the system under action C may move to Q_2 or M. If the system lies in M, then under action B may stay in M or move to any other region. If the system lies in Q_2, then under action C will move to Q_5 or M. If it moves to Q_5, then under action A it will move to Q_1 or M. It can be seen that the important tail of trajectories takes the form (A,C,C). Therefore the condition $\phi_1 \phi_1''^2 \leq 1$ (the eigenvalue of matrix AC^2) ensures the stability of the system.

Similarly if $\phi_1 < 0$, $0 \leq \phi_1''$, then $\phi_1'' \phi_1^2 \leq 1$ is the stability condition.

Case 3 : $\phi_1 < 0$, $\phi_1'' < 0$. Similar to case 1 the system is stable if $\phi_1 \phi_1'' \leq 1$.

Lemma 3.1.1.2.2: Let Z_n satisfy (3.1.1.2.6). Then $\{Z_n\}$ is Lagrange stable if and only if one of the following conditions holds:

i) $0 < \phi_1 \leq 1$, $0 < \phi_1'' \leq 1$,

ii) ($\phi_1 = 0$, $|\phi_1''| \leq 1$) or ($\phi_1'' = 0$, $|\phi_1| \leq 1$),

iii) $\phi_1 < 0$, $0 \leq \phi_1''$ and $\phi_1^2 \phi_1'' \leq 1$,

iv) $\phi_1'' < 0$, $0 \leq \phi_1$ and $\phi_1'^2 \phi_1 \leq 1$,

v) $\phi_1 < 0$, $\phi_1'' < 0$ and $\phi_1 \phi_1'' \leq 1$.

Example (3.1.1.2.2): The system

$$X_n = \begin{cases} -1.5X_{n-1} & \text{if } X_{n-2} < -1.0 \\ -8.0X_{n-1} + 3.0X_{n-2} & \text{if } -1.0 \leq X_{n-2} \leq 2.0 \\ 0.4X_{n-1} & \text{if } X_{n-2} > 2.0 , \end{cases}$$

is stable with a limit cycle with a long period . In fact this system satisfies the condition (iii) of lemma 3.1.1.2.2.

The system

$$X_n = \begin{cases} -1.5X_{n-1} & \text{if } X_{n-2} < -1.0 \\ 0.4X_{n-1} - 0.6X_{n-2} & \text{if } -1.0 \leq X_{n-2} \leq 2.0 \\ 0.6X_{n-1} & \text{if } X_{n-2} > 2.0 , \end{cases}$$

is not stable since $\phi_1 \phi_1'^2 = 1.35 > 1$.

The results of Lemma 3.1.1.2.1 and 3.1.1.2.2 can be easily generalized to a more general case.

Lemma 3.1.1.2.3: The system

$$X_n = a_o^{(i)} + \sum_{j=1}^{p} a_j^{(i)} X_{n-j} \quad \text{if } r_{i-1} < X_{n-d} \leq r_i ,$$

where $\{-\infty = r_o < r_1 < \cdots < r_{l-1} < r_l = \infty\}$ is an ordered partition of R, $d \leq p$, $a_o^{(i)}$, $a_j^{(i)}$ are real constant, is Lagrange stable if $\sum_{j=1}^{p} |a_j^{(1)}|$ and $\sum_{j=1}^{p} |a_j^{(l-1)}| < 1$.

In fact these two conditions control the behaviour of the system in the two extreme regions and push it towards the origin.

Example (3.1.1.2.3) The system

$$X_n = \begin{cases} 0.2+0.3X_{n-1}-0.1X_{n-2}+0.3X_{n-3} & \text{if } X_{n-3} < -50.0 \\ 2.0+1.5X_{n-1}-4.0X_{n-2}+3.0X_{n-3} & \text{if } -50.0 \le X_{n-3} \le 100.0 \\ -0.1+0.2X_{n-1}-0.3X_{n-2}+0.3X_{n-3} & \text{if } X_{n-3} > 100.0 \end{cases}$$

is stable with possibly a chaotic behaviour. Fig. 3.1.1.2.7 shows the scatter plot of X_n vs X_{n-1} of this system.

3.2. Bilinear Models

Granger and Andersen (1978) and Subba Rao (1981) proposed a special class of non-linear models known as bilinear time series models. This type of models has been extensively discussed in control theory to describe input-output relationships for a deterministic non-linear system (See for example Mohler 1973).

The general bilinear autoregressive moving average model of order (p,q,P,Q) (abbreviated as BL(p,q,P,Q) by Subba Rao and Gabr(1984) is given by

$$X_t = \sum_{i=1}^{P} \alpha_i X_{t-i} + \sum_{j=0}^{q} \beta_j e_{t-j} + \sum_{k=1}^{Q}\sum_{l=1}^{P} \gamma_{kl} X_{t-l} e_{t-k} , \tag{3.2.1}$$

where $\{e_t\}$ are independently, identically distributed random variables with mean zero and variance σ_e^2 and $\beta_o = 1$. The completely bilinear model is given by (3.2.1) with $p = q = 0$, so that

$$X_t = \sum_{k=1}^{Q}\sum_{l=1}^{P} \gamma_{kl} X_{t-l} e_{t-k} + e_t . \tag{3.2.2}$$

If $\gamma_{kl} = 0$ for all $k > l$, the model is called *superdiagonal* and if $\gamma_{kl} = 0$ for all $k \neq l$, the model is said to be *diagonal*. Finally if $\gamma_{kl}=0$ for $k < l$, the model is called *subdiagonal*. Stationarity and invertibility of simple bilinear models has been discussed by Quinn (1982) and Tuan and Tran (1981). Gabr and Subba Rao (1981) have discussed the estimation and prediction of subset bilinear time series models and Liu (1985) has studied theoretical properties of some more general bilinear models BL$(p,q,r,1)$. Wang, An and Tong (1983) have studied the distribution of simple stationary bilinear processes. Subba Rao and Gabr (1984) gives a comprehensive account of this class of models.

Let X_t be a discrete parameter time series satisfying the difference equation

$$X_t = \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} + (\beta_o + \beta_1 X_{t-1} + \cdots + \beta_q X_{t-q}) e_t \,, \tag{3.2.3}$$

where $\{e_t\}$ is a sequence of independent and identically distributed random variables with $E(e_t)=0$ and $Var(e_t)=\sigma_e^2$ and $\beta_o = 1$. The model (3.2.3) is a special case of the bilinear model, which we will refer to as *Markov Bilinear Stochastic Process*. It is more convenient to study the properties of the model (3.2.3), such as stationarity, invertibility, etc, via an equivalent state-space representation.

The vector form of the model (3.2.3) can be written as

$$\begin{aligned} \mathbf{X}_t &= \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{X}_{t-1} e_t + \mathbf{C} e_t \\ X_t &= \mathbf{H}' \mathbf{X}_t \,, \end{aligned} \tag{3.2.4}$$

where

$$\mathbf{A} = \begin{bmatrix} a_1 & a_2 & a_3 & \cdots & a_p \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} = \left[\begin{array}{c|c} a_1 & a_2 \cdots a_p \\ \hline 0 & \\ \vdots & I \\ 0 & \end{array}\right],$$

$$\mathbf{B} = \begin{bmatrix} b_1 & b_2 & \cdots & b_q & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ . & . & \cdots & . & \cdots & . \\ . & . & \cdots & . & \cdots & . \\ . & . & \cdots & . & \cdots & . \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} = \left[\begin{array}{ccc|c} b_1 & b_2 & \cdots\ b_q & 0 \\ \hline 0 & & & 0 \end{array}\right],$$

$$\mathbf{X}_t = (X_t, X_{t-1}, \cdots, X_{t-p+1})' \,, \ \mathbf{C} = (1,0,...,0)' \,,$$

and $\mathbf{H} = (1,0,\cdots,0)'$. We also assume that $e_t \sim N(0,\sigma_e^2)$.

Let

$$\begin{aligned} \mu_t &= E(\mathbf{X}_t) \,, \\ \mathbf{V}_t &= E(\mathbf{X}_t \mathbf{X}_t') \,. \end{aligned}$$

Taking expectation on both sides of

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{B}\mathbf{X}_{t-1} e_t + \mathbf{C} e_t \,,$$

one can get

$$\mu_t = \mathbf{A}\mu_{t-1} .$$

Hence a sufficient condition for $\{X_t\}$ be asymptotically stationary of the first order (i.e $E(X_t)$ is independent of t as $t \to \infty$) is

$$\rho(\mathbf{A}) = \max_{1 \le i \le p} \{\lambda_i(\mathbf{A})\} < 1 , \tag{3.2.5}$$

where $\lambda_1(\mathbf{A}), \cdots, \lambda_p(\mathbf{A})$ are the eigenvalues of the matrix **A**. If this condition is satisfied then

$$\mu = \lim_{t \to \infty} \mu_t = \mathbf{0} .$$

Note that the condition (3.2.5) is the same as the stationarity condition of AR part of the model (3.2.3).

Following Subba Rao (1984) and Neudecker (1969), the sufficient condition for the model (3.1.3) to be second order stationary is

$$\rho[\mathbf{A} \otimes \mathbf{A} + \mathbf{B} \otimes \mathbf{B} \sigma_e^2] < 1 ,$$

where $\mathbf{A} \otimes \mathbf{A}$ is the *kronecker* product which is of order $p^2 \times p^2$.

The covariance of X_t and X_{t-k} can be calculated as

$$\gamma_k = E(X_t, X_{t-k}) = \alpha_1 \gamma_{k-1} + \cdots + \alpha_p \gamma_{k-p} , \tag{3.2.6}$$

which is identical with the covariance of AR(p) part of the model (3.2.3).

Multiplying both sides of equation (3.2.3) by X_t and taking expectation, one can get

$$E(X_t^2) = \alpha_1 \gamma_1 + \cdots + \alpha_p \gamma_p + E(X_t e_t) + \beta_1 E(X_t X_{t-1} e_t) + \cdots + \beta_q E(X_t X_{t-q} e_t) ,$$

or

$$E(X_t^2) = \sum_{i=1}^{p} \alpha_i \gamma_i + E(X_t e_t) + \sum_{j=1}^{q} \beta_j E(X_t X_{t-j} e_t) .$$

It can be shown that

$$E(X_t e_t) = \sigma_e^2 ,$$
$$E(X_t X_{t-1} e_t) = \sigma_e^2 (\beta_1 \gamma_o + \beta_2 \gamma_1 + \cdots + \beta_q \gamma_{q-1}) ,$$
$$E(X_t X_{t-2} e_t) = \sigma_e^2 (\beta_1 \gamma_1 + \beta_2 \gamma_o + \cdots + \beta_q \gamma_{q-2}) ,$$
$$\vdots$$
$$E(X_t X_{t-q} e_t) = \sigma_e^2 (\beta_1 \gamma_{q-1} + \beta_2 \gamma_{q-2} + \cdots + \beta_q \gamma_o) .$$

Hence

$$\begin{aligned} \gamma_o = {} & \alpha_1 \gamma_1 + \cdots + \alpha_p \gamma_p + \sigma_e^2 + \sigma_e^2 [\beta_1 (\beta_1 \gamma_o + \beta_2 \gamma_1 + \cdots + \beta_q \gamma_{q-1}) \\ & + \beta_2 (\beta_1 \gamma_1 + \beta_2 \gamma_o + \cdots + \beta_q \gamma_{q-2}) + \cdots + \beta_q (\beta_1 \gamma_{q-1} + \beta_2 \gamma_{q-2} + \cdots + \beta_q \gamma_o)] \\ = {} & \alpha_1 \gamma_1 + \cdots + \alpha_p \gamma_p + \sigma_e^2 \gamma_o (\beta_1^2 + \beta_2^2 + \cdots + \beta_q^2) \end{aligned}$$

$$+\sigma_e^2\beta_1(\beta_2\gamma_1+\beta_3\gamma_2+\cdots+\beta_q\gamma_{q-1})+\cdots+$$
$$\sigma_e^2\beta_q(\beta_1\gamma_{q-1}+\beta_2\gamma_{q-2}+\cdots+\beta_{q-1}\gamma_1)$$
$$=\frac{\sum_{i=1}^{p}\alpha_i\gamma_i+\sigma_e^2\beta_1(\beta_2\gamma_1+\cdots+\beta_q\gamma_{q-1})+\cdots+\sigma_e^2\beta_q(\beta_1\gamma_{q-1}+\cdots+\beta_{q-1}\gamma_1)}{1-\sigma_e^2(\beta_1^2+\beta_2^2+\cdots+\beta_q^2)}.$$

As the autocovariance function of X_t is equal to autocovariance function of AR part of the model, therefore the spectrum of X_t is equal to the spectrum of AR part.

3.3. Exponential Autoregressive Models

An exponential autoregressive model of order p EXPAR(p) for a univariate time series $\{X_t, t=0,\pm1,\cdots\}$, which was first announced in Ozaki and Oda (1978), takes the form

$$X_t=\sum_{j=1}^{p}A_{t,j}X_{t-j}+e_t\,, \tag{3.3.1}$$

where $\{e_t\}$ is a sequence of independent identically distributed random variables with zero mean and constant variance σ_e^2 and e_t independent of X_s , $s<t$. Here

$$A_{t,j}=b_j+c_j\exp(-gX_{t-1}^2)\,,$$

b_j's , c_j's and $g\geq0$, being constants. Ozaki (1982) gives a comprehensive account of the development of EXPAR models.

The physical idea seems to be that the impulse response $\{A_{t,j}\}$, at time t or equivalently the instantaneous frequency response, is controlled by the rise and fall of X_{t-1}^2, interpreted as energy at time $t-1$. (See Tong (1987a)).

The model (3.3.1) has a symmetric distribution about the origin, which can be easily deduced from Theorem 1, Tong (1983) page 16. The model (3.3.1) was extended by Ozaki (1982) to

$$X_t=(\alpha_1+f_1(X_{t-1})e^{-X_{t-1}^2})X_{t-1}+\cdots+(\alpha_p+f_p(X_{t-1})e^{-X_{t-1}^2})X_{t-p}+e_t \tag{3.3.2}$$

with the object of giving a more sophisticated specification of the dynamics of the characteristic roots of AR model using Hermite type polynomials $f_i(x_{t-1})e^{-x_{t-1}^2}$ $(i=1,2,...,p)$ where

$$f_i(x_{t-1})=\pi_o^{(i)}+\pi_1^{(i)}x_{t-1}+\cdots+\pi_{r_i}x_{t-1}^{r_i}\quad(i=1,2,...,p)\,. \tag{3.3.3}$$

The limitation of symmetry distribution of model (3.3.1) may be obviated if a model of type (3.3.2) is employed where the order r_i for some i of the Hermite polynomials $f_i(x_{t-1})$ are odd integers.

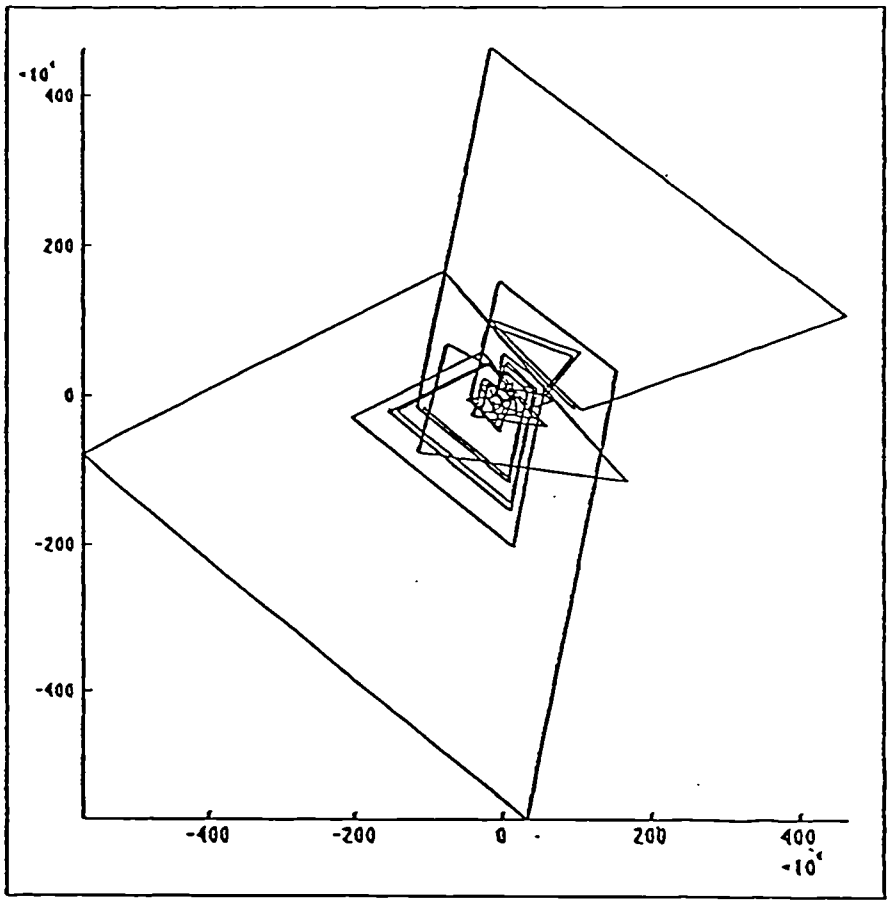


Fig. 3.1.1.2.7 : Scatter diagram of example 3.1.1.2.3. X_n (vertical axis) vs X_{n-1} (horizonal axis) for n=1500,...,1600, with X_1=X_2=X_3=0.

CHAPTER FOUR

SOME STRUCTURAL PROPERTIES OF LINEAR AND NONLINEAR MODELS

To be able to use time series models in practice, we must be able to fit the models to data and estimate the parameters. Computational procedures for determining parameters for various non-linear model classes are studied by many authors. Often these are based on a least squares or a maximum likelihood type criterion.

The method of conditional least squares (CLS) seems to be popular among users of SETAR models, EXPAR models and more generally non-linear autoregressive models appropriately parametrized. The method is essentially one of minimizing the sum of squares Σe_t^2 . The paper by Kilmko and Nelson (1978) and Tjostheim (1986) provide most of the theoretical framework for sampling properties of the estimates. For SETAR models, strong consistency Chan (1987), asymptotic normality and rates of convergence of parameters estimates are available (See e.g. Chan and Tong (1986a)).

Let $\{X_t ; t = 1,2,...\}$ be a stochastic process defined in a probability space $(\Omega, \mathrm{F}, P_\theta)$ whose distribution depends on a (column) vector $\theta = (\theta_1, \cdots, \theta_p)'$ of unknown parameters, which lies in some open set A of the Euclidean p-space. The following notations are adapted:

$\theta^o = (\theta_1^o, \cdots, \theta_p^o)$ denotes the 'true' value of θ ;

$E_\theta(.)$ denotes the expectation where θ is the parameter ;

$E_\theta(. \mid .)$ denotes the conditional expectation with θ being the parameter ;

$\{\mathrm{F}\}_{t=0}^{\infty}$ denotes the sequence of sub-sigma fields with F_t generated by an arbitrary subset of $\{X_1, X_2, \cdots, X_t\}$, $t \geq 1$, and with F_o equal to the trivial sigma field.

Assume that $E|X_t| < \infty$, t=1,2,.... Define the function $g(.,.)$ by

$$g(\theta, \mathbf{F}_t) = E_\theta(X_{t+1} \mid \mathbf{F}_t), \quad t \geq 0 .$$

Given a set of observations X_t, t=1,2,...,N, we minimize the conditional sum of squares

$$S_N(\theta) = \sum_{t=0}^{N-1} [X_{t+1} - g(\theta, \mathbf{F}_t)]^2 , \tag{4.1}$$

with respect to θ to obtain the estimates. That is we solve the 'least-squares' equation

$$\frac{\partial S_N(\theta)}{\partial \theta_i} = 0 , \; i=1,2,...,p .$$

It can be shown that for non-linear autoregressive models, maximizing the conditional likelihood function is equivalent to minimizing the conditional sum of squares of errors, under the normality assumption of e_t.

The error sum of squares function $S_N(\theta)$ is a function of the elements of θ only; the data provide the numerical coefficients in $S_N(\theta)$ and these are fixed for many specific estimation problem. In the *parameter space*, that is, the p-dimensional Euclidean space of $\theta_1, \cdots, \theta_p$, the function $S_N(\theta)$ can be represented by the contours of a surface.

If the $g(\theta, \mathbf{F}_t)$ were linear in the θ's, the surface contours would be ellipsoidal and would have a single local (and so a single global) minimum height, $S_N(\hat{\theta})$, at the location defined by the least-square estimator $\hat{\theta}$. The error terms are assumed to be normal.

If the $g(\theta, \mathbf{F}_t)$ is non-linear, the contours may not be ellipsoidal but tend to be irregular and some times 'banana-shaped', perhaps with several local minima and perhaps with more than one global minimum, that is, the minimum height may be obtained at more than one θ-location.

The precise shape and orientation of the $S_N(\theta)$ contours depend on the model and the data. When the contours surrounding the least-squares estimator $\hat{\theta}$ are greatly elongated, and many possible θ-value are 'nearly as good as' $\hat{\theta}$ in the sense that their $S_N(\theta)$ bowl height values are close to $S_N(\hat{\theta})$, the problem is said to be ill-conditioned and $\hat{\theta}$ may be difficult to obtain computationally.

Two obvious ways of examining the error sum of squares surface $S_N(\theta)$ are often overlooked; they can be particularly useful when an iterative procedure (most non-linear time series), beginning from chosen values, does not satisfactorily converge.

The first of these is to select a grid of points in the space of the parameters $(\theta_1,\theta_2,\cdots,\theta_p)$ and to evaluate the error sum of squares function at every point of the grid. These values will provide some idea of the form of the error sum of squares surface and may reveal, for example, that multiple minima are possible. In any case, the grid points at which the smallest error sum of squares is found can be used as the starting point of an iterative parameter estimation procedure, or a reduced grid can be examined in the best neighbourhood, to obtain a better starting point.

The second possibility is to draw the error sum of squares contours in any particular region of the parameter space in which difficulty in convergence occurs or in which additional information would be helpful. This is usually straightforward when only one or two parameters are involved. When there are more than two parameters, two-dimensional slices of the contours can be obtained for selected values of all but two of the parameters, and a composite picture can be built up.

All iterative procedures require initial values $\theta_{1o},\theta_{2o},\cdots,\theta_{po}$, of the parameters $\theta_1,\theta_2,\cdots,\theta_p$, to be selected. All available prior information should be used to make these starting values as reliable as they possibly can be. Also if multiple minima exist or if there are several local minima in addition to an absolute minimum, poor starting values may result in convergence to an unwanted stationary point of the error sum of squares surface. This unwanted point may have parameter values which are physically impossible or which do not provide true minimum value of $S_N(\theta)$. As suggested above, a preliminary evaluation of $S_N(\theta)$ at a number of grid points in the parameter space is often useful.

4.1. Moving Average

In the maximum likelihood estimation of a moving average process non-invertible estimates frequently appear, both with actual data and in simulation studies. Kang (1975) showed how noninvertibility occurs in the moving average of order one and indicated why it should be expected with positive probability. The positive probability that an estimated moving average process be non-invertible is studied for maximum likelihood estimation of a univariate process by Anderson (1986).

4.1.1. Exact Maximum Likelihood Estimation of MA(1)

Let $\{X_t\}$ be a stochastic process defined by

$$X_t = \theta e_{t-1} + e_t \,, \; t = 0, \pm 1, \cdots \,, \qquad 4.1.1.1$$

where $\{e_t\}$ is a sequence of unobservable random variables with the properties

$$E(e_t) = 0 \,, \; V(e_t) = \sigma_e^2 < \infty \,, \; E(e_t e_s) = 0 \; t \neq s \,.$$

If $|\theta| < 1$, we can invert (4.1.1.1) to obtain the autoregressive representation of order infinity

$$e_t = X_t - \theta X_{t-1} + \theta^2 X_{t-2} + \cdots \,. \qquad 4.1.1.2$$

The autoregressive representation is important because it may be used for prediction.

It is easy to show that the lag one autocorrelation of process (4.1.1.1) is

$$\rho = \frac{\theta}{1+\theta^2} \,.$$

The other autocorrelations are zero. If θ is replaced by its reciprocal,

$$\rho = \frac{\frac{1}{\theta}}{1+(\frac{1}{\theta})^2} = \frac{\theta}{1+\theta^2} \,,$$

the autocorrelation is unchanged. The exact likelihood function of the parameters (θ, σ_e^2) in the first order moving average model (4.1.1.1), where $e_t \sim N(0, \sigma_e^2)$ is

$$L(\theta, \sigma_e^2 \mid \mathbf{x}) = (2\pi\sigma_e^2)^{-\frac{1}{2}N} \mid \Omega_\theta \mid^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma_e^2} \mathbf{x}' \Omega_\theta^{-1} \mathbf{x}) \,,$$

where $\mathbf{x}' = (X_1, X_2, \cdots, X_N)$ and Ω_θ is an $N \times N$ band matrix with elements $w_{ii} = 1+\theta^2$, $w_{i,i+1} = w_{i+1,1} = -\theta$ and $w_{ij} = 0$, $i \neq j \pm 1$. (Cryer and Ledolter, 1981). Maximizing over σ_e^2 one can obtain the concentrated likelihood function

$$L(\theta \mid \mathbf{x}) \propto \{g(\theta)\}^{-\frac{1}{2}N} \,,$$

where

$$g(\theta) = \mid \Omega_\theta \mid^{\frac{1}{N}} \mathbf{x}' \Omega_\theta^{-1} \mathbf{x} \,.$$

The matrix Ω_θ may be represented as

$$\Omega_\theta = \begin{bmatrix} 1+\theta^2 & & & & & & & & & \\ -\theta & 1+\theta^2 & & & & & & & & \\ 0 & -\theta & 1+\theta^2 & & & & & & & \\ 0 & 0 & -\theta & 1+\theta^2 & & & & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\theta & 1+\theta^2 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1+\theta^2 \end{bmatrix}$$

One can check that

$$\Omega_\theta = \theta^2 \Omega_{1/\theta} .$$

Hence

$$\begin{aligned} g(\frac{1}{\theta}) &= |\Omega_{1/\theta}|^{\frac{1}{N}} \mathbf{x}'\Omega_{1/\theta}^{-1}\mathbf{x} \\ &= |\theta^{-2}\Omega_\theta|^{\frac{1}{N}} \mathbf{x}'[\theta^{-2}\Omega_\theta]^{-1}\mathbf{x} \\ &= |\Omega_\theta| \mathbf{x}'\Omega_\theta\mathbf{x} \\ &= g(\theta) . \end{aligned}$$

(Chan private communication).

It shows that $g(\theta)$ has two equal local minimum on $\mathbf{R}$, at $\hat{\theta}$ and $\frac{1}{\hat{\theta}}$. This is related to the two identical MA(1) processes with parameters θ and $\frac{1}{\theta}$.

For N=2 the maximum likelihood estimates can be found by minimizing

$$g(\theta) = (1+\theta^2+\theta^4)^{-\frac{1}{2}} \{(1+\theta^2)(x_1^2+x_2^2)+2\theta x_1 x_2\} .$$

The graph of $g(\theta)$ is given by Fig. 4.1.1.1.

4.1.2. Conditional Maximum Likelihood Estimation

Box and Jenkins (1970) have developed a simple procedure for evaluating an approximate likelihood function numerically by first determining the $\{e_t\}$ recursively, as follows. Given observations $X_1, X_2, \cdots, X_N$ from the model (4.1.1.1) we may first set (as starting value) $e_o = 0$. We then have,

$$e_1 = X_1$$
$$e_2 = X_2 - \theta X_1$$
$$e_3 = X_3 - \theta e_2 = X_3 - \theta X_2 + \theta^2 X_1$$
$$\vdots$$
$$e_n = X_n - \theta e_{n-1} .$$

Under the assumption that the $e_t's$ are normal, the conditional likelihood estimates are thus obtained by minimizing $S_N(\theta) = \sum_{t=1}^{N} e_t^2$, and may therefore regarded also as conditional least squares estimates. Note, however, that $S_N(\theta)$ is no longer a quadratic function of the parameters.

For $N=3$, $S_N(\theta)$ is a polynomial of degree four and its derivative $S_3'(\theta)$ is a polynomial of degree 3 in θ. The cubic polynomial $S_3'(\theta)$ may have three real roots. Suppose that $\mu_1 < \mu_2 < \mu_3$ denote the real roots. Since the coefficient of θ^4 in $S_3(\theta)$ is positive, $S_3(\theta)$ has two local minimum at $\hat{\theta}_1 = \mu_1$ and $\hat{\theta}_2 = \mu_3$. Unfortunately the exact relation between $S_3(\hat{\theta}_1)$ and $S_3(\hat{\theta}_3)$ is not available. The graph of $S_3(\theta)$ for $X_1 = -1$, $X_2 = 1.2$ and $X_3 = 1.6$ is given by Fig. (4.1.2.1).

4.2. Exponential Autoregressive Models

Consider the family of Exponential Autoregressive (EXPAR) Models

$$X_t = \sum_{j=1}^{p} \{\alpha_j + f_j(X_{t-1}) \exp(-\gamma X_{t-1}^2)\} X_{t-j} + e_t , \qquad (4.2.1)$$

where $\alpha's$ are real constants and $0 \le \gamma \le \infty$, $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean zero and constant variance, and e_t is independent of X_s, $s < t$ and

$$f_j(x) = \sum_{i=1}^{r_j} \pi_i^{(j)} x ,$$

where $\pi_i^{(j)}'s$ are real constants.

The simultaneous estimation of parameters $\{p, \gamma, (\alpha_i, \pi_i^{(j)})\}$,i=1,2,...,p, and j=1,2,...,r_i, is essentially a nonlinear optimization procedure, involving all the computational difficulties inherent in such a procedure. This problem may be overcome by fixing the parameter γ at one of a grid of values and estimating the order, p, and the corresponding α_i, π_i parameters. The problem then becomes one of fitting a linear regression of X_t on the series $\{X_s ; s < t\}$ and $\{e^{-\gamma X_{t-1}^2} X_s ; s < t\}$. The order p can be estimated by using AIC or other criterion for non-linear time series. The models fitted for each γ may then also be compared

using one of the criterion (AIC), to find the 'best' model over all γ. The above method was proposed by Haggan and Ozaki (1981), who assumed that $\{X_t\}$ is a mean-deleted series.

In model (4.2.1) it is obvious that if $\gamma=\infty$, the model is linear. Thus, the parameter γ is a key factor affecting the non-linearity of the model and we would naturally pay special attention to its estimation. Our experience, and we understand also of Dr. Valerie Haggan (now Dr. Valerie Haggan - Ozaki) - private communication with Professor Howell Tong, strongly suggests that the estimation of γ can encounter difficulties, Chan et al (1988).

We may illustrate the type of difficulties by reference to the likelihood function based on 100 observations artificially generated by the following model

$$X_t = \{b+(c+dX_{t-1})\ exp(-gX_{t-1}^2)\}X_{t-1}+e_t\ , \qquad (4.2.2)$$

where $e_t \sim N(0,1)$, $b=0.6$, $c=-0.6$, $d=g=1$. In our simulation studies, we use NAG routine G05CBF(4) to initialize the random number generator G05DDF to generate a fixed sequence of normal random numbers with mean zero and variance equal to one. The value of X_1 is set to be zero and the first 1499 observations of X_t are discarded to remove the affect of the starting point. The contours of negative log likelihood function, which is proportional to $S_N(\theta)=\Sigma e_t^2$, of model (4.2.2) are given by Figs. 4.2.1 - 4.2.3 . The contours of $(b,c),(b,d),(c,d)$ are ellipsoidal, for fixed $g=1$, and it confirms that when the scale parameter g is known, the parameters can be estimated efficiently.

A family of contours of (d,g) when $b=0.6$, $c=-0.6$ over different grids of points are plotted to show the difficulties of estimation of parameters. An isometric projection of $S_N(\theta)$ is also included, Fig. 4.2.7 . As it can be seen, most of the curves are open from above along the g axis, and near the minimum, they are elongated. On considering the height of curves, the flatness of likelihood surface is quite clear. Fig. 4.2.6 reveals the existence of two local minima.

The likelihood surface of (g,d) in Figs. 4.2.6 reveals the difficulty in the elongated shape of the contours. Shown in Table (4.2.1) is the theoretical information matrix of the four parameters based on numerical integration with respect to the stationary distribution of the model which may be obtained by numerical techniques. (Details will be given in chapter five).

Table 4.2.1

(Model 4.2.2)

Information Matrix

	b	c	d	g
b	1.5588			
c	0.1861	0.0802		
d	0.0568	0.0142	0.0516	
g	0.0210	0.0158	–0.0464	0.0764

The eigenvalues of information matrix are 0.0098, 0.0603, 0.1125, and 1.5844 .

The eigenvalues range from 0.0098 to 1.5844 (ratio of maximum to minimum $\doteq$ 161.67) and the determinant is approximately 1.05×10^{-4}, which clearly high light the near-singularity of the information matrix. The inverse of information matrix is given by Table 4.2.2 .

Table 4.2.2

(Model 4.2.1)

The Inverse of Information Matrix

0.8988			
–1.8952	22.9870		
–0.7408	–17.7297	62.8325	
–0.3035	–15.0124	42.0293	41.8091

The variances of b , c , d , g are 0.8988, 22.9870, 62.8325, 41.8091 and the covariance of (c , d), (c , g) and (d , g) are -17.7297, -15.0124 and 42.0293 respectively.

The likelihood surface of (g,d) shown in Figs. 4.2.5 and 4.2.6 reveals the difficulty in the elongated shape of the contours.

The general message is that corresponding to some non-linear time series models, the information matrix may be ill-conditioned. We now explain that the ill-conditioning is closely connected with the relative variation of π_g as g varies over [0,∞] , where π_g denotes the stationary distribution with parameter g . For simplicity of discussion, we consider a simple example.

EXAMPLE 4.2.1: Let $\{X_t\}$ satisfy the EXPAR model

$$X_t = T_g(X_{t-1}) + \varepsilon_t \ , \tag{4.2.3}$$

where $\varepsilon_t \sim N(0,1)$, and

$$T_g(x) = \{0.2 - 0.2exp(-gx^2)\}x \ .$$

It is clear that T_g is odd and increasing for $g \geq 0$ and $|T_g(x)| \leq |T_{g'}(x)|$ for $0 \leq g \leq g' \leq +\infty$, $\forall x \in \mathbf{R}$. For any two probability measures μ and ν on $\mathbf{R}$, we write

$$\mu \lhd \nu \ \textit{iff} \ \forall c > 0 \ , \ \mu[-c, c] \geq \nu[-c, c] \ ,$$

i.e. μ is more concentrated around the origin than ν.

Now, using Proposition 2 of Hõgnãs (1986, p. 207), we may deduce that $g \leq g'$ implies that $\pi_g \lhd \pi_{g'}$, where π_g denotes the stationary distribution of model (4.2.3) with g as the parameter. For $g = 0$, $X_t = e_t$ and $g = \infty$, $X_t = 0.2X_{t-1} + e_t$. Hence, for $g \in [0, +\infty]$, it holds that

$$N(0,1) = \pi_o \lhd \pi_g \lhd \pi_{+\infty} = N(0, 0.96^{-1}) \ .$$

Thus, the infinite range of g only corresponds to a very small variation of the stationary distributions. This lies at the heart of the ill-condition of the information matrix.

We approximate a family of stationary marginal probability density functions of model (4.2.3) and (4.2.2) over the finite range g, $g \in [0, 20]$ and $[0.5, 15]$ respectively, to show that the large range variation of g only corresponds to a small variation of the stationary distribution. The results are given in Fig. 4.2.8 and Fig. 4.2.9 .

Let

$$X_t = \{a + be^{-gX_{t-1}^2}\}X_{t-1} + e_t \ ,$$

where $e_t \sim N(0, 1)$, $a = 0.2$, $b = -0.2$ and $g = 2.5$. The condition $|a| + |b| < 1$, ensure the asymptotic normality of $(\hat{a}, \hat{b}, \hat{g})$. Tjostheim (1986). The information matrix, $E(H(\theta))$, is calculated by employing the numerical integration technique, where $\theta = (a, b, g)'$, $H(\theta) = \sigma^{-2}\left[\frac{\partial^2 S_N(\theta)}{\partial\theta_i \partial\theta_j}\right]$, $i, j = 1,2,3$,

$$S_N(\theta) = \sum_{t=2}^{N} (X_t - aX_{t-1} - bX_{t-1}e^{-gX_{t-1}^2})^2 \ .$$

The information matrix, its eigenvalues and the inverse of information matrix are given by

Information Matrix

	a	b	g
a	2.0744		
b	0.1348	0.0541	
g	0.0136	0.0030	0.0003

The eigenvalues of the information matrix are 0.0001, 0.0452, and 2.0834 . We test the validity of the above by simulation with a random sample of size 10000 . The eigenvalues of the simulated information matrix are 0.0001, 0.0452, and 2.0796 .

The Inverse of Information Matrix

0.6860		
0.0342	41.4955	
−31.4385	−416.5032	8923.5768

The determinant $\doteq 9.41\times10^{-6}$ and the eigenvalues range from 0.0001 to 2.0834 (ratio of maximum to minimum $\doteq 20834$, which high light the near-singularity of the information matrix. The entries of the inverse of information matrix clearly show that the variances of $\hat{b}$ and $\hat{g}$ are substationally large and g is highly correlated with b.

The above analysis then suggests that we should examine the likelihood function of EXPAR models fitted to real data.

EXAMPLE 4.2.2 (An EXPAR model for the Canadian lynx data).

Ozaki (1982, Equ. 3.1) has reported the following model,

$$X_t = \{0.138+(0.316+0.982X_{t-1})e^{-\hat{\gamma}X_{t-1}^2}\}X_{t-1} - \{0.437+(0.659+1.260X_{t-1})e^{-\hat{\gamma}X_{t-1}^2}\}X_{t-2}+\varepsilon_t \ , \qquad (4.2.4)$$

where $\varepsilon_t \sim N(0,\hat{\sigma}^2)$, $\hat{\gamma}=3.89$, $\hat{\sigma}^2=4.327 \times 10^{-2}$, and $\{X_t\}$ are the mean deleted logarithmically transformed lynx data. The parameter estimates were supposed to be obtained by the method of least squares i.e. by minimizing $\Sigma\varepsilon_t^2$. Checking the above parameter estimates, we have discovered that there must be a misprint in the recorded value of 0.138 because with the above stated parameter values, the nor-

malized residual sum of squares differs substantially from the value of 4.327 x 10^{-2} given for $\hat{\sigma}^2$. In fact, our non-linear optimization gives 1.167 in lieu of 0.138 with all other parameters estimates in agreement with those given in (4.2.4). We therefore study the revised EXPAR model

$$X_t = \{\underset{(0.216)}{1.167}+(\underset{(0.486)}{0.316}+\underset{(1.152)}{0.982}X_{t-1})e^{-\hat{\gamma}X_{t-1}^2}\}X_{t-1}$$
$$-\{\underset{(0.268)}{0.437}+(\underset{(0.360)}{0.659}+\underset{(1.146)}{1.260}X_{t-1})e^{-\hat{\gamma}X_{t-1}^2}\}X_{t-2}+\varepsilon_t \ , \qquad (4.2.5)$$

where $\varepsilon_t \sim N(0,\hat{\sigma}^2)$, $\hat{\gamma}=\underset{(3.68)}{3.89}$, $\hat{\sigma}^2 = 4.327 \text{ x } 10^{-2}$. The misprint might explain the peculiar second order moments of model (4.2.4) reported by Lim (1987). Taking model (4.2.5) as the true model, a sample of 10000 observations is generated by using a simulation technique and the sample autocovariance function of the lynx data (1821-1934) and the simulated data are given in Fig. 4.2.10 . Using model (4.2.5) and repeating the same simulation technique leading to the likelihood function shown in Figs. 4.2.11-4.2.14, we may trace the likelihood function of the following model

$$X_t = \{a_o+(a_1+a_2X_{t-1})e^{-gX_{t-1}^2}\}X_{t-1}$$
$$+\{b_o+(b_1+b_2X_{t-1})e^{-gX_{t-1}^2}\}X_{t-2}+\varepsilon_t \ , \qquad (4.2.6)$$

where $\varepsilon_t \sim N(0,\sigma^2)$.

Fig 4.2.12 clearly reveals similar difficulty for model (4.2.5) as that experienced by model (4.2.2) and shown in Fig. 4.2.6 . Shown in Table (4.2.3) is the theoretical information matrix of the parameters of model (4.2.6) based on numerical integration with respect to the stationary distribution of the model which may be obtained by numerical techniques. The eigenvalues range from 0.0029 to 14.2046 . The determinant is approximately 4.2×10^{-5} and the ratio of the maximum eigenvalue to the minimum eigenvalue is 4831.49

Table 4.2.3

(EXPAR MODEL FOR LYNX DATA)

Information Matrix $\times\hat{\sigma}^2$

	a_o	a_1	a_2	b_o	b_1	b_2	g
a_o	7.6136						
a_1	1.2113	0.5055					
a_2	–0.0238	–0.0161	0.0880				
b_o	5.9868	0.9862	–0.0067	7.6136			
b_1	0.9862	0.4218	–0.1200	2.2380	1.4088		
b_2	–0.0067	–0.0120	0.0714	–0.0257	–0.0246	0.1209	
g	0.0818	0.0205	0.0007	0.1575	0.0554	0.0132	0.0102

The eigenvalues of the information matrix are 0.0029, 0.0333, 0.1340, 0.1771, 0.5682, 2.2406 and 14.2046

Table 4.2.4

The Inverse of Information Matrix $\div\hat{\sigma}^2$

1.0812						
–1.6790	5.4589					
1.6233	–1.6337	30.6559				
–1.1926	1.6008	–2.5461	1.6539			
1.2808	–2.5112	1.1916	–1.6981	3.0016		
–1.9960	2.5199	–24.1376	3.1008	–1.3201	30.3614	
8.6096	–11.7149	52.0432	–13.7589	6.3107	–67.1161	312.7660

Not surprisingly, Table 4.2.3 reveals the near-singularity of the information matrix associated with the parameters a's , b's and g. Similarly to example 4.2.1, the parameters a_1 , a_2 , b_1 , b_2 , and g have a large variance and the scale parameter g is highly correlated with the others.

It should also be mentioned that EXPAR models of the form (4.2.1) give in general $E(X_t)\neq 0$, unless all the $f_j's$ are even functionals and e_t has a symmetric distribution. (See e.g. Pemberton and Tong, 1981). Therefore, it violate the least squares principle to estimate the unknown parameters by minimizing Σe_t^2

when the data are mean deleted. However, fortunately for the present example, the affect does not seem very serious and we shall therefore ignore it here and in the next section.

4.2.1. Reparametrization

With the benefit of hindsight, perhaps we should have anticipated problems because of the 'exponential regression' nature of the EXPAR models. Analogy may be made with ordinary regression involving exponential type functions (See, e.g., Draper and Smith, 1981, p. 489). The analogy then naturally suggests reparametrization as a way out of the difficulty. Specifically, we may replace $exp(-gX_{t-1}^2)$ by $exp(-g(X_{t-1}^2-\mu_2'))$, c by $\tilde{c}=c\ exp(-g\mu_2')$ and d by $\tilde{d}=d\ exp(-g\mu_2')$ where

$$\mu_2' = EX_{t-1}^2 \ .$$

We shall refer to the above replacement by the term 'centering' and the reparametrized model as the 'centered model'. Typically μ_2' is estimated by the obvious second sample moment.

Comparing Fig. 4.2.1.1-4.2.1.4 for the centered version of model (4.2.2) with Figs. 4.2.1-4.2.6 ,we can see substantial improvement. The estimate of g is now much closer to the true value. ($\overline{X^2}=1.643$)

Table 4.2.1.1

(Model 4.2.2 centered)

Information Matrix

	b	$\tilde{c}$	$\tilde{d}$	g
b	1.5281			
$\tilde{c}$	0.8586	1.7164		
$\tilde{d}$	0.2563	0.2987	1.0973	
g	0.0223	0.0750	−0.2139	0.0765

The eigenvalues of the information matrix are 0.0227, 0.7601, 1.0456 and 2.5899 .

Table 4.2.1.2

(Model 4.2.2 centered)

The Inverse of Information Matrix

0.9217			
–0.4197	1.0798		
–0.1607	–0.8314	2.9510	
–0.3060	–3.2607	9.1112	41.8258

The visual improvement is confirmed by the calculations recorded in Table 4.2.1.2 . Now the determinant has increased to 4.67×10^{-2} and the eigenvalues range from 0.0277 to 2.5899 (ratio $\doteq$114.09). Note that the information matrix of the centered model is calculated by a simulation method based on a simulated data of size 10000 .

Substantial improvement may also be gained by centering the EXPAR model (4.2.5) for the Canadian lynx data as can be seen by comparing Figs. 4.2.11-4.2.12 with Figs. 4.2.1.5 and Table 4.2.4 with Table 4.2.1.4 . ($\overline{X^2}=0.30908$, and typically $\tilde{a}_1=a_1\ exp(-g\mu_2')$.)

Table 4.2.1.3

(CENTERED EXPAR MODEL FOR LYNX DATA)

Information Matrix $\times\hat{\sigma}^2$

	a_o	$\tilde{a}_1$	$\tilde{a}_2$	b_o	$\tilde{b}_1$	$\tilde{b}_2$	g
a_o	8.0011						
$\tilde{a}_1$	4.6131	7.2820					
$\tilde{a}_2$	–0.1157	–0.2301	1.2814				
b_o	6.3353	3.7912	–0.0501	8.0024			
$\tilde{b}_1$	3.7912	6.1543	–0.1917	8.6580	20.7964		
$\tilde{b}_2$	–0.0501	–0.1917	1.0474	–0.1340	–0.4225	1.8025	
g	0.0849	0.0786	0.0025	0.1628	0.2153	0.0518	0.0106

The eigenvalues of information matrix are 0.0033, 0.4563, 0.5278, 2.5940, 4.0232, 9.2703 and 30.3014. The determinant $\doteq$ 2.33 .

Table 4.2.1.4

(CENTERED EXPAR MODEL FOR LYNX DATA)

The Inverse of Information Matrix $\div\hat{\sigma}^2$

1.0177						
−0.4142	0.3659					
0.4095	−0.1094	2.0841				
−1.1174	0.3927	−0.6319	1.5372			
0.3123	−0.1642	0.0694	−0.4094	0.1949		
−0.4955	0.1643	−1.6236	0.7658	−0.0738	2.0323	
8.0807	−2.8764	13.2925	−12.8756	1.3912	−17.0938	301.5600

Note, in particular, the substantional reduction in the Var(typically $\tilde{a}_1$) and the covariances of estimates. The information matrix of the centered model is obtained by using the simulation method.

A family of stationary marginal probability density functions of model (4.2.5) over the finite range of g , $g\in[0.5\ ,\ 7]$ is approximated by numerical integration technique and the result is given in Fig. 4.2.1.6 . It can be seen that for large variation of g, there are a small change in the stationary distribution.

4.3. Bilinear Models

A discrete parameter time series $\{X_t\}$ that satisfies the difference equation

$$X_t+\sum_{i=1}^{P} a_iX_{t-i}+a_o = e_t+\sum_{l=1}^{P}\sum_{k=1}^{Q} b_{lk}X_{t-l}e_{t-k}\ , \tag{4.3.1}$$

where $\{e_t\}$ are independent and each e_t is distributed $N(0\ ,\ \sigma_e^2)$, is said to be a bilinear process of order $BL(p\ ,0\ ,P\ ,Q)$. Proceeding as in Subba Rao (1981), we can show that maximizing the likelihood function of $(X_{m_1},X_{m_1+1},\cdots,X_N)$, with $m_1=p+1$, p being the maximum order of the AR model fitted to the data, is the same as minimizing the function $S_N(\theta)=\sum_{t=m_1}^{N} e_t^2$ with respect to the parameters. The maximization of the likelihood is then typically performed numerically either by the Newton Raphson algorithm or any other method. This approach is adopted in bilinear models. e.g. See Subba Rao and Gabr (1984).

The parameter estimates so obtained are consistent (op. cit). However, the asymptotic distribution of the estimates is not available. One difficulty may be due to the lack of moment property of bilinear models,

e.g. See Tong (1987a) and Granger and Andersen (1978, page 40). Another difficulty may be intimately connected with the invertibility problem of bilinear models, e.g. See Tjostheim (1986, page 259). The invertibility problem, except for some rather special cases, seems to be very complicated.

The exact likelihood function of any class of bilinear models does not seem to be possible to obtained, therefore we consider the conditional likelihood function of the simple diagonal bilinear model of the form

$$X_t = bX_{t-1}e_{t-1}+e_t\ ,\ t=0,\pm1,\dots\ , \tag{4.3.2}$$

where $e_t \sim N(0\ ,\ \sigma_e^2)$. Since maximizing the likelihood function is equivalent to minimizing the function $S_N(b)=\Sigma e_e^2$, we may adopt the following method:

Given the observations $X_1, X_2, \cdots, X_N$, from the model (4.3.2), we first set $e_o=0$, then we have

$$\begin{aligned} e_1 &= X_1 \\ e_2 &= X_2 - bX_1e_1 \\ &= X_2 - bX_1^2 \\ &\vdots \\ e_n &= X_N - bX_{N-1}e_{N-1} \end{aligned}$$

and $S_N(b)=\sum_{t=1}^{N} e_t^2$. Obviously $S_N(b)$ is not a quadratic function in b . Similarly to MA(1), suppose that 3 observations are available, the function $S_N(b)$ is given by

$$S_N(b)=\sum_{t=1}^{3} e_e^2 = X_1^2+(X_2-bX_1^2)^2+(X_3-bX_2^2+b^2X_1^2X_2)^2\ ,$$

which is a polynomial of degree four in b and under some conditions on X_1 , X_2 , X_3 , $S_N(b)$ has three real roots, such that the smallest and the largest one of which correspond to the two local minima. Although the simulation study show that $\hat{b}_1$ is inside the invertibility region and $\hat{b}_2$ outside, the exact relation between $\hat{b}_1$ and $\hat{b}_2$ is not clear. When $N>3$, $S_N(b)$ is a polynomial of degree greater than four in b, therefore $S_N(b)$ may have several local minima. Shown in Figs. 4.3.1-4.3.12 are the negative of the likelihood function of diagonal and super diagonal bilinear models. Some simulation study show that when b the parameter of the model is inside the invertibility region, the likelihood function has a unique minima, but when the simulated model is not invertible, the likelihood function has more than one local minimum.

As the asymptotic distribution of the estimates of bilinear models is not available, we employed the Monte Carlo method to investigate the effect of the variance of the noise term on the variance of estimates. For both the diagonal bilinear model

$$X_t = -0.4X_{t-1}e_{t-1} + e_t ,$$

and the linear AR(1) process

$$X_t = -0.4X_{t-1} + e_t ,$$

where $e_t \sim N(0, \sigma_e^2)$, 100 replications each of length 50 and 100 of length 100 were generated using NAG routine G05DDF, which was initialized by NAG routine G05CCF. For each replication, start-up values were set to zero and the first 1499 observations discarded to remove transient effects. For each realization the error sum of squares $S_N(\theta)$ is calculated and an estimate of b is found by searching over a grid of points for which the bilinear model is stationary and invertible (in the bilinear case) and the linear model is stationary (in the linear case). The results are given in Table 4.3.1

Table 4.3.1

σ_e	Linear model				Bilinear model			
	N=50		N=100		N=50		N=100	
	$E(\hat{b})$	$V(\hat{b})$	$E(\hat{b})$	$V(\hat{b})$	$E(\hat{b})$	$V(\hat{b})$	$E(\hat{b})$	$V(\hat{b})$
1.0	-0.404	0.0173	-0.390	0.0088	-0.383	0.0077	-0.389	0.0027
0.80	-0.364	0.0186	-0.392	0.0113	-0.380	0.0130	-0.391	0.0052
0.60	-0.416	0.0168	-0.390	0.0091	-0.423	0.0190	-0.403	0.0119
0.40	-0.405	0.0143	-0.385	0.0080	-0.342	0.0353	-0.397	0.0228
0.20	-0.371	0.0215	-0.394	0.0074	-0.402	0.1823	-0.414	0.1082
0.10	-0.408	0.0195	-0.401	0.0087	-0.350	0.6448	-0.369	0.3505
0.05	-0.420	0.0152	-0.407	0.0068	-0.303	2.1609	-0.322	1.2671
0.01	-0.330	0.0174	-0.390	0.0073	-0.841	23.9600	-0.236	20.2834

The general message is that decreasing the variance of error term e_t leads to an increasing of the variance of the estimates of the bilinear term.

Let $\{X_t\}$ be a discrete parameter time series satisfying the difference equation

$$X_t=\sum_{i=1}^{p} a_i X_{t-i}+\sum_{j=1}^{q} b_j X_{t-j}e_t+e_t\,, \tag{4.3.3}$$

where the e_t's are independent and identically distributed $N(0\,,\,\sigma_e^2)$. The method of conditional least squares can not be applied for estimating the parameters of the model (4.3.3), because some of the parameters are absent in the expression of the conditional mean of the observations. Since the e_t's are normal, we may use the conditional maximum likelihood method to estimate the parameters $a_1\,,\,a_2\,,\,\cdots\,,\,a_p$, $b_1\,,\,b_2\,,\,\cdots\,,\,b_q$ and σ_e^2.

It may be shown that the Jacobian of the transformation is given by $\left|\frac{\partial e_t}{\partial X_s}\right| = |J|$ where

$$J=\begin{bmatrix} \frac{\partial e_{k+1}}{\partial X_{k+1}} & 0 & 0 & \cdots & 0 \\ \frac{\partial e_{k+2}}{\partial X_{k+1}} & \frac{\partial e_{k+2}}{\partial X_{k+2}} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{\partial e_N}{\partial X_{k+1}} & \frac{\partial e_N}{\partial X_{k+2}} & \frac{\partial e_N}{\partial X_{k+3}} & \cdots & \frac{\partial e_N}{\partial X_N} \end{bmatrix} = \prod_{t=k+1}^{N} \frac{\partial e_t}{\partial X_t}$$

$$= \prod_{t=k+1}^{N} \frac{1}{|\,1+b_1X_{t-1}+\cdots+b_qX_{t-q}\,|}$$

and $k=\max\{p\,,\,q\}$. The joint probability density function of $X_{k+1},\ldots,X_N$ given $X_1,\cdots,X_k$ may be written as

$$f(X_{k+1},\ldots,X_N|X_1,\cdots,X_k)=\left(\frac{1}{\sigma_e\sqrt{2\pi}}\right)^{N-k}|J|\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=k+1}^{N}\left(\frac{X_t-\sum_{i=1}^{p}a_iX_{t-i}}{1+\sum_{j=1}^{q}b_jX_{t-j}}\right)^2\right).$$

For given σ_e^2, maximizing the likelihood function is equivalent to minimizing the function

$$S_N(\theta)=\sum_{t=k+1}^{N} 2\sigma_e^2 \ln|1+b_1X_{t-1}+\cdots+b_qX_{t-q}|+\left(\frac{X_t-\sum_{i=1}^{p}a_iX_{t-i}}{1+\sum_{j=1}^{q}b_jX_{t-j}}\right)^2 .$$

The maximum likelihood estimate of σ_e^2 is given by $\hat{\sigma}_e^2=\frac{1}{N-k}S_N(\hat{\theta})$, where $\hat{\theta}$ is the maximum likelihood estimate of $\theta=(a_1,\cdots,b_1,\cdots,b_q)'$. Some examples are given below to illustrate the effect of σ_e^2 on the variance of the estimates of bilinear parameters.

Example 1 : Consider the simple bilinear model

$$X_t=(1+bX_{t-1})e_t\ ,\ t=0,\pm1,\cdots\ ,$$

where $e_t\sim N(0,\sigma_e^2)$. The log likelihood function of this model is given by

$$L(b)=\frac{1-N}{2}\ln\sigma_e^2+(1-N)\ln\sqrt{2\pi}-\sum_{t=2}^{N}\ln|1+bX_{t-1}|-\frac{1}{2\sigma_e^2}\sum_{t=2}^{N}\left(\frac{X_t}{1+bX_{t-1}}\right)^2 .$$

Following Basawa and Scott (1980), in general $E(\xi_N(b))=I_N(b)$, where $I_N(b)$ is the usual (mean) Fisher information and

$$\xi_N(b)=\sum_{k=1}^{N}E\left\{\left(\frac{d\log L_k(b)}{db}-\frac{d\log L_{k-1}(b)}{db}\right)^2|\mathbf{F}_{k-1}\right\},$$

where $L_k(b)$ is the likelihood function based on the observations $X_1,\cdots,X_k$ and $\mathbf{F}_k=\sigma(X_1,\cdots,X_k)$, with $\mathbf{F}_o$ the trivial σ-field.

Here,

$$\begin{aligned}\frac{d\log L_k(b)}{db}-\frac{d\log L_{k-1}(b)}{db}&=-\sum_{t=2}^{k}\frac{X_{t-1}}{1+bX_{t-1}}+\frac{1}{\sigma_e^2}\sum_{t=2}^{k}\frac{e_t^2X_{t-1}}{1+bX_{t-1}}\\&\quad+\sum_{t=2}^{k-1}\frac{X_{t-1}}{1+bX_{t-1}}-\frac{1}{\sigma_e^2}\sum_{t=2}^{k-1}\frac{X_{t-1}}{1+bX_{t-1}}e_t^2\\&=-\frac{X_{k-1}}{1+bX_{k-1}}+\frac{e_k^2}{\sigma_e^2}\cdot\frac{X_{k-1}}{1+bX_{k-1}}\ ,\end{aligned}$$

and

$$\begin{aligned}\xi_N(b)&=\sum_{k=2}^{N}E\left\{\frac{X_{k-1}}{1+bX_{k-1}}\left(\frac{e_k^2}{\sigma_e^2}-1\right)|\mathbf{F}_{k-1}\right\}^2\\&=\sum_{k=2}^{N}\left(\frac{X_{k-1}}{1+bX_{k-1}}\right)^2E\left(\frac{e_k^2}{\sigma_e^2}-1\right)^2\\&=2\sum_{k=2}^{N}\left(\frac{X_{k-1}}{1+bX_{k-1}}\right)^2 .\end{aligned}$$

Thus

$$I_N(b) = E(\xi_N(b)) = 2\sum_{k=2}^{N} E(\frac{X_{k-1}}{1+bX_{k-1}})^2$$

$$= 2(N-1)E(\frac{X_{t-1}}{1+bX_{t-1}})^2 .$$

Here

$$X_t = (1+bX_{t-1})e_t \text{ , } all\ t .$$

Now $e_t^2 = (\frac{X_{t-1}}{1+bX_{t-1}})^2$ and $I_N(b) = 2(N-1)\sigma_e^2$, $I_N^{-1}(b) = \frac{1}{2(N-1)\sigma_e^2}$. For fixed σ_e^2, the estimate of b , $\hat{b}$, is efficient but, as $\sigma_e^2 \to 0$ the Fisher information $I_N(b)$ tends to zero and $V(\hat{b}) \to \infty$.

Example 2: Let

$$X_t = aX_{t-1} + bX_{t-1}e_t + e_t \text{ , } t=0,\pm1,\ldots, \quad e_t \sim N(0, \sigma_e^2) .$$

The log likelihood function is proportional to

$$L(a, b) \propto \sum_{t=2}^{N} \ln|1+bX_{t-1}| - \frac{1}{2\sigma_e^2}\sum_{t=2}^{N}\left(\frac{X_t - aX_{t-1}}{1+bX_{t-1}}\right)^2 .$$

and it is not difficult to show that the inverse of the information matrix Z, has the following form

$$Z = \begin{bmatrix} \frac{N-1}{\sigma_e^2}E(\frac{X_{t-1}}{1+bX_{t-1}})^2 & 0 \\ 0 & 2(N-1)E(\frac{X_{t-1}}{1+bX_{t-1}})^2 \end{bmatrix}^{-1}$$

$$= \frac{1}{N-1} \cdot \frac{\sigma_e^2}{E(\frac{X_{t-1}}{1+bX_{t-1}})^2} \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2\sigma_e^2} \end{bmatrix}$$

Tong (1981) has proved that $E(X_t^n)$ does not exist for all integer n if $b \neq 0$. However, for small b we may approximate the function $\frac{X_{t-1}}{1+bX_{t-1}}$ by its Taylor expansion about zero. Using this approximation, we can show that

$$E(\frac{X_{t-1}}{1+bX_{t-1}})^2 \approx E(X_t^2) = \sigma_X^2 = \frac{\sigma_e^2}{1-a^2-b^2\sigma_e^2} .$$

Substituting the above approximation, the information matrix Z reduces to

$$Z \approx \frac{1-a^2-b^2\sigma_e^2}{2(N-1)} \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{\sigma_e^2} \end{bmatrix} .$$

Obviously $\hat{a}$ and $\hat{b}$ are uncorrelated and $V(\hat{b}) \to \infty$ as $\sigma_e^2 \to 0$.

In order to examine the relation between the variance of the parameter estimates and σ_e^2 by simulation, we consider the bilinear models

$$X_t = a_1 X_{t-1} + (1 + bX_{t-1})e_t \,, \; a_1 = -0.3 \,, \qquad \text{Model I}$$

and

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + (1 + bX_{t-1})e_t \,, \; a_1 = 0.7 \,, \; a_2 = -0.6 \,, \; a_3 = 0.3 \,, \qquad \text{Model II}$$

where $e_t \sim N(0, \sigma_e^2)$ and $b = 0.4$. For estimating the variance of the parameters, 100 replications each of length 250 using a simulation technique are generated. For each realization, the simplex method is employed to maximize the log likelihood function. The sample mean and sample variance are taken as approximations of the mean and variance of the parameter estimates. The results are given in Table 4.3.2

Table 4.3.2

Model I

σ_e	$E(\hat{a}_1)$	$V(\hat{a}_1)$	$E(\hat{b})$	$V(\hat{b})$
1.0	-0.299	0.0009	0.391	0.0003
0.1	-0.281	0.0036	0.304	0.1303
0.01	-0.314	0.0036	0.213	11.8762
0.001	-0.313	0.0037	-4.239	1093.8762

Model II

σ_e	$E(\hat{a}_1)$	$V(\hat{a}_1)$	$E(\hat{a}_2)$	$V(\hat{a}_2)$	$E(\hat{a}_3)$	$V(\hat{a}_3)$	$E(\hat{b})$	$V(\hat{b})$
1.0	0.696	0.0017	-0.600	0.0004	0.298	0.0008	0.380	0.0015
0.1	0.695	0.0031	-0.607	0.0038	0.297	0.0034	0.292	0.1050
0.01	0.694	0.0029	-0.603	0.0039	0.295	0.0034	0.721	9.6659

These results reveal the conflicting aspect of bilinearity and RSS: if you want bilinearity, you must allow RSS to be of reasonable size.

4.3.1. Invertibility

It is well known that one way to obtain a unique identification of linear Gaussian moving average model is by imposing the invertibility condition. It is also well known that the condition has a bearing on prediction. The major source of difficulties is the complexity of the likelihood function, which tends to have several local maxima.

The situation with a more complicated case like a bilinear time series model is expected to be even more complex. Consider the univariate class of time series models which takes the form

$$X_t = f(X_{t-j}, e_{t-j}, j=1,2,...,p) + e_t \,. \qquad (4.3.1.1)$$

where e_t is pure white noise, so that e_t and e_s are independent for all $t \neq s$, and is an unobserved input to the system. If this type of model is to be used for forecasting and provided the function $f(.)$ does contain some e_{t-j}, $j \geq 1$, it is necessary to be able to estimate the e_t sequence from the observed X_t's. A natural way of doing this is to assume values for as many initial, or starting up e's as are necessary, i.e. $\hat{e}_{-j}$, $j=0,1,...,(p-1)$, say. Assuming now that X_{-j}, $j=0,1,\cdots,(p-1)$ are also observed, then e_1 can be estimated directly from the assumed known generating function (4.3.1.1) by

$$\hat{e}_1 = X_1 - f(X_{1-j}, \hat{e}_{1-j}, j=1,\cdots,p)\,,$$

and then $\hat{e}_2$ is estimated by

$$\hat{e}_2 = X_2 - f(X_{2-j}, \hat{e}_{2-j}\ ,\ j=1,2,...,p)\ ,$$

and so forth. Thus, using the assumed starting up values, an iterative procedure can be used to estimate all $\hat{e}_t$, based on the observed $X_t's$ and the generating model. The error that will almost certainly arise from such a procedure will be denoted $\varepsilon_t = e_t - \hat{e}_t$.

The model (4.3.1.1) is said to be invertible if

$$E(\varepsilon_t^2) \to 0 \ \ as\ t \to \infty\ , \tag{4.3.1.2}$$

which assumed that the correct generating formula is known completely, both in form and the actual parameter values.

If the parameter values in the model are not known exactly, as would occur if they were estimated from an earlier finite length of data for instance, (4.3.1.2) might be replaced by

$$E(\varepsilon_t^2) \to c\ ,\ as\ \ t \to \infty\ .$$

where $c < \infty$ is some constant.

It would therefore seem always prudent to check the invertibility of bilinear models fitted to real data, especially if prediction is part of the principal purpose of the exercise. We adopt the definition of invertibility introduced by Granger and Anderson (1978, P. 69). We propose the following practical procedures for checking invertibility of non-linear time series models because analytic conditions of invertibility are usually impractical except for the simplest situations. To describe the method, we consider the model (4.3.1.1) which has been fitted to data $X_1, \cdots, X_N$ conditional on $X_o = X_{-1} = \cdots = X_{-p} = e_o = e_{-1} = \cdots = e_{-p} = 0$. We can *extend* the sample by using the fitted model and a simulated random sample of $\varepsilon_{n+1}, \varepsilon_{n+2}, \cdots, \varepsilon_{n+m}$ (m say equal 1000) from the fitted distribution of e_t. Replacing e_t by $\hat{\varepsilon}_t$ in (4.3.1.1)) and setting $\hat{\varepsilon}_o = \hat{\varepsilon}_{-1} = \cdots = \hat{\varepsilon}_{-p} = 0$, we may obtain $\hat{\varepsilon}_t$, $t=1,2,...,n+m$, where $\hat{\varepsilon}_j = e_j$ for $j=-p,...,n$. Thus, a practical check for invertibility is to calculate the sample mean of $(\varepsilon_t - \hat{\varepsilon}_t)^2$. Explosive tendency suggests non-invertibility.

Figs. 4.3.1.1 and 4.3.1.2 suggest that the bilinear models fitted by Gabr & Subba Rao (1981, equations 5.3 and 5.7) to the Wolf's sunspot numbers and the Canadian lynx data are likely to be non-invertible. As a check on this practical method, it has been applied to the linear model which is fitted to lynx data by Gabr and Rao (1981). The result is given in Fig. 4.3.1.3 . Since there is no theory to tell us how to restrict

our search over surfaces like that shown in Fig. 4.5.8 to invertible bilinear models only, one possibility is to adopt a trial-and-error method. However, this may involve horrendous calculations.

4.4. SETAR Models

Chan (1987) has shown that under suitable regularity conditions, the least squares estimate of a stationary ergodic Threshold Autoregressive model are strongly consistent. In particular, both the threshold and the delay parameters can be consistently estimated by the method of least squares. Chan (1988) has proved that for SETAR(2;p,p) the asymptotic distribution of $\hat{r}$ is related to a *Compound Poisson Process*. Also he has shown that for all sufficiently large N, $\hat{d}_N = d$ with probability one. The sampling properties of the other parameters are studied by Tong (1983), Chan and Tong (1984) and (1986a).

To assess some properties of the conditional error sum of squares, $S_N(\theta)$, of the class of SETAR models, we consider the simple SETAR(2;1,1) model

$$X_t = \begin{cases} a_o + a_1 X_{t-1} + e_t & \text{if } X_{t-1} \le r \\ b_o + b_1 X_{t-1} + e_t & \text{if } X_{t-1} > r \end{cases}, \qquad (4.4.1)$$

where $e_t \sim N(0, 1)$, $a_o = 0.7$, $a_1 = -0.5$, $b_o = -1.8$, $b_1 = 0.7$ and $r = 0$. Since the estimation of threshold is quite important and moreover, SETAR is linear in each piece, therefore we only plot the contour of the surface of $S_N(\theta)$ for a_1 and b_1 versus r. The result is given in Fig. 4.4.1 and Fig. 4.4.2 . An isometric projection of $S_N(\theta)$ is also given in Fig. 4.4.3 . As it can be seen, there exist a dip around the $\hat{r}$ and $\hat{a}_1$ in the isometric projection. Also all contour curves from contour key 78 to 145 are closed and ellipsoidal. In the next section we will study the function $S_N(\theta)$ of SETAR models which are fitted to real data in more details.

4.5. Real Data

In previous sections we have described some difficult aspects of non-linear time series modelling. This section considers some univariate non-linear models which have been applied to Canadian lynx data. A collection of contours of the likelihood surface of each model are obtained to reveal some features of the fitted models likelihood functions.

Tong (1983) has reported two SETAR models which were fitted for two different periods to the log transformation of the Canadian lynx data. The first SETAR(2;5,2) model, which will be referred as SETAR

model I, is fitted to the first 100 observations. This model is given by

$$X_t = \begin{cases} a_o + \sum_{i=1}^{5} a_i X_{t-i} + e_t^{(1)} & \text{if } X_{t-2} \le r \\ b_o + \sum_{i=1}^{2} b_i X_{t-i} + e_t^{(2)} & \text{if } X_{t-2} > r \end{cases} \qquad \text{(Model I)}$$

where

$$a_o = 0.77\,,\ a_1 = 1.06\,,\ a_2 = -0.20\,,\ a_3 = 0.16\,,\ a_4 = -0.43\,,\ a_5 = 0.18\,,$$
$$b_o = 2.25\,,\ b_1 = 1.47\,,\ b_2 = -1.20\,,\ r = 3.05\,.$$

The second model, SETAR(2;7,2), which is fitted to 114 observations and will be referred as model II, takes the form

$$X_t = \begin{cases} a_o + \sum_{i=1}^{7} a_i X_{t-i} + e_t^{(1)} & \text{if } X_{t-2} \le r \\ b_o + \sum_{i=1}^{2} b_i X_{t-i} + e_t^{(2)} & \text{if } X_{t-2} > r \end{cases} \qquad \text{(Model II)}$$

where

$$a_o = 0.546\,,\ a_1 = 1.032\,,\ a_2 = -0.173\,,\ a_3 = 0.171\,,\ a_4 = -0.431\,,\ a_5 = 0.332\,,$$
$$a_6 = -0.284\,,\ a_7 = 0.210\,,\ b_o = 2.632\,,\ b_1 = 1.492\,,\ b_2 = -1.324\,,\ b_3\,,\ r = 3.116\,.$$

Gabr and Subba Rao (1981) have fitted the following subset bilinear model to the log transformation of the first 100 observations of lynx data. Their fitted model may be written as

$$\begin{aligned} X_t + a_o + a_1 X_{t-1} + a_2 X_{t-2} + a_3 X_{t-3} + a_4 X_{t-4} + a_9 X_{t-9} + a_{12} X_{t-12} \\ = b_{39} X_{t-3} e_{t-9} + b_{99} X_{t-9} e_{t-9} + b_{62} X_{t-6} e_{t-2} + b_{11} X_{t-1} e_{-1} \\ + b_{27} X_{t-2} e_{t-7} + b_{42} X_{t-4} e_{t-2} + e_t\,, \end{aligned}$$

where

$$a_o = -1.486292\,,\ a_1 = -0.77227\,,\ a_2 = 0.091527\,,\ a_3 = -0.083073\,,\ a_4 = 0.261493\,,$$
$$a_9 = -0.225585\,,\ a_{12} = 0.245841\,,\ b_{39} = -0.7893\,,\ b_{99} = 0.4798\,,\ b_{62} = 0.3902\,,\ b_{11} = 0.1326\,,$$
$$b_{27} = 0.07944\,,\ b_{42} = -0.3212\,.$$

Results are give in Figs. 4.5.1-4.5.28 .

As far as the experiment goes, it seems that the estimation of threshold is not difficult and since in practice only a finite nominations for parameter r exist, therefore it is easy to obtain the least square estimation of threshold. The contour plots of the likelihood function of SETAR models which were fitted to lynx

data, do not exhibit any unexpected pattern in the likelihood function of those fitted models.

The contour plots of the bilinear model which was fitted to lynx data reveal the complexity and difficulty of estimation of coefficient of bilinear terms and the existence of more than one local maxima in the likelihood function. In the contour plot of $(\hat{b}_{99}, \hat{b}_{11})$, it seems that the global minimum is near point (0.8,0.7) whereas the maximum likelihood estimates of these two parameters which are reported by Gabr and Subba Rao (1981), are $\hat{b}_{99}=0.4798$ and $\hat{b}_{11}=0.1326$ which correspond to another local minima. Since the estimation procedure which has been used by Gabr and Subba Rao (1981, page 159) is based on the Newton Raphson iterative method with initial estimates of the coefficient of bilinear terms equal to zero, therefore it is not surprising that the iterative procedure converges to a local minimum nearest to this initial point. There thus is no guaranteey of a global minimum. Similar problems arise in estimation of coefficients of the other bilinear terms. In practice it seems that the likelihood plots may give some vital information about the fitted model which might ultimately lead to an improvement .

likelihood function of MA(1) with N=3.

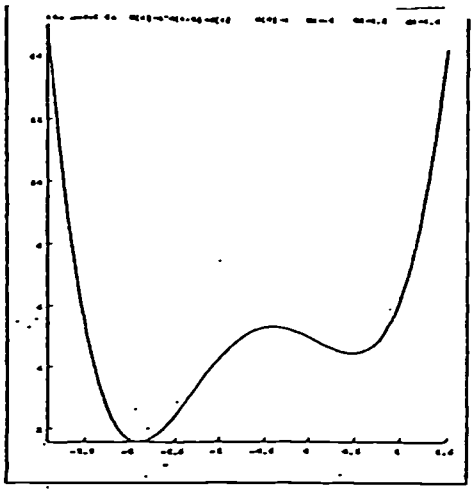


Fig. 4.1.2.1

Exact likelihood function of MA(1) with N=2.

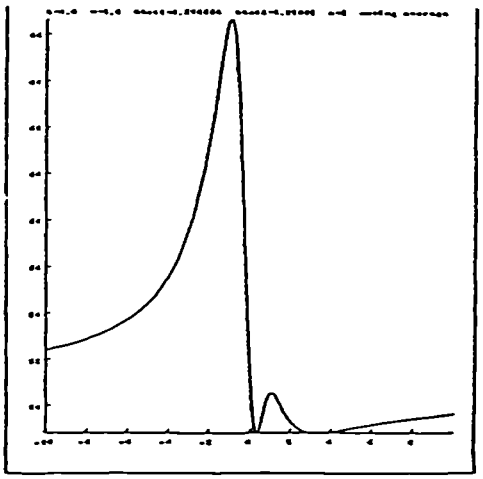
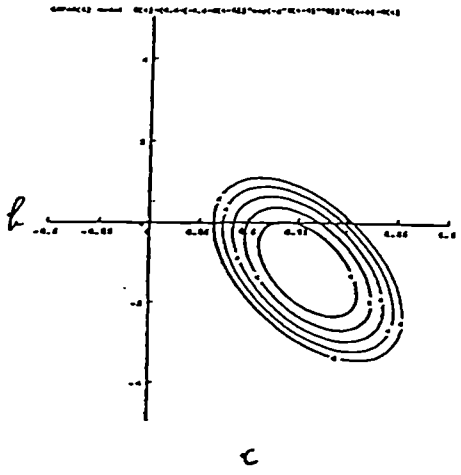
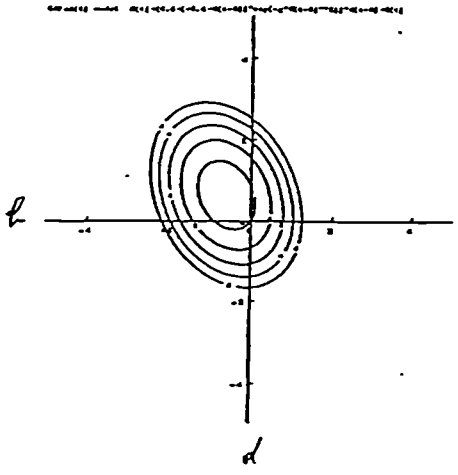
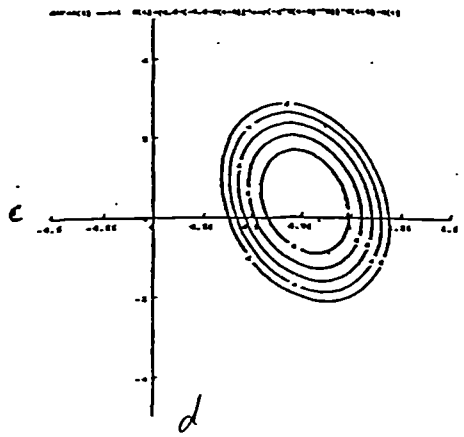


Fig. 4.1.1.1

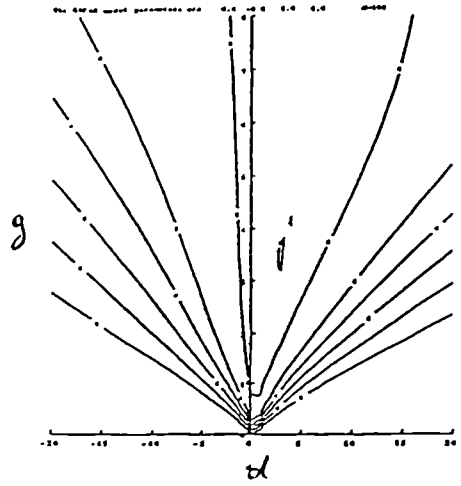




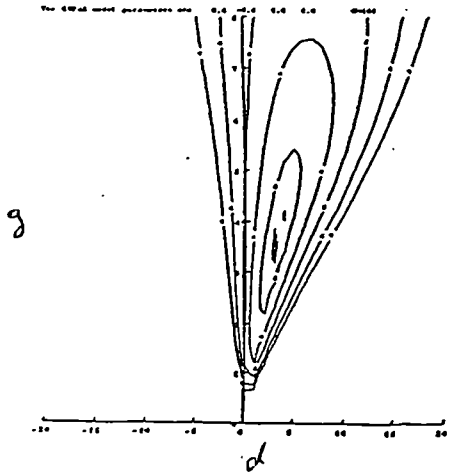


CONTOUR KEY	
1	80.00
2	85.00
3	90.00
4	95.00
5	100.00

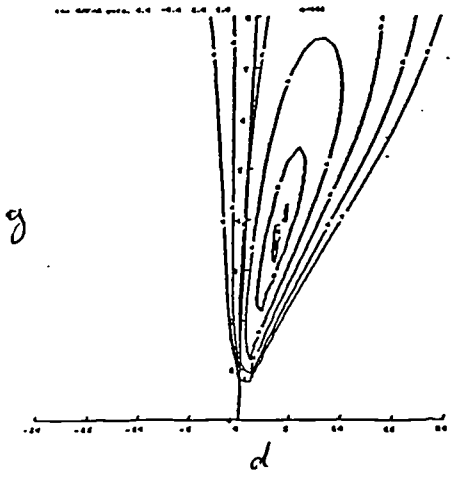
Fig. 4.2.1-4.2.3 : Negative of log likelihood function of model (4.2.2) for (b,c), (b,d) and (c,d).



CONTOUR KEY	
1	75.20
2	80.00
3	100.00
4	120.00
5	150.00
6	200.00
7	300.00



CONTOUR KEY	
1	75.17
2	75.20
3	75.50
4	76.50
5	78.00
6	80.00
7	82.00



CONTOUR KEY	
1	75.15
2	75.21
3	75.50
4	76.50
5	78.00
6	80.00
7	82.00

Fig. 4.2.4-4.2.6 : Negative log likelihood function of model (4.2.2) for (d,g).

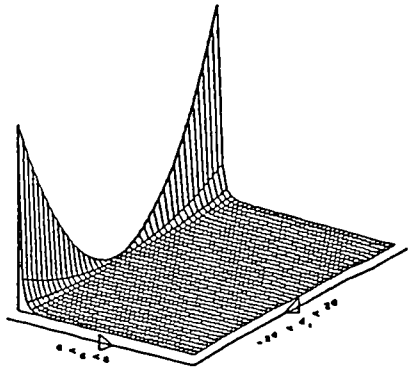


Fig. 4.2.7 : Isometric projection of $S_N(\theta)$.

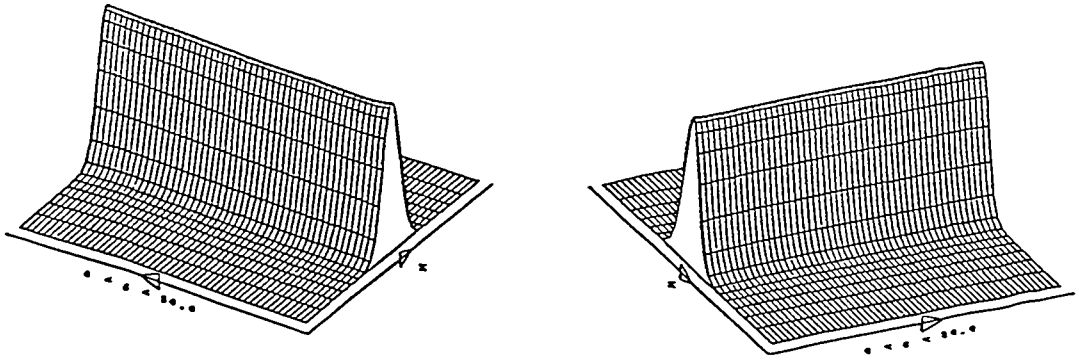


Fig. 4.2.8 : Family of stationary marginal p.d.f. of model (4.2.2) for $g \in [0,20]$.

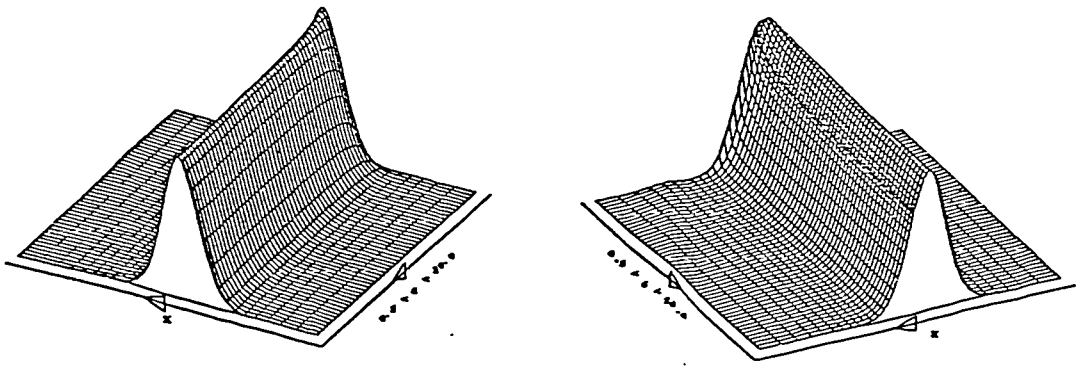


Fig. 4.2.9 : Family of stationary marginal p.d.f. of model (4.2.2) for $g \in [0.5,15]$.

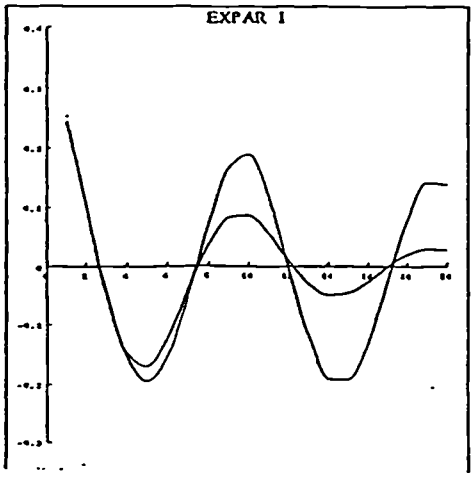
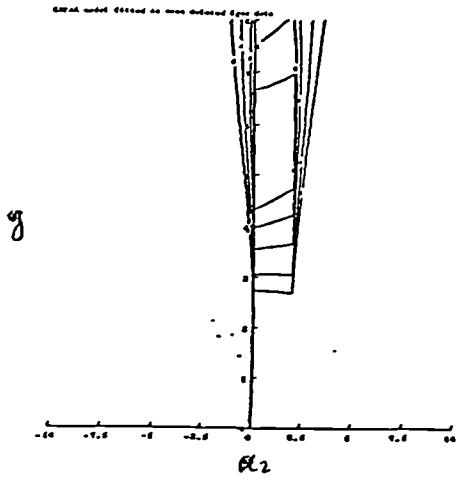
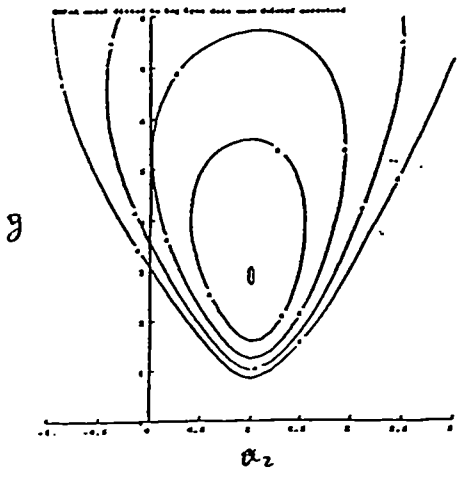


Fig. 4.2.10 : ACF of Lynx data and model (4.2.5).



CONTOUR KEY	
1	5.25
2	5.30
3	5.40
4	5.60
5	5.80



CONTOUR KEY	
1	4.80
2	5.00
3	5.20
4	5.40
5	5.60

Fig. 4.2.11-4.2.12 : Negative log likelihood function of model (4.2.5) for $a_2 \in (-20, 20)$ and $a_2 \in (-2, 2)$.

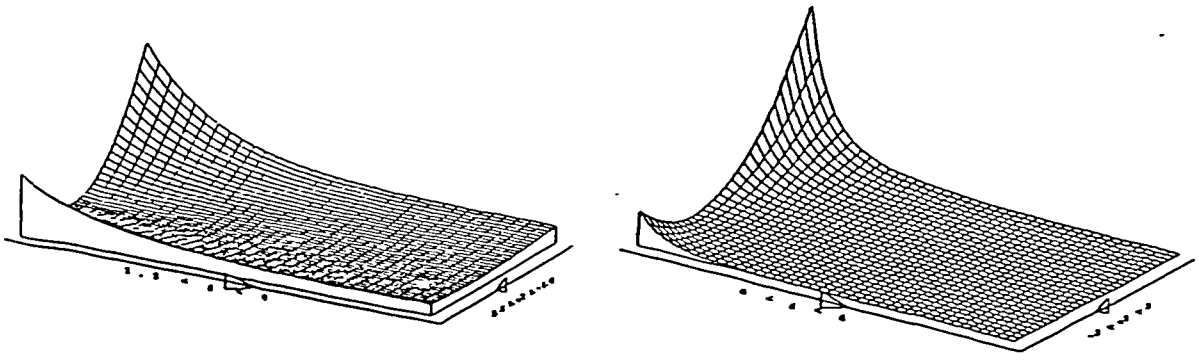


Fig. 4.2.13-4.2.14 : isometric projections of negative log likelihood function of model (4.2.5).

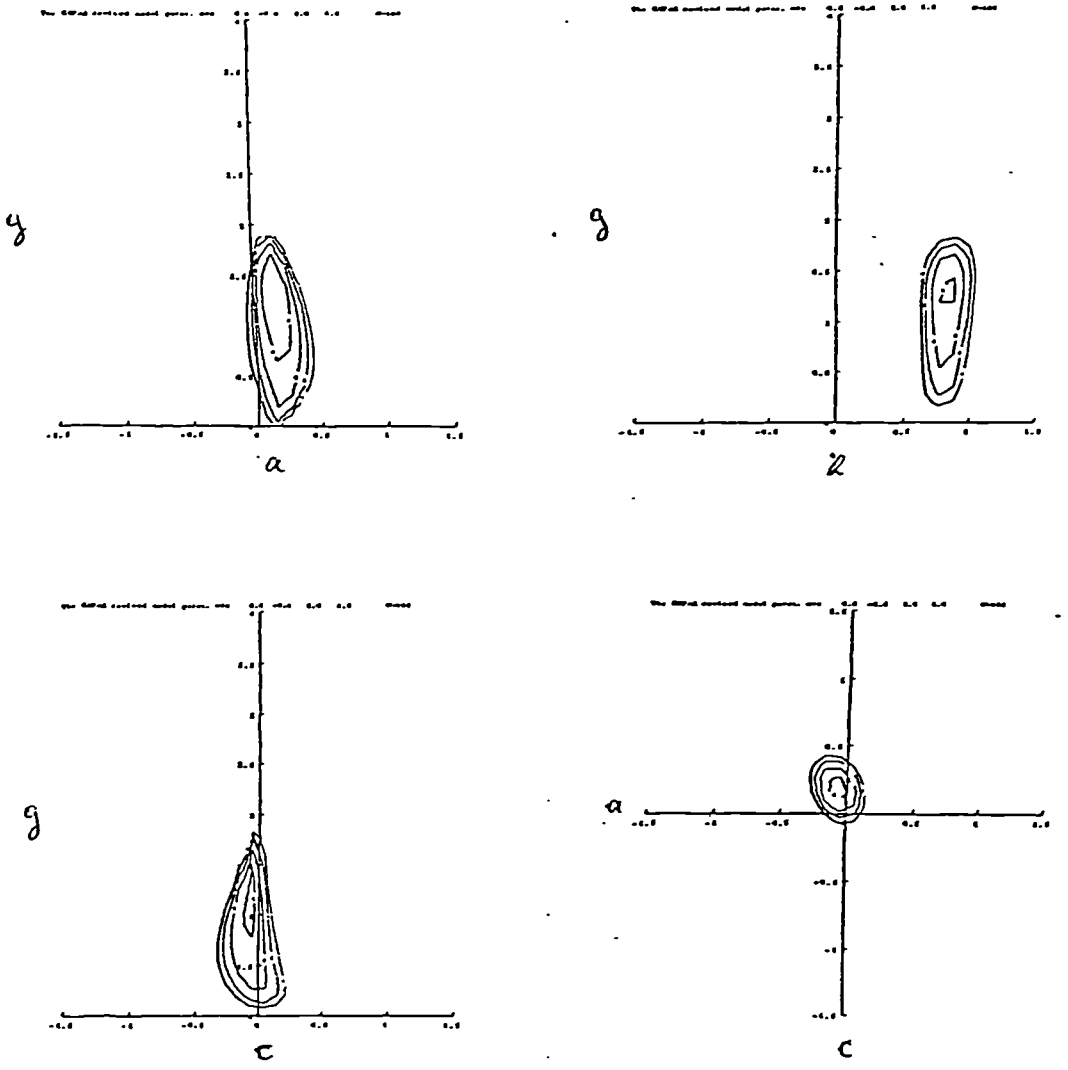


Fig. 4.2.1.1-4.2.1.4 : Centred versions of model (4.2.2) with Figs. 4.2.1-4.2.4.

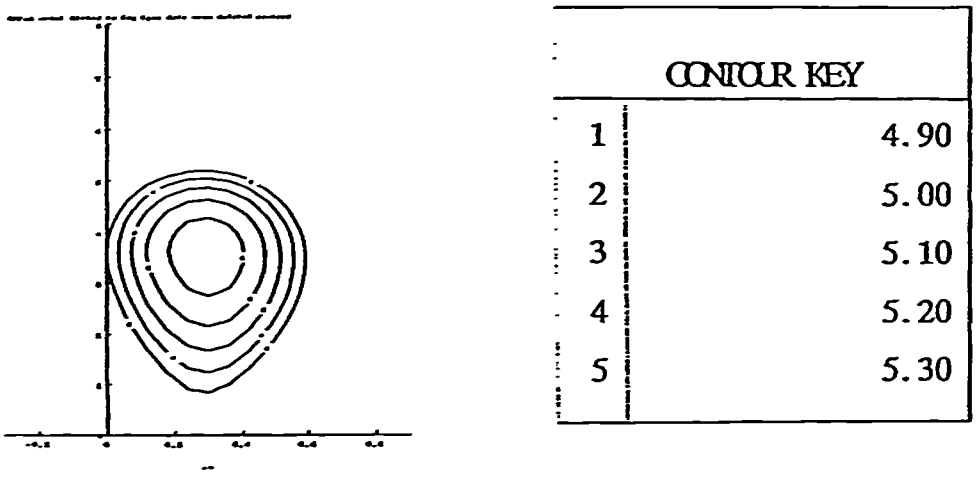
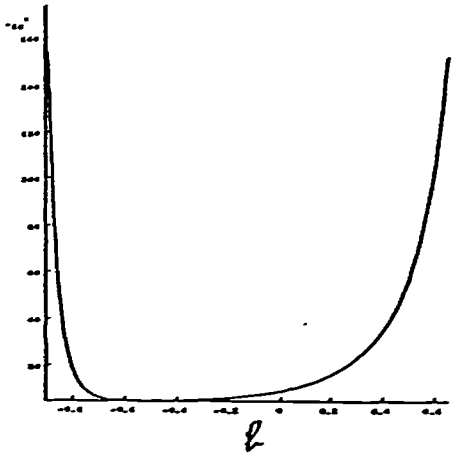
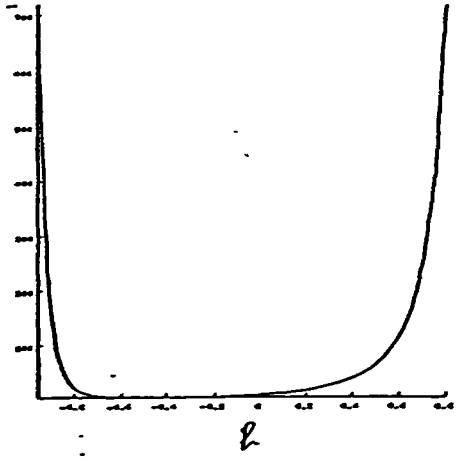


Fig. 4.2.1.5 : Centred version of model (4.2.5) with Fig. 4.2.12.

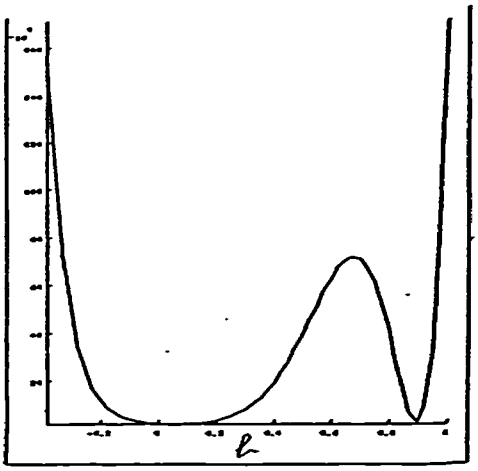
1- $b=-0.4$, $i=j=1$, $N=50$.



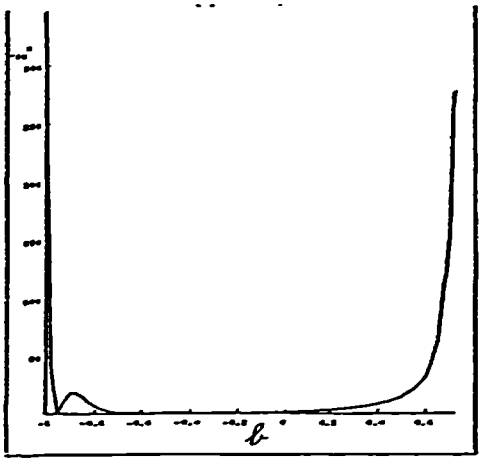
2- $b=-0.4$, $i=j=1$, $N=70$.



3- $b=0.9$, $i=j=1$, $N=50$.



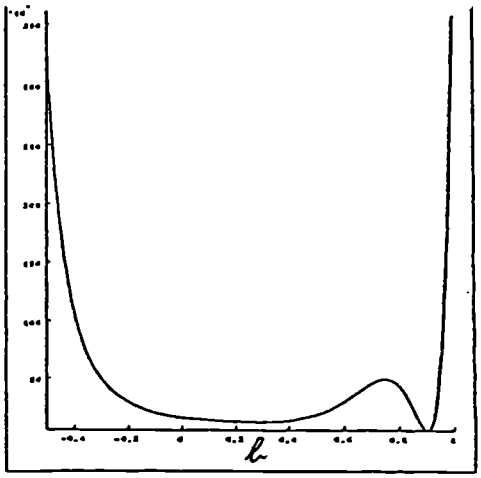
4- $b=-0.95$, $i=j=1$, $N=50$.



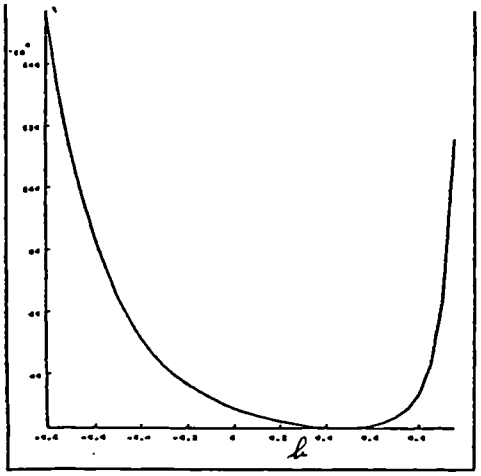
Figs. 4.3.1-4.3.4

Fig. 4.3.1-4.3.10 : Negative log likelihood function of model $X_t=bX_{t-i}e_{t-j}+e_t$.

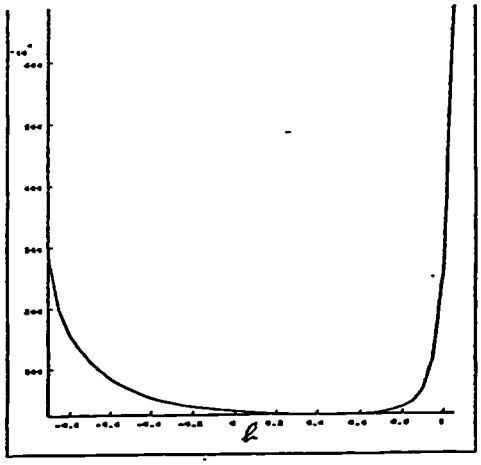
5- b=0.9 , i=j=1 , N=70.



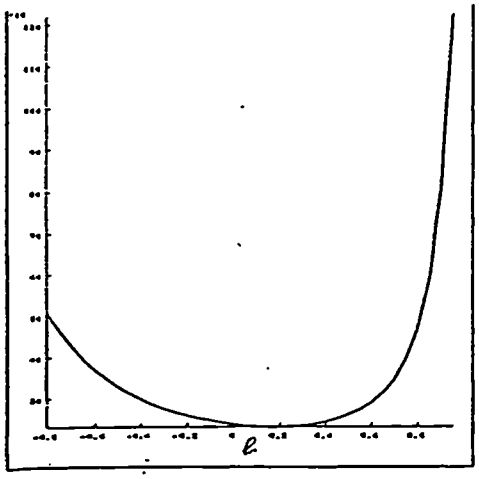
6- b=0.5 , i=j=1 , N=250.



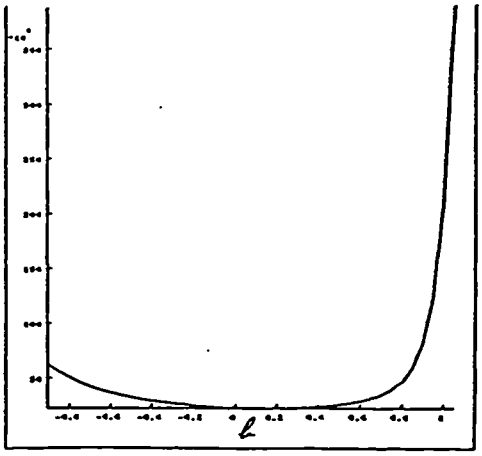
7- b=0.5 , i=1 , j=2 , N=250.



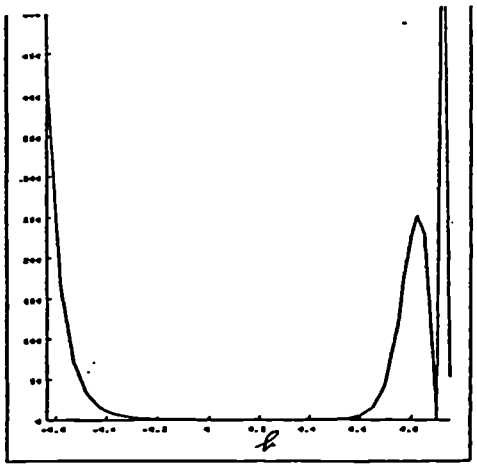
8- b=0.2 , i=1 , j=2 , N=250.



9- b=0.2 , i=2 , j=1 , N=250.

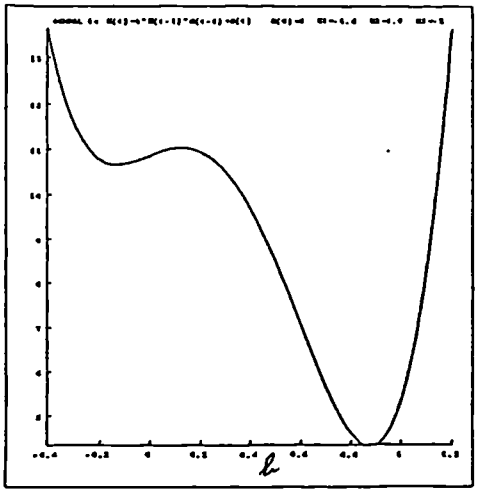


10- b=0.9 , i=2 , j=1 , N=250.



Figs. 4.3.5-4.3.10

Fig. 4.3.1-4.3.10 : Negative log likelihood function of model $X_t=bX_{t-i}e_{t-j}+e_t$.



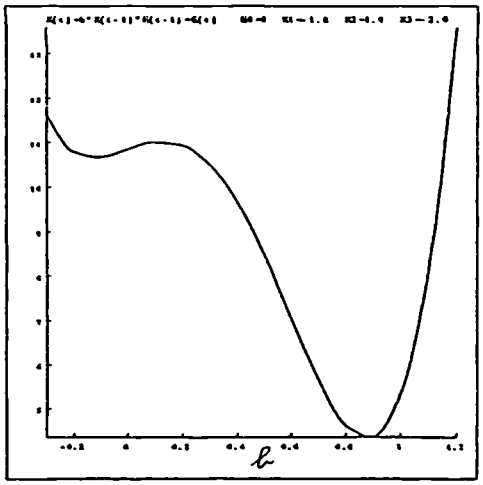
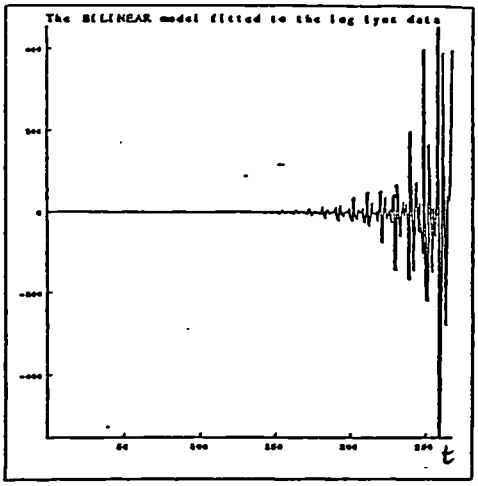
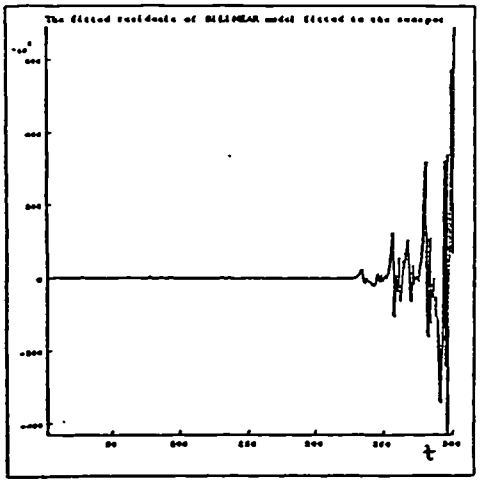


Fig. 4.3.11-4.3.12 : Negative log likelihood function with N=3.





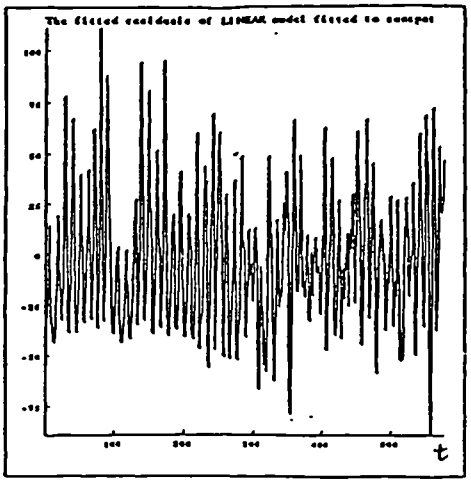


Fig. 4.3.1.1-4.3.1.3 : Time plots of residuals of bilinears which are fitted to the lynx and sunspot numbers and that of linear model which is fitted to sunspot numbers.

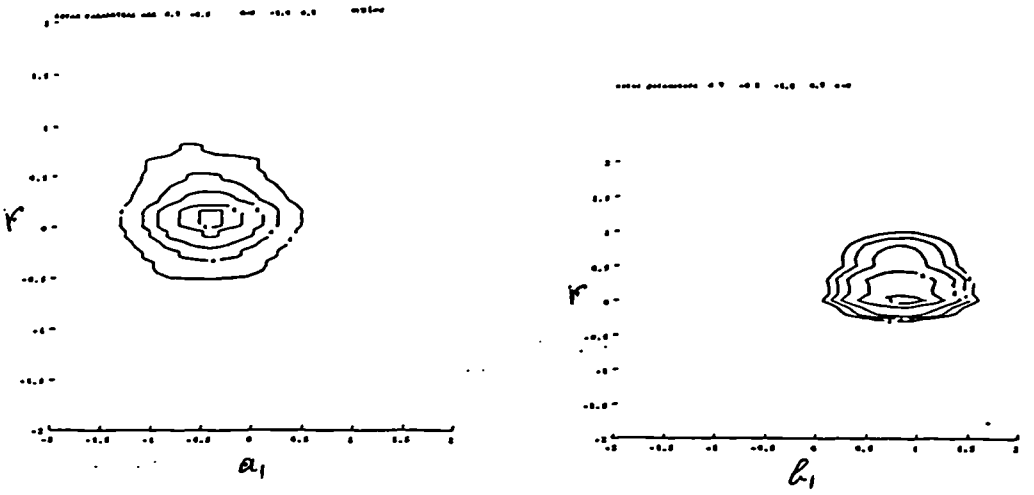


Fig. 4.4.1-4.1.2 : Negative log likelihood function of SETAR model (4.4.1) for (a_1, r) and (b_1, r).

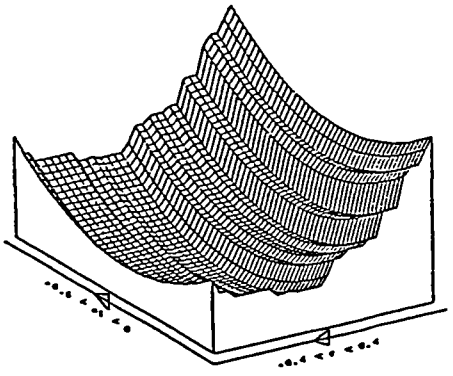


Fig. 4.4.3 : Isometric projection of negative log likelihood function of model (4.4.1)

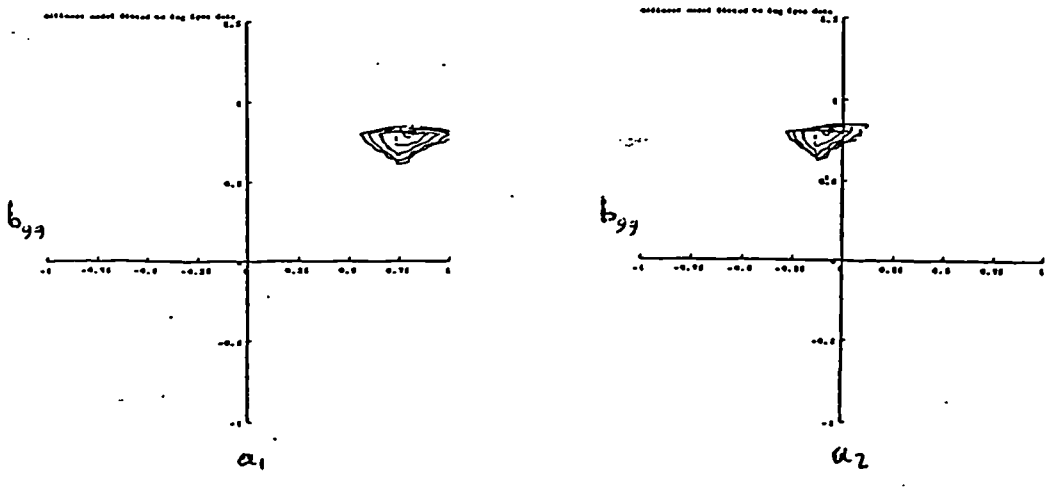
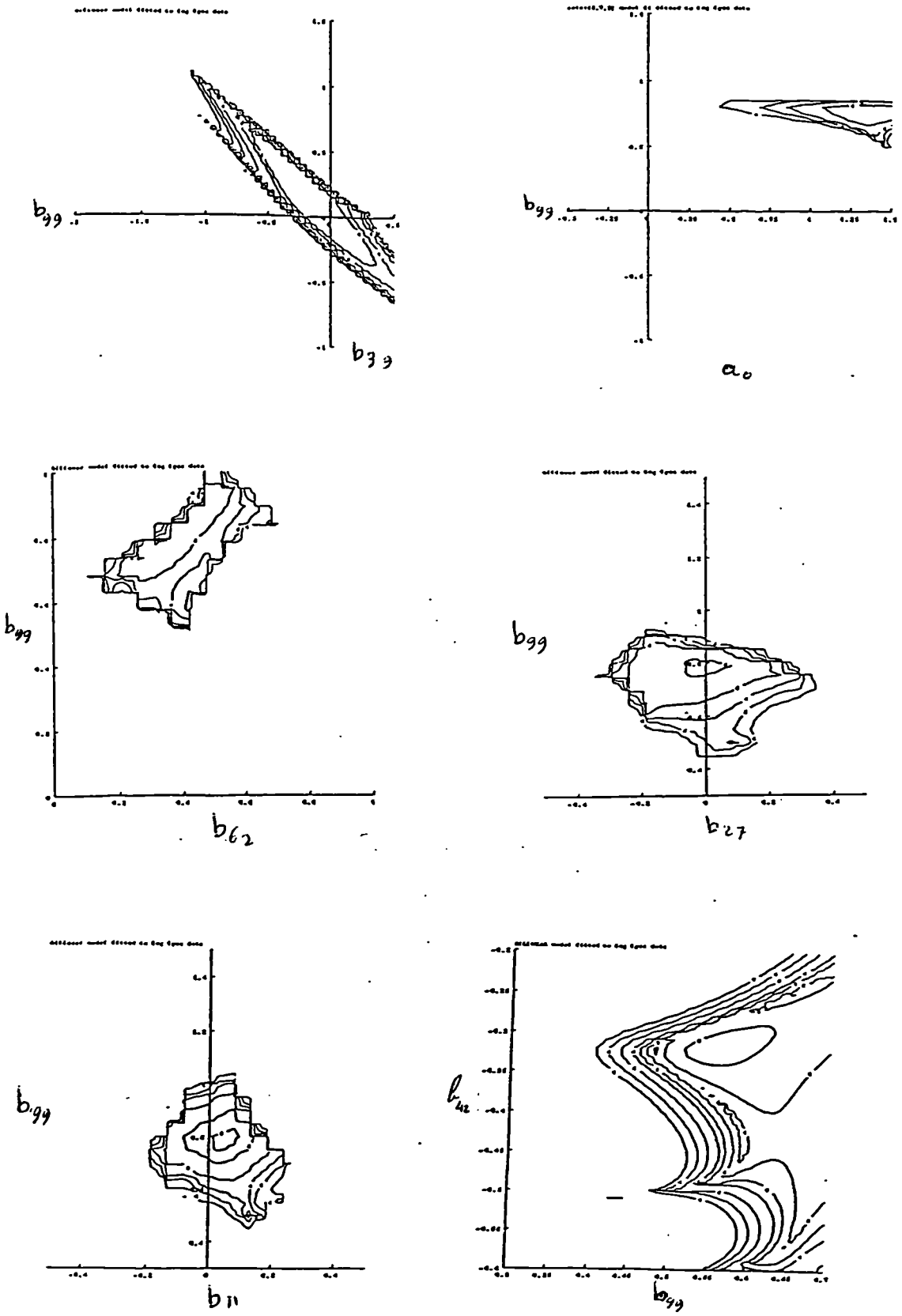
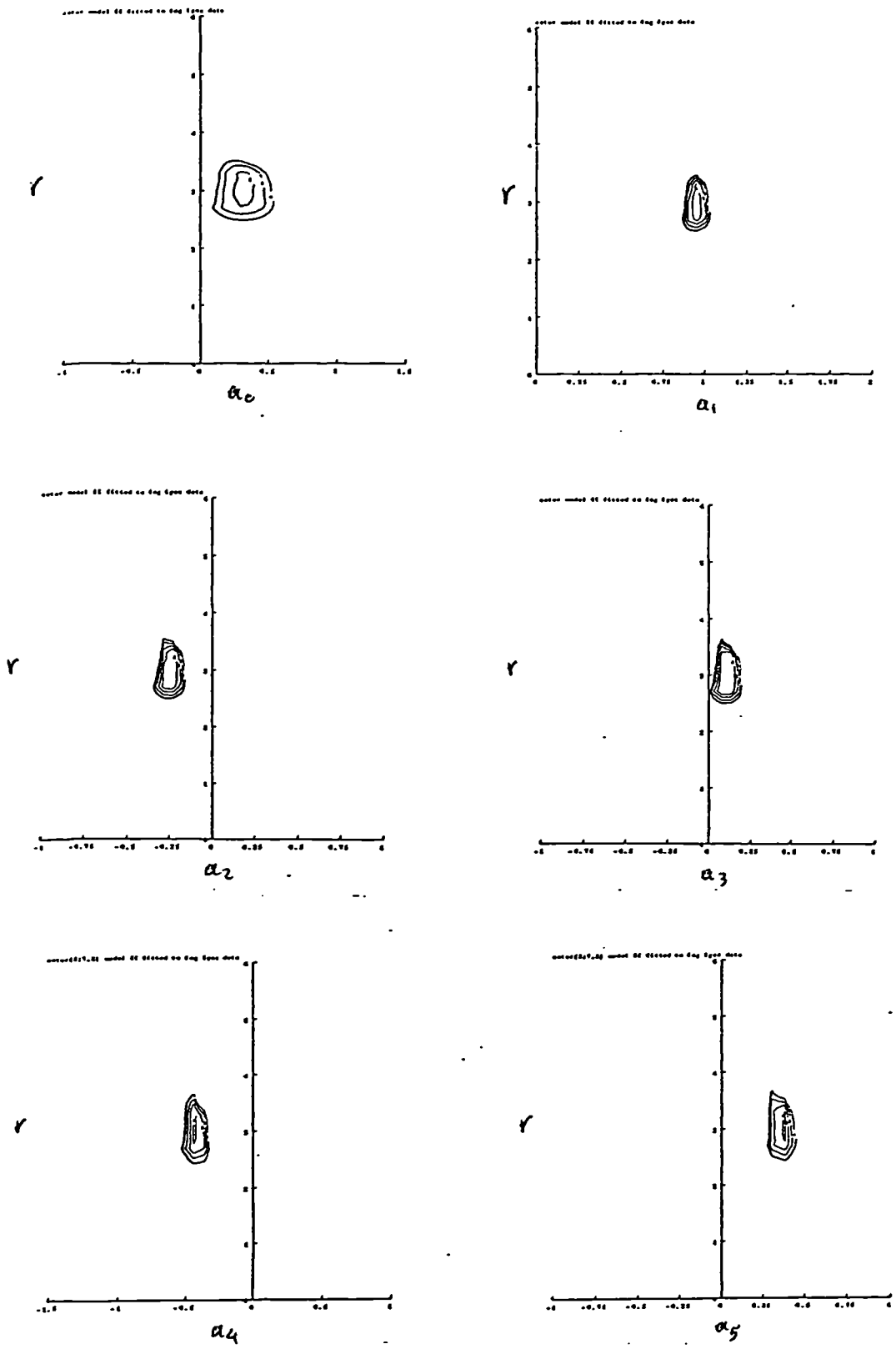


Fig. 4.5.1-4.5.8 : Negative log likelihood function of bilinear fitted to lynx data



Figs. 4.5.3-4.5.8 (Bilinear)

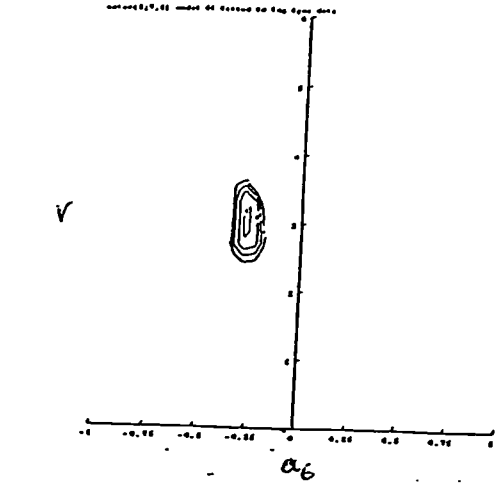
Fig. 4.5.1-4.5.8 : Negative log likelihood function of bilinear fitted to lynx data

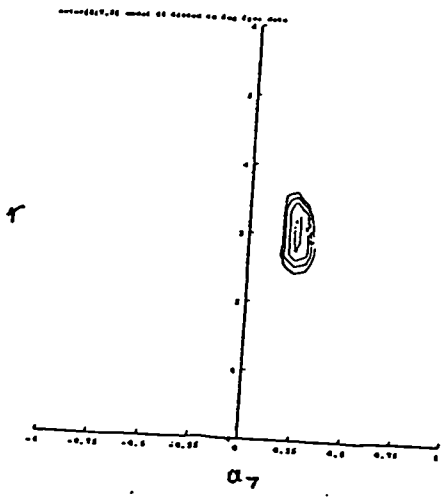


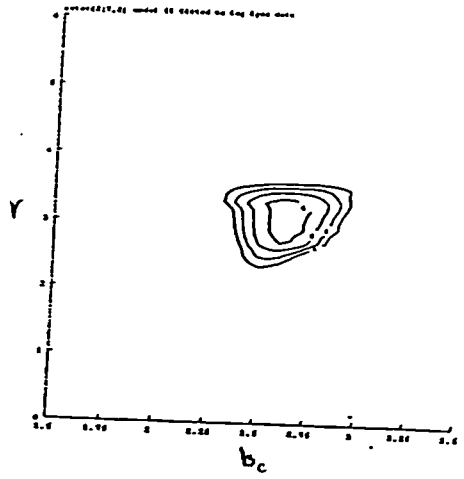
Figs. 4.5.9-4.5.14 (SETAR)

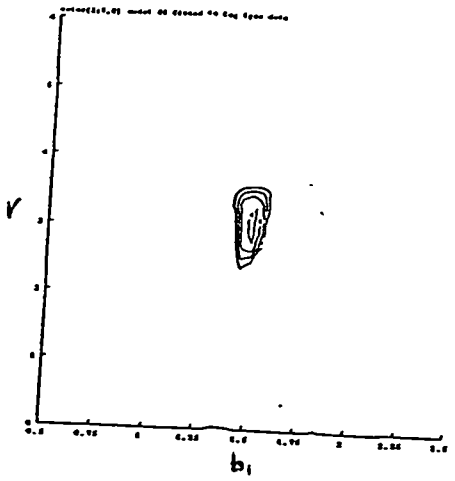
Fig. 4.5.9-4.5.26 : Negative log likelihood functions of SETAR models fitted to lynx and sunspot data.

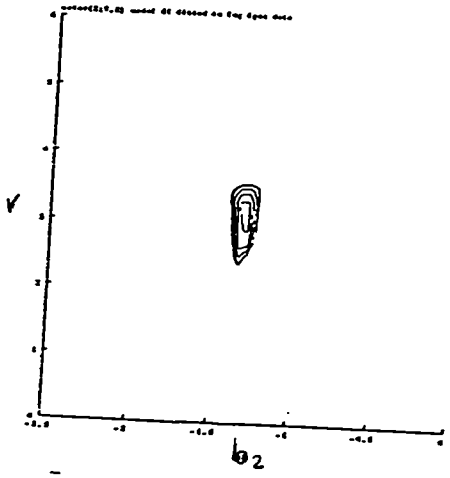
Fig. 4.5.9-4.5.26 : Negative log likelihood functions of SETAR models fitted to lynx and sunspot data.



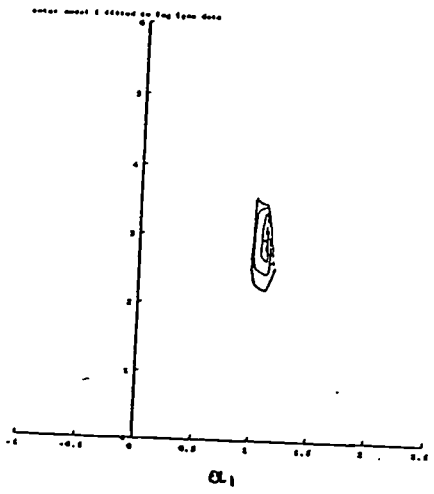


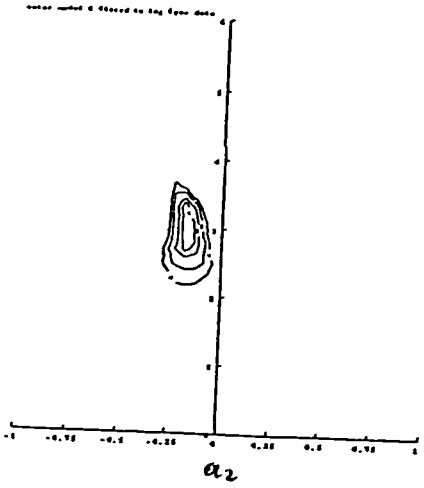


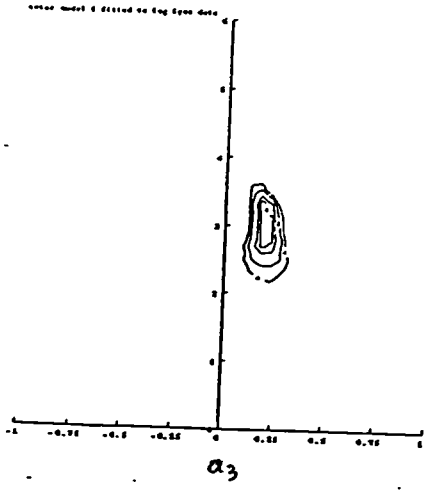


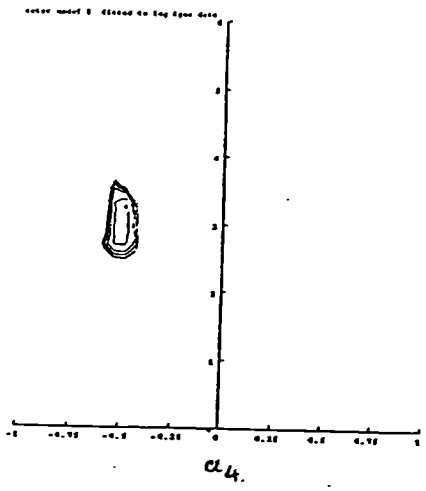


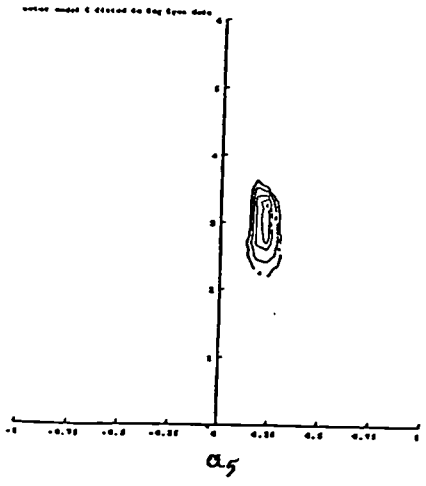
Figs. 4.5.15-4.5.20 (SETAR)

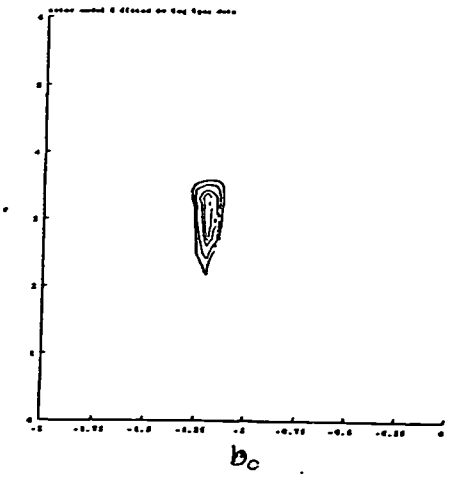






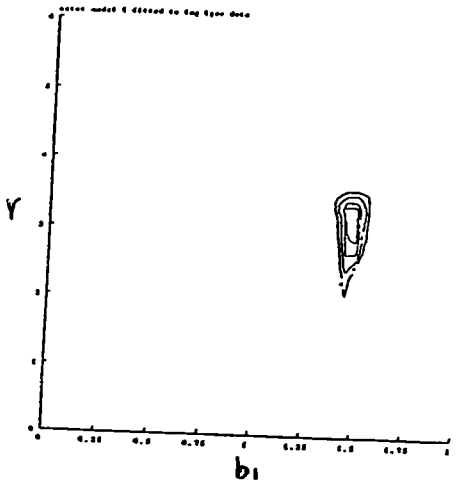






Figs. 4.5.21-4.5.26 (SETAR)

Fig. 4.5.9-4.5.26 : Negative log likelihood functions of SETAR models fitted to lynx and sunspot data.



Figs. 4.5.27-4.5.28 (SETAR)

CHAPTER FIVE

STATIONARY PROBABILITY DENSITY FUNCTION OF NON-LINEAR AUTOREGRESSIVE PROCESSES

In this chapter we examine the stationary probability density function of certain non-linear, discrete time series. In particular, we examine numerical solutions for a recursive relation for computing the m-step-ahead conditional density of a non-linear autoregressive model which is obtained by using the Chapman-Kolmogorov formula. We approximate the stationary marginal probability density function of the model by the m-step-ahead conditional density, for sufficiently large m.

Although the classical autoregressive linear model is very useful as a first order approximation to stochastic discrete time phenomena, the generating mechanism of most physical phenomena is essentially non-linear. Jones (1978) has considered the first order non-linear autoregerssive model, NLAR(1),

$$X_t = \lambda(X_{t-1}) + e_t \,, \; t = 0, \pm 1, \ldots, \tag{5.1}$$

where $\lambda(.)$ is a fixed real function of a real argument and $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean zero and constant variance.

The problem of evaluating the stationary marginal distribution of the model (5.1) explicitly is extremely involved and usually intractable. Jones (1978) has proposed three methods for evaluating numerically the stationary distribution of the model (5.1). All the procedures are based on power-series expansions. However, they are rather complicated for practical use.

One type of non-linear discrete time parameter models is the SETAR model for which the simplest first order SETAR is of the following form

$$X_t = \begin{cases} \alpha X_{t-1} + e_t & \text{if } X_{t-1} \le r \\ \beta X_{t-1} + e_t & \text{if } X_{t-1} > r \,, \end{cases} \tag{5.2}$$

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

constant variance. Generally it is difficult to obtain the explicit form of the marginal stationary distribution of X_t (if it exists). A necessary and sufficient condition for ergodicity of model (5.2) and hence the existence the stationary marginal distribution is that $\alpha<1$, $\beta<1$, $\alpha\beta<1$, (Petruccelli and Woolford 1984). Andel et al (1984) have proposed a method for evaluating the stationary marginal density in non-linear processes and have managed to obtain the explicit analytic solution for a special subclass of (5.2) namely

$$X_t=-\alpha|X_{t-1}|+e_t\,, \tag{5.3}$$

where $\alpha\in(0,1)$ and $e_t\sim N(0,1)$. The stationary marginal density function is shown to be

$$f(x)=[2(1-\alpha^2)/\pi]^{\frac{1}{2}}\ exp[-\frac{1}{2}(1-\alpha^2)x^2]\Phi(-\alpha x)\,, \tag{5.4}$$

where $\Phi(.)$ is the distribution function of the standard normal.

Let $\{e_t\}$ in model (5.3) be independent Cauchy variables with density $\frac{1}{\pi(1+x^2)}$. Andel and Barton (1986) have proved that for this case the process $\{X_t\}$ has a unique stationary distribution with the density

$$f(x)=\frac{2A}{\pi^2}\Big[-\{4A^2x^2+(1-A^2+x^2)^2\}^{-1}\{x\ln A^{-2}(1+x^2)+(A^2-1+x^2)arctgx\}+ \tag{5.5}$$

$$(2A)^{-1}\{(1+A)^2+x^2\}^{-1}(1+A)\pi\Big]\,,$$

where $A=\frac{\alpha}{1-\alpha}$.

Chan and Tong (1986b) have considered the integral equation in f

$$f(x)=\int_{R^n}K(x-\tilde{T}(y))f(y)dy\,,$$

where K is a probability density function and $\tilde{T}:R^n\to R^n$. They have shown that the success of the method of Andel et al (1984,1986) in fact depends on the symmetry of the autoregressive function in (5.3), which is linear over the non-negatives and defined over the negative by assigning the function values taken by the mirror images. They have developed a systematic method and generalized it to higher dimension by exploiting the symmetry of the autoregressive function defined by a compact group.

Denote the conditional density of X_m given X_o by f_m. On letting m tend to infinity, f_m will converge to the stationary marginal density function of X_t if $\{X_t\}$ is ergodic. The Chapman-Kolmogorov formula gives a recursive relation to compute f_m. In this chapter we use the numerical integration method to

calculate the m-step-ahead conditional density f_m and we compare the numerical approximation of the stationary density function $f(x)$ with the theoretical probability density function in the bench-mark cases, namely cases studied by Andel at al (1984,1986).

A general univariate NLAR(p) process may be written as

$$X_t = \lambda(\mathbf{X}_{t-1}) + e_t \,, \tag{5.6}$$

where $\mathbf{X}_{t-1} = (X_{t-1}, \ldots, X_{t-p})'$, $\{e_t\}$ is a sequence of independent and identically distributed scalar random variables with mean zero and constant variance and e_t is independent of X_s for all $s<t$. The conditional density function of X_t given $\mathbf{X}_{t-1}$ may be written as

$$f_{X_t|\mathbf{X}_{t-1}}(x|\mathbf{X}_{t-1}=\mathbf{y}) = K(x-\lambda(\mathbf{y})) \,, \tag{5.7}$$

where $K(.)$ is the probability density function of the random variable e_t. The joint density of $(X_t, \mathbf{X}_{t-1})$ can be obtained simply as

$$f_{X_t,\mathbf{X}_{t-1}}(x,\mathbf{y}) = K(x-\lambda(\mathbf{y}))h(\mathbf{y}) \,, \tag{5.8}$$

where $h(.)$ denotes the joint density of $(X_{t-1}, \ldots, X_{t-p})'$. By integrating with respect to $\mathbf{y}$ over R^p on both sides of (5.8), the stationary marginal density function of X_t may be obtained as

$$f(x) = \int_{R^p} K(x-\lambda(\mathbf{y}))h(\mathbf{y})d\mathbf{y} \,. \tag{5.9}$$

For the self-exciting threshold autoregressive model SETAR($l;k_1,\ldots,k_l$) of the form

$$X_t = b_0^{(j)} + \sum_{i=1}^{k_j} b_i^{(j)} X_{t-i} + e_t^{(j)} \quad \text{if } X_{t-d} \in (r_{j-1}, r_j] \,, \tag{5.10}$$

for $j=1,2,\ldots,l$, where $-\infty=r_0<r_1<\cdots<r_l=\infty$, it can be shown that

$$f(x) = \sum_{j=1}^{l} \int_{r_{j-1}}^{r_j} K_j(x-\lambda_j(\mathbf{y}))h(\mathbf{y})d\mathbf{y} \,, \tag{5.11}$$

where $K_j(.)$ denotes the probability density function of $e_t^{(j)}$ and $\lambda_j(.)$ denotes the autoregressive function in the j-th regime. The integral equations (5.9) and (5.11) do not readily admit analytic solutions.

5.1. Numerical Solution

As the numerical solution is commonly the only solution, we will review briefly some of the numerical methods and then employ the Chapman-Kolmogorov relation to obtain a simple recursive formula for a numerical solution of integral equation (5.9). Without loss of generality, first we restrict our discussion to

the first order case, i.e. NLAR(1).

5.1.1. Series Expansion

Jones (1976,1978) has derived power series expansions for obtaining the probability density function and moments of a stationary discrete time series model (5.1). The method involves consideration of the family of time series indexed by the parameter β

$$X_t(\beta) = a+bX_{t-1}(\beta)+\beta[\lambda(X_{t-1}(\beta))-a-bX_{t-1}(\beta)]+e_t ,$$

where a is an arbitrary fixed constant and $|b|<1$. The resulting series expansion in powers of β usually requires b to be zero for practical calculation, with truncation. The value of the constant a affects the convergence of the series and so its value has to be chosen carefully. Some difficulties have been experienced with convergence when $\lambda(.)$ is piecewise linear and the variance of e_t is "small" (Pemberton, 1985).

5.1.2. Iterative Numerical Quadrature Method

Pemberton (1985) has approximated the solution of integral equation (5.11) through the iteration

$$f_{n+1}(x) = C_n \sum_{j=1}^{l} \int_{r_{j-1}}^{r_j} f_n(y)K_j(x-\lambda_j(y))dy , \qquad (5.1.2.1)$$

with some suitable starting function $f_o(x)$, normalized so that

$$\int_{-\infty}^{\infty} f_o(x)dx = 1 .$$

The multiplying factor C_n is a normalization constant, and is necessary because the numerical evaluation of the integral requires a finite range of integration. It is mentioned by Pemberton (1985) that convergence is not always achieved. The starting function $f_o(x)$ can be the standard normal density.

5.2. Conditional Density Approach

The idea of approximating the stationary distribution as the limit of conditional distribution was reported by Jones (1976). In practice it would appear that it is not so easy to use his results and we therefore propose an alternative. (See Moeanaddin and Tong , 1989).

Since we have a Markov chain over R, we may recall the Chapman-Kolmogorov relation

$$f(x_{t+m}|x_t) = \int_{-\infty}^{\infty} f(x_{t+m}|x_{t+1})f(x_{t+1}|x_t)dx_{t+1} , \qquad (5.2.1)$$

where $f(x_s|x_t)$ denotes the conditional probability density function of X_s given $X_t=x_t$. Let $K(.)$ denote the probability density function of e_t and let $f_m(x)$ denote the conditional density of X_{t+m} given X_t. Equation (5.2.1) immediately gives

$$f_m(x) = \int_{-\infty}^{\infty} f_{m-1}(y)K(x-\lambda(y))dy \ . \tag{5.2.2}$$

Equation (5.2.2) gives, in particular a recursive formula for conditional density with

$$f_1(x) = f_{X_{t+1}|X_t}(x|X_t=y) = K(x-\lambda(y)).$$

In evaluating the improper integrals of the form (5.2.2), we are faced with two difficulties: (i) The discontinuity of λ and (ii) the integration over R. In case (i) the integral can be split up over disjoint intervals, over each of which the intervals λ is continuous. In case (ii), by employing a suitable method, the integral can be approximated efficiently. We adopt the following procedure: it is well known that if f is integrable over an interval, then there exist some points $x_1,x_2,...,x_n$ with corresponding weights $w_1,w_2,\cdots,w_n$ such that $\int f(x)dx \approx \sum_{k=1}^{n} w_k f(x_k)$ so that it is exact for functions f of class P_{2n-1}. (P_n denotes the class of all functions that are polynomial of degree $\leq n$.) Needless to say, when numerical integration is employed, care must be taken to handle the accumulation of rounding errors.

In practice, one may employ some Gauss-type formula to generate an appropriate set of points $x_1,\cdots,x_n$ with corresponding weights $w_1,\cdots,w_n$ and by using the recursive formula (5.2.2) with $f_1(x)=K(x-\lambda(y))$, a sequence of conditional densities $f_m(x)$, $m=2,3,...,$ can be calculated. As m increases, the conditional density f_m converges to f, the unique stationary marginal probability density function of X_t, if the latter exists. Convergence is deemed to have been achieved when $|f_m-f_{m-1}|<\varepsilon$ for a small positive value of ε, e.g. $\varepsilon=10^{-6}$.

In our study, NAG routine D01BBF is employed to generate an appropriate set of points $x_1,\cdots,x_n$ with corresponding weights $w_1,\cdots,w_n$. We normalized the conditional density to unity at each step, to avoid accumulating errors, and also we check the integration of conditional density, before normalization, over R in each step, as a measure of accuracy. If the evaluated integral is not close to one (i.e. with error $\geq 10^{-3}$), then we would change the parameters of the NAG routines to generate another set of points and so on until a more adequate set is found.

5.3. Application

As a check on the method, it has been applied to obtain the unique stationary marginal density of the linear AR(1) process

$$X_t = \alpha X_{t-1} + e_t \,, \; |\alpha| < 1 \,, \tag{5.3.1}$$

and the results are given in Table (5.3.1).

We repeated the exercise with the SETAR(2;1,1) process

$$X_t = -\alpha |X_{t-1}| + e_t \,, \; 0 < \alpha < 1 \,, \tag{5.3.2}$$

where $e_t \sim N(0,1)$, for different values of α, and different number of points. Since the theoretical stationary probability density function and rth moment for model of the form (5.3.1) and (5.3.2) are available, in Table (5.3.1) and Table (5.3.2) we compare the approximation of rth moments (r=1,2,3,4) based on the conditional density approach with $\varepsilon < 10^{-6}$ (column marked approx.) with the theoretical (column marked theo.). Using the same points $x_1, \cdots, x_n$ and $w_1, \cdots, w_n$ and the theoretical probability density function $f(x)$, we may also perform numerical integration to evaluate $\int_{-\infty}^{\infty} x^r f(x) dx$. Results are listed in the column marked numerical .

Table 5.3.1

n	α	$E(X_i)$			$E(X_i^2)$			run time
		approx.	numerical	theo.	approx.	numerical	theo.	(in sec.)
12	-0.9	8.6416×10^{-4}	1.9377×10^{-8}	0.0	5.26547	5.26316	5.26316	7.50
24	-0.9	-1.4769×10^{-8}	-9.7981×10^{-9}	0.0	5.26316	5.26316	5.26316	9.23
48	-0.9	6.0558×10^{-8}	-1.4428×10^{-8}	0.0	5.26316	5.26315	5.26316	16.70
64	-0.9	-1.2280×10^{-10}	-6.7862×10^{-9}	0.0	5.26315	5.26316	5.26316	15.41
12	-0.5	1.4903×10^{-8}	2.1202×10^{-8}	0.0	1.33333	1.33333	1.33333	2.45
24	-0.5	-4.4961×10^{-8}	-9.0807×10^{-10}	0.0	1.33333	1.33333	1.33333	2.82
48	-0.5	-2.0900×10^{-8}	3.6395×10^{-9}	0.0	1.33333	1.33333	1.33333	5.48
64	-0.5	-1.0532×10^{-8}	3.4370×10^{-9}	0.0	1.33333	1.33333	1.33333	6.32
12	-0.2	2.7223×10^{-9}	7.0360×10^{-9}	0.0	1.04167	1.04167	1.04167	2.05
24	-0.2	1.8319×10^{-9}	1.5711×10^{-8}	0.0	1.04167	1.04167	1.04167	2.19
48	-0.2	1.5378×10^{-8}	5.0135×10^{-9}	0.0	1.04167	1.04167	1.04167	3.65
64	-0.2	1.5165×10^{-9}	5.5102×10^{-9}	0.0	1.04167	1.04167	1.04167	4.96

Table 5.3.1 (continue)

n	α	$E(X_i^3)$			$E(X_i^4)$			run time
		approx.	numerical	theo.	approx.	numerical	theo.	(in sec.)
12	-0.9	1.3692×10^{-3}	-1.1456×10^{-7}	0.0	83.15470	83.10154	83.10245	7.50
24	-0.9	-1.7757×10^{-7}	1.8475×10^{-7}	0.0	83.10223	83.10243	83.10245	9.23
48	-0.9	1.8262×10^{-7}	8.0142×10^{-8}	0.0	83.10185	83.10229	83.10245	16.70
64	-0.9	-1.0245×10^{-6}	2.9038×10^{-7}	0.0	83.10244	83.10246	83.10245	15.41
12	-0.5	-5.8058×10^{-8}	-2.9490×10^{-8}	0.0	5.33173	5.33180	5.33333	2.45
24	-0.5	3.2773×10^{-9}	-1.5317×10^{-8}	0.0	5.33333	5.33333	5.33333	2.82
48	-0.5	-1.1381×10^{-7}	7.4208×10^{-8}	0.0	5.33333	5.33333	5.33333	5.48
64	-0.9	2.4576×10^{-9}	-9.6766×10^{-8}	0.0	5.33333	5.33333	5.33333	6.32
12	-0.2	7.4203×10^{-8}	4.2992×10^{-9}	0.0	3.25521	3.25521	3.25521	2.05
24	-0.2	-6.1182×10^{-9}	-2.1336×10^{-8}	0.0	3.25521	3.25521	3.25521	2.19
48	-0.2	2.1607×10^{-8}	-3.4305×10^{-8}	0.0	3.25521	3.25521	3.25521	3.65
64	-0.2	-1.1060×10^{-7}	4.1696×10^{-8}	0.0	3.25521	3.25521	3.25521	4.96

Table 5.3.2

n	α	$E(X_t)$			$E(X_t^2)$			run time
		approx.	numerical	theo.	approx.	numerical	theo.	(in sec.)
24	0.9	1.64855	1.64742	1.64742	5.27112	5.26316	5.26316	7.79
48	0.9	1.64742	1.64742	1.64742	5.26314	5.26311	5.26316	14.98
96	0.9	1.64742	1.64742	1.64742	5.26316	5.26316	5.26316	36.11
128	0.9	1.64742	1.64742	1.64742	5.26312	5.26315	5.26316	75.89
24	0.5	0.46066	0.46066	0.46066	1.33333	1.33333	1.33333	3.50
48	0.5	0.46066	0.46066	0.46066	1.33333	1.33333	1.33333	4.90
96	0.5	0.46066	0.46066	0.46066	1.33333	1.33333	1.33333	12.76
128	0.5	0.46066	0.46066	0.46066	1.33333	1.33333	1.33333	16.55
24	0.2	0.16287	0.16287	0.16287	1.04167	1.04167	1.04167	2.86
48	0.2	0.16287	0.16287	0.16287	1.04167	1.04167	1.04167	5.03
96	0.2	0.16287	0.16287	0.16287	1.04167	1.04167	1.04167	9.39
128	0.2	0.16287	0.16287	0.16287	1.04167	1.04167	1.04167	10.80

Table 5.3.2 (continue)

n	α	$E(X_t^3)$			$E(X_t^4)$			run time
		approx.	numerical	theo.	approx.	numerical	theo.	(in sec.)
24	0.9	19.04422	18.98873	18.98874	83.40329	83.10229	83.10242	7.79
48	0.9	18.98840	18.98809	18.98874	83.09683	83.09399	83.10242	14.98
96	0.9	18.98893	18.98874	18.98874	83.10239	83.10245	83.10242	36.11
128	0.9	18.98837	18.98861	18.98874	83.09853	83.10078	83.10242	75.89
24	0.5	1.68908	1.68908	1.68908	5.33333	5.33333	5.33333	3.50
48	0.5	1.68908	1.68908	1.68908	5.33333	5.33333	5.33333	4.90
96	0.5	1.68908	1.68908	1.68908	5.33333	5.33333	5.33333	12.76
128	0.5	1.68908	1.68908	1.68908	5.33333	5.33333	5.33333	16.55
24	0.2	0.50217	0.50217	0.50217	3.25521	3.25521	3.25521	2.86
48	0.2	0.50217	0.50217	0.50217	3.25521	3.25521	3.25521	5.03
96	0.2	0.50217	0.50217	0.50217	3.25521	3.25521	3.25521	9.39
128	0.2	0.50217	0.50217	0.50217	3.25521	3.25521	3.25521	10.80

For model (5.3.2), $E(X_t^2)$ is given by

$$E(X_t^2) = (1-\alpha^2)^{-1}(1-\frac{2\alpha^2}{\pi})+[-(\frac{2}{\pi})^{1/2}\alpha(1-\alpha^2)^{-1/2}]^2$$

$$= \frac{1}{1-\alpha^2},$$

which is equal to the $E(X_t^2)$ of the linear model (5.3.1). We can also check that $E(X_t^4)$'s of these two models are equal for fixed α.

As further evidence of the efficiency of the conditional density method, we have computed the density and moments of the process considered by Jones (1976). This is of the form $X_t = \lambda(X_{t-1})+e_t$, where $e_t \sim N(0,1)$ and

$$\lambda(x) = \begin{cases} 0.5+x & \text{if } -1.5 < x \le 1 \\ -1 & \textit{otherwise.} \end{cases} \tag{5.3.3}$$

The results are given in Table (5.3.3).

Table 5.3.3

n	$E(X_t)$	$E(X_t^2)$	$E(X_t^3)$	$E(X_t^4)$	Skewness	Kurtosis	run-time (in sec.)
16	-0.18206	1.68061	-0.62693	7.87811	0.13195	-0.14362	5.78
20	-0.18232	1.68340	-0.65533	7.81880	0.11957	-0.18397	4.19
28	-0.18216	1.68417	-0.65196	7.83041	0.12083	-0.17972	4.58
40	-0.18216	1.68416	-0.65198	7.83036	0.12082	-0.17974	5.52
56	-0.18216	1.68416	-0.65198	7.83036	0.12082	-0.17974	9.64

Jones' values are: mean=-0.18216, variance=1.65098, skewness=0.12082 and kurtosis= -0.17974.

In order to investigate the affect of the probability density function of e_t on the conditional density method, it has been applied to obtain the unique stationary marginal density function of the SETAR(2;1,1) process

$$X_t = -0.2 \mid X_{t-1} \mid + e_t, \tag{5.3.4}$$

where e_t is Cauchy variable with density $\frac{1}{\pi(1+x^2)}$. Since the theoretical probability density function of model (5.3) with Cauchy innovations has a very heavy tail ($f(\pm 500) > 10^{-6}$), therefore we expect that the accuracy of the method in this case should not be as good as the previous experiments i.e. $e_t \sim N(0, \sigma^2)$. Although convergence is not achieved, we approximate the stationary probability density function of model (5.3.4) by 30-steps-ahead conditional density, with the error $\Sigma_i(\hat{f}(x_i)-f(x_i))^2 \cong 2.5\times10^{-6}$. The approximated and the theoretical stationary marginal probability density function of model (5.3.4) are shown by Fig. 5.3.1

. It seems that the two curves are almost identical. (See Table 5.3.4 for comparison of the approximated values i.e. $\hat{f}(x_i)$ and the theoretical values i.e. $f(x_i)$ of this experiment) . However, regardless of some difficulties in obtaining an adequate set of points $\{x_1, x_2, \cdots, x_n\}$ for evaluating the conditional densities numerically, it seems that the method is able to provide a good approximation of the stationary probability density function of a process when the kernel, $K(.)$, is not near zero outside the finite range of integration i.e. It is not close to zero on intervals $(-\infty, a)$ and $(b, , \infty)$ for moderate values of a and b , say $a \,\&\, b < 100$

Table 5.3.4

x_i	$f(x_i)$	$f(x_i)$	l error l	x_i	$f(x_i)$	$f(x_i)$	l error l
-1179.047852	0.000000	0.000000	0.000000	-300.525299	0.000004	0.000005	0.000001
-1095.159302	0.000000	0.000000	0.000000	-283.718109	0.000004	0.000006	0.000002
-1028.360107	0.000000	0.000000	0.000000	-267.533508	0.000005	0.000007	0.000002
-970.755676	0.000000	0.000001	0.000001	-251.951996	0.000005	0.000008	0.000003
-919.291016	0.000000	0.000001	0.000001	-236.955719	0.000012	0.000009	0.000003
-872.374756	0.000000	0.000001	0.000001	-222.528183	0.000008	0.000010	0.000002
-829.043030	0.000000	0.000001	0.000001	-208.654175	0.000009	0.000011	0.000002
-788.655396	0.000001	0.000001	0.000000	-195.319656	0.000014	0.000013	0.000001
-750.758301	0.000001	0.000001	0.000000	-182.511612	0.000014	0.000014	0.000000
-715.015625	0.000001	0.000001	0.000000	-170.218002	0.000013	0.000017	0.000004
-681.168945	0.000001	0.000001	0.000000	-158.427597	0.000022	0.000019	0.000003
-649.014404	0.000001	0.000001	0.000000	-147.130051	0.000017	0.000022	0.000005
-618.387329	0.000001	0.000001	0.000000	-136.315689	0.000033	0.000026	0.000007
-589.152222	0.000001	0.000001	0.000000	-125.975525	0.000025	0.000030	0.000005
-561.196045	0.000001	0.000002	0.000001	-116.101212	0.000030	0.000036	0.000006
-534.422974	0.000001	0.000002	0.000001	-106.684982	0.000049	0.000042	0.000007
-508.751160	0.000001	0.000002	0.000001	-97.719589	0.000048	0.000051	0.000003
-484.109711	0.000001	0.000002	0.000001	-89.198318	0.000052	0.000061	0.000009
-460.436890	0.000002	0.000002	0.000000	-81.114906	0.000062	0.000074	0.000012
-437.678467	0.000002	0.000003	0.000001	-73.463539	0.000078	0.000090	0.000012
-415.786530	0.000002	0.000003	0.000001	-66.238823	0.000101	0.000111	0.000010
-394.718384	0.000002	0.000003	0.000001	-59.435753	0.000131	0.000137	0.000006
-374.435944	0.000002	0.000003	0.000001	-53.049698	0.000170	0.000173	0.000003
-354.904907	0.000003	0.000004	0.000001	-47.076378	0.000245	0.000220	0.000025
-336.094360	0.000003	0.000004	0.000001	-41.511848	0.000290	0.000283	0.000007
-317.976318	0.000003	0.000005	0.000002	-36.352493	0.000387	0.000370	0.000017

Table 5.3.4 (continue)

x_i	$f(x_i)$	$f(x_i)$	| error |	x_i	$f(x_i)$	$f(x_i)$	| error |
-31.594995	0.000519	0.000491	0.000028	-4.382050	0.025372	0.025323	0.000049
-27.236332	0.000702	0.000663	0.000039	-4.299705	0.026289	0.026244	0.000045
-23.273764	0.000917	0.000911	0.000006	-4.213091	0.027308	0.027266	0.000042
-19.704826	0.001285	0.001275	0.000010	-4.122414	0.028438	0.028399	0.000039
-16.527319	0.001785	0.001819	0.000034	-4.027888	0.029690	0.029654	0.000036
-13.739298	0.002594	0.002641	0.000047	-3.929739	0.031076	0.031042	0.000034
-11.339070	0.003867	0.003889	0.000022	-3.828199	0.032607	0.032578	0.000029
-9.325185	0.005833	0.005760	0.000073	-3.723508	0.034302	0.034279	0.000023
-7.696431	0.008509	0.008452	0.000057	-3.615915	0.036179	0.036162	0.000017
-6.451829	0.012085	0.011986	0.000099	-3.505675	0.038261	0.038247	0.000014
-5.590613	0.015968	0.015869	0.000099	-3.393050	0.040569	0.040557	0.000012
-5.112079	0.018951	0.018872	0.000079	-3.278307	0.043130	0.043120	0.000010
-4.998263	0.019781	0.019708	0.000073	-3.161718	0.045972	0.045962	0.000010
-4.990850	0.019837	0.019764	0.000073	-3.043559	0.049129	0.049118	0.000011
-4.977533	0.019939	0.019866	0.000073	-2.924111	0.052638	0.052623	0.000015
-4.958341	0.020086	0.020014	0.000072	-2.803657	0.056540	0.056518	0.000022
-4.933317	0.020281	0.020209	0.000072	-2.682483	0.060878	0.060847	0.000031
-4.902522	0.020525	0.020453	0.000072	-2.560876	0.065700	0.065658	0.000042
-4.866028	0.020819	0.020748	0.000071	-2.439124	0.071058	0.071003	0.000055
-4.823923	0.021167	0.021097	0.000070	-2.317517	0.077007	0.076937	0.000070
-4.776305	0.021570	0.021501	0.000069	-2.196343	0.083602	0.083514	0.000088
-4.723289	0.022032	0.021964	0.000068	-2.075889	0.090897	0.090792	0.000105
-4.664998	0.022556	0.022490	0.000066	-1.956441	0.098943	0.098820	0.000123
-4.601573	0.023147	0.023084	0.000063	-1.838282	0.107783	0.107642	0.000141
-4.533163	0.023809	0.023750	0.000059	-1.721693	0.117445	0.117289	0.000156
-4.459931	0.024548	0.024494	0.000054	-1.606950	0.127938	0.127766	0.000172

Table 5.3.4 (continue)

x_i	$f(x_i)$	$f(x_i)$	|error|	x_i	$f(x_i)$	$f(x_i)$	|error|
-1.494325	0.139238	0.139054	0.000184	0.022467	0.252410	0.252191	0.000219
-1.384085	0.151284	0.151088	0.000196	0.041659	0.250185	0.249968	0.000217
-1.276492	0.163963	0.163757	0.000206	0.066683	0.247122	0.246908	0.000214
-1.171801	0.177104	0.176890	0.000214	0.097478	0.243121	0.242910	0.000211
-1.070261	0.190473	0.190251	0.000222	0.133972	0.238080	0.237875	0.000205
-0.972112	0.203774	0.203546	0.000228	0.176077	0.231918	0.231718	0.000200
-0.877586	0.216663	0.216428	0.000235	0.223695	0.224585	0.224392	0.000193
-0.786909	0.228765	0.228526	0.000239	0.276711	0.216085	0.215900	0.000185
-0.700295	0.239713	0.239471	0.000242	0.335002	0.206481	0.206304	0.000177
-0.617950	0.249181	0.248936	0.000245	0.398427	0.195905	0.195737	0.000168
-0.540069	0.256925	0.256679	0.000246	0.466837	0.184548	0.184389	0.000159
-0.466837	0.262813	0.262565	0.000248	0.540069	0.172649	0.172501	0.000148
-0.398427	0.266836	0.266589	0.000247	0.617950	0.160474	0.160337	0.000137
-0.335002	0.269109	0.268863	0.000246	0.700295	0.148289	0.148161	0.000128
-0.276711	0.269848	0.269604	0.000244	0.786909	0.136337	0.136219	0.000118
-0.223695	0.269341	0.269099	0.000242	0.877586	0.124824	0.124717	0.000107
-0.176077	0.267907	0.267669	0.000238	0.972112	0.113911	0.113812	0.000099
-0.133972	0.265872	0.265637	0.000235	1.070261	0.103704	0.103614	0.000090
-0.097478	0.263539	0.263307	0.000232	1.171801	0.094268	0.094186	0.000082
-0.066683	0.261172	0.260943	0.000229	1.276492	0.085626	0.085551	0.000075
-0.041659	0.258990	0.258763	0.000227	1.384085	0.077770	0.077702	0.000068
-0.022467	0.257165	0.256941	0.000224	1.494325	0.070671	0.070609	0.000062
-0.009150	0.255825	0.255602	0.000223	1.606950	0.064285	0.064228	0.000057
-0.001737	0.255054	0.254831	0.000223	1.721693	0.058558	0.058506	0.000052
0.001737	0.254686	0.254464	0.000222	1.838282	0.053435	0.053387	0.000048
0.009150	0.253888	0.253667	0.000221	1.956441	0.048859	0.048815	0.000044

Table 5.3.4 (continue)

x_i	$f(x_i)$	$f(x_i)$	| error |	x_i	$f(x_i)$	$f(x_i)$	| error |
2.075889	0.044773	0.044733	0.000040	4.723289	0.011382	0.011371	0.000011
2.196343	0.041127	0.041090	0.000037	4.776305	0.011157	0.011146	0.000011
2.317517	0.037872	0.037838	0.000034	4.823923	0.010959	0.010949	0.000010
2.439124	0.034965	0.034933	0.000032	4.866028	0.010789	0.010779	0.000010
2.560876	0.032366	0.032337	0.000029	4.902522	0.010645	0.010635	0.000010
2.682483	0.030041	0.030013	0.000028	4.933317	0.010526	0.010516	0.000010
2.803657	0.027958	0.027932	0.000026	4.958341	0.010430	0.010420	0.000010
2.924111	0.026090	0.026066	0.000024	4.977533	0.010358	0.010348	0.000010
3.043559	0.024412	0.024390	0.000022	4.990850	0.010308	0.010298	0.000010
3.161718	0.022904	0.022883	0.000021	4.998263	0.010280	0.010270	0.000010
3.278307	0.021546	0.021526	0.000020	5.112079	0.009870	0.009861	0.000009
3.393050	0.020323	0.020303	0.000020	5.590613	0.008388	0.008380	0.000008
3.505675	0.019219	0.019200	0.000019	6.451829	0.006446	0.006439	0.000007
3.615915	0.018222	0.018204	0.000018	7.696431	0.004641	0.004636	0.000005
3.723508	0.017321	0.017304	0.000017	9.325185	0.003233	0.003229	0.000004
3.828199	0.016506	0.016490	0.000016	11.339070	0.002228	0.002226	0.000002
3.929739	0.015769	0.015754	0.000015	13.739298	0.001542	0.001540	0.000002
4.027888	0.015101	0.015087	0.000014	16.527319	0.001079	0.001078	0.000001
4.122414	0.014498	0.014484	0.000014	19.704826	0.000767	0.000766	0.000001
4.213091	0.013952	0.013938	0.000014	23.273764	0.000555	0.000554	0.000001
4.299705	0.013458	0.013445	0.000013	27.236332	0.000408	0.000407	0.000001
4.382050	0.013013	0.013000	0.000013	31.594995	0.000305	0.000305	0.000000
4.459931	0.012611	0.012599	0.000012	36.352493	0.000231	0.000231	0.000000
4.533163	0.012250	0.012239	0.000011	41.511848	0.000178	0.000178	0.000000
4.601573	0.011927	0.011916	0.000011	47.076378	0.000139	0.000139	0.000000
4.664998	0.011638	0.011627	0.000011	53.049698	0.000110	0.000110	0.000000

Table 5.3.4 (continue)

x_i	$f(x_i)$	$f(x_i)$	|error|	x_i	$f(x_i)$	$f(x_i)$	|error|
59.435753	0.000088	0.000088	0.000000	394.718384	0.000002	0.000002	0.000000
66.238823	0.000071	0.000071	0.000000	415.786530	0.000002	0.000002	0.000000
73.463539	0.000058	0.000058	0.000000	437.678467	0.000002	0.000002	0.000000
81.114906	0.000047	0.000047	0.000000	460.436890	0.000001	0.000001	0.000000
89.198318	0.000039	0.000039	0.000000	484.109711	0.000001	0.000001	0.000000
97.719589	0.000033	0.000033	0.000000	508.751160	0.000001	0.000001	0.000000
106.684982	0.000028	0.000028	0.000000	534.422974	0.000001	0.000001	0.000000
116.101212	0.000023	0.000023	0.000000	561.196045	0.000001	0.000001	0.000000
125.975525	0.000020	0.000020	0.000000	589.152222	0.000001	0.000001	0.000000
136.315689	0.000017	0.000017	0.000000	618.387329	0.000001	0.000001	0.000000
147.130051	0.000015	0.000015	0.000000	649.014404	0.000001	0.000001	0.000000
158.427597	0.000013	0.000013	0.000000	681.168945	0.000001	0.000001	0.000000
170.218002	0.000011	0.000011	0.000000	715.015625	0.000001	0.000001	0.000000
182.511612	0.000009	0.000009	0.000000	750.758301	0.000001	0.000001	0.000000
195.319656	0.000008	0.000008	0.000000	788.655396	0.000001	0.000001	0.000000
208.654175	0.000007	0.000007	0.000000	829.043030	0.000000	0.000000	0.000000
222.528183	0.000006	0.000006	0.000000	872.374756	0.000000	0.000000	0.000000
236.955719	0.000006	0.000006	0.000000	919.291016	0.000000	0.000000	0.000000
251.951996	0.000005	0.000005	0.000000	970.755676	0.000000	0.000000	0.000000
267.533508	0.000004	0.000004	0.000000	1028.360107	0.000000	0.000000	0.000000
283.718109	0.000004	0.000004	0.000000	1095.159302	0.000000	0.000000	0.000000
300.525299	0.000004	0.000003	0.000001	1179.047852	0.000000	0.000000	0.000000
317.976318	0.000003	0.000003	0.000000				
336.094360	0.000003	0.000003	0.000000				
354.904907	0.000003	0.000003	0.000000				
374.435944	0.000002	0.000002	0.000000				

We now illustrate results concerning the stationary and conditional densities etc.

Example 1: Let $\{X_t\}$ satisfy the SETAR(2;1,1) model

$$X_t = \begin{cases} 1.5-0.9X_{t-1}+e_t & \text{if } X_{t-1}\le 0 \\ -0.4-0.6X_{t-1}+e_t & \text{if } X_{t-1}>0\,, \end{cases} \tag{5.3.5}$$

where $e_t \sim N(0,1)$. This model has a limit cycle of period 2 at $C=\{-2.826,4.043\}$. The stationary marginal

probability density function of model (5.3.5) is shown in Fig. 5.3.2 when $\sigma^2=1$ and in Fig. 5.3.3 when $\sigma^2=4$ and in Fig. 5.3.4 when e_t is a Cauchy variable with density $\frac{1}{\pi(1+x^2)}$.

Also we plot the conditional density of X_{t+m} given $X_t=1$ for $m=1,2,3,4,5$ and $\sigma^2=4$ to highlight the non-normality and complexity of these densities for $m>2$. Obviously the conditional density of X_t given X_{t-1} is normal. The results are shown in Figs 5.3.5-5.3.9 . The joint probability density function of (X_t,X_{t-1}) may be found by using

$$f_{t+1,t}(x,y)=K_j(x-\lambda_j(y))f(y)\,,\ y\in(r_{j-1},r_j]$$

and since the process is Markovian, this can be extended to obtain the joint density of $(X_t,X_{t-1},\cdots,X_{t-k})$, $k\geq 1$. The joint density function of (X_t,X_{t-1}) for $\sigma^2=1$ is given in Figs 5.3.10 and the joint density function of (X_t,X_{t-2}) for $\sigma^2=1$ is given in Fig. 5.3.11 .

Example 2: Let

$$X_t=\left\{0.1+(0.1-20.0X_{t-1})e^{-X_{t-1}^2}\right\}X_{t-1}+e_t\,, \tag{5.3.6}$$

where $e_t\sim N(0,1)$. The stationary probability density function of this model is given by Fig. 5.3.12 .

Let

$$X_t=\lambda(X_{t-1},X_{t-2})+e_t\,,\ t=0,\pm1,..., \tag{5.3.7}$$

where $\lambda(.,.)$ is a fixed real function of real arguments and $\{e_t\}$ is a sequence of independent and identically distributed random variables with probability density function $K(.)$. Denote the conditional density of X_{t+m},X_{t+m-1} given X_{t-1},X_{t-2} by $f_m(x,y)=f_{X_{t+m},X_{t+m-1}|X_{t-1},X_{t-2}}(x\,|\,x_{t-1}x_{t-2})$. It can be shown, by arguments similar to those of order one case, that the following recursive formula holds

$$f_{m+1}(x,y)=\int_R f_m(y,z)K(x-\lambda(y,z))dz\,, \tag{5.3.8}$$

and

$$f_{m+1}(x)=\int_R f_{m+1}(x,y)dy\,,$$

where $f_1(x,y)=f_{X_{t+1},X_t|X_{t-1}=x_{t-1},X_{t-2}=x_{t-2}}(x,y\,|\,X_{t-1}=x_{t-1},X_{t-2}=x_{t-2})$

$$=K(x-\lambda(y,x_{t-1}))K(y-\lambda(x_{t-1},x_{t-2}))\,.$$

Example 3: As an application of our method for order two case, we consider the following SETAR(2;2,2)

model

$$X_t = \begin{cases} 0.62+1.25X_{t-1}-0.43X_{t-2}+e_t & \text{if } X_{t-2} \le 3.25 \\ 2.25+1.52X_{t-1}-1.24X_{t-2}+e_t & \text{if } X_{t-2} > 3.25 , \end{cases} \tag{5.3.9}$$

where $\{e_t\}$ is a sequence of independent and identically distributed random variables $N(0,0.0503)$, Tong (1983). Our numerical method gives the joint densities of (X_t, X_{t-1}), (X_t, X_{t-2}), (X_t, X_{t-3}) which are given by Figs. 5.3.13-5.3.15 and the conditional expectations of $(X_t | X_{t-j})$, j=1,2,3 which are given by Figs 5.3.16-5.3.18 .

5.4. Accelerating by Matrix Squaring

In the above experiments we have found that a judicious choice of the set of grid points x_i may increase the accuracy and improve the convergence of the method. Moreover, when the model has a limit point or its autoregressive function is continuous, convergence is achieved after at most 10 steps. When the model has a limit cycle and the variance of e_t is small the convergence rate is slower. In all our experiments convergence is always achieved.

However, by comparing the numerical integration of (5.2.1) with the numerical eigenvalue problem, it is clear that we may consider adopting the matrix-squaring method (See, eg. Wilkinson, p. 615) from the latter. Specifically, instead of iterating (5.2.1) over $m=1,2,3,....$, we may calculate $f(x_{t+2^m} | x_t)$ for $m=0,1,2,....$ Intermediate conditional densities may be calculated in an obvious economical way. For example, we may calculate $f(x_{t+6} | x_t)$ by .

$$f(x_{t+6} | x_t) = \int_{-\infty}^{\infty} f(x_{t+6} | x_{t+2}) f(x_{t+2} | x_t) \, dx_{t+2} ,$$

which involved $f(x_{t+2^2} | x_t)$ and $f(x_{t+2^1} | x_t)$ already calculated. The number of iterations is nominally reduced from m to $\log_2 m$. However, our experience suggests that whether there is any reduction in the computing time depends to a large extent on the ratio $|\lambda_1/\lambda_2|$, where $|\lambda_1|$ is the largest absolute value of the eigenvalues and $|\lambda_2|$ the next largest. Here, the eigenvalues correspond to the 'transition matrix' whose i-j th entry is $f(x_{t+1} | x_t)$ with x_t set equal to ξ_i and x_{t+1} set equal to ξ_j, $\{\xi_1, \xi_2, \cdots, \xi_N\}$ being the points selected for the numerical integration described in §5.2 . Our observations are as expected by analogy with the situation in the numerical eigenvalue problem. (Wilkinson, op.cit.) That is, there a substantial saving in computing time if $|\lambda_1/\lambda_2|$ is close to unity. Otherwise there need not be any saving at all. We illustrate our

observations in Table 5.4 and 5.5 below.

Table 5.4

Numerical Evaluation of stationary Density of the AR(1) Model

$$X_t = 0.5X_{t-1}+e_t ,$$

where $e_t \sim N(0,1)$.

N	$\lvert\lambda_1\rvert$	$\lvert\lambda_2\rvert$	$\frac{\lvert\lambda_1\rvert}{\lvert\lambda_2\rvert}$	run time without matrix squaring	run time with matrix squaring
24	3.896	0.938	4.1	1.4	1.3
32	5.542	1.303	4.2	2.0	2.4
64	9.190	2.184	4.2	4.4	11.9

Table 5.5

Numerical Evaluation of the Stationary Density of the

$$X_t = \begin{cases} 1.5-0.9X_{t-1}+e_t & \text{if } X_{t-1} \le 0 \\ -0.4-0.6X_{t-1}+e_t & \text{if } X_{t-1} > 0 , \end{cases}$$

where $e_t \sim N(0,1)$.

N	$\lvert\lambda_1\rvert$	$\lvert\lambda_2\rvert$	$\frac{\lvert\lambda_1\rvert}{\lvert\lambda_2\rvert}$	run time without matrix squaring	run time with matrix squaring
24	1.496	0.973	1.5	76.9	1.2
32	2.482	1.544	1.6	85.0	2.5
64	5.038	3.063	1.6	160.0	20.9

However, we have observed in the above experiments that

(i) the variance of e_t and the number of points, N, do not affect the ratio of $\lvert\lambda_1/\lambda_2\rvert$ in the linear case. However, in the SETAR case N does not affect the ratio of $\lvert\lambda_1/\lambda_2\rvert$ but the variance of e_t does.

(ii) In the linear case, the run time with $Var(e_t)=0.01$, are approximately the same as those with $Var(e_t)=1$.

(iii) In the non-linear case, when the $Var(e_t)$ is reduced the method without matrix squaring does not converge but the method with matrix squaring converges whether $Var(e_t)=1$ or $Var(e_t)=0.01$ and the run time are similar.

(iv) In the non-linear case with $Var(e_t)=0.01$, the ratio $|\lambda_1/\lambda_2|$ is approximately equal to one.

: Stationary p.d.f of model (5.3.4). . Stationary p.d.f. of model (5.3.5) for σ_e^2=

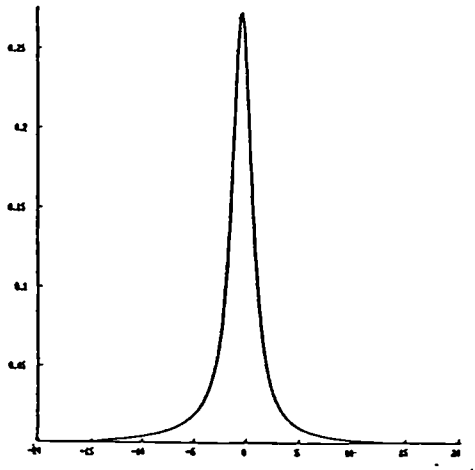


Fig. 5.3.1

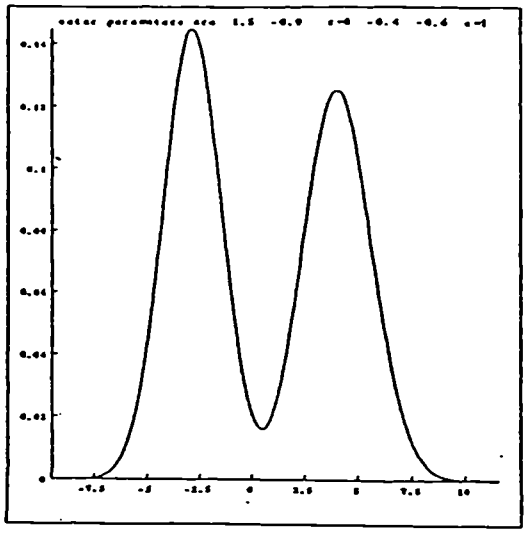


Fig. 5.3.2

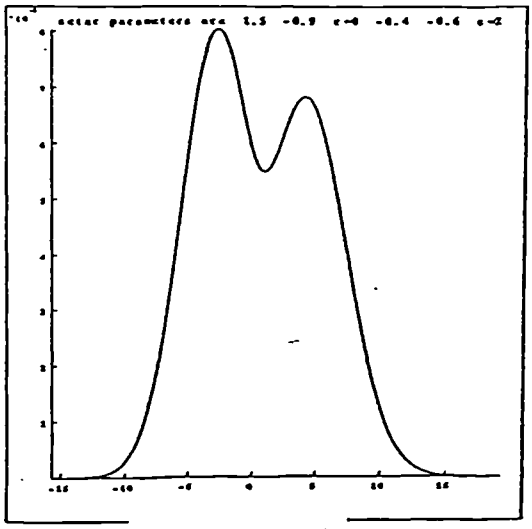


Fig. 5.3.3

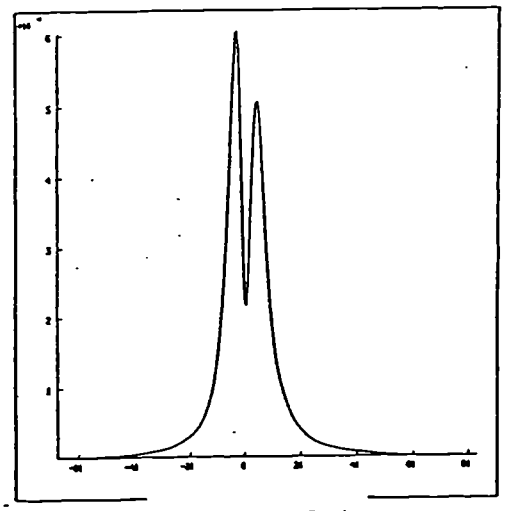
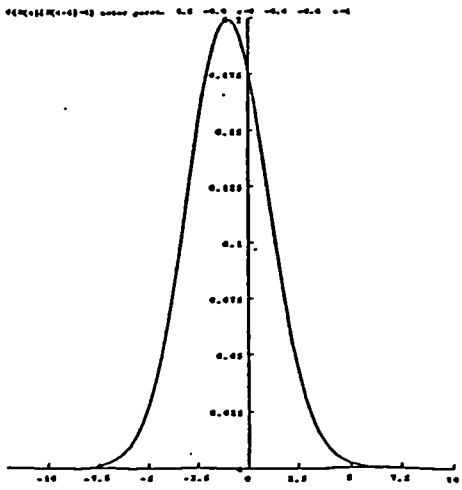
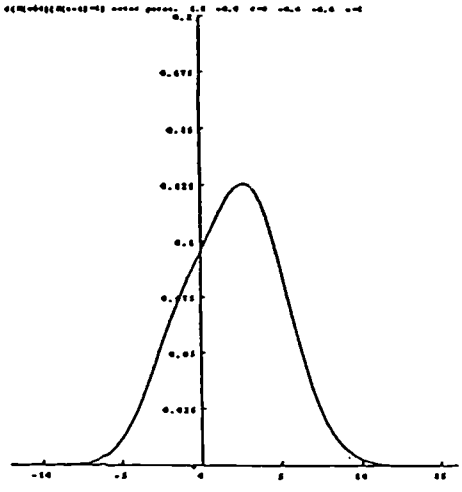


Fig. 5.3.4

Cauchy innovation.

: Stationary p.d.f. of model (5.3.5) for σ_e^2=4. Stationary p.d.f. of model (5.3.5) with

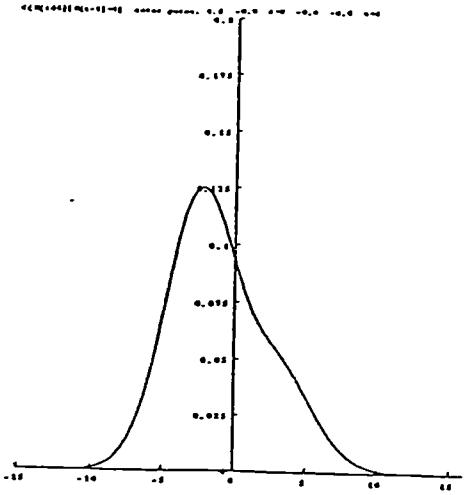


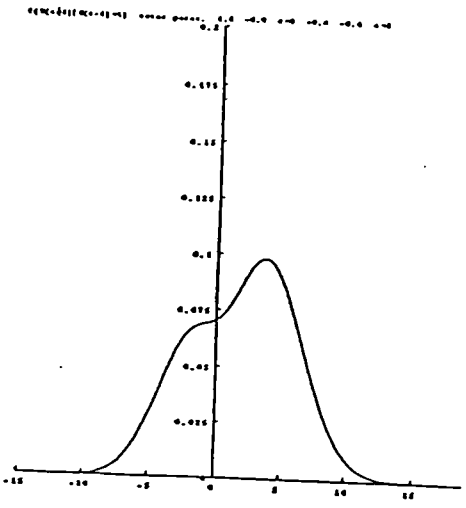


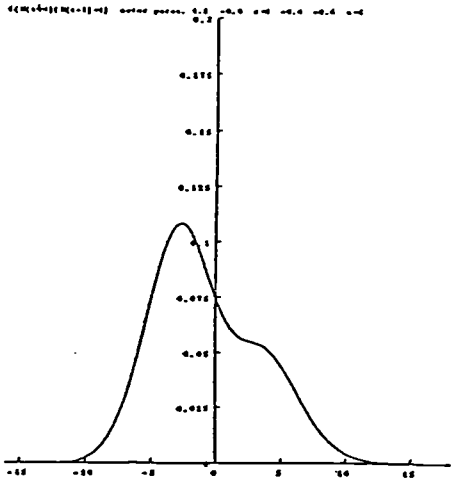
Figs. 5.3.5-5.3.6

Fig. 5.3.5 : $f(X_t \mid X_{t-1}=1)$ of model (5.3.5).

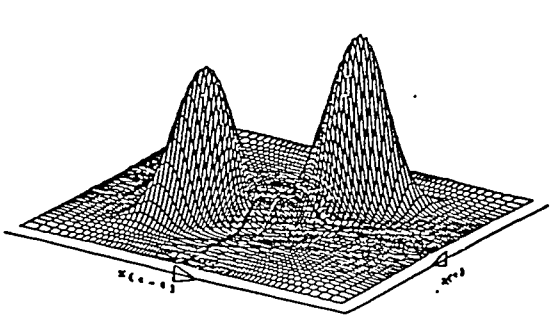
Fig. 5.3.6 : $f(X_t \mid X_{t-2}=1)$ of model (5.3.5).

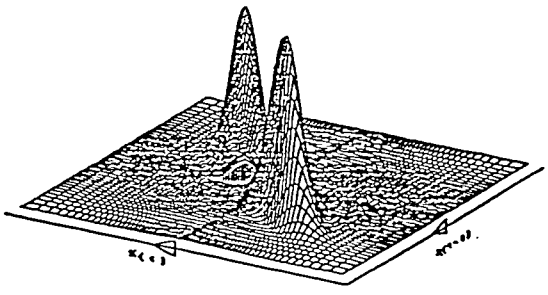






Figs. 5.3.7-5.3.9





Figs. 5.3.10

Fig. 5.3.7 : $f(X_t \mid X_{t-3}=1)$ of model (5.3.5).

Fig. 5.3.8 : $f(X_t \mid X_{t-4}=1)$ of model (5.3.5).

Fig. 5.3.9 : $f(X_t \mid X_{t-5}=1)$ of model (5.3.5).

Fig. 5.3.10 : $f(X_t , X_{t-1})$ of model (5.3.5).

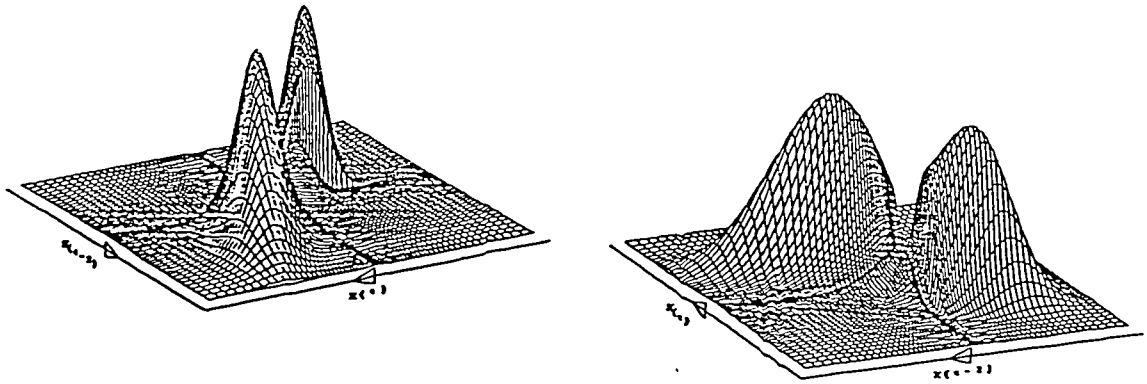


Fig. 5.3.11 : $f(X_t , X_{t-2})$ of model (5.3.5).

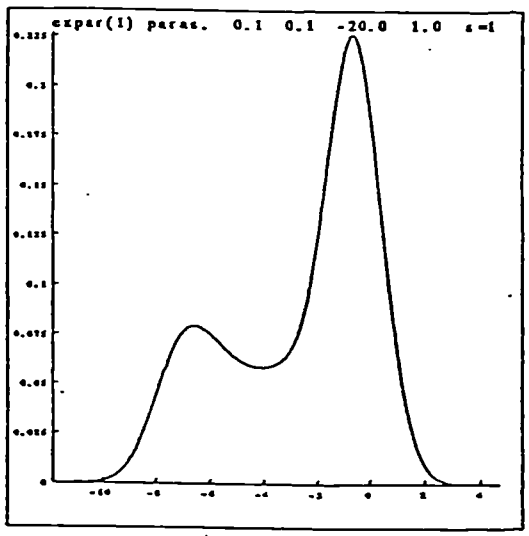


Fig. 5.3.12 : Stationary p.d.f. of model (5.3.6).

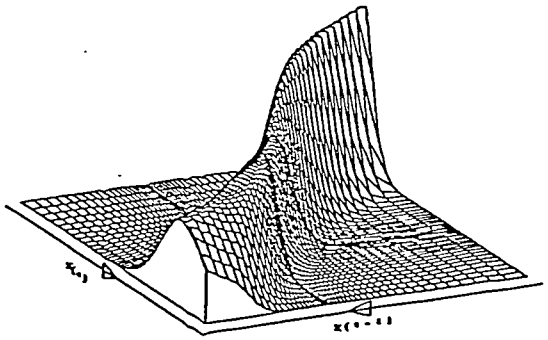
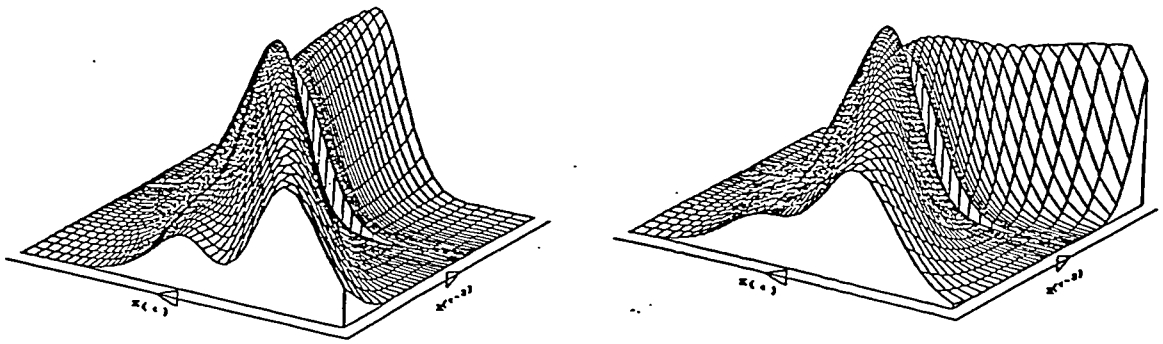


Fig. 5.3.13 : $f(X_t , X_{t-1})$ of model (5.3.9).



Figs. 5.3.14-5.3.15

Fig. 5.3.14 : $f(X_t , X_{t-2})$ of model (5.3.9).

Fig. 5.3.15 : $f(X_t , X_{t-3})$ of model (5.3.9).

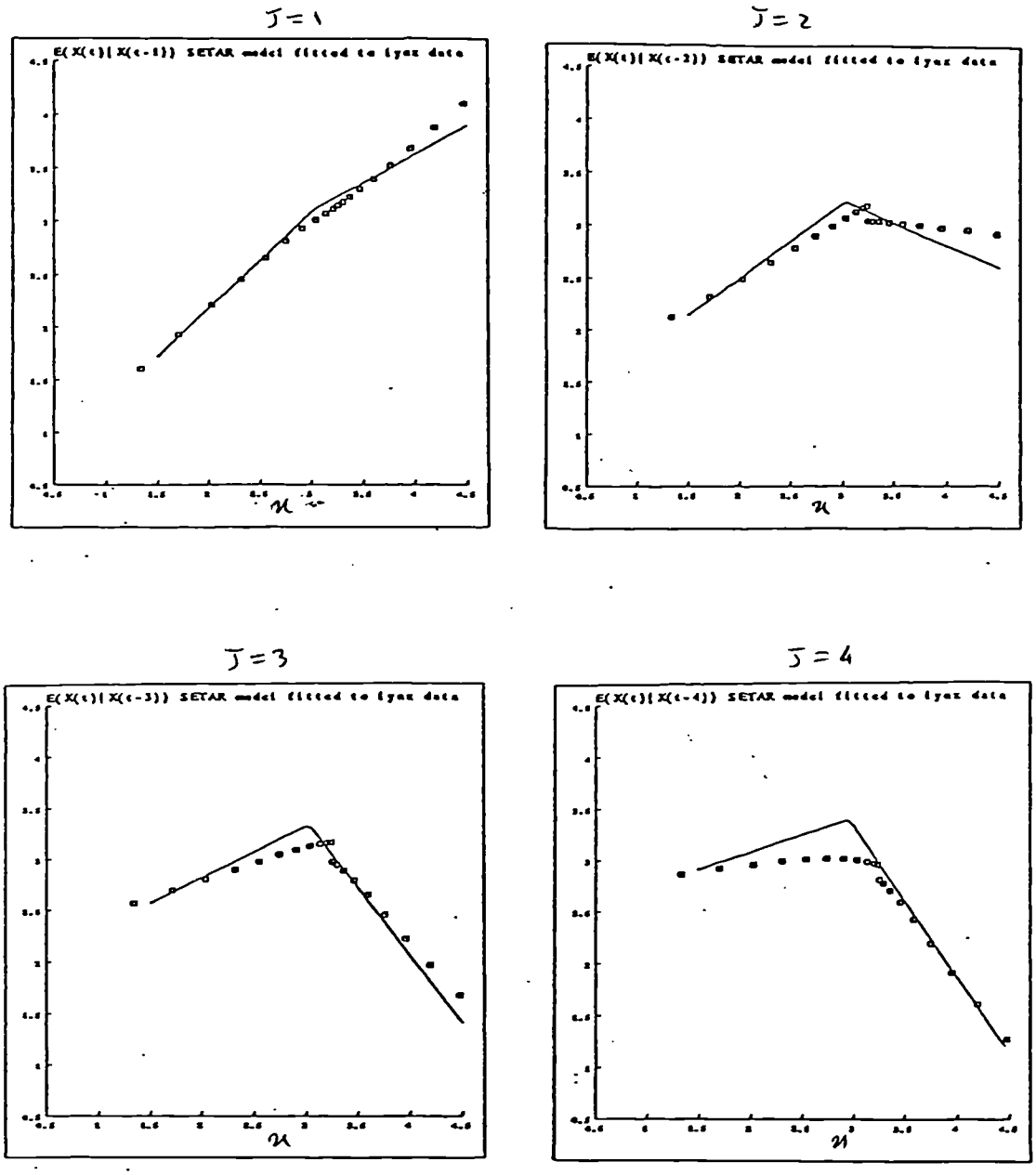


Fig 5.3.16-5.3.18 : (□) exact and (solid line) Tong's approximation values of $E(X_t \mid X_{t-j}=x)$, $j=1,2,3,4$ of model (5.3.9).

CHAPTER SIX

CONDITIONAL MEAN AND CONDITIONAL VARIANCE IN TIME SERIES ANALYSIS

Graphical methods should form a very important part of identification and diagnostic checking of non-linear time series models. Graphical methods found useful in building linear time series models should also be used in building non-linear time series models. (See Tong 1987a).

The importance of non-parametric lag regression in non-linear time series modelling, has been emphasised by Tong (1983). In this chapter the lag regression function of non-linear autoregressive models and bilinear models are considered. Post-model examination of observed non-parametric lag regression versus the fitted non-parametric lag regressions of some models which were fitted to the real data are also included. The numerical integration technique is employed to obtain the exact lag regression function of some simple non-linear autoregressive models.

The problem of evaluating the regression functions of some of the non-linear models explicitly is extremely involved and usually intractable. It seems that the non-parametric estimates of regression functions can provide valuable information. (See e.g. Robinson P.M., 1983).

Non-parametric estimates of the regression function of X_t on $X_t \pm j$, ($j=1,2...,p$, say) provide some information about the non-linearity of the time series. Let $m_j(x)$ denote $E(X_t \mid X_{t+j}=x)$. It is suggested that the estimates of these $m_j(x)$'s, denoted by $\hat{m}_j(x)$'s, may be useful for the identification of the delay parameter in SETAR models. (See Tong, 1983).

Let $\{\delta_N(z)\}$ be a sequence of non-negative functions of z, of total area unity. As $N \to \infty$, $\delta_N(z)$ is to tend to the Dirac delta function. Such a sequence may be realized by introducing a fixed function $k(z) \geq 0$, $\int k(z)dz = 1$ and a B_N so that the family is $\{B_N^{-1} k(\frac{z}{B_n})\}$, i.e. $\delta_N(z) = B_N^{-1} k(\frac{z}{B_N})$, with $B_N \to 0$ as

$N \to \infty$. For convenience, we further require $\int zk(z)dz = 0$. Given the observations $\{x_1, x_2, \cdots, x_N\}$, a kernel estimate of $m_j(x)$ may be given by, for $j = 1,2,...,p$, $p << N$,

$$\hat{m}_{j,N}(x) = \frac{\sum_{t=1}^{N-J} x_t \delta_N(x - x_{t+j})}{\sum_{t=1}^{N-j} \delta_N(x - x_{t+j})}$$

and

$$\hat{m}_{-j,N}(x) = \frac{\sum_{t=j+1}^{N} x_t \delta_N(x - x_{t-j})}{\sum_{t=j+1}^{N} \delta_N(x - x_{t-j})} .$$

The second suffix, namely N, emphasizes the dependence of the estimate on the sample size. A simple form of $k(z)$ which we use in our numerical study is given by

$$k(z) = \begin{cases} 1 - |z| & \text{for } |z| \le 1 \\ 0 & \textit{otherwise} . \end{cases}$$

In practice, estimates corresponding to the two ends of the horizontal axis generally tend to exhibit greater sample fluctuations. In all our non-parametric estimates, we choose the standard deviation of observations as the smoothing parameter B_N.

6.1. Gaussian Processes

Consider a Gaussian process, $\{X_t\}$. For any admissible subset of two elements $\{r, s\}$, the joint probability distribution of (X_r, X_s) is a bivariate normal distribution. Hence the conditional probability distribution of X_s given X_r is normal and the conditional expected value and the conditional variance of X_s given $X_r = x$ are linear in x and constant respectively. In linear Gaussian processes, it is well known that processes are time-reversible i.e. ($E(X_s | X_r = x) = E(X_r | X_s = x)$) for any r and s. (See Weiss, 1975).

A real process, $\{X_t\}$, may be considered the transformation of a white noise process, $\{e_t\}$, by an unknown function, say $\Gamma(B)$. If the unknown function is linear and $\{X_t\}$ is stationary and Gaussian, then the regression function of X_s on X_r, for any s and r, of the process should appear linear and the conditional variances should be constant. Thus, if the plot of $E(X_s | X_r = x)$ and $Var(X_s | X_r = x)$ against x does not appear linear, then the linearity assumption of the function is questionable. Also, the two-dimensional histogram of X_r and X_s, for any r and s, of an observed process $\{X_t\}$ gives some preliminary indication about

the joint probability distribution of X_r and X_s.

6.2. Bilinear Models

As far as the bilinear models are concerned, obtaining analytic expression for the conditional mean and the conditional variance of these models is not always easy except for some special cases. In order to get some idea about the behaviour of regression functions and the variance of X_t given X_{t-j} , $j=1,2,...$, we will study some theoretical properties of $E(X_t|X_{t-j})$ and $Var(X_t|X_{t-j})$ of some sub-classes of bilinear models. In all cases the non-parametric estimates are also considered.

6.2.1. Superdiagonal Bilinear Models

In this section the regression function of the easiest class of bilinear models, which takes the following general form, is considered.

$$X_t = \sum_{\substack{k=1 \\ k<l}}^{Q} \sum_{l=2}^{P} \beta_{kl} e_{t-k} X_{t-l} + e_t \,, \tag{6.2.1.1}$$

where e_t is strict white noise, and as a special case the model

$$X_t = \beta X_{t-3} e_{t-2} + e_t \,, \tag{6.2.1.2}$$

will be used throughout this section as a particular example.

Assuming $E(e_t)=0$, it follows that $E(X_t)=0$ for all t. Obviously the conditional expectation of X_t given X_{t-j} of model (6.2.1.1) are equal to zero for all $j>k$. The regression functions, $E(X_t|X_{t-j}=x)$ for $j=1,2,...,k$, are involved with the terms $E(X_{t-l}|X_{t-j}=x)$ and $E(e_{t-k}|X_{t-j}=x)$ which because of the dependence of X_{t-l} with X_{t-j} and e_{t-k} with X_{t-j}, cannot be evaluated explicitly.

In order to examine the properties of regression functions of model (6.2.1.1) approximately, we generate a sample of size 5000 observations using a simulation technique, of model (6.2.1.2) with $\beta=0.5$ and $\beta=-0.5$, and calculate $\hat{m}_{-j}$, $j=1,2,3,4$. The results are given in Figs. (6.2.1.1). It seems that regardless of the sign of β, the expectations of X_t given X_{t-j} , $j=1,2,3,4$ are all very close to zero and none of them shows any obvious regular pattern.

Obtaining the conditional variance of X_t given X_{t-j} for the model (6.2.1.1) is more complicated. However, for model (6.2.1.2), $\hat{m}_{-j}=0$, $j=3$. Therefore

$$\begin{aligned} Var(X_t \mid X_{t-3}) &= E(X_t^2 \mid X_{t-3}=x) \\ &= \beta^2 x^2 E(e_{t-2}^2 \mid X_{t-3}=x)+E(e_t^2 \mid X_{t-3}=x) \\ &= (1+\beta^2 x^2)\sigma_e^2 , \end{aligned} \tag{6.2.1.3}$$

which is a quadratic function in x. The results of non-parametric estimates of conditional variances of model (6.2.1.2) are given by Figs. 6.2.1.2. The agreement of the non-parametric estimate of $Var(X_t \mid X_{t-3})$ with formula (6.2.1.3) is quite remarkable. The graph of $Var(X_t \mid X_{t-2}=x)$ exhibits some non-linearity in x which may be related to the dependence of X_{t-3} on X_{t-2}.

6.2.2. Diagonal Models

In this section a group of models are considered which are possibly more interesting than the superdiagonal models. Let

$$X_t = \sum_{k=1}^{P} \beta_k e_{t-k} X_{t-k} + e_t , \tag{6.2.2.1}$$

where e_t is strict white noise. As an especial case the model

$$X_t = \beta X_{t-3} e_{t-3} + e_t \tag{6.2.2.2}$$

will be used throughout this section. Suppose

$$X_t = \beta X_{t-k} e_{t-k} + e_t , \; k \geq 1 \tag{6.2.2.3}$$

then X_t can be written as

$$\begin{aligned} X_t &= \beta[\beta X_{t-2k} e_{t-2k} + e_{t-k}] e_{t-k} + e_t \\ &= \beta^2 X_{t-2k} e_{t-2k} e_{t-k} + \beta e_{t-k}^2 + e_t . \end{aligned}$$

Hence

$$\begin{aligned} E(X_t \mid X_{t-j}=x) &= \beta^2 E(X_{t-2k} e_{t-2k} \mid X_{t-j}=x) E(e_{t-k} \mid X_{t-j}=x) \\ &\quad + \beta E(e_{t-k}^2 \mid X_{t-j}=x) \end{aligned}$$

which is equal to $\beta\sigma_e^2$ for $j > k$.

For $j=k$ we have

$$\begin{aligned} E(X_t \mid X_{t-k}=x) &= E(\beta X_{t-k} e_{t-k} \mid X_{t-k}=x) \\ &= \beta x E(e_{t-k} \mid X_{t-k}=x) \\ &= \beta x E[X_{t-2k} - \beta X_{t-2k} e_{t-2k} \mid X_{t-k}=x] \\ &= \beta x [x - \beta E(X_{t-2k} e_{t-2k} \mid X_{t-k}=x] \\ &= \beta x^2 - \beta^2 x E(X_{t-2k} e_{t-2k} \mid X_{t-k}=x) \end{aligned}$$

which is a quadratic function in x when we approximate the term $E(X_{t-2k} e_{t-2k} \mid X_{t-k}=x)$ by

$$\beta E(X_{t-3k}e_{t-3k}\,|\,X_{t-k}=x)E(e_{t-2k}\,|\,X_{t-k}=x)+E(e_{t-2k}^2\,|\,X_{t-k}=x)$$

which for β small, is approximately equal to $E(e_{t-2k}^2\,|\,X_{t-k}=x)\approx C>0$,*say*.

For $k>j$, $E(X_t\,|\,X_{t-j}=x)$ involves the term $E(e_{t-k}^2\,|X_{t-j}=x)$ which it seems could be approximated by σ_e^2 and the term $\beta^2E(e_{t-k}\,|\,X_{t-j}=x)$ which is close to zero. However, the non-parametric estimates of regression functions of model (6.2.2.3) show that $E(X_t\,|\,X_{t-j}=x)\approx\beta\sigma_e^2$ for $k\neq j$ and is a quadratic function in x for $k=j$. Figs. 6.2.2.1 show the non-parametric estimates of regression functions of model (6.2.2.3) with $\beta=0.5$ and -0.5.

With similar arguments, one can show that the conditional variance of $X_t\,|\,X_{t-j}=x$, j=1,2,4,5 are approximately constant. Figs. (6.2.2.2) show the non-parametric estimates of these conditional variances. It can be seen that $Var(X_t\,|\,X_{t-3}=x)$ is a non-linear (quadratic) function in x.

6.2.3. Subdiagonal Models

The subdiagonal models which will be considered in this section are those of the form

$$X_t=\beta X_{t-k}e_{t-l}+e_t\,, \tag{6.2.3.1}$$

where $k<l$, e_t is strict white noise, and a special case is

$$X_t=\beta X_{t-2}e_{t-3}+e_t\,, \tag{6.2.3.2}$$

which will be used through this section as a particular example. Assuming $E(e_t)=0$, it is easy to show that $E(X_t)=0$ for all k , $l>0$.

The conditional expectation of X_t given $X_{t-j}=x$ for model (6.2.3.2) takes the form

$$\begin{aligned}E(X_t\,|\,X_{t-j}=x)&=\beta E(X_{t-2}e_{t-3}\,|\,X_{t-j}=x)\\&=\beta E((\beta X_{t-4}e_{t-5}+e_{t-2})e_{t-3}\,|\,X_{t-j}=x)\\&=\beta^2E(X_{t-4}e_{t-5}\,|\,X_{t-j}=x)E(e_{t-3}\,|\,X_{t-j}=x)\\&+\beta E(e_{t-2}\,|\,X_{t-j}=x)E(e_{t-3}\,|\,X_{t-j}=x)\\&=0\text{ for }j>3\,.\end{aligned}$$

For $j\leq3$, the exact value of regression function does not seem to be easy to obtain. However, for model (6.2.3.2), the non-parametric approximation of regression functions for $\beta=0.5$ and -0.5 show that $E(X_t\,|\,X_{t-j}=x)\approx0$ for all $j>0$.

If we accept that the conditional mean of the model (6.2.3.1) is equal to zero, then the variance of X_t

given $X_{t-j}=x$ is equal to $E(X_t^2 \mid X_{t-j}=x)$. For $j=2$ we have

$$\begin{aligned} Var(X_t \mid X_{t-2}=x) &= \beta^2 E(X_{t-2}^2 e_{t-3}^2 \mid X_{t-2}=x)+\sigma_e^2 \\ &= \beta^2 x^2 E(e_{t-3}^2 \mid X_{t-2}=x)+\sigma_e^2 , \end{aligned}$$

which is a non-linear function in x. The non-parametric estimates of regression functions and conditional variances of model (6.2.3.2) are given by Figs. 6.2.3.1 and 6.2.3.2 respectively.

6.3. Non-linear Autoregressive Models

The class of non-linear autoregressive models has a very close connection with the analysis of cyclical data. A general univariate NLAR(p) process may be written as

$$X_t = \lambda(\mathbf{X}_{t-1})+e_t , \tag{6.3.1}$$

where $\mathbf{X}_{t-1}=(X_{t-1}, \cdots, X_{t-p})'$, $\{e_t\}$ is a sequence of independent and identically distributed scalar random variables with mean zero and constant variance and e_t is independent of X_s for all $s < t$.

The conditional density function of X_{t+1} given $\mathbf{X}_t$ can be written as

$$f_{X_{t+1} \mid \mathbf{X}_t}(x \mid \mathbf{X}_t=\mathbf{y}) = k(x-\lambda(\mathbf{y})) , \tag{6.3.2}$$

where $k(.)$ denotes the probability density function of e_t. Obviously the conditional mean and the conditional variance of X_{t+1} given $\mathbf{X}_t=\mathbf{y}$ are equal to $\lambda(\mathbf{y})$ and σ_e^2 respectively.

Since we have a Markov chain over $\mathbf{R}^p$, we may recall the Chapman-Kolmogorov relation

$$f(x_{t+m} \mid \mathbf{x}_t) = \int_{-\infty}^{\infty} f(x_{t+m} \mid \mathbf{x}_{t+1}) f(x_{t+1} \mid \mathbf{x}_t) dx_{t+1} \tag{6.3.3}$$

where $f(x_s \mid \mathbf{x}_t)$ denotes the conditional probability density function of X_s given $\mathbf{X}_t=\mathbf{x}_t$. Let $f_m(x)$ denote the conditional density function of X_{t+m} given $\mathbf{X}_t$. Equation (6.3.3) immediately gives

$$f_m(x) = \int_{-\infty}^{\infty} f_{m-1}(\mathbf{y}) k(x-\lambda(\mathbf{y}) d\mathbf{y} . \tag{6.3.4}$$

Equation (6.3.4) gives, in particular a recursive formula for evaluating the conditional density with $f_1(x) = k(x-\lambda(\mathbf{y}))$.

By using the results of Chapter Five, we may employ the numerical integration method to obtain the conditional mean and conditional variance of X_{t+m} given $\mathbf{X}_t$ for $m \geq 2$. Let us first restrict our discussion to the first order case:

$$X_t = \lambda(X_{t-1}) + e_t \,, \; t = 0, \pm 1, \cdots \,,$$

where $\{e_t\}$ is a sequence of independent and identically distributed normal random variables with $E(e_t) = 0$ and $Var(e_t) = \sigma_e^2$ all t. Since $\lambda(.)$ is a non-linear function in its argument, therefore $E(X_{t+1} \mid X_t = x) = \lambda(x)$ may takes any non-linear shape. In order to compare the regression functions of SETAR(2;1,1) and EXPAR(1) models, we conduct some experiments.

Example 1: Let $\{X_t\}$ satisfy the EXPAR(1) model

$$X_t = \lambda(X_{t-1}) + e_t \,,$$

where $e_t \sim N(0, 1)$ and

$$\lambda(x) = (a + b\, e^{-\gamma x^2}) x \,.$$

It is clear that for large values of $|x|$, $e^{-\gamma x^2}$ is close to zero, and $\lambda(x)$ is approximately equal to ax which is a linear function in x. For all x small, $\lambda(x)$ is non-linear in x. The conditional mean of X_{t+m} given X_t, for m small, is a linear function in the tails and a non-linear function around the origin. Obviously this non-linearity depends on $\lambda(.)$ and it is well known that the conditional mean and the conditional variance tend to the mean and variance of the process as $m \to \infty$.

The regression functions and the conditional standard deviation of X_t given $X_{t-j} = x$, $j = 1,2,3,4,5$ of three EXPAR(1) models with

$$\lambda(x) = -0.9x - 0.95x e^{-x^2} \,. \tag{6.3.5}$$

and

$$\lambda(x) = -0.5x + 0.8x e^{-x^2} + 2x^2 e^{-x^2} \tag{6.3.6}$$

and

$$\lambda(x) = 0.85x + 0.35x e^{-x^2} \tag{6.3.7}$$

are given by Figs. 6.3.1 , 6.3.2 and 6.3.3 respectively. These three examples show the varieties of regression functions which may be produced by those simple EXPAR(1) models.

Example 2: Let $\{X_t\}$ satisfy the SETAR(2;1,1) model

$$X_t = \lambda(X_{t-1}) + e_t$$

where $e_t \sim N(0, 1)$ and

$$\lambda(x) = \begin{cases} a_o + a_1 x & \text{if } x \le r \\ b_o + b_1 x & \text{if } x > r \,. \end{cases}$$

Similar to Example one, $E(X_t \mid X_{t-1} = x) = \lambda(x)$ which is obviously a non-linear function in x. Usually, $\lambda(x)$ is the union of two discontinuous line segments. The regression function and the conditional standard deviation of X_t given X_{t-j}, $j = 1,2,3,4,5$ of SETAR(2;1,1) models with

$$\lambda(x) = \begin{cases} 0.7-0.5x & \text{if } x \le 0 \\ -1.8+0.7x & \text{if } x > 0, \end{cases} \tag{6.3.8}$$

and

$$\lambda(x) = \begin{cases} 0.5x & \text{if } x \le 0 \\ -0.5x & \text{if } x > 0, \end{cases} \tag{6.3.9}$$

are given by Figs. 6.3.4 and 6.3.5 respectively.

Example 3: As an application of our numerical method for the order two case, we consider the following SETAR(2;2,2) model

$$X_t = \begin{cases} 0.62+1.25X_{t-1}-0.43X_{t-2}+e_t & \text{if } X_{t-2} \le 3.25 \\ 2.25+1.52X_{t-1}-1.24X_{t-2}+e_t & \text{if } X_{t-2} > 3.25, \end{cases} \tag{6.3.10}$$

where $\{e_t\}$ is a sequence of independent and identically distributed normal random variables with mean zero and variance 0.0503 . Tong (1983, pp 102-106) has given a rough-and ready calculation of $E(X_t \mid X_{t-j})$, $j = 1,2,3,4$. Our numerical approximation of expectation of X_t given X_{t-j}, $j = 1,2,3,4$ (Figs. 6.3.6), may be compared with Tong's rough calculations.

6.4. Real Data

In the past, various authors have used the annual record of the numbers of Canadian Lynx trapped in the Mackenzie river district of North West Canada for the years 1821-1934 and different periods of Wolf's annual sunspot numbers to illustrate their time series models, both linear and non-linear.

A comparative study of various univariate time series models which have been applied to the lynx data, has been done by Lim (1987). This section only considers the comparison of non-parametric estimates of regression functions and conditional variances of some of different models which are fitted to lynx and sunspot data.

A detailed statistical analysis of the lynx data was first attempted by Moran (1953). In order to reduce the asymmetry of the original data, he initially applied a logarithmic transformation to the raw data, X_t.

The transformation $Y_t = \log_{10} X_t$ was thereafter adopted by many other authors. All models discussed in this section use this transformation. The AR(2) model which was obtained by Moran is given by

$$X_t = 1.0549 + 1.4101X_{t-1} - 0.7734X_{t-2} + e_t \qquad (6.4.1)$$

where $V(e_t) = 0.0459$.

Tong (1977b) fitted an AR(11) model based on the Akaike's Information Criterion (AIC). The chosen model is given by

$$\begin{aligned} X_t = {} & 1.13X_{t-1} - 0.51X_{t-2} + 0.23X_{t-3} - 0.29X_{t-4} + 0.14X_{t-5} \\ & - 0.14X_{t-6} + 0.08X_{t-7} - 0.04X_{t-8} + 0.13X_{t-9} + 0.19X_{t-10} - 0.31X_{t-11} + e_t , \end{aligned} \qquad (6.4.2)$$

with $V(e_t) = 0.0437$. A type of TAR model, self-exciting threshold autoregressive model was fitted to the first 100 observations of the log-transformation of lynx data . (See Tong, 1983). The fitted model is given by

$$X_t = \begin{cases} 0.77 + 1.06X_{t-1} - 0.20X_{t-2} + 0.16X_{t-3} \\ \quad - 0.43X_{t-4} + 0.18X_{t-5} + e_t^{(1)} & \text{if } X_{t-2} \le 3.05 \\ 2.25 + 1.47X_{t-1} - 1.20X_{t-2} + e_t^{(2)} & \text{if } X_{t-2} > 3.05 , \end{cases} \qquad (6.4.3)$$

where $V(e_t^{(1)}) = 0.0302$, $V(e_t^{(2)}) = 0.0564$ and the pooled variance is equal to 0.0415. For the full set of lynx data, a SETAR(2;7,2) has been obtained. (See Tong, 1983). The model can be written as

$$X_t = \begin{cases} 0.55 + 1.03X_{t-1} - 0.17X_{t-2} + 0.17X_{t-3} - 0.43X_{t-4} \\ \quad 0.33X_{t-5} - 0.28X_{t-6} + 0.21X_{t-7} + e_t^{(1)} & \text{if } X_{t-2} \le 3.116 \\ 2.63 + 1.49X_{t-1} - 1.32X_{t-2} + e_t^{(2)} & \text{if } X_{t-2} > 3.116 \end{cases} \qquad (6.4.4)$$

where $V(e_t^{(1)}) = 0.0258$, $V(e_t^{(2)}) = 0.0505$ (pooled var=0.0360).

Gabr and Subba Rao (1981) have fitted another class of non-linear model, a bilinear model, to the lynx data. By using the log-transformation of the first 100 observations, they have reported the following subset bilinear model.

$$\begin{aligned} X_t = {} & 1.486292 + 0.77227X_{t-1} - 0.091572X_{t-2} + 0.083073X_{t-3} \\ & - 0.261493X_{t-4} + 0.225585X_{t-9} - 0.24841X_{t-12} - 0.7893X_{t-3}e_{t-9} \\ & + 0.4798X_{t-9}e_{t-9} + 0.3902X_{t-6}e_{t-2} + 0.1326X_{t-1}e_{t-1} \\ & + 0.07944X_{t-2}e_{t-7} - 0.3212X_{t-4}e_{t-2} + e_t \end{aligned} \qquad (6.4.5)$$

where $V(e_t)=0.0223$.

The Exponential Autoregressive Model was another class of non-linear model for the lynx data. Ozaki (1982) found that the almost symmetric series generated by an EXPAR model developed by Haggan and Ozaki (1981) was unsatisfactory. He then has fitted two exponential autoregressive models to the full set log-transformation of lynx data. The first one which could reproduce the asymmetric limit cycle structure of the lynx data may be written † as

$$X_t = \{1.1668+(0.316+0.982X_{t-1})e^{-3.89X_{t-1}^2}\}X_{t-1} - \{0.437;(0.659+1.260X_{t-1})e^{-3.89X_{t-1}^2}\}X_{t-2}+e_t \quad (6.4.6)$$

where $V(e_t)=0.04327$.

The second EXPAR model with smaller variance of fitted residuals, which was believed to be more appropriate for forecasting is given by

$$X_t = -0.481X_{t-1}-0.247X_{t-2}+0.318X_{t-3}+0.23X_{t-4} + 0.352X_{t-5}+0.096X_{t-6}-0.085X_{t-7}-0.289X_{t-8} - 0.181X_{t-9}+Y_t, \quad (6.4.7)$$

$$Y_t = \{1.514+(0.480-3.332Y_{t-1}-0.610Y_{t-1}^2+8.906Y_{t-1}^3)e^{-3.89Y_{t-1}^2}\}Y_{t-1} + \{-0.902+(-0.228+0.923Y_{t-1}+0.193Y_{t-1}^2-4.216Y_{t-1}^3)e^{-3.89Y_{t-1}^2}\}Y_{t-2}+e_t,$$

where $V(e_t)=-0.03153$.

Gabr and Subba Rao (1981) for their first illustration have considered the annual sunspot numbers for the years 1700-1955, giving 256 observations. Their linear and bilinear models which are fitted to the first 221 observations are as follows

a) AR model

The fitted model to the mean deleted observations is given by

$$X_t = 1.2163X_{t-1}-0.4670X_{t-2}-0.1416X_{t-3}+0.1691X_{t-4}-0.1473X_{t-5} + 0.0543X_{t-6}-0.0534X_{t-7}+0.0667X_{t-8}+0.1129X_{t-9}+e_t \quad (6.4.8)$$

where $V(e_t)=199.27$.

† Note : The first coefficient of this model has been corrected in Chapter Four.

b) Subset bilinear model

The reported subset bilinear model is

$$\begin{aligned} X_t &= 6.8860+1.5012X_{t-1}-0.767X_{t-2}+0.1152X_{t-9} \\ &\quad -0.01458X_{t-2}e_{t-1}+0.006312X_{t-8}e_{t-1}-0.007152X_{t-1}e_{t-3} \\ &\quad +0.006047X_{t-4}e_{t-3}+0.003619X_{t-1}e_{t-6}+0.00433X_{t-2}e_{t-4} \\ &\quad +0.001782X_{t-3}e_{t-2}+e_t\,, \end{aligned} \tag{6.4.9}$$

where $V(e_t)=124.33$.

As far as the SETAR models are concerned, different SETAR models have been fitted to the different period of raw and transformed sunspot data. In this section we will only consider the SETAR(2;3,11) which is fitted to the raw sunspot numbers of period 1700-1920, and is given by (See Tong, 1983)

$$X_t = \begin{cases} 11.97+1.71X_{t-1}-1.26X_{t-2}+0.43X_{t-3}+e_t^{(1)} & \text{if } X_{t-3}\le 36.6 \\ 7.84+0.73X_{t-1}-0.04X_{t-2}-0.20X_{t-3}+0.16X_{t-4} & \\ -0.22X_{t-5}+0.02X_{t-6}+0.15X_{t-7}-0.24X_{t-8} & \\ 0.31X_{t-9}-0.37X_{t-10}+0.38X_{t-11}+e_t^{(2)} & \text{if } X_{t-3}>36.6\,, \end{cases} \tag{6.4.10}$$

where $V(e_t^{(1)})=254.64$, $V(e_t^{(2)})=66.80$, pooled var=153.7.

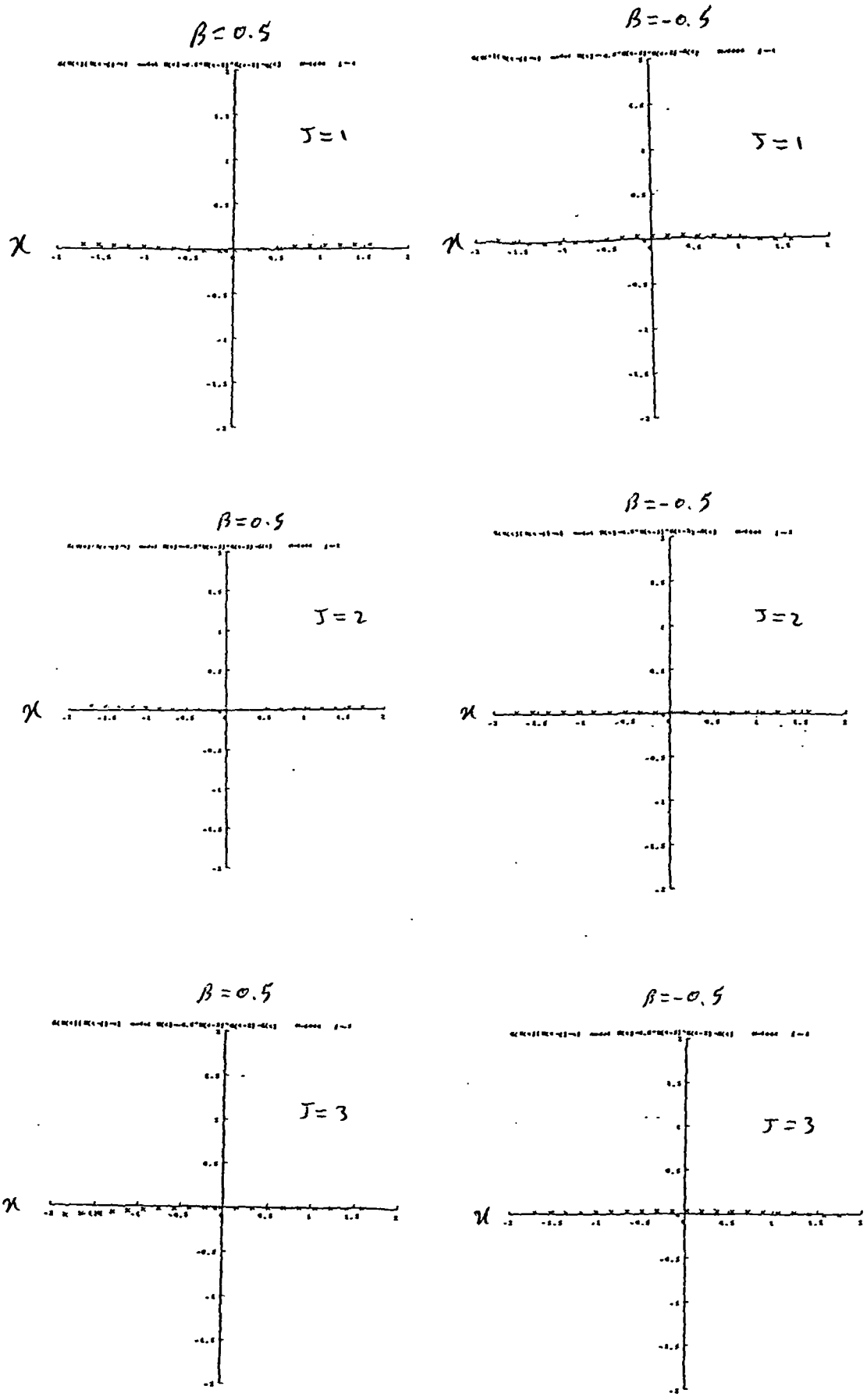
In order to examine the post-model of observed non-parametric lag regression versus the fitted non-parametric lag regression of those models which have been fitted to the sunspot and lynx data, we generate a sample of size 10000 observations using a simulation technique, of the fitted models and estimate

$$E(X_t\,|\,X_{t\pm j})\,,\ V(X_t\,|\,X_{t\pm j})\,,\ j=1,2,...,11.$$

The Figs. 6.4.1 (the non-parametric estimates of conditional mean and conditional variance of bilinear for lynx data), 6.4.2 (those of SETAR models) and 6.4.3 (those of EXPAR models) show that the SETAR models compare very well with the others for lynx data. Also, it can be seen that EXPAR1 is better than EXPAR2 and perhaps the bilinear model is the worst model for this particular data set. It should be noted that the performance of the corrected EXPAR1 model is much better than uncorrected model which have been reported by Lim (1987).

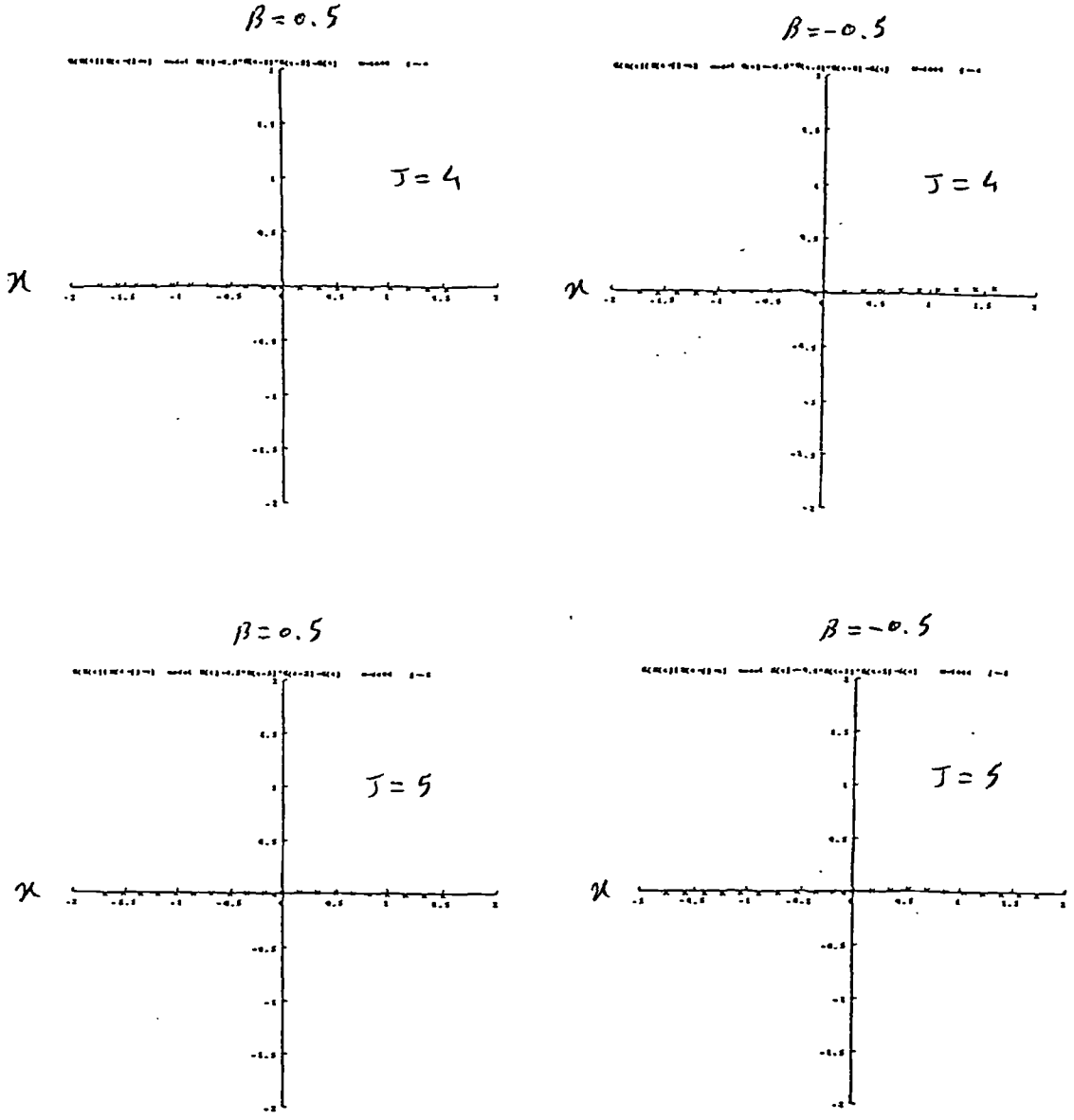
Comparing the Figs 6.4.5 (conditional mean and conditional variance of bilinear model which is fitted to sunspot) with Figs. 6.4.6 (those of SETAR model), we can see that the performance of SETAR

model is superior to that of bilinear model for raw sunspot data .

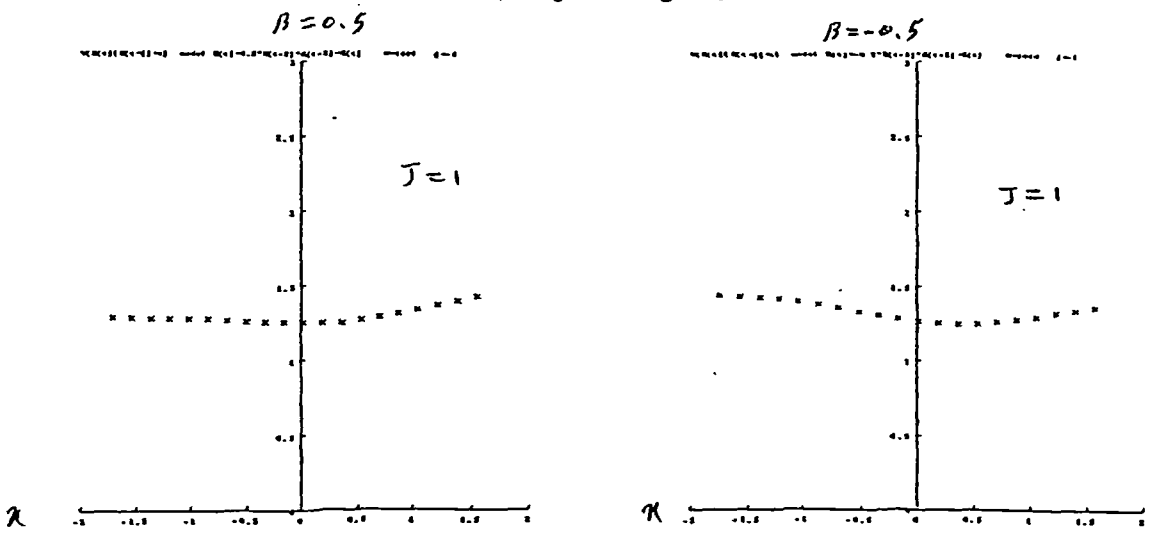


Figs. 6.2.1.1 $E(X_t|X_{t-j}=x)$, $j=1,2,3$ model 6.2.1.2

$$X_t=\beta X_{t-3}e_{t-2}+e_t$$

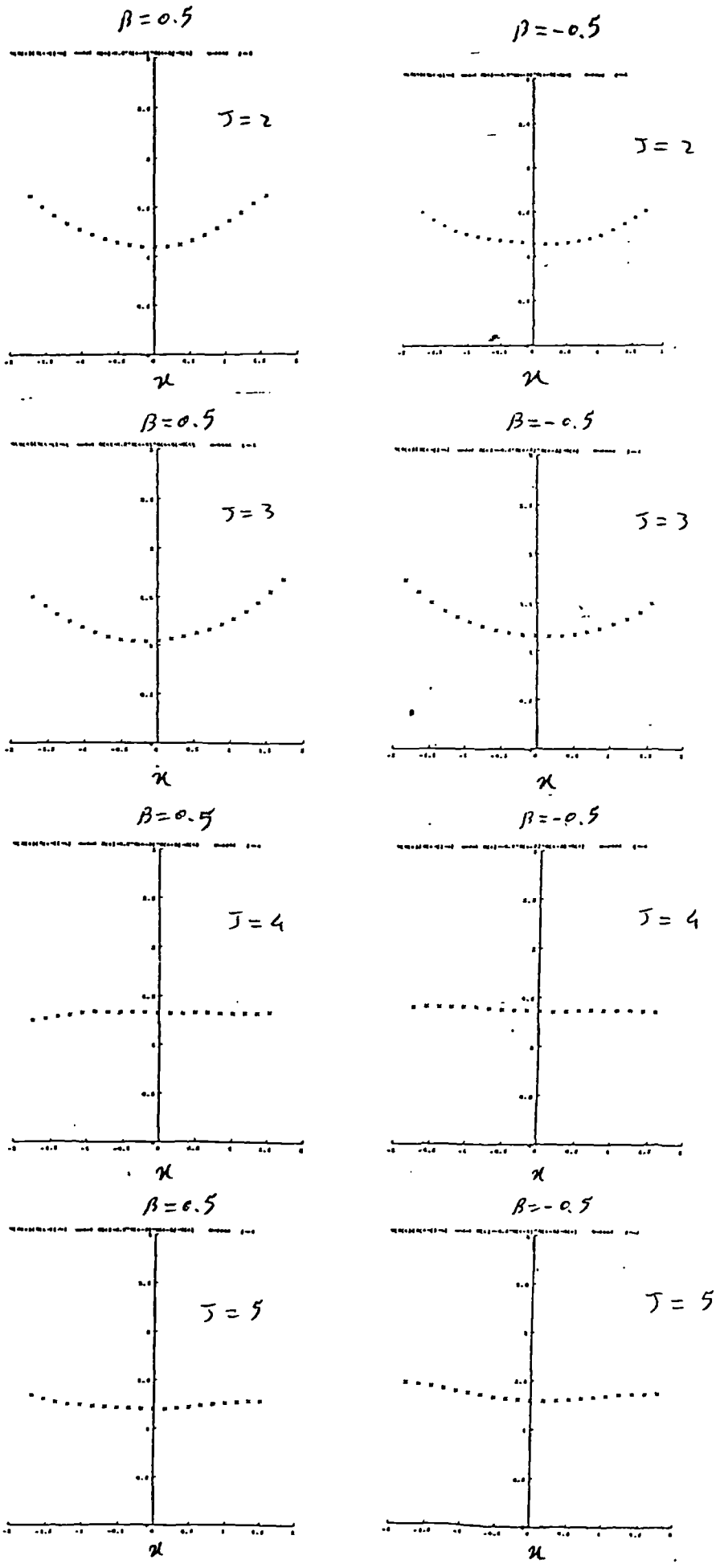


Figs. 6.2.1.1 $E(X_t|X_{t-j}=x)$, j=4,5 model 6.2.1.2 $X_t=\beta X_{t-3}e_{t-2}+e_t$



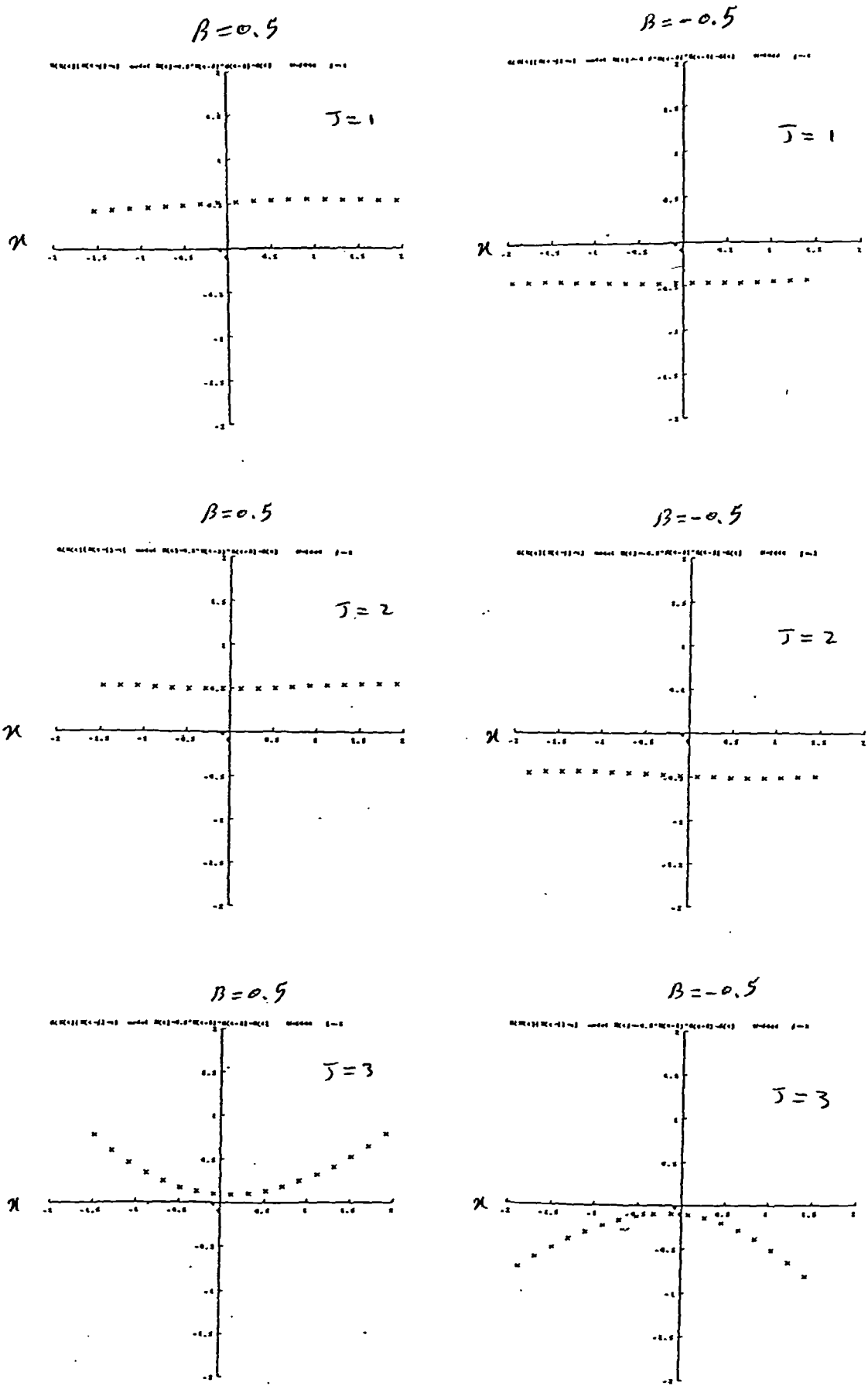
Figs. 6.2.1.2 $V(X_t|X_{t-j}=x)$, j=1 model 6.2.1.2

$X_t=\beta X_{t-3}e_{t-2}+e_t$



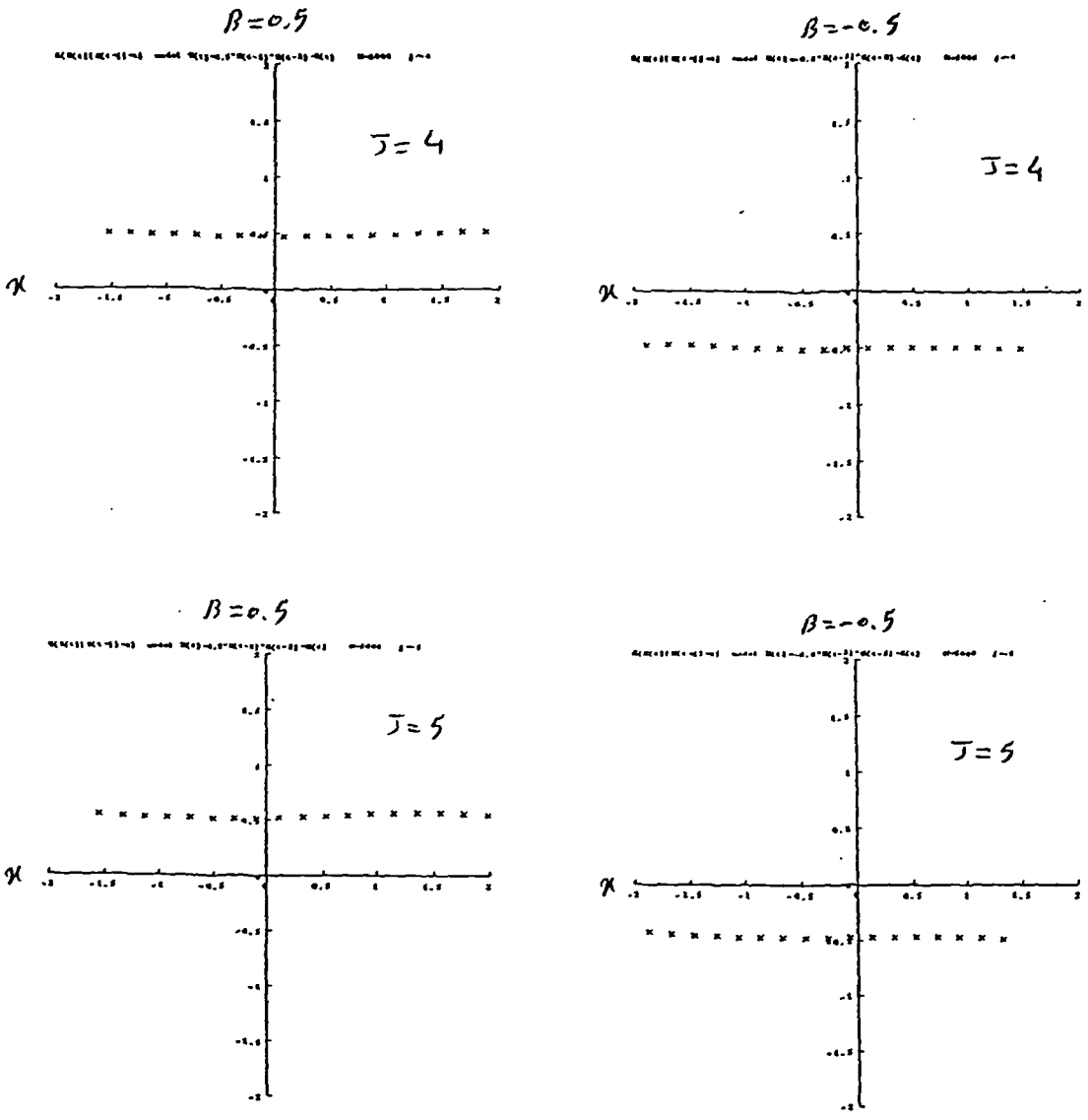
Figs. 6.2.1.2 $V(X_t|X_{t-j}=x)$, $j=2,3,4,5$ model 6.2.1.2

$X_t=\beta X_{t-3}e_{t-2}+e_t$

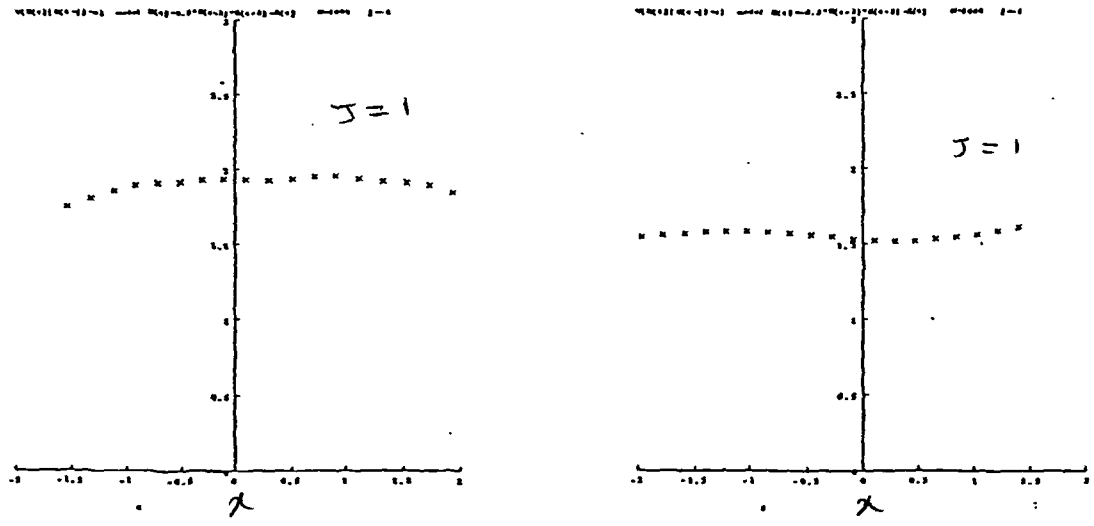


Figs. 6.2.2.1 $E(X_t|X_{t-j}=x)$, j=1,2,3 model 6.2.2.2

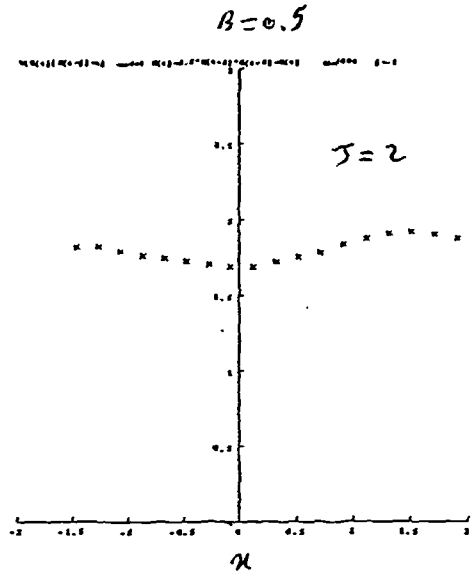
$$X_t=\beta X_{t-3}e_{t-3}+e_t$$

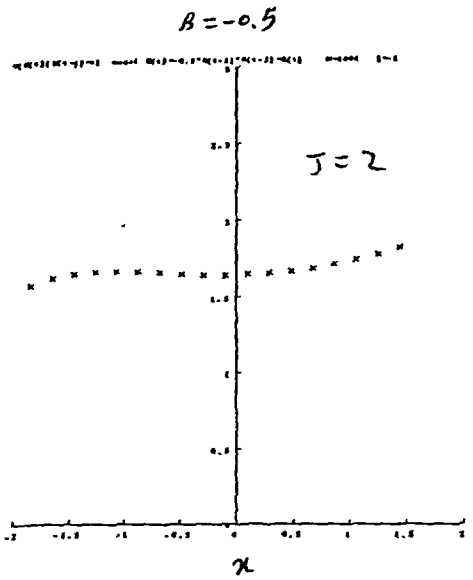


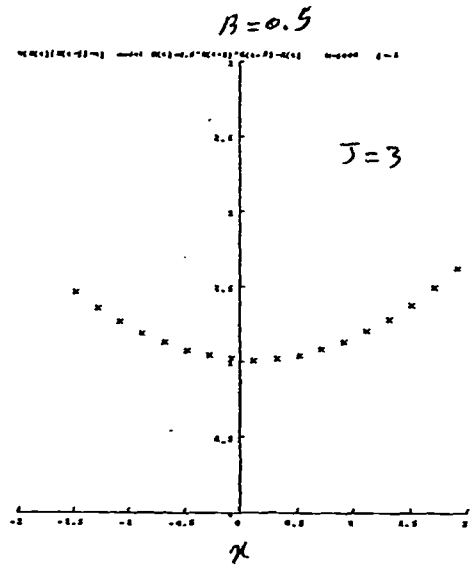
Figs. 6.2.2.1 $E(X_t|X_{t-j}=x)$, j=4,5 model 6.2.2.2 $X_t=\beta X_{t-3}e_{t-3}+e_t$

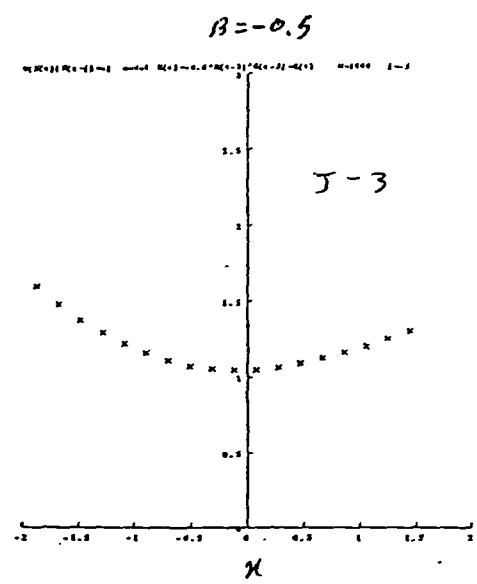


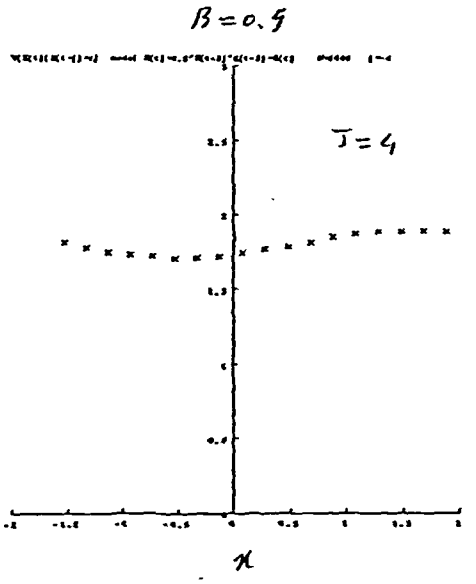
Figs. 6.2.2.2 $V(X_t|X_{t-j}=x)$, j=1 model 6.2.2.2

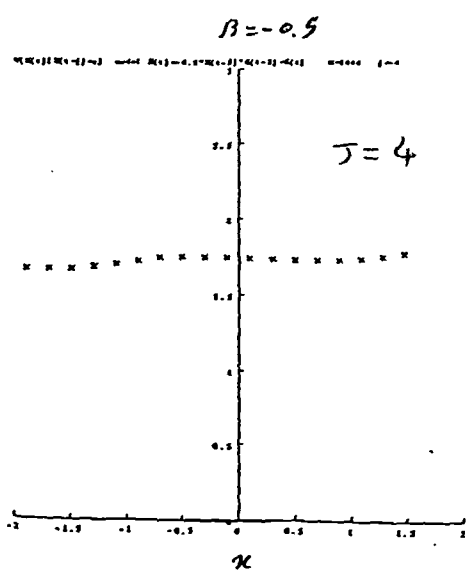






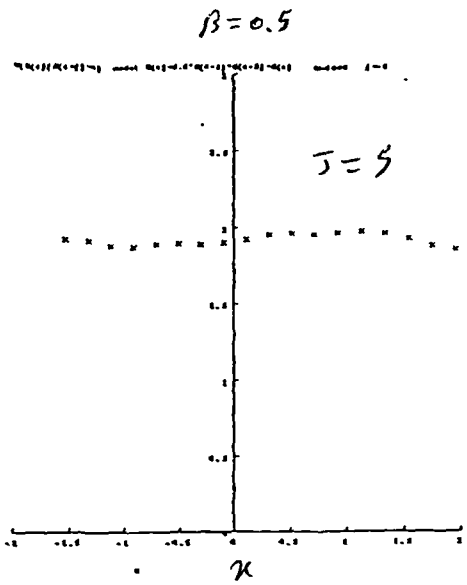


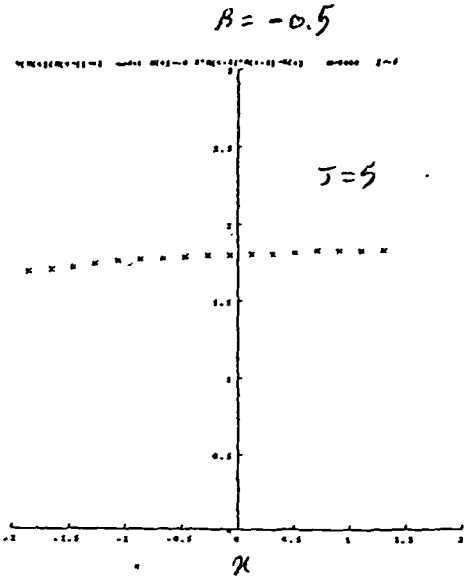




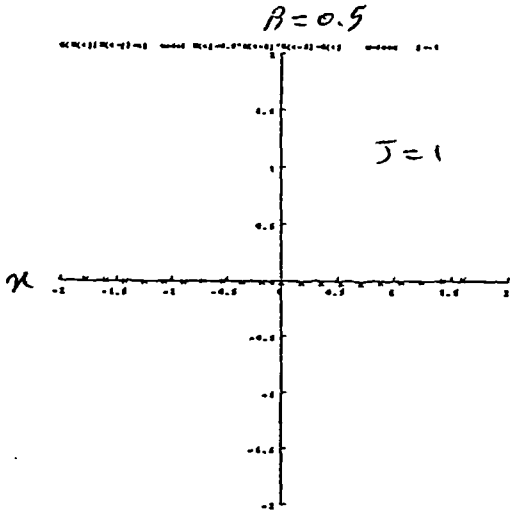
Figs. 6.2.2.2 $V(X_t|X_{t-j}=x)$, $j=2,3,4$ model 6.2.2.2

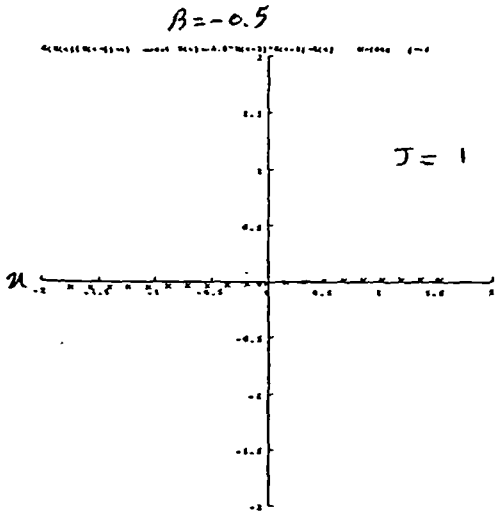
$$X_t=\beta X_{t-3}e_{t-3}+e_t$$

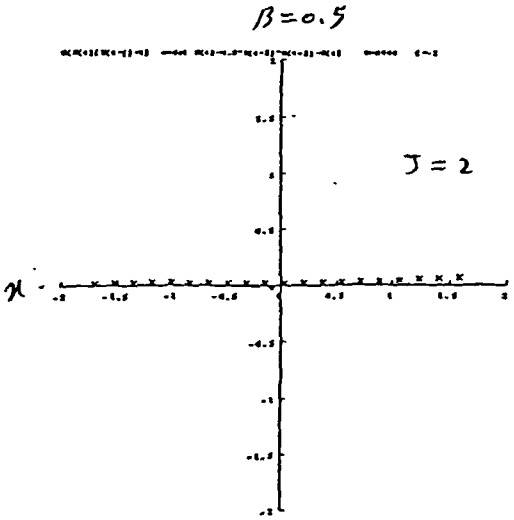


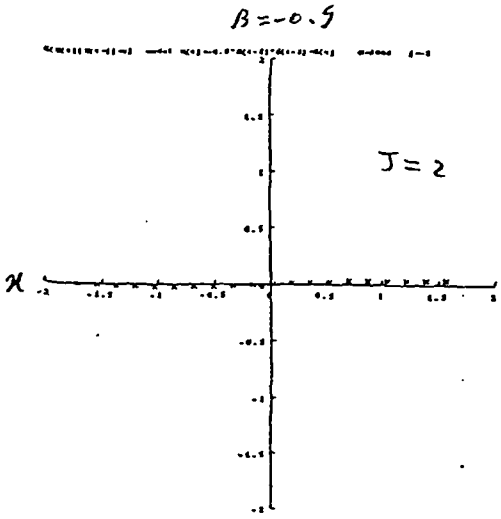


Figs. 6.2.2.2 $V(X_t|X_{t-j}=x)$, j=5 model 6.2.2.2 $X_t=\beta X_{t-3}e_{t-3}+e_t$



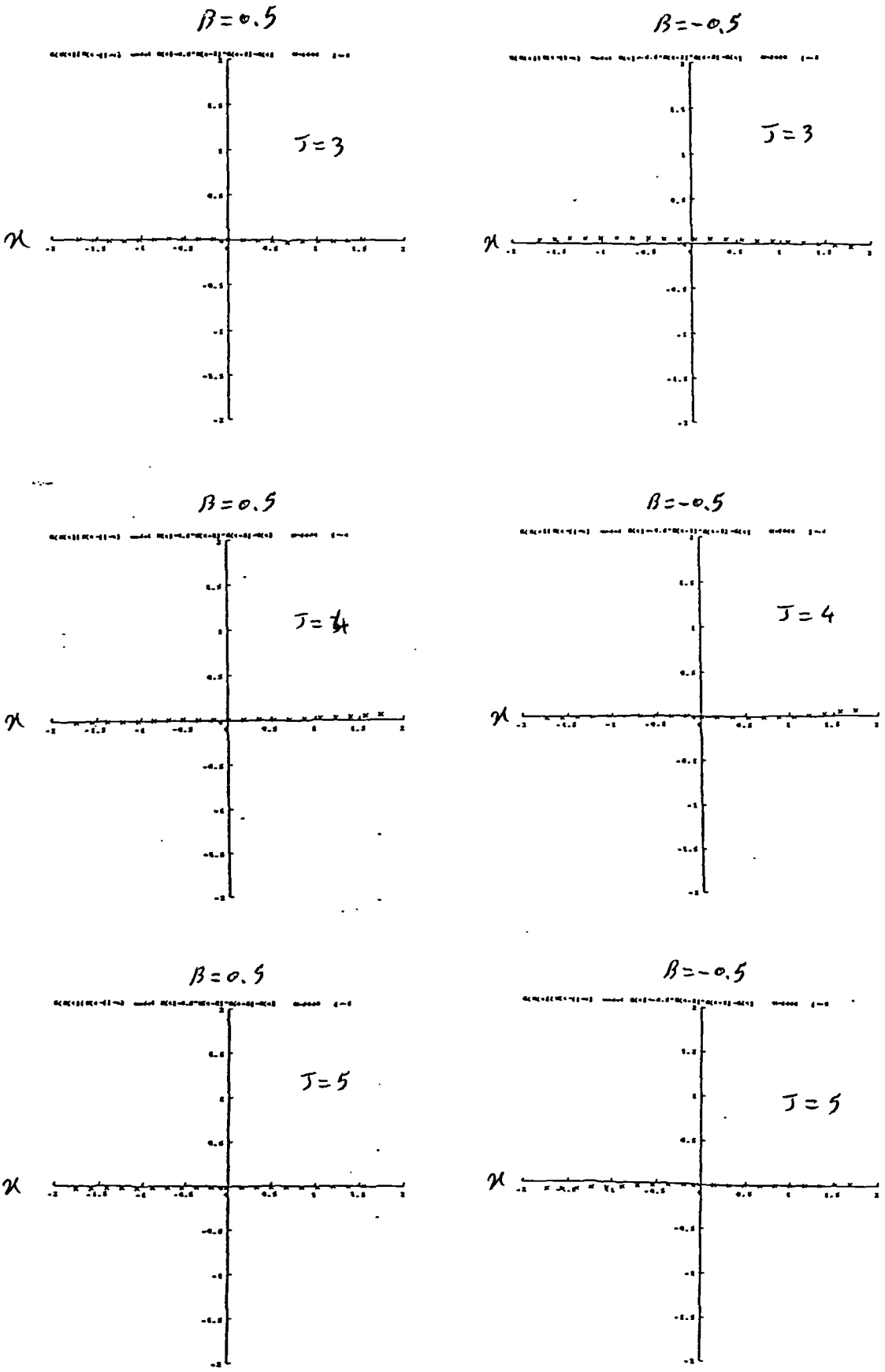






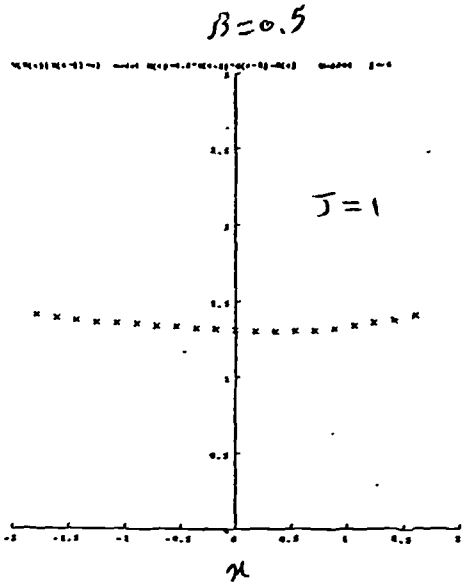
Figs. 6.2.3.1 $E(X_t|X_{t-j}=x)$, j=1,2 model 6.2.3.2

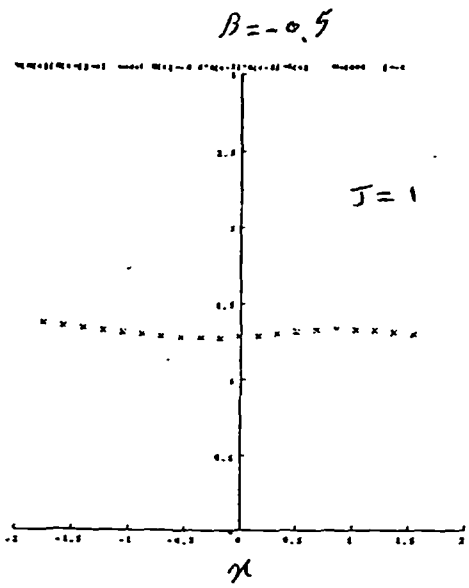
$X_t=\beta X_{t-2}e_{t-3}+e_t$

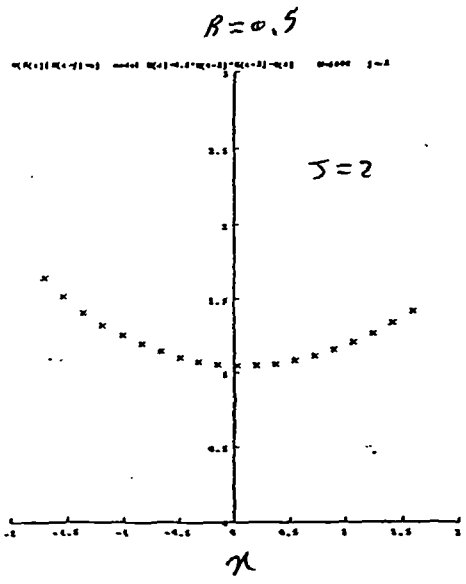


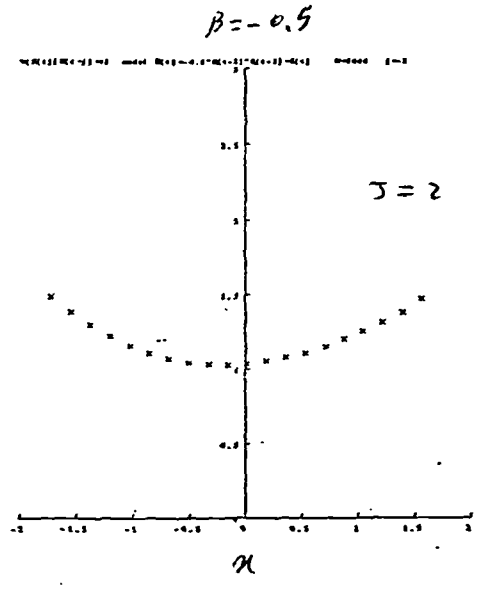
Figs. 6.2.3.1 $E(X_t|X_{t-j}=x)$, j=3,4,5 model 6.2.3.2

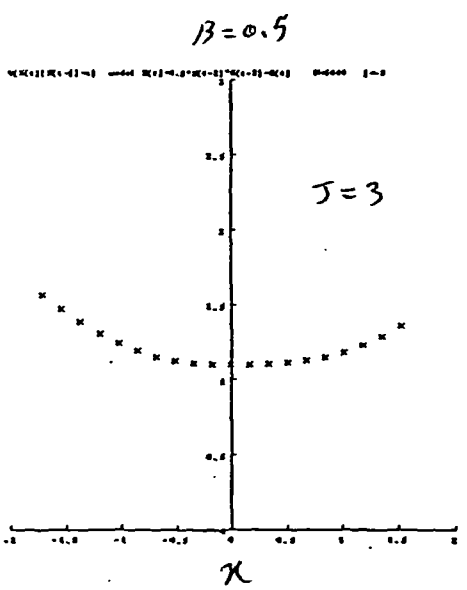
$$X_t=\beta X_{t-2}e_{t-3}+e_t$$

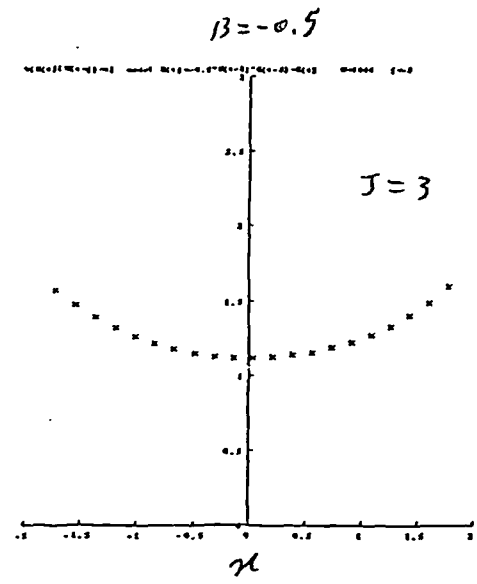






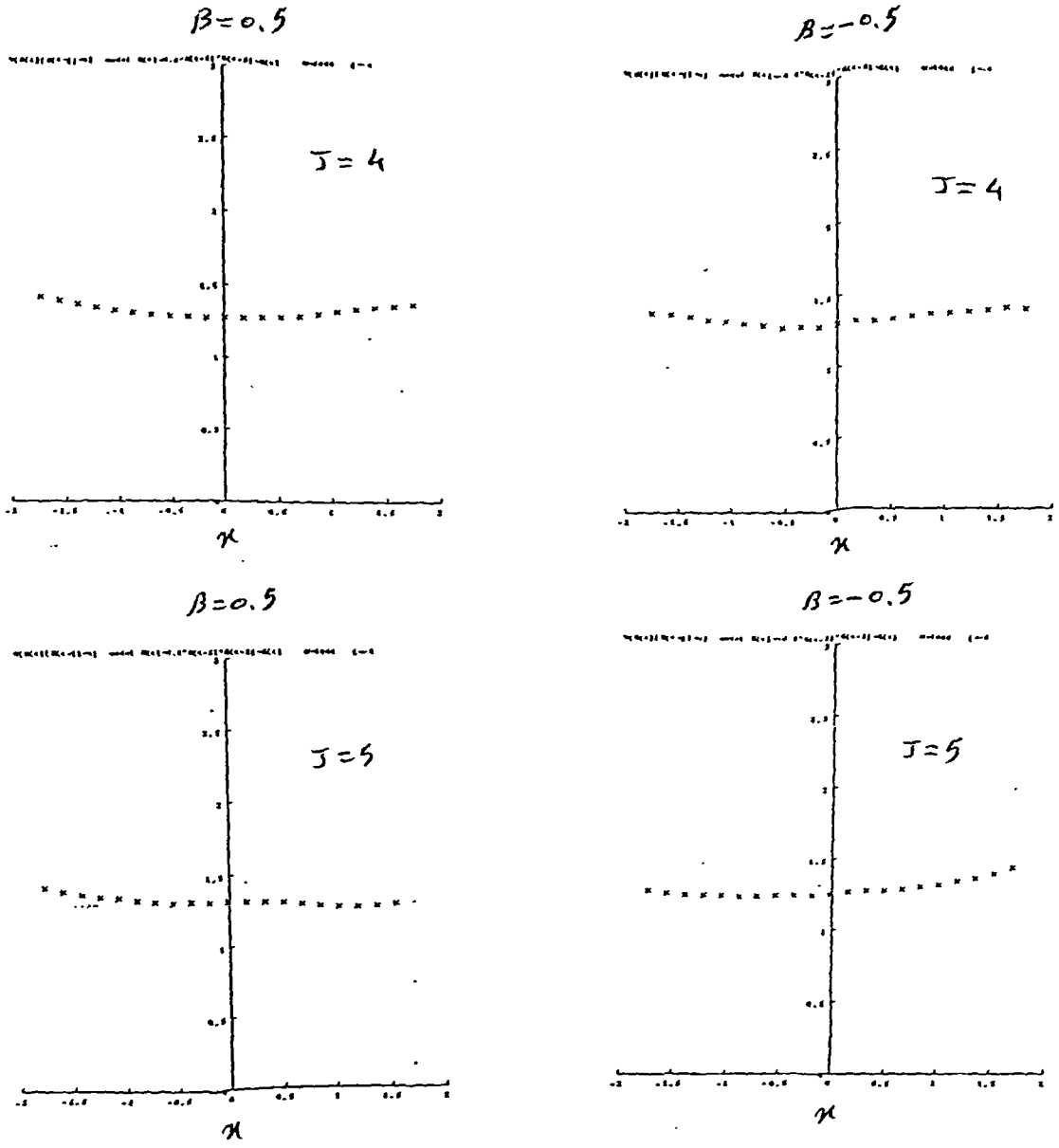






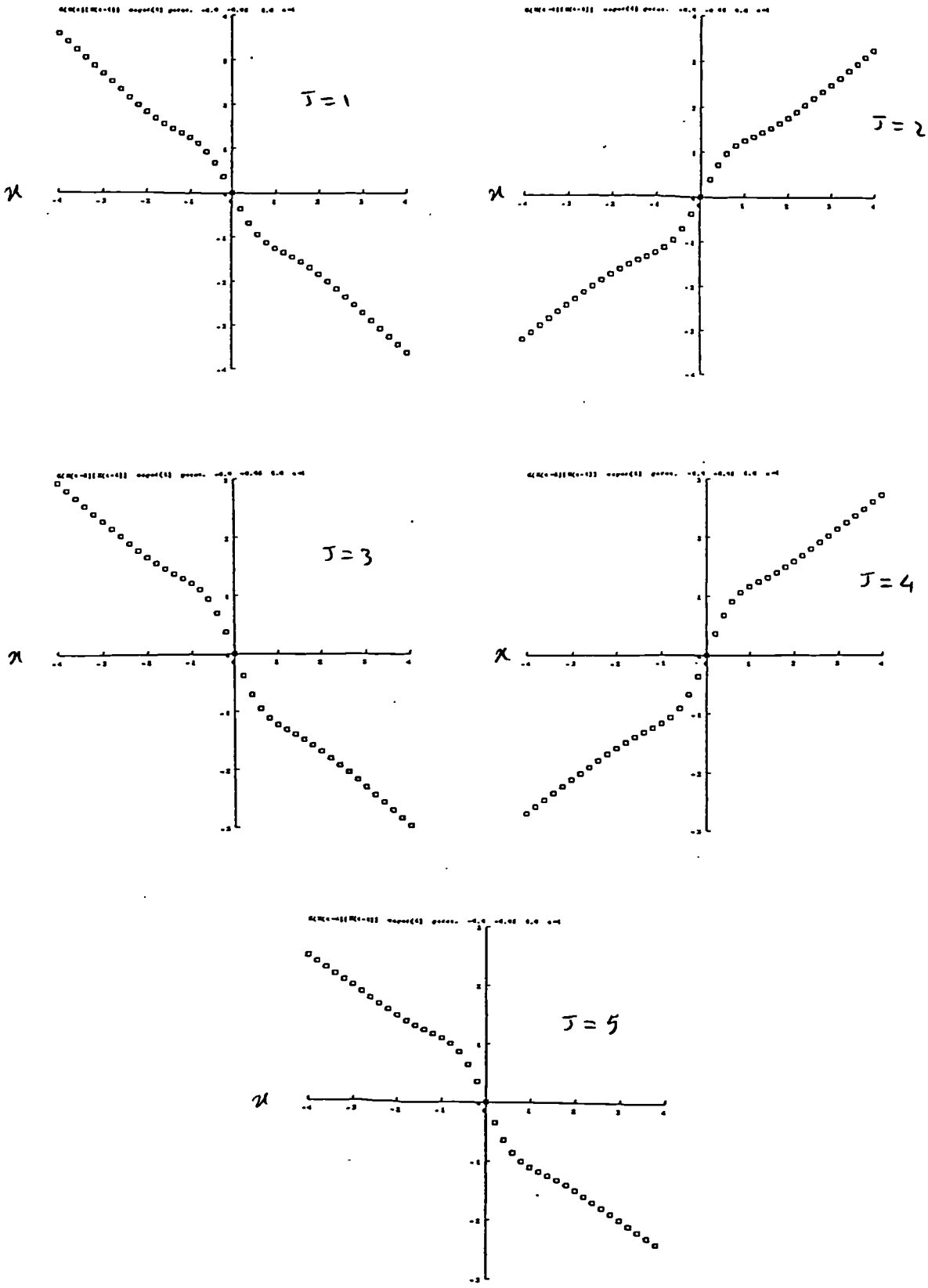
Figs. 6.2.3.2 $V(X_t|X_{t-j}=x)$, $j=1,2,3$ model 6.2.3.2

$$X_t=\beta X_{t-2}e_{t-3}+e_t$$



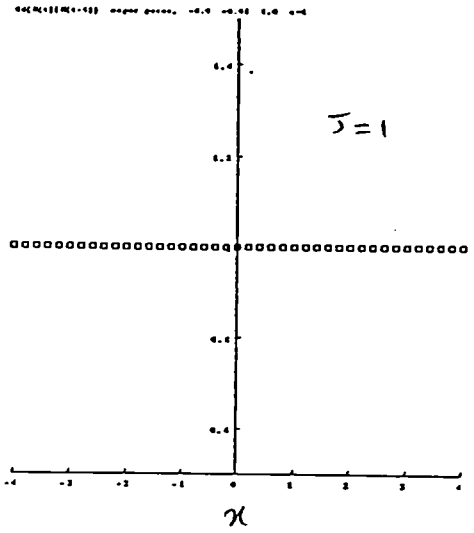
Figs. 6.2.3.2 $V(X_t|X_{t-j}=x)$, $j=4,5$ model 6.2.3.2

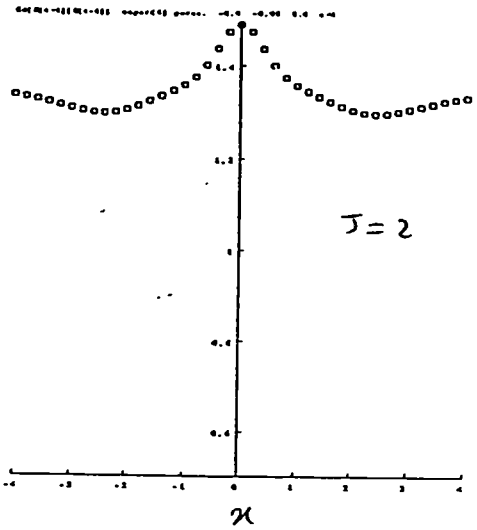
$$X_t=\beta X_{t-2}e_{t-3}+e_t$$

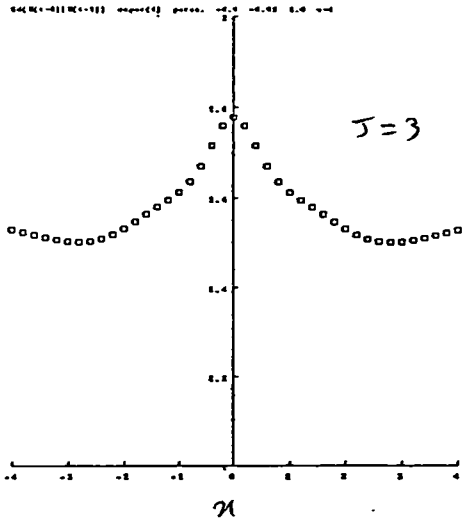


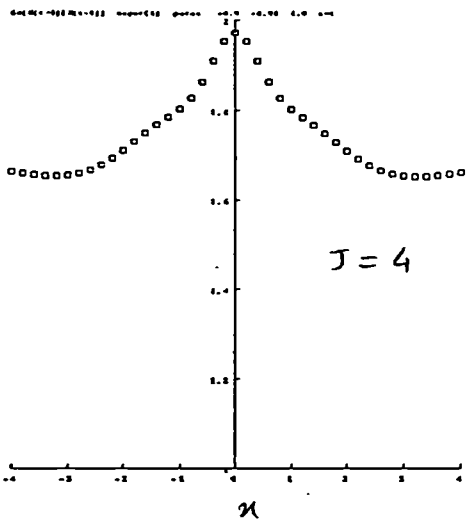
Figs. 6.3.1 $E(X_t | X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.5

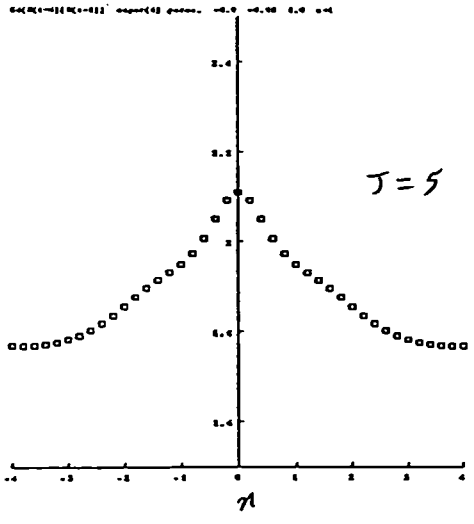
$$X_t = -0.9X_{t-1} - 0.95X_{t-1}e^{-x_{t-1}^2} + e_t$$





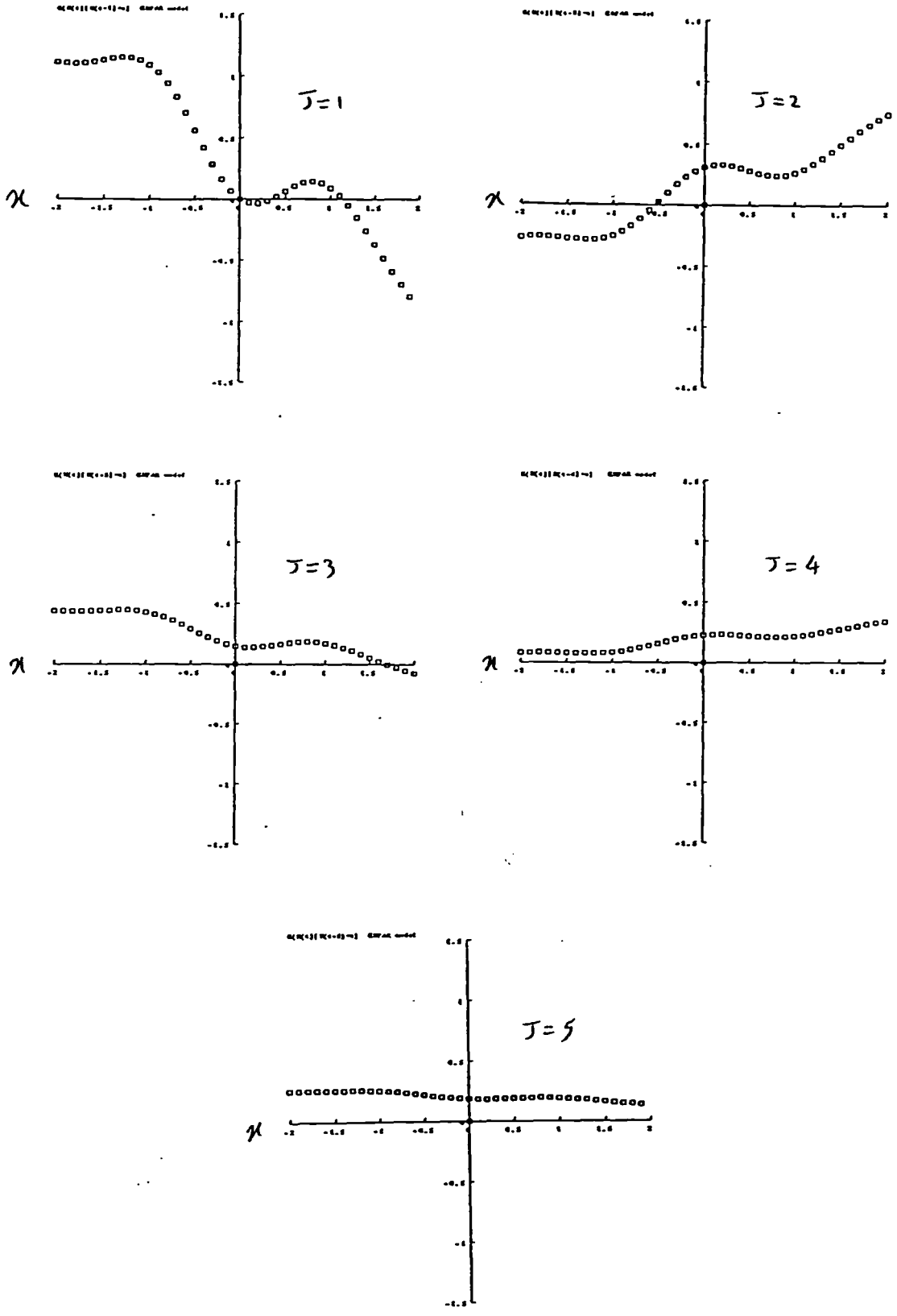






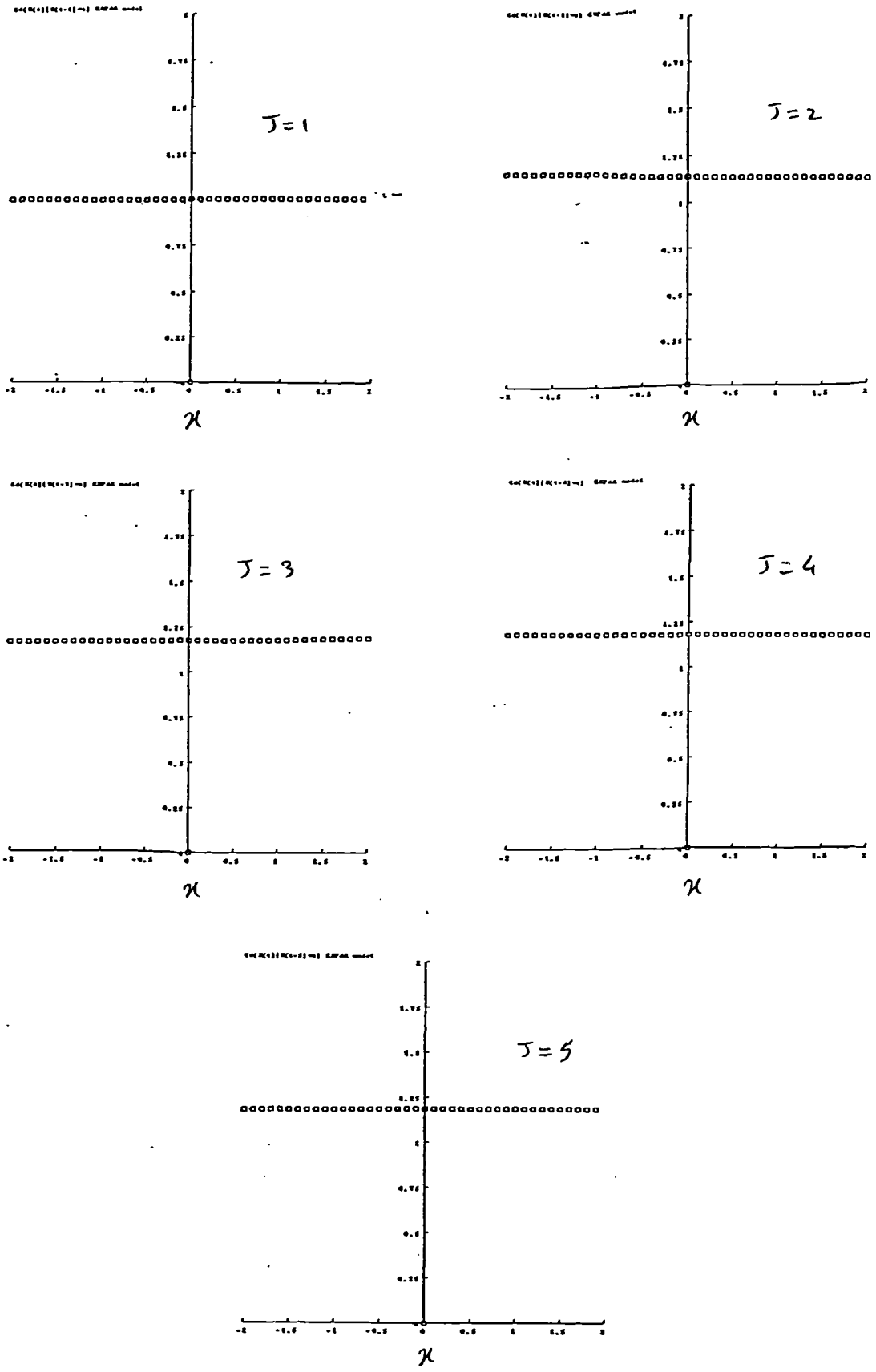
Figs. 6.3.1 $V(X_t \mid X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.5

$$X_t=-0.9X_{t-1}-0.95X_{t-1}e^{-x_{t-1}^2}+e_t$$



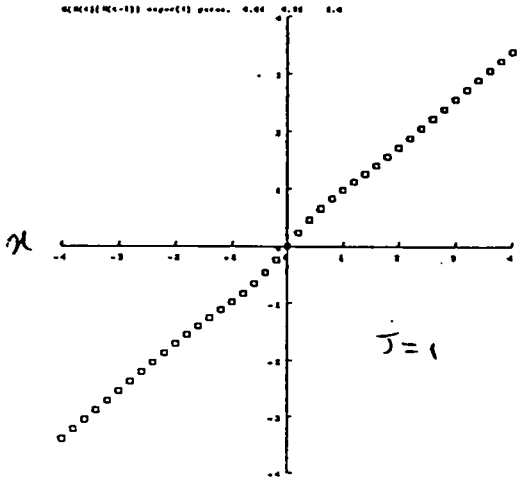
Figs. 6.3.2 $E(X_t|X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.6

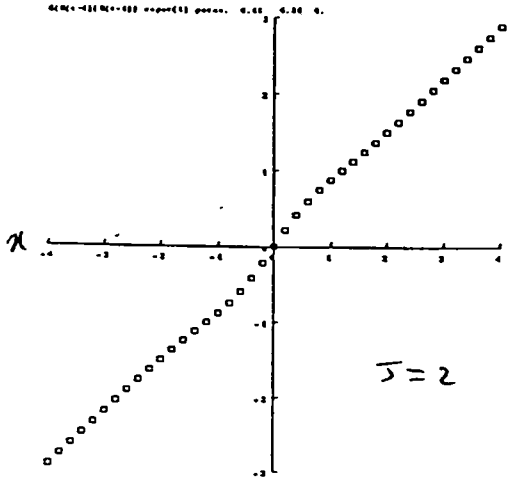
$$X_t=-0.5X_{t-1}+0.8X_{t-1}e^{-x_{t-1}^2}+2X_{t-1}^2e^{-x_{t-1}^2}+e_t$$

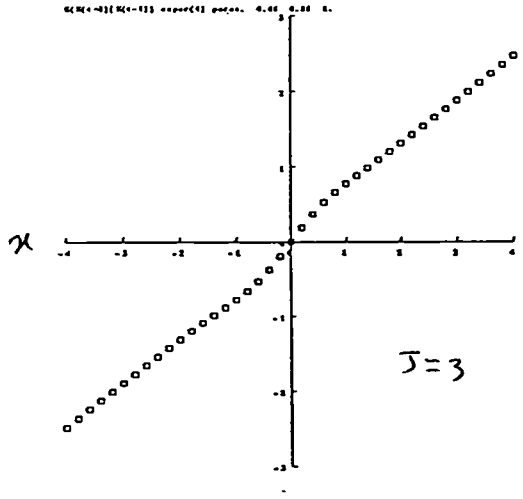


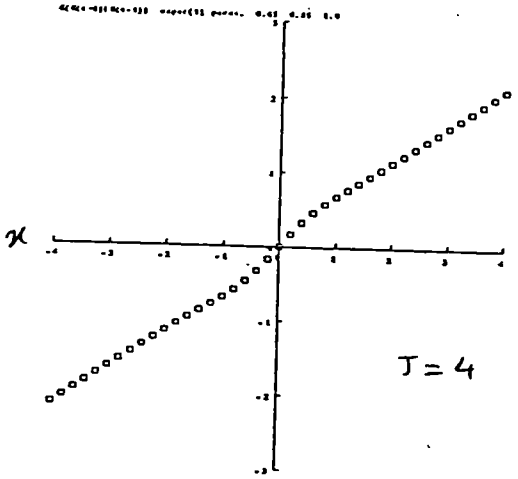
Figs. 6.3.2 $V(X_t|X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.6

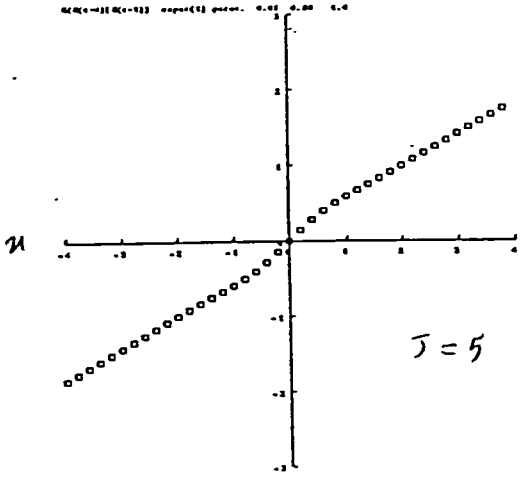
$$X_t=-0.5X_{t-1}+0.8X_{t-1}e^{-X_{t-1}^2}+2X_{t-1}^2e^{-X_{t-1}^2}+e_t$$





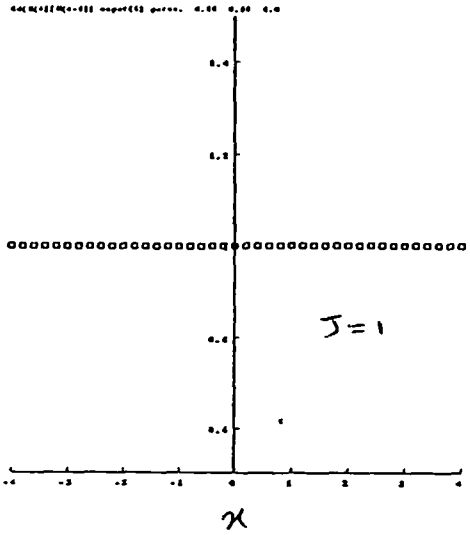


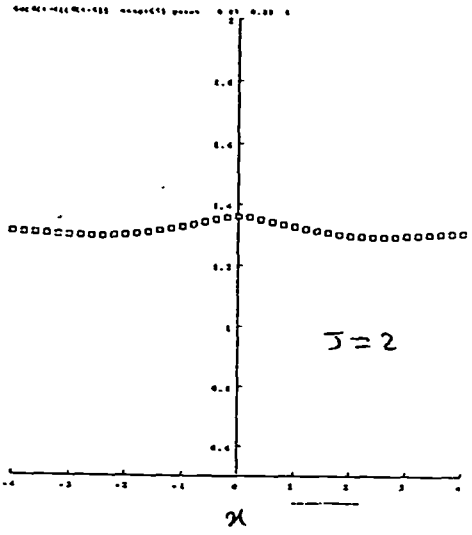


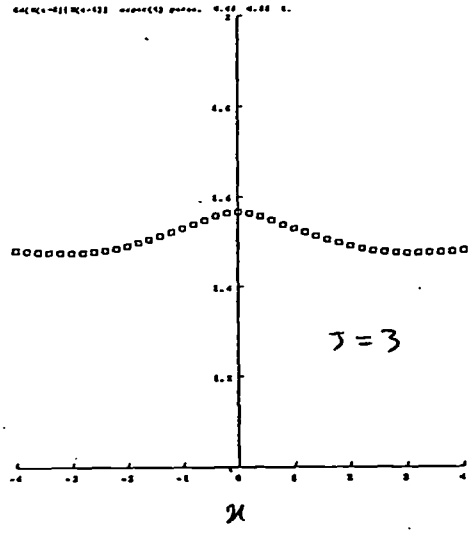


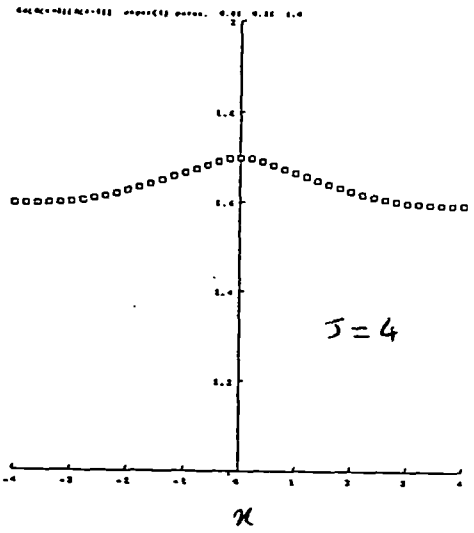
Figs. 6.3.3 $E(X_t | X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.7

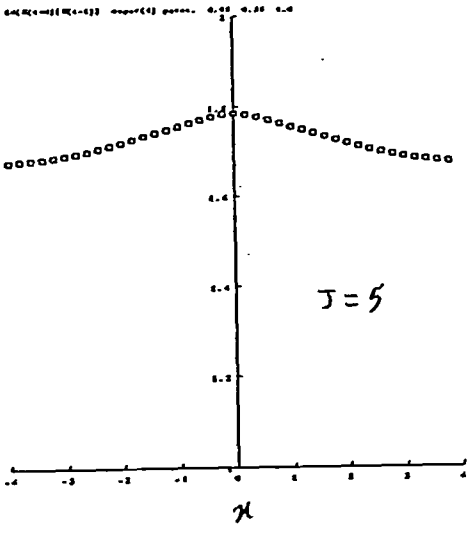
$$X_t=0.85X_{t-1}+0.3X_{t-1}e^{-X_{t-1}^2}+e_t$$





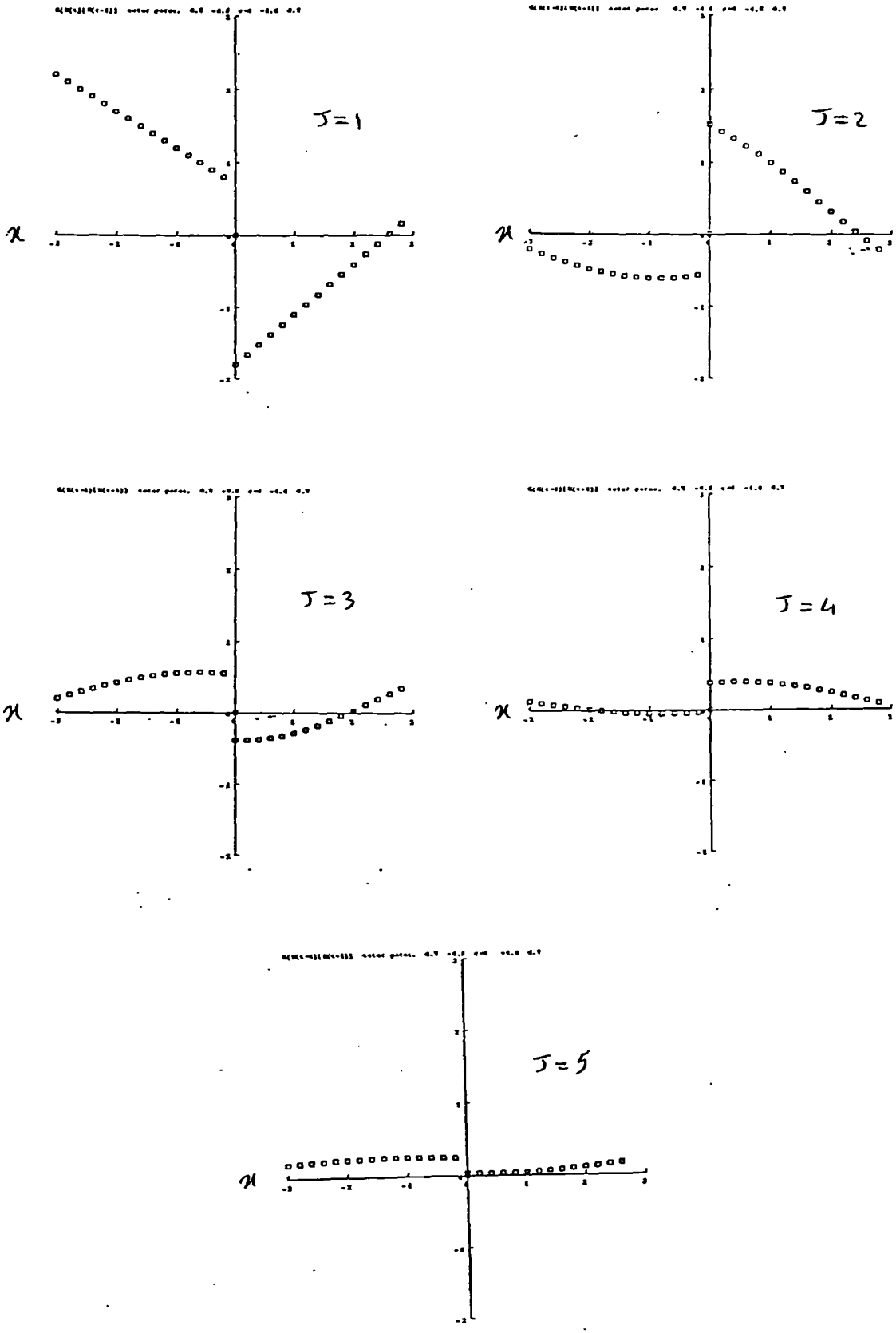






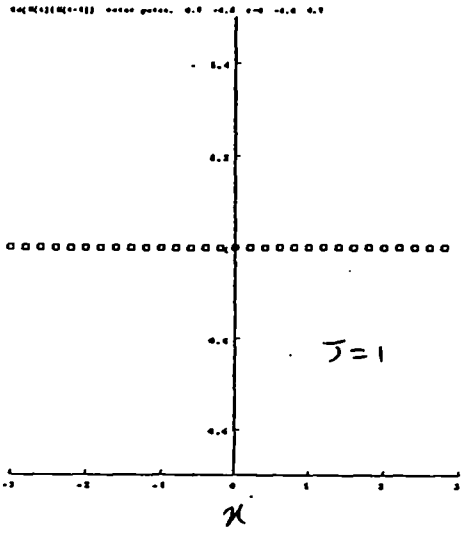
Figs. 6.3.3 $V(X_t \mid X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.7

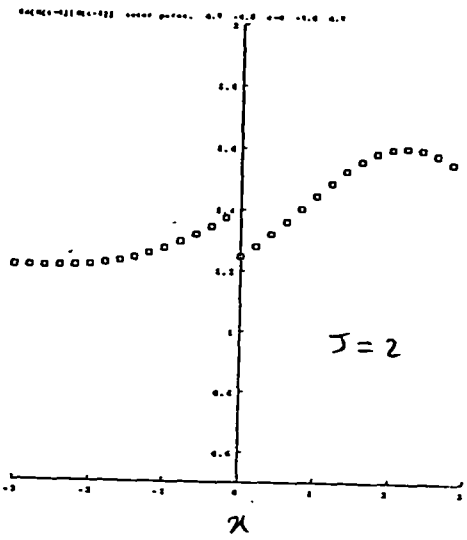
$$X_t=0.85X_{t-1}+0.3X_{t-1}e^{-X_{t-1}^2}+e_t$$

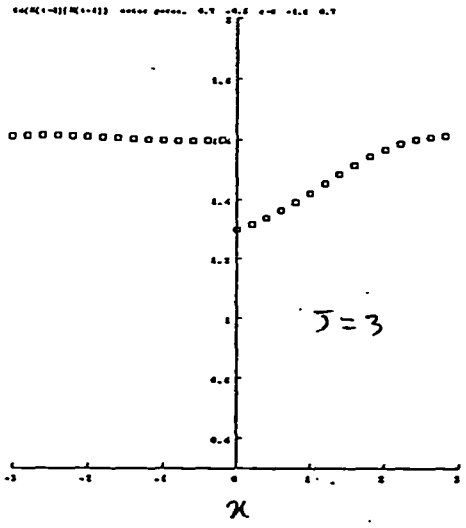


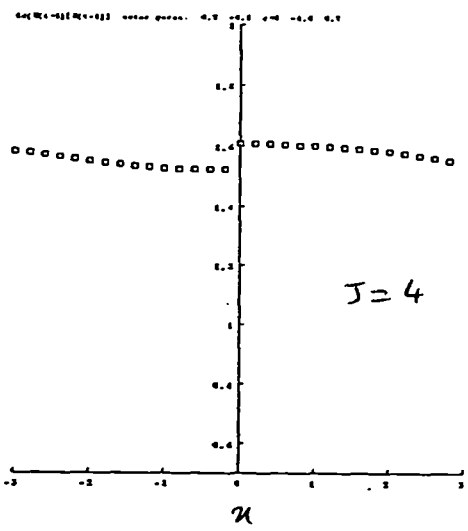
Figs. 6.3.4 $E(X_t|X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.8

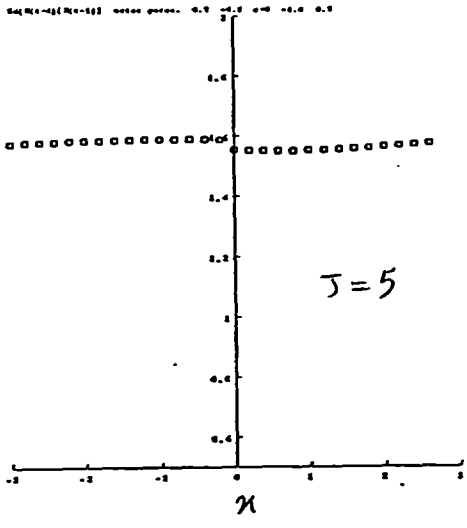
$$X_t=\begin{cases}0.7-0.5X_{t-1}+e_t & \text{if } X_t\le 0\\ -1.8+0.7X_{t-1}+e_t & \text{if } X_t>0\end{cases}$$





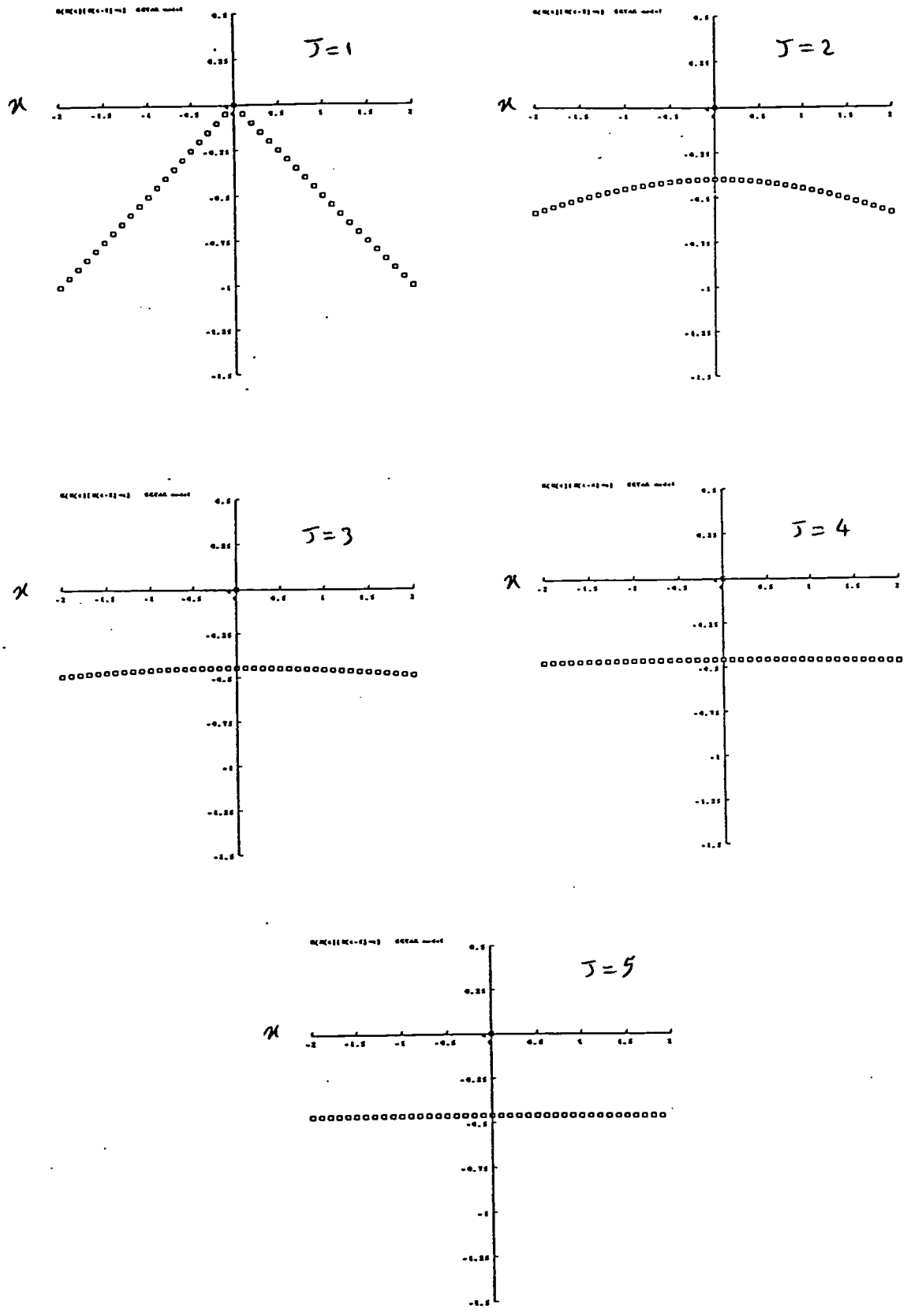






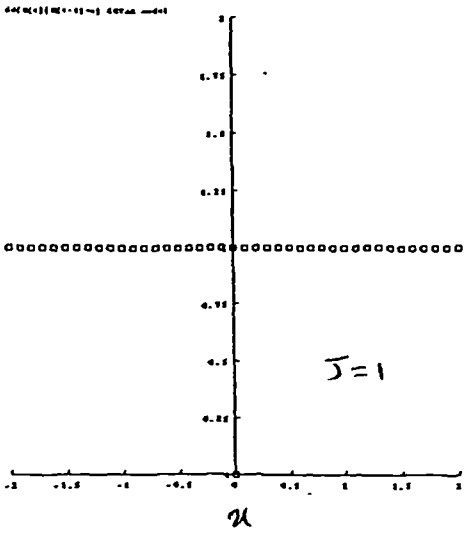
Figs. 6.3.4 $V(X_t|X_{t-j}=x)$, j=1,2,3,4,5 model 6.3.8

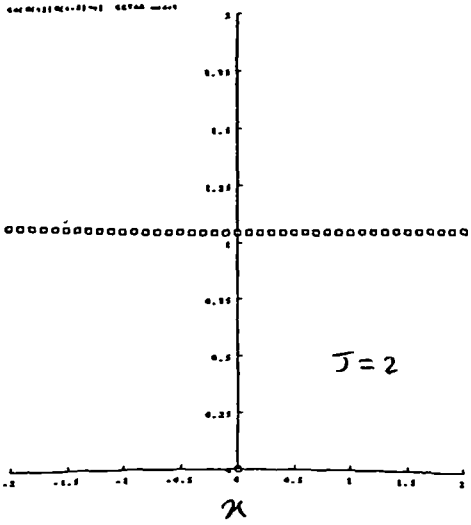
$$X_t=\begin{cases}0.7-0.5X_{t-1}+e_t & \text{if } X_t\leq 0\\ -1.8+0.7X_{t-1}+e_t & \text{if } X_t>0\end{cases}$$

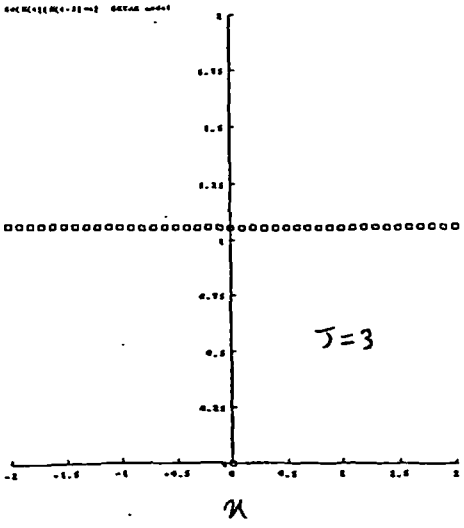


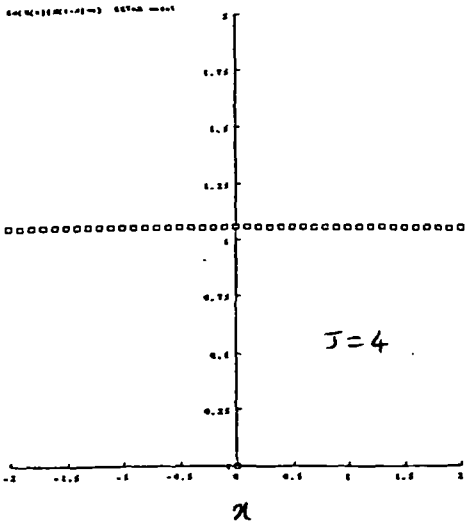
Figs. 6.3.5 $E(X_t|X_{t-j}=x)$, j=1,2,3,4,5 model 6.3.9

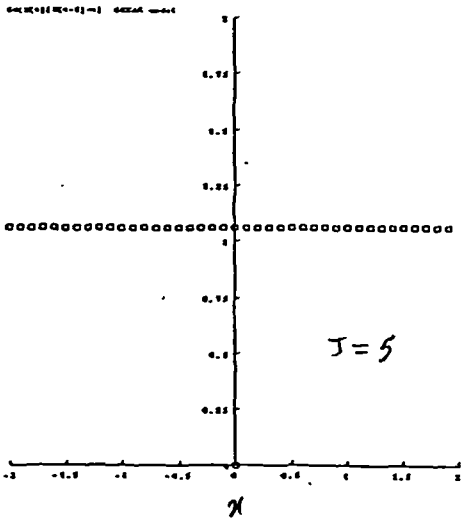
$$X_t=-0.5|X_{t-1}|+e_t$$





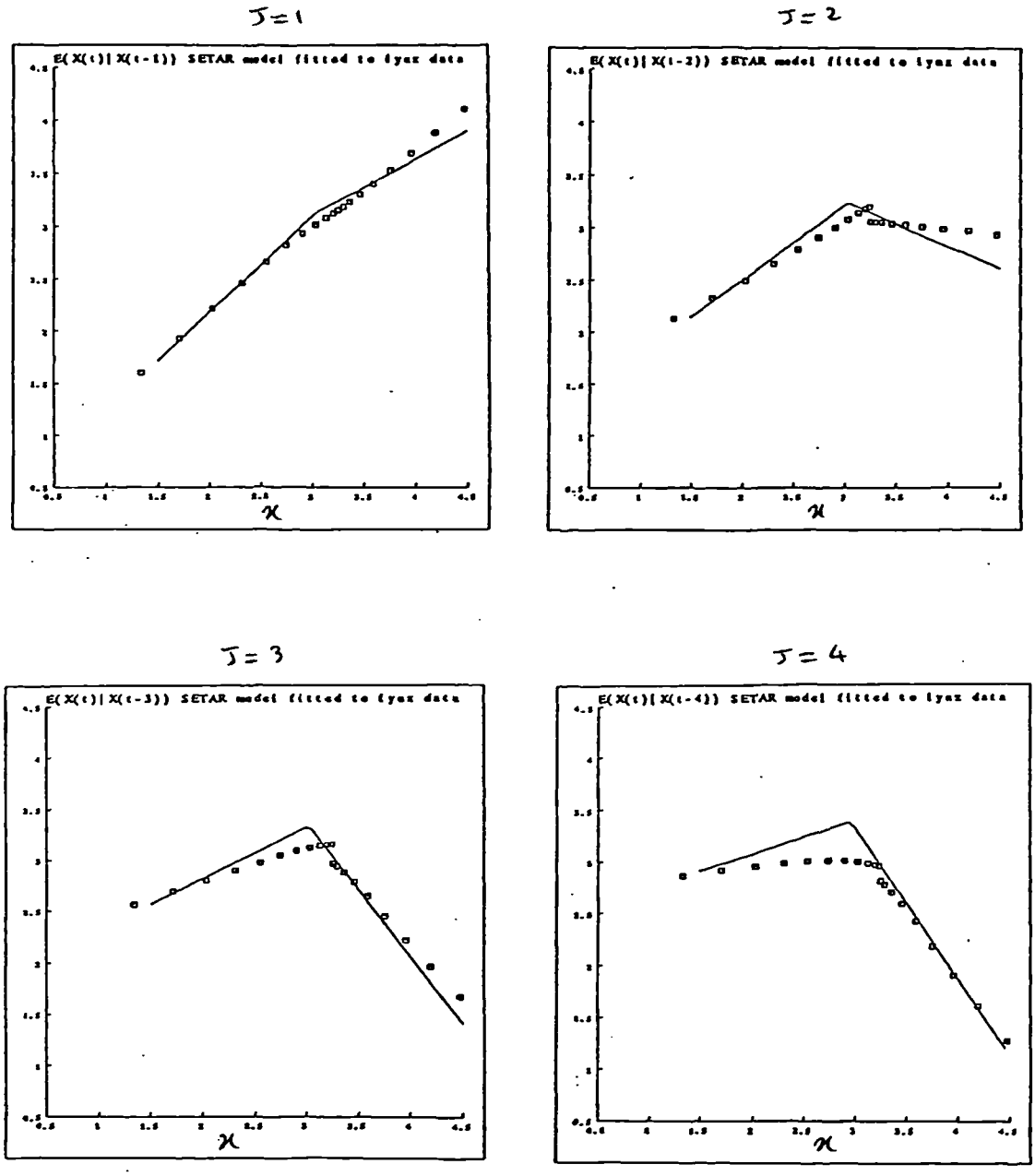






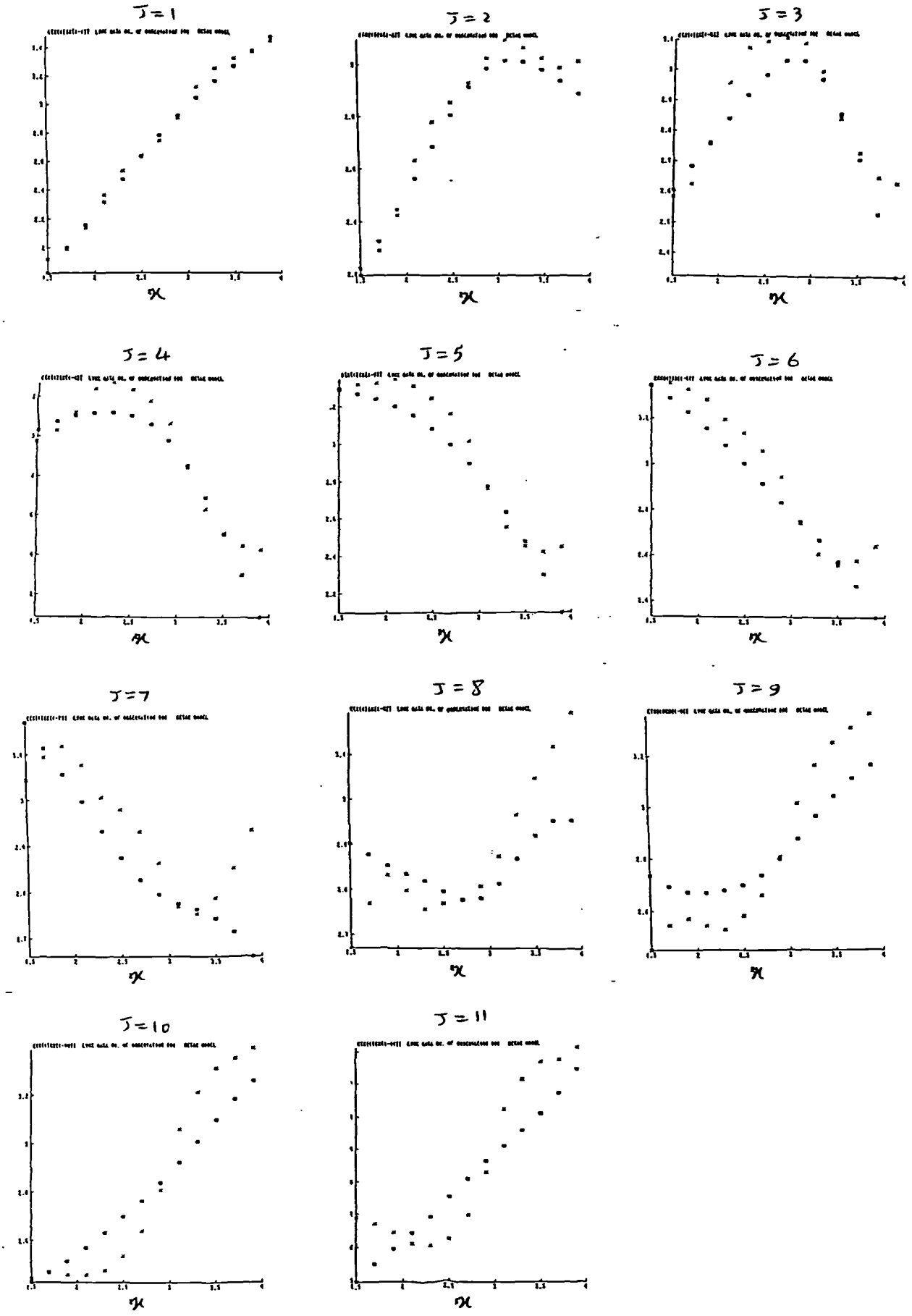
Figs. 6.3.5 $V(X_t \mid X_{t-j}=x)$, $j=1,2,3,4,5$ model 6.3.9

$$X_t = -0.5 \mid X_{t-1} \mid + e_t$$



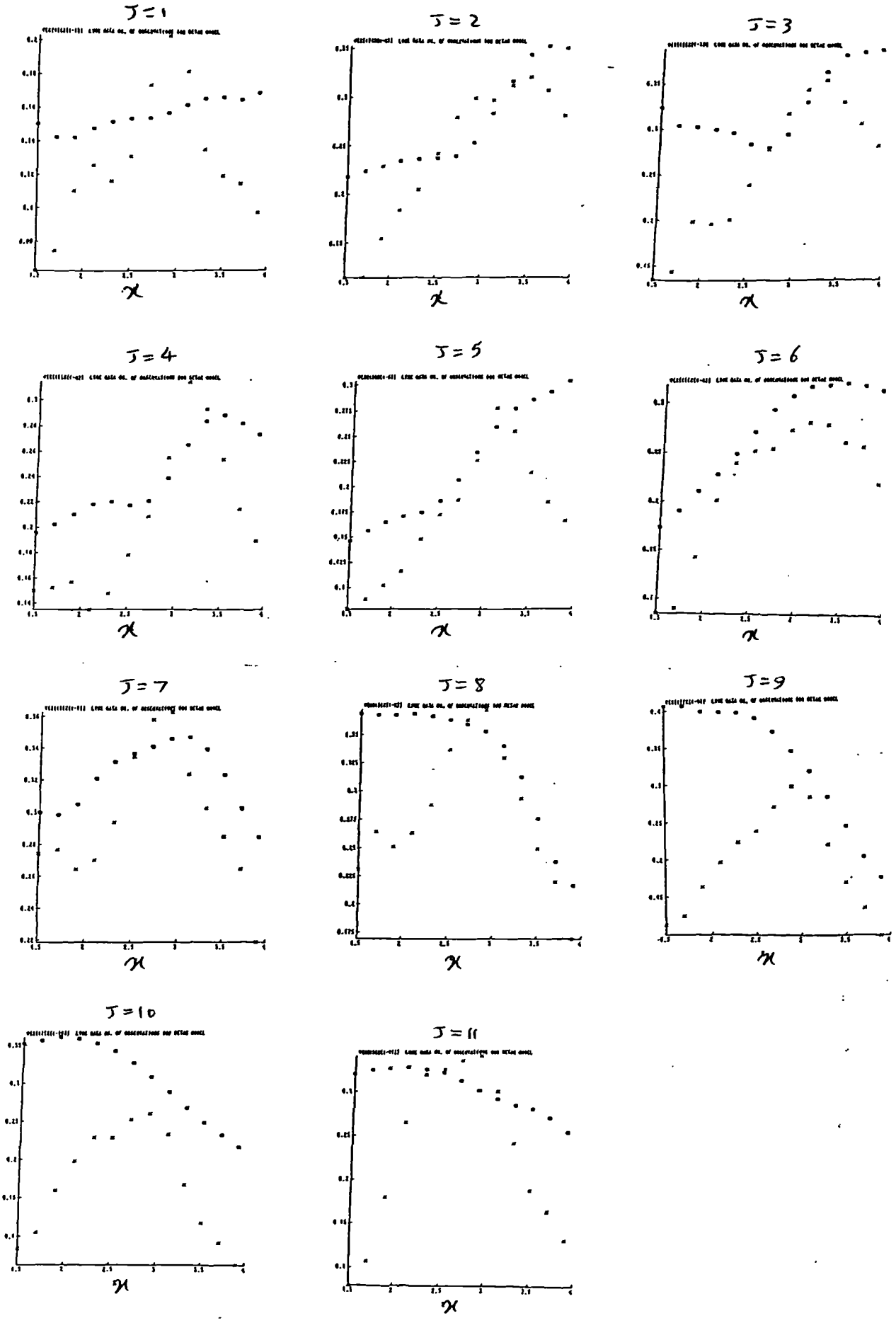
Figs. 6.3.6 $E(X_t|X_{t-j}=x)$, $j=1,2,3,4$ model 6.3.10

$$X_t=\begin{cases}0.62+1.25X_{t-1}-0.43X_{t-2}+e_t & \text{if } X_{t-2}\le 3.25\\ 2.25+1.52X_{t-1}-1.24X_{t-2}+e_t & \text{if } X_{t-2}>3.25\end{cases},$$



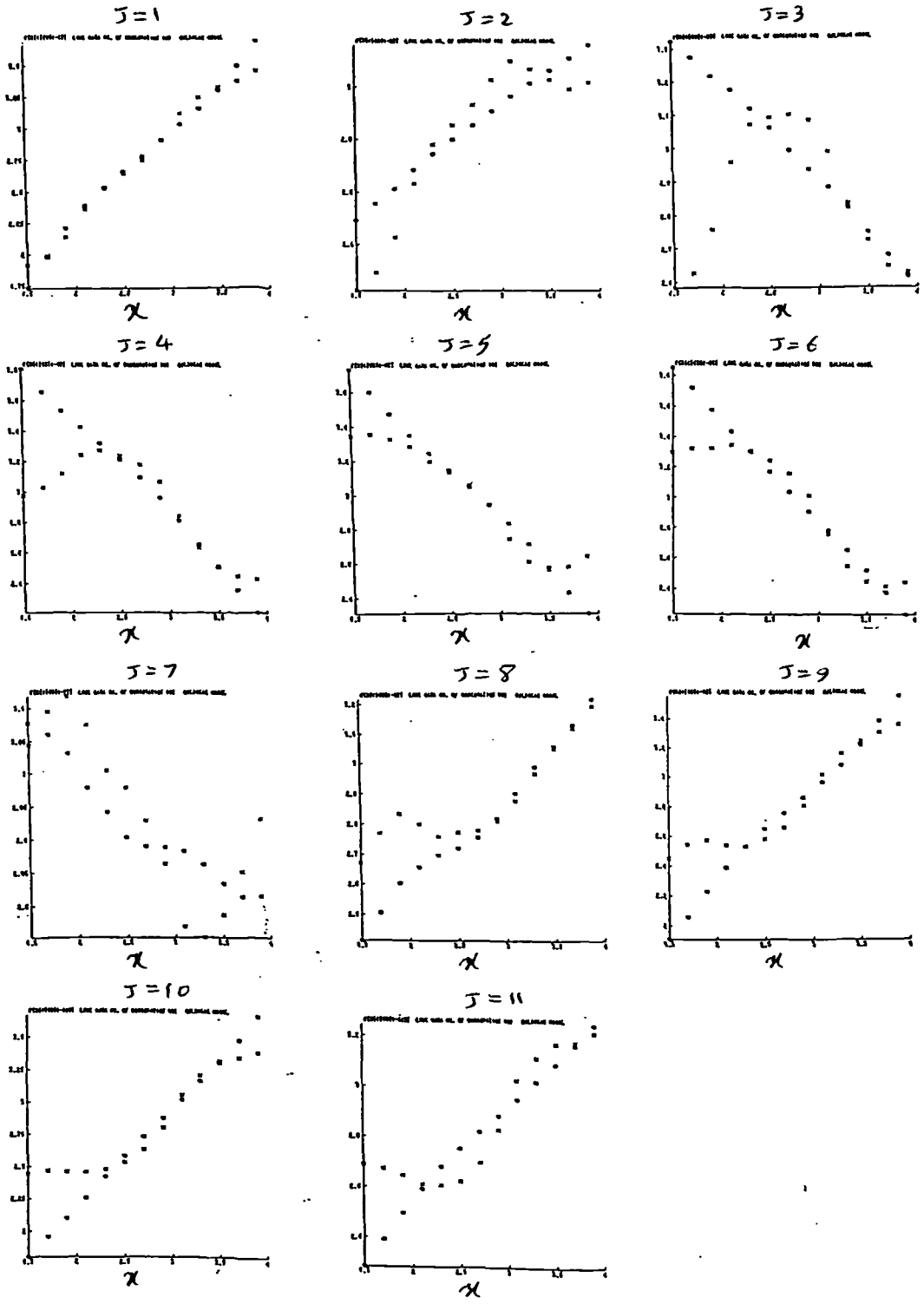
Figs. 6.4.2 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.3

SETAR(2;5,2) which is fitted to lynx data.



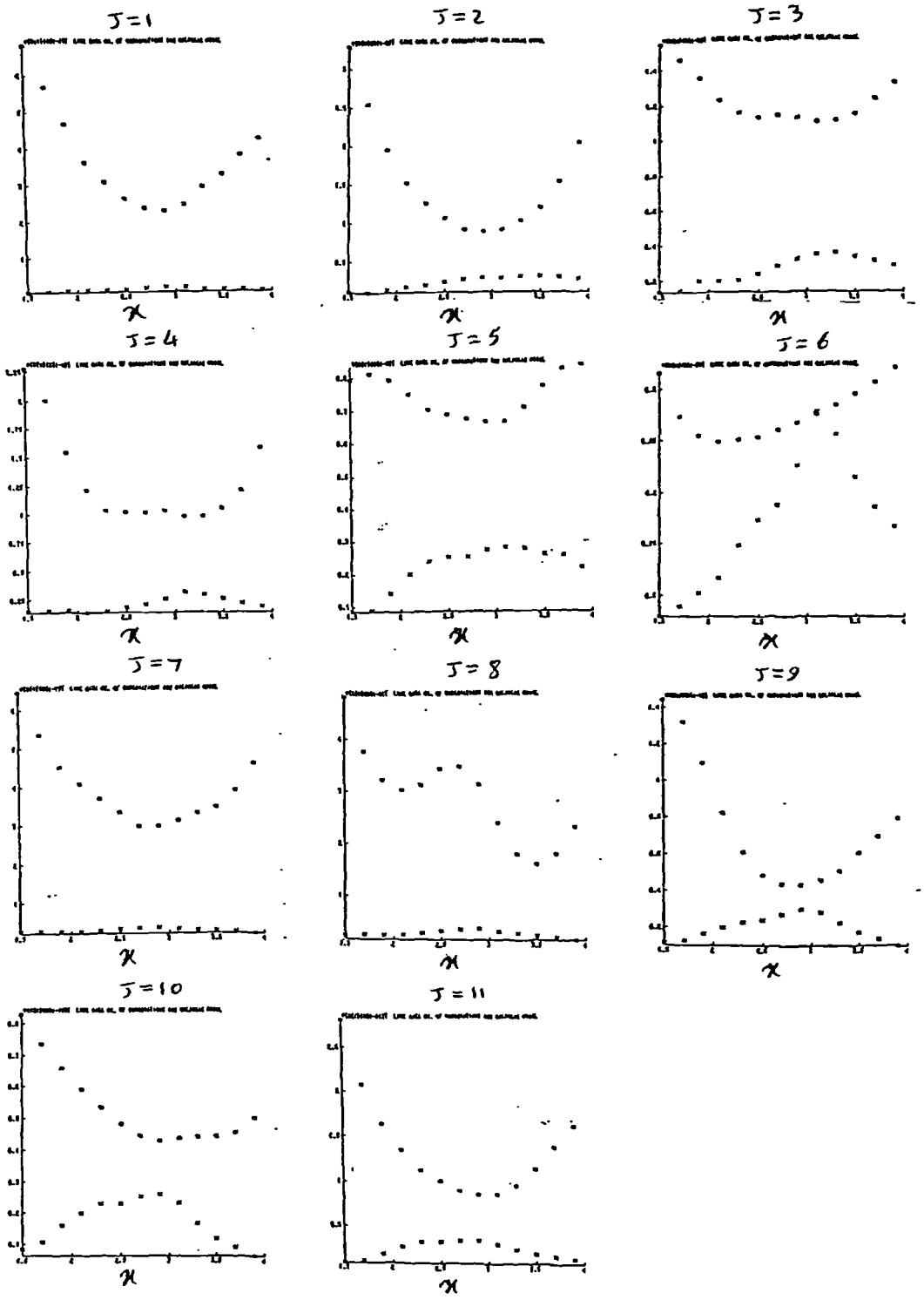
Figs. 6.4.2 $V(X_t \mid X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.3

SETAR(2;5,2) which is fitted to lynx data.



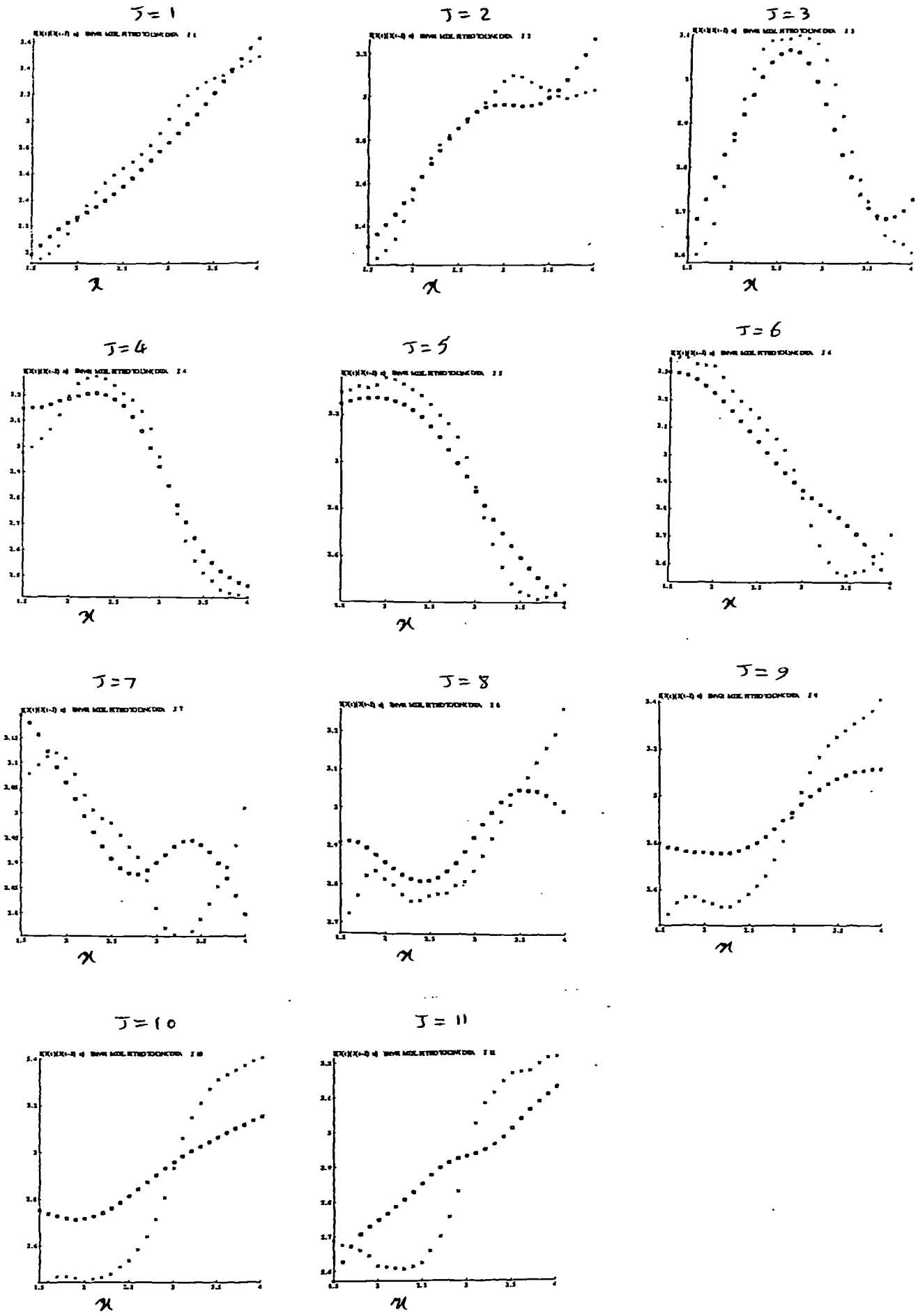
Figs. 6.4.1 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.5

bilinear model which is fitted to lynx data.



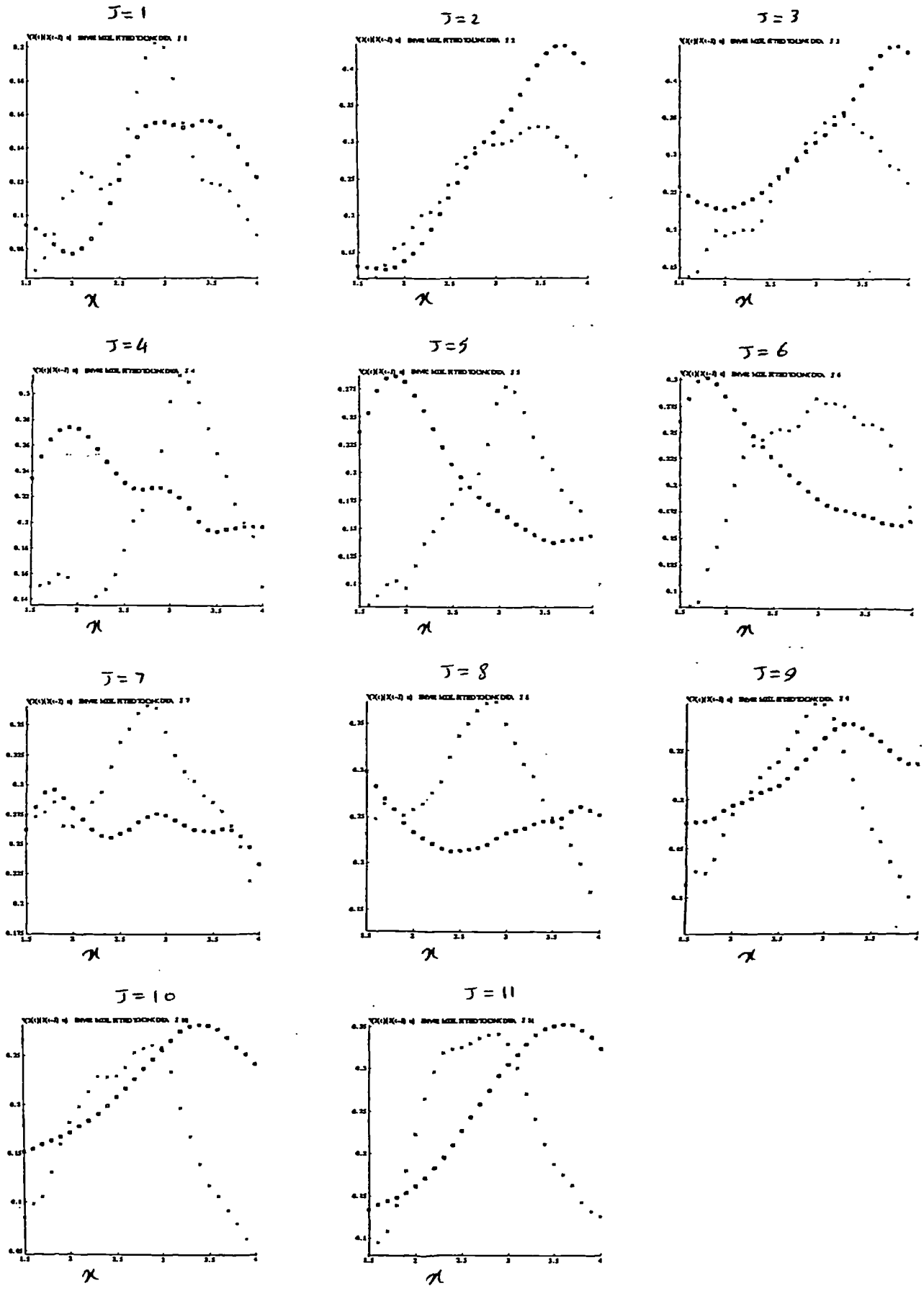
Figs. 6.4.1 $V(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.5

bilinear model which is fitted to lynx data.



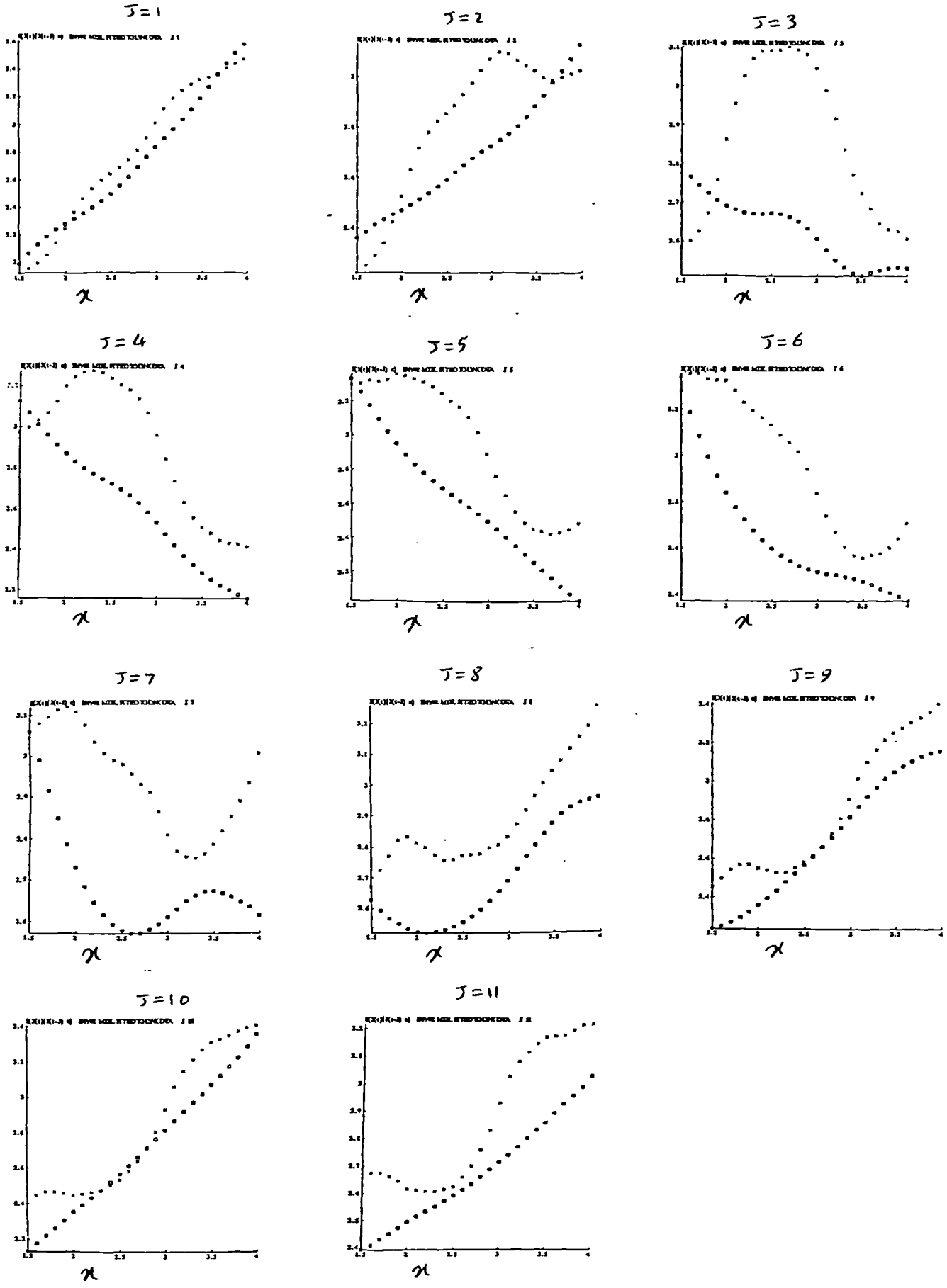
Figs. 6.4.3 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.6

EXPAR model which is fitted to lynx data.



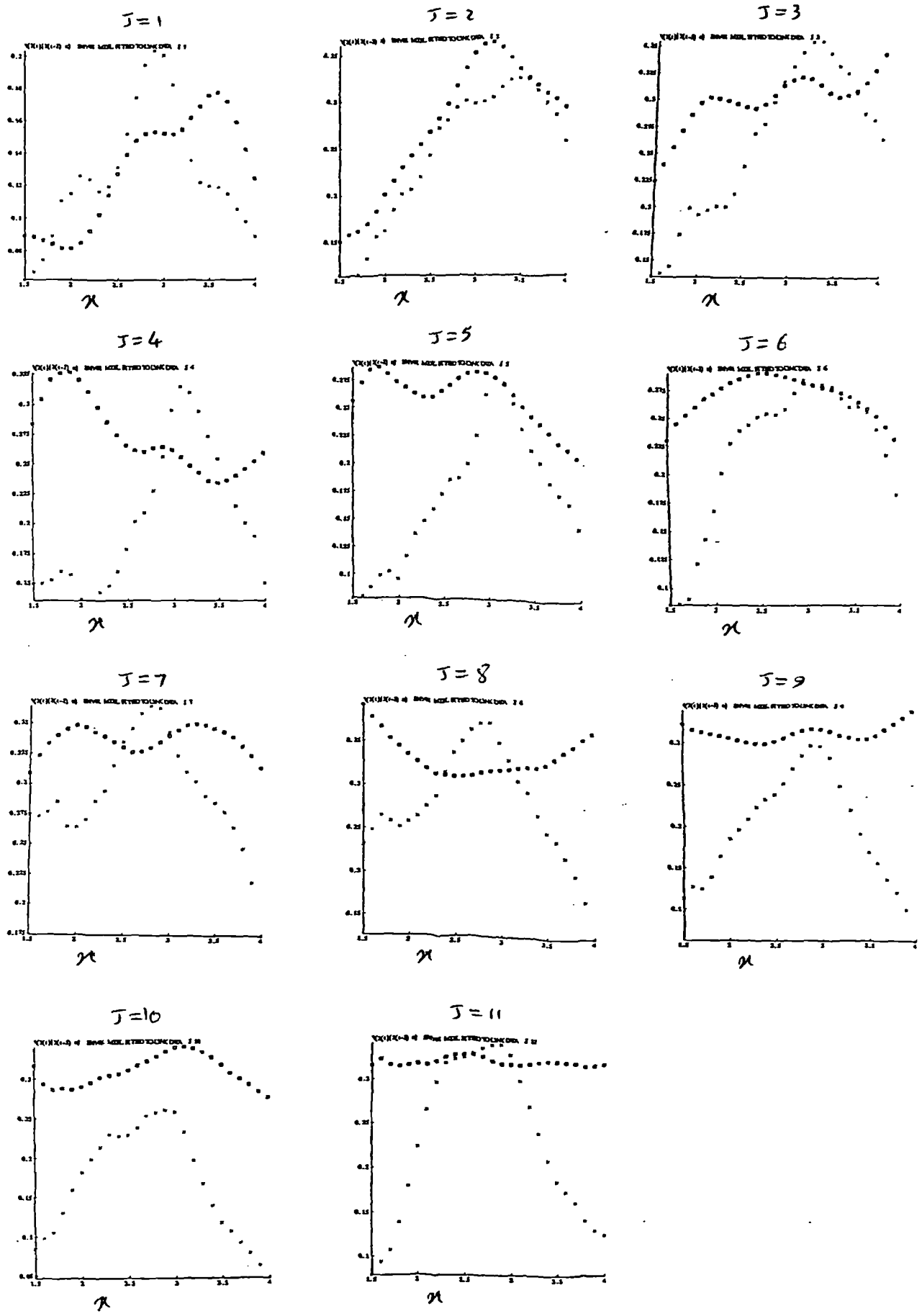
Figs. 6.4.3 $V(X_t \mid X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.6

EXPAR model which is fitted to lynx data.



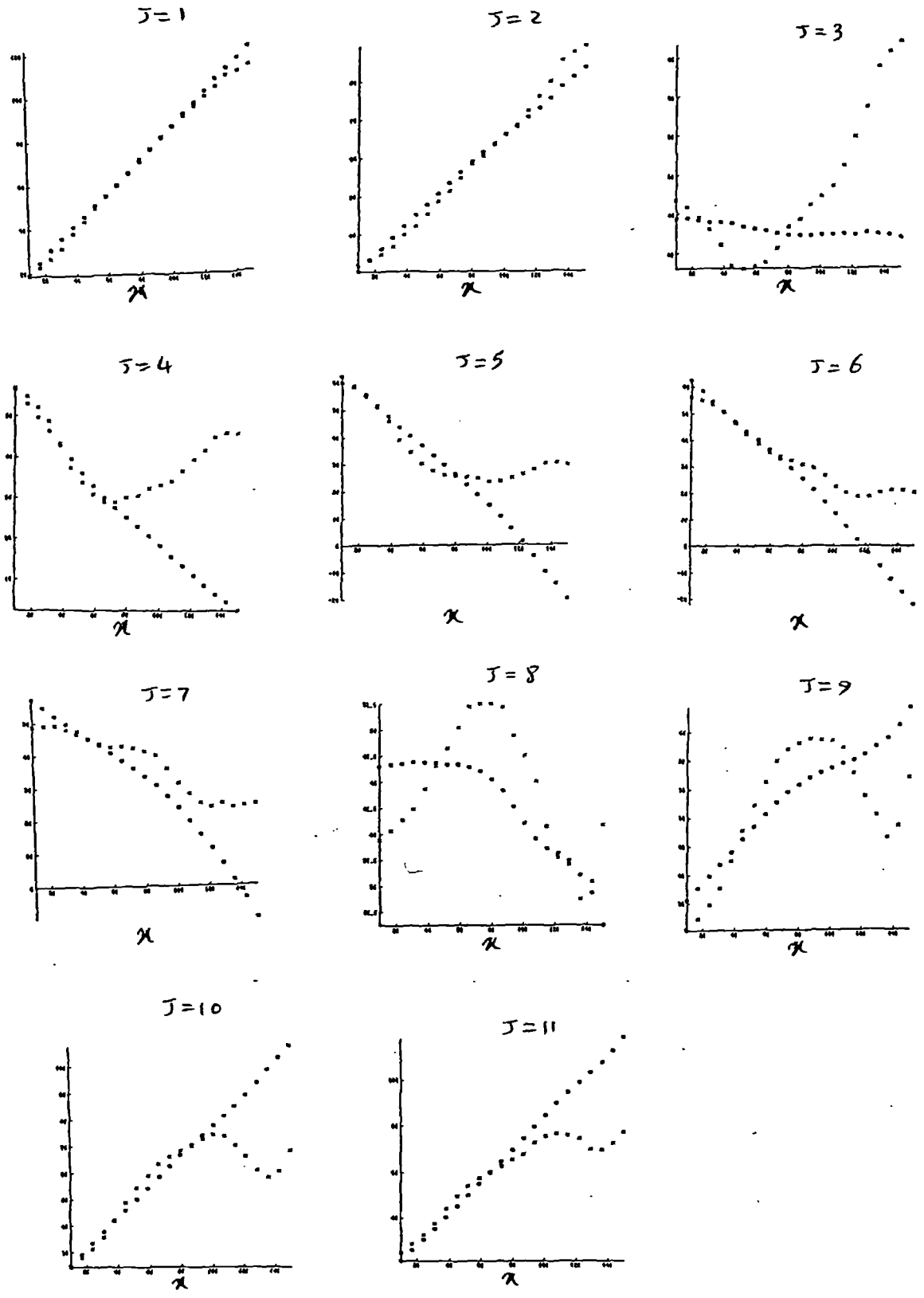
Figs. 6.4.4 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.7

Second EXPAR model which is fitted to lynx data.



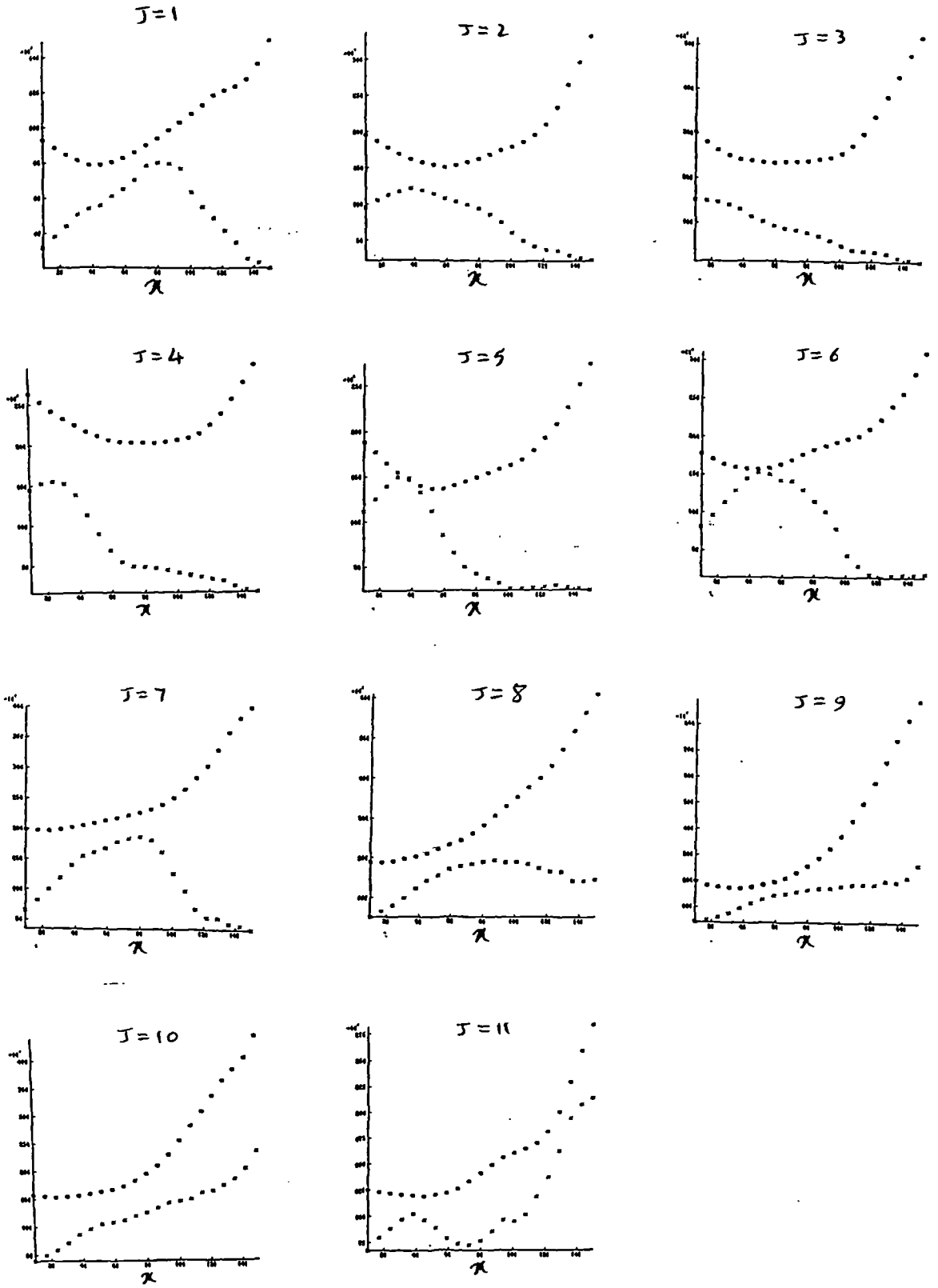
Figs. 6.4.4 $V(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.7

Second EXPAR model which is fitted to lynx data.



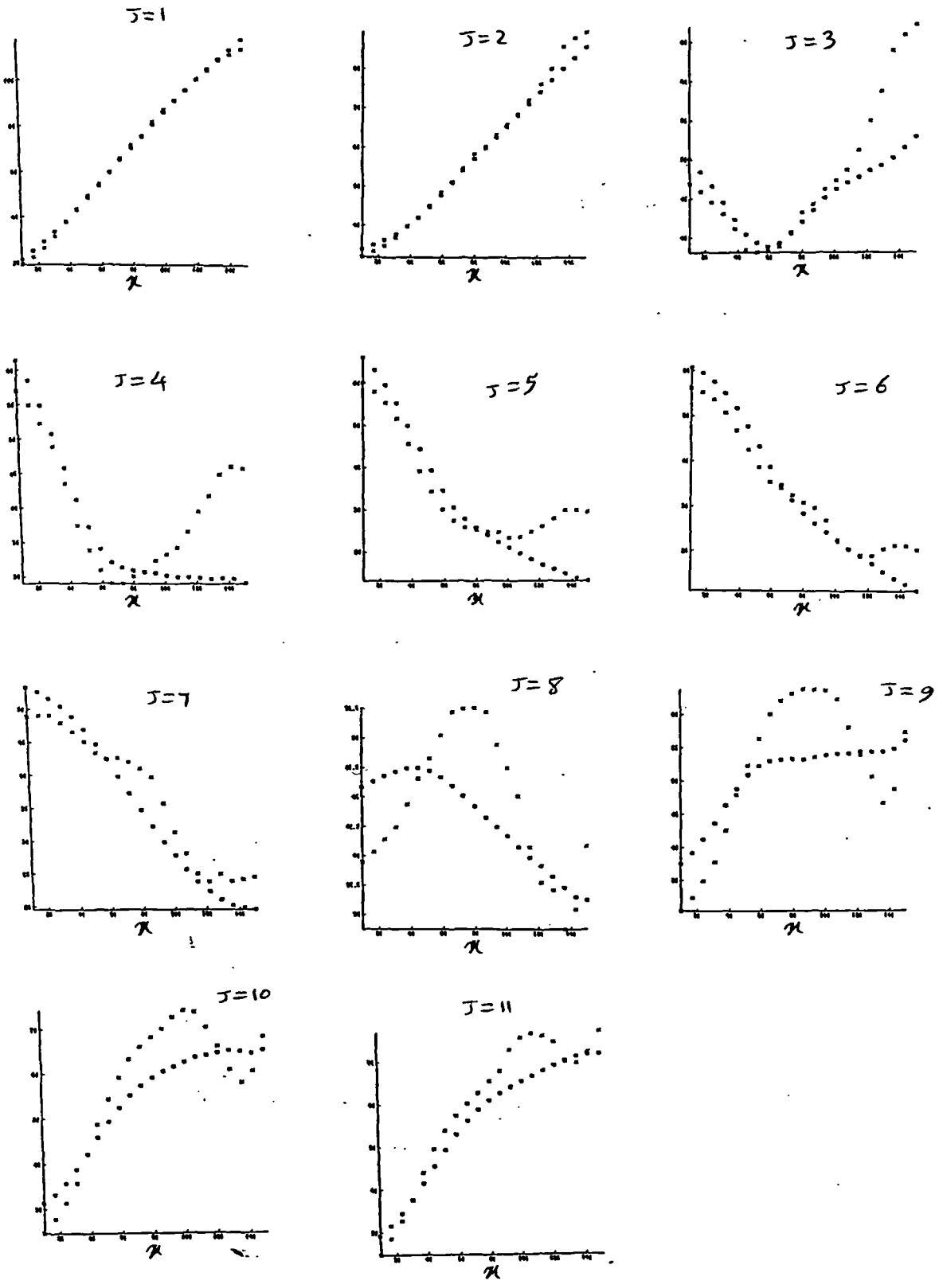
Figs. 6.4.5 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.9

Bilinear model which is fitted to sunspot data.



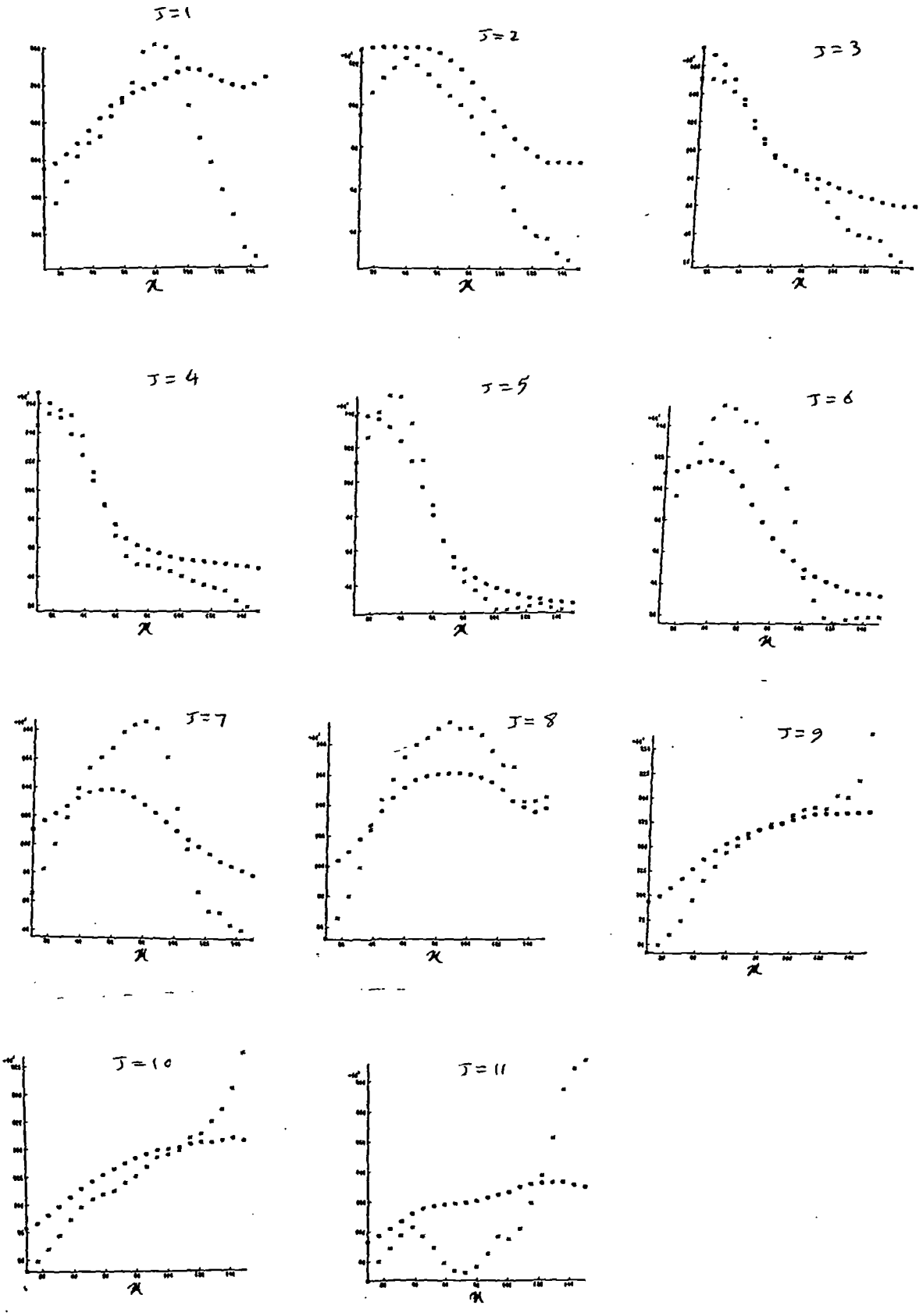
Figs. 6.4.5 $V(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.9

Bilinear model which is fitted to sunspot data.



Figs. 6.4.6 $E(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.10

SETAR(2;3,11) which is fitted to sunspot data.



Figs. 6.4.6 $V(X_t|X_{t-j}=x)$, $j=1,\ldots,11$ model 6.4.10

SETAR(2;3,11) which is fitted to sunspot data.

CHAPTER SEVEN

A COMPARISON OF LIKELIHOOD RATIO TEST AND CUSUM AND TSAY'S TESTS FOR THRESHOLD AUTOREGRESSION

As the theory and application of non-linear time series analysis continues to evolve, the need of tests for non-linearity will grow. If, as appears to be the case, there is a need for non-linear modelling in time series analysis, it certainly follows that there is a need for methods to ascertain when a series is non-linear.

Recently there has been a growing interest in studying the problem of testing threshold non-linearity in time series. One of the proposed tests is the Portmanteau test of non-linearity of Petruccelli and Davies (1986); it is based on ordered autoregression and predictive residuals. Tsay (1987)-considered a variant of the test. Earlier, K.S. Chan and Tong suggested a likelihood ratio approach in a Research Workshop in Singapore in 1984. The details have been circulated in the form of an unpublished technical report from the Chinese University of Hong Kong and some details are available in K.S. Chan and Tong (1986a). A first implementation of the idea was reported in W.S. Chan and Tong (1986). Also, they give a detailed discussion of several frequency and time domain tests for general non-linearity.

In this chapter we investigate some tests designed to detect self exciting threshold autoregressive (SETAR) type non-linearity.

7.1. The Likelihood Ratio Test for Threshold Autoregression

Consider the following threshold model, i.e. the SETAR(2;p,p) model

$$X_t = \begin{cases} b_o^{(1)} + \sum_{i=1}^{p} b_i^{(1)} X_{t-i} + e_t & \text{if } X_{t-d} \le r \\ b_o^{(2)} + \sum_{i=1}^{p} b_i^{(2)} X_{t-i} + e_t & \text{if } X_{t-d} > r \,, \end{cases} \tag{7.1.1}$$

where p and d are known positive integers, r is an unknown parameter usually called the threshold

parameter, $\{e_t\}$ is a sequence of independent identically distributed random variables with zero mean and variance $\sigma^2 < \infty$, and e_t is independent of X_s , $s < t$. Given a set of observations $X_1, X_2, \cdots, X_N$, K.S. Chan and Tong (1986a) consider testing the hypothesis

$$H_o\colon\ b_i^{(1)} = b_i^{(2)}\ ,\ \text{for}\, i = 0,1,2,\ldots,p\ ,$$

against

$$H_1\colon\ b_i^{(1)} \neq b_i^{(2)}\ ,\ \text{for}\ \textit{some}\ i\ ,\ i = 0,1,2,\ldots,p\ .$$

H_1 states that the generating mechanism is non-linear and is piecewise linear. Under H_o, the nuisance parameter, r , is absent.

Assume that e_t is normally distributed; then the likelihood ratio test statistic λ_r is given by

$$\lambda_r = \left\{\hat{\sigma}^2_{(NL,r)}/\hat{\sigma}^2_{(L)}\right\}^{\frac{(N-p)}{2}} \tag{7.1.2}$$

where N denotes the sample size, $\hat{\sigma}^2_{(NL,r)}$ is the usual average residual sum of squares under H_1 for fixed r and $\hat{\sigma}^2_{(L)}$ is that under H_o.

If r is known, then under H_o, $-2\ln\lambda_r$ is asymptotically χ^2_{p+1} (K.S. Chan and Tong, 1986a). However, in practice, the threshold parameter r is seldom known and the likelihood ratio test statistic becomes

$$\lambda' = \left\{\hat{\sigma}^2_{(NL,\hat{r})}/\hat{\sigma}^2_{(L)}\right\}^{\frac{(N-p)}{2}} \tag{7.1.3}$$

where $\hat{r}$ *is the least squares estimate of* r. The asymptotic distribution of λ' *is no longer* χ^2. To use the likelihood ratio test in practice, W.S. Chan and Tong (1986) have resorted to the computing-intensive Monte Carlo method to obtain approximate tail areas of the null distribution for the likelihood ratio test statistic λ'.

Inherent in the likelihood ratio approach is a fundamental difficulty in that the threshold parameter is a nuisance parameter which is absent under the null hypothesis (of linearity). This invalidates the standard asymptotic theory. The two implementations of the likelihood ratio approach represent two different pragmatic methods of obtaining a crude null distribution. A major disadvantage of the W.S. Chan-Tong implementation is the computation time which is much greater than the other time-domain tests for non-linearity, although it may still be argued that with modern fast computers the problem should not be exaggerated.

Fortunately, much theoretical progress has been made in the past year and there are now available theoretical results: *The asymptotic null distribution may be identified with the first passage probability of an m-dimensional Gaussian process* (Chan and Tong 1988). If m is equal to one, *explicit* null distribution may be tabulated via a Brownian bridge result. If m is greater than one, unfortunately the first passage probability problem is not yet resolved by the probabilists to the extent that explicit null distributions may be tabulated. In such cases, an approximation of the tail probability of the null distribution may be obtained using a Monte Carlo technique.

It is therefore of interest to compare the likelihood ratio test using the newly available asymptotic results with a non-likelihood ratio approach such as the modified Petruccelli-Davies' test and Tsay's test.

Definition 7.1.1: A stochastic process $\{X_t(\omega): t \geq 0\}$ defined on a probability space $(\Omega,\mathbf{F},P)$ is called a *Brownian motion* or a *Wiener process* if it satisfies:

a. $X_o(\omega) = 0$ for almost all ω and

b. the system $\{X_t(\omega): t \geq 0\}$ is Gaussian on $(\Omega,\mathbf{F},P)$, and for any t and h with $t+h > 0$, $X_{t+h}(\omega)-X_t(\omega)$ has expectation 0 and variance $|h|$.

The parameter t denotes time and usually extends over the interval $[0, 1]$ or $[0, \infty]$.

Take a Brownian motion which begins at the point x, such motion being realized by $X_t(\omega)+x$. We wish to "deform" it so that it passes through a fixed point y at a fixed instant t_o of time. In other words we are interested in the modified motion of a Brownian particle constrained at the extreme points of the interval $[0, t_o]$.

Definition 7.1.2: A *Brownian bridge* or *tied down Brownian motion* is a Gaussian process whose mean μ_t and covariance $\gamma_{t,s}$ are of the form

$$\mu_t = E(X_t) = \frac{t_o - t}{t_o} x + \frac{t}{t_o} y ,$$

$$\gamma_{t,s} = E\{[X_t - E(X_t)]\,[X_s - E(X_s)]\}$$

$$= s\frac{t_o - t}{t_o} , \; s < t .$$

A Brownian bridge depends on the choice $[0, t_o]$ of the time interval, the starting point x and the terminal point y. (See T. Hida 1980).

Recently, K.S. Chan and Tong (1988) have proved that the null distribution is given by the first passage probability of a higher dimensional Gaussian process. Moreover, for the special cases of (i) SETAR(2;0,0) and (ii) SETAR(2;1,1) without the intercepts, (i.e. $b_o^{(i)}=0$, *all* i), the asymptotic distribution of the likelihood ratio test may be tabulated. (See Chan and Tong 1988 for detail). This greatly facilitates the use of the test.

In practice, we may proceed as follows: First, the time series data $X_1, X_2, \cdots, X_N$ are sorted in ascending order, giving $X_{(1)} \le X_{(2)} \le \cdots \le X_{(N)}$. For each $r = X_{(i)}$, the likelihood ratio test statistic

$$\lambda_r = N\left\{(\hat{\sigma}^2_{(L)} - \hat{\sigma}^2_{(NL,r)})/\hat{\sigma}^2_{(NL,r)}\right\} \tag{7.1.4}$$

can be calculated. At level α, H_o will be rejected if $\lambda_r > C$ for some r, where C is the critical value determined from the first passage probability of a 1-dimensional Brownian bridge. To use the tabulated values, r is allowed to vary from $X_{(\nu_1)}$ to $X_{(\nu_2)}$, $\nu_1 < \nu_2$, where $X_{(\nu_1)}$ and $X_{(\nu_2)}$ are such that $P_{H_o}(X_t \le X_{(\nu_1)}) = p$ and $P_{H_o}(X_t \le X_{(\nu_2)}) = q$, ($q > p$). In all our comparative study to be described later, we set $p = 25\,\%$ and $q = 75\,\%$.

7.2. Tests Based on Predictive Residuals

Consider a SETAR(l;p,p) model with delay parameter d and thresholds $-\infty = r_o < r_1 < \cdots < r_l = \infty$,

$$X_t = b_o^{(j)} + \sum_{i=1}^{p} b_i^{(j)} X_{t-i} + e_t\,, \quad r_{j-1} < X_{t-d} \le r_j\,, \tag{7.2.1}$$

where $t = \max\{\, d+1\,,\, p+1 \,\}, \ldots, N$, and $\{e_t\}$ a sequence of independent identically distributed random variables with mean zero and variance $\sigma^2 < \infty$. It is more convenient to rewrite the model (7.2.1) in terms of an ordered autoregression.

Let $h = \max\{\, 1\,,\, p+1-d \,\}$ and $X_{(i)}$ denote the ith smallest observation among $X_h, \ldots, X_{N-d}$, $i = 1,2,\ldots,N-d-h+1$. Suppose $m_o = 0$, $m_l = N$ and m_j, $j = 1,\ldots,l-1$, are such that $X_{(m_j)} \le r_j < X_{(m_j+1)}$. Model (7.2.1) may be rewritten as

$$X_{(i)+d} = b_o^{(j)} + \sum_{\nu=1}^{p} b_\nu^{(j)} X_{(i)+d-\nu} + e_{(i)+d}\,, \tag{7.2.2}$$

$$i = m_{j-1}+1, \ldots, m_j\,,\ j = 1,\ldots,l\,.$$

Under the null hypothesis that $b_i^{(j)} = b_i^{(k)}$ for *all* $j \ne k$ and for *each* i, i.e. model (7.2.1) is linear, we can write the ordered autoregression (7.2.2) as

$$\mathbf{Y} = \mathbf{Xb}+\mathbf{e} \qquad (7.2.3)$$

where $\mathbf{Y}$ is a column vector containing $X_{(i)+d}$, $i=1,...,N-d-h+1$, $\mathbf{X}$ is an $(N-d-h+1)\times(p+1)$ design matrix and $\mathbf{b}$ is a column vector of unknown parameters $b_o^{(1)}$, $b_1^{(1)}$, $\cdots$, $b_p^{(1)}$, $\mathbf{e}$ is a column vector of noise terms.

Let $r_{\min}$ denote a positive integer $\ll N$. For each $r \geq r_{\min}$, an AR(p) (p is the fixed order) model of the form (7.2.3) is fitted and z_{r+1}, the one step ahead standardized forecast error (i.e. the predictive residual) is computed successively for $r = r_{\min}, r_{\min}+1, ..., N-d-h$.

Petruccelli and Davies (1986) form the CUSUMS

$$Z_r = \sum_{i=r_{\min}+1}^{r} z_i \,, \quad r = r_{\min}+1, \ldots, N-d-h+1$$

and from them they construct the P statistic,

$$P = \max_{r_{\min}+1 \leq r \leq N^*+r_{\min}} |Z_r| / \sqrt{N^*}$$

where $N^* = N-d-h+1-r_{\min}$.

More recently, Petruccelli (1988) has improved the test and a *reverse CUSUM* test results. Let

$$W_r = \sum_{i=1}^{r} z_{N^*+r_{\min}+1-i} \,, \quad r = 1,2,...,N^* \,.$$

In order to provide greater sensitivity to deviations in W_r for small r, he has used boundaries of the form $ar+b$, and H_o is rejected if $|W_r| > ar+b$ for some r, $1 \leq r \leq N^*$.

For a level α test, he has suggested that choosing

$$b = \left[\sqrt{(-(N^*/2)\ln(\alpha/2))}\right]/2 \quad \text{and} \quad a = \sqrt{(-\ln(\alpha/2)/2N^*)}$$

gives satisfactory results. We follow this suggestion in our comparative study in §4.

The second test based on one step ahead forecast errors is that of Tsay (1987). He suggests performing the regression

$$\mathbf{z} = \tilde{\mathbf{X}}\beta+\varepsilon \,,$$

where $\mathbf{z}$ is the vector of standardized one step ahead forecast errors with entries z_i, $i=r_{\min}+1, \cdots, N^*+r_{\min}$, and $\tilde{\mathbf{X}}$ is the design matrix of the form (7.2.3) with the first $r_{\min}$ rows deleted, and computing the usual test statistic $\hat{F}$ for testing $H_o : \beta = \mathbf{0}$,

$$\hat{F}_{\upsilon_1,\upsilon_2} = \frac{N^*-p-1}{p+1} \cdot \frac{\Sigma z_i^2 - \Sigma \hat{\varepsilon}_i^2}{\Sigma \hat{\varepsilon}_i^2} .$$

He has shown that under H_o , $\hat{F}$ is asymptotically distributed as $F_{\upsilon_1,\upsilon_2}$ where $\upsilon_1 = p+1$, $\upsilon_2 = N^*-p-1$.

7.3. Empirical Critical Value

For the special cases of (i) SETAR(2;0,0) and (ii) SETAR(2;1,1) without the intercepts (i.e. $b_o^{(i)} = 0$, *all i*) in formula (7.2.1), the asymptotic distribution of the likelihood ratio test is tabulated. (See Chan and Tong 1988). For models of the forms (i) and (ii), the asymptotic 5% point to the likelihood ratio statistic is 7.84 . An uncritical use of the standard likelihood ratio theory would refer the test statistic to the 5% point of a χ_1^2 (i.e. 3.84) . The correct value of 7.84 is closer to the 5% point of a χ_3^2 distribution ! In practice for those models outside the special cases (i) and (ii), we need the empirically determined 5% point.

The similarity of the 5% point of the cases (i) and (ii) with the χ_3^2 suggests that the empirical 5% point of the model (7.2.1) with intercept is close to the 5% point of χ_{3p}^2 and without intercept is near to the 5% point of $\chi_{3(p-1)}^2$.

In order to examine this suggestion, for each following model, 100 replications each of length 100, using a simulation technique is generated. For each realization, the likelihood ratio test statistic (7.1.4) is calculated. At level 5% , linearity is rejected if $\lambda_r > C$ for some r, where C is the critical value determined from the χ_{3p}^2 distribution and r is allowed to vary from $X_{(v_1)}$ to $X_{(v_2)}$, where $X_{(i)}$'s are ascending sorted data, $X_{(v_1)}$ and $X_{(v_2)}$ are such that $P_{H_o}(X_t \le X_{(v_1)}) = 0.25$ and $P_{H_o}(X_t \le X_{(v_2)}) = 0.75$. The results are given by Table 7.3.1 .

The models are:

$$Model\ 1: X_t = 0.4X_{t-1} - 0.3X_{t-2} + e_t$$

$$Model\ 2: X_t = -0.4e_{t-1} + 0.3e_{t-2} + e_t$$

$$Model\ 3: X_t = (0.3 - 0.8exp(-X_{t-1}^2))X_{t-1} + e_t$$

$$Model\ 4: X_t = 0.5 - 0.4X_{t-1} + 0.4X_{t-1}e_{t-1} + e_t$$

$$Model\ 5: X_t = \begin{cases} 1 - 0.5X_{t-1} + e_t & \text{if } X_{t-1} \le 0 \\ -1 - 0.5X_{t-1} + e_t & \text{if } X_{t-1} > 0 \end{cases}$$

$$Model\ 6: X_t = \begin{cases} 2 + 0.5X_{t-1} + e_t & \text{if } X_{t-1} < 1 \\ 0.5 - 0.4X_{t-1} + e_t & \text{if } X_{t-1} \ge 1 \end{cases}$$

$$Model\ 7: X_t = -0.4e_{t-1} + 0.3e_{t-2} + 0.5e_{t-2}e_{t-1} + e_t$$

$$Model\ 8: X_t = -0.3e_{t-1} + 0.2e_{t-2} + 0.4e_{t-1}e_{t-2} - 0.25e_{t-1}^2 + e_t$$

$$Model\ 9: X_t = 0.4X_{t-1} - 0.3X_{t-2} + 0.5X_{t-1}e_{t-1} + 0.8e_{t-1} + e_t$$

$$Model\ 10: X_t = -0.5X_{t-1} + e_t$$

$$Model\ 11: X_t = 1.0 + 0.5X_{t-1} + e_t$$

$$Model\ 12: X_t = e_t$$

$$Model\ 13: X_t = \begin{cases} 1 + e_t & \text{if } X_{t-1} \le 0 \\ -1 + e_t & \text{if } X_t > 0 \end{cases}$$

$$Model\ 14: X_t = 1.0 + 0.5X_{t-1} - 0.75X_{t-2} + e_t$$

$$Model\ 15: X_t = \begin{cases} 1 + 0.5X_{t-1} + e_t & \text{if } X_{t-1} \le 0 \\ -1 + X_{t-1} + e_t & \text{if } X_t > 0 \end{cases}$$

$$Model\ 16: X_t = \begin{cases} -1 - 0.5X_{t-1} + e_t & \text{if } X_{t-1} \le 0 \\ 1 - X_{t-1} + e_t & \text{if } X_t > 0 \end{cases}$$

$$Model\ 17: X_t = \begin{cases} -0.5 + 0.3X_{t-1} + e_t & \text{if } X_{t-1} \le 0.2 \\ 1 + 0.3X_{t-1} + e_t & \text{if } X_{t-1} > 0.2 \end{cases}$$

$$Model\ 18: X_t = 0.5X_{t-1} - 0.3X_{t-5} + e_t$$

$$Model\ 19: X_t = 2.0 + 0.5X_{t-1} - 0.8X_{t-4} + 0.3X_{t-5} + e_t$$

$$Model\ 20:\ X_t = \begin{cases} 1+0.5X_{t-1}-0.6X_{t-2}+0.6X_{t-4}+0.3X_{t-5}+e_t & \text{if } X_{t-1}\le 0 \\ -2+0.5X_{t-1}-0.8X_{t-4}+0.5X_{t-5}+e_t & \text{if } X_t > 0 \end{cases}$$

Judging by the results with respect to Table 7.3.1 and models 1 , 2 , 10 , 11 , 12 , 14 ,18 and 19 , the test for all p, seems to detect the linear models satisfactorily , although frequencies as high as 98 , 99 and 100 are slightly higher than expected. The performance of the test to detect the non-linearity of models 3 , 4 , 5 , 8 , 13 , 15 , 16 , 17, decrease regularly as the order p increase. These suggest that the degree of freedom of null distributions may not be exactly $3p$ for all p and sample size 100 and $d = 1$, therefore we adopt as the null distribution of the test χ^2 with 3 , 6 , 8 , 10 , 12 and 14 degree of freedom for $p = 0,1,2,3,4,5$, and repeat the same experiments in order to examine the adequacy of this adoption. The results of frequency (per cent) of correct decision where a nominal 5% significance level is used in each of the 100 replications, are given in Table 7.3.2 . Examining the performance of tests to detect the non-linearity of models 4 , 12 , 14 , 16 and 17 suggests that the likelihood ratio statistics to the 5% point may be chosen as 7.81 , 12.6 , 14.5 , 17.0 , 19.5 and 22.0 . Under the same conditions, we repeat our experiments on 20 models. The results are given in Table 7.3.3. Note that columns 7 to 11 in tables 7.3.1-7.3.3 represents the percentages of correct decisions.

Table 7.3.1

Model	original type *	p				
		1	2	3	4	5
1	L	98	97	97	100	99
2	L	97	100	100	99	100
3	NL	21	07	12	03	02
4	NL	86	79	76	56	56
5	NL	97	78	63	52	39
6	NL	100	100	100	100	100
7	NL	24	46	33	28	20
8	NL	64	61	54	49	52
9	NL	8	73	53	55	60
10	L	92	99	100	98	96
11	L	98	99	100	98	99
12	L	96	99	96	97	100
13	NL	100	98	93	87	83
14	L	99	98	98	99	98
15	NL	100	99	99	97	92
16	NL	100	99	94	92	90
17	NL	72	69	61	43	34
18	L	98	99	99	100	98
19	L	97	96	100	100	100
20	NL	87	99	98	100	100

* Note: L= Linear model and NL= Non-linear model.

Table 7.3.2

Model	original type	*p*				
		1	2	3	4	5
1	L	98	95	98	96	96
2	L	95	98	95	96	92
3	NL	18	15	17	19	11
4	NL	86	77	83	76	74
5	NL	99	93	85	72	73
6	NL	100	100	100	100	100
7	NL	25	54	46	42	45
8	NL	56	67	65	65	63
9	NL	11	65	71	79	70
10	L	98	96	96	96	99
11	L	99	93	96	100	96
12	L	98	95	98	96	96
13	NL	100	99	100	96	92
14	L	98	97	96	96	91
15	NL	99	100	100	99	96
16	NL	98	99	99	97	93
17	NL	77	72	77	65	50
18	L	99	100	96	98	94
19	L	97	98	98	98	96
20	NL	76	100	100	100	100

Table 7.3.3

Model	original type	p				
		1	2	3	4	5
1	L	97	99	91	94	99
2	L	100	94	97	94	97
3	NL	19	17	14	8	19
4	NL	89	91	84	82	83
5	NL	97	96	85	81	79
6	NL	100	100	100	100	100
7	NL	15	48	49	53	52
8	NL	60	74	65	67	72
9	NL	10	73	86	70	76
10	L	95	93	95	98	96
11	L	97	97	95	95	93
12	L	97	95	95	91	98
13	NL	100	98	98	96	93
14	NL	98	98	97	91	92
15	NL	100	100	100	100	99
16	NL	99	100	98	98	97
17	NL	84	82	75	70	69
18	L	97	99	97	96	94
19	L	99	97	99	97	94
20	NL	87	99	100	100	100

Examining the results of tests with respect to the models 7 , 8 , 9 , and 20 in Tables 7.3.1 , 7.3.2 and 7.3.3 , shows the importance of order p higher than one in improvement of the performance of the likeli-

hood ratio test in detecting the non-linearity. Finally, the following adaptation seems to be applicable.

In practice one can detect the non-linearity of a time series by applying the likelihood ratio test on data set for some values p (e.g. p=1,2) and $d \leq p$. Then one can say the data is a non-linear time series data set if at least one of the different combination of p and d shows the non-linearity of the series.

We employed the above suggestion and applied the likelihood ratio test (with intercept) for $p=1$, $d=1$ and $p=2$, $d=1$, 2 to some real data sets. The nominal 5 per cent significance is used. The empirical 5% level critical values for likelihood ratio tests are 12.6 and 15.6 (χ^2_6 and χ^2_9). The results are given in Table 7.3.4 .

Table 7.3.4

data	data transformation	*LR* *	order and delay **	reverse CUSUM
Lynx, $N = 114$	*RAW*	NL	$p=1, d=1$	NL
	$\log_{10}$	NL	$p=2, d=1$	L
Sunspot, N=256	*RAW*	NL	$p=2, d=1$	L
	square root	NL	$p=2, d=2$	L
Blowfly (first part), N=126	*RAW*	NL	$p=2, d=1$	L
	square root	NL	$p=2, d=1$	L
	$\log_{10}$	NL	$p=1, d=1$	L
Blowfly (last part), N=82	*RAW*	L		L
	square root	L		L
	$\log_{10}$	L		L

* LR: Likelihood ratio test.

** Order and delay: These are the order and delay with which the test reject the hypothesis of linearity.

Not surprisingly the results of Table 7.3.4 column 3 are exactly the same as Table 8 in W.S Chan and

Tong(1986).

A Monte Carlo study has been done by W.S Chan and Tong (1986) based on 100 replications of each the nine models listed as models 1-9 with a minor change in model 7 (i.e. The term $e_{t-1}e_{t-2}$ is replaced by e_te_{t-2} to make the model 7 identical to W.S. Chan and Tong's model). Each replication has a sample size 204 and in each case, e_t–$N(0,1)$.

If the choice of χ^2_{3p} for $p=1$, 2 as the 5% critical value of likelihood ratio test is adequate, then we will expect to get approximately the same result as W.S. Chan and Tong (1986). A Monte Carlo experimentation along the same line as W.S Chan and Tong (1986) with the same models (i.e Models 1-9), gives similar results and it tends to support the adequacy of this choice. Frequencies in per cent of correct decision with nominal significance level at 5 per cent with critical value equal to 12.6 and 15.6 for $p=1$ and $p=2$ are given in Table 7.3.5.

Table 7.3.5

	model and numbers	original type*	CT^{**} test	LR^{***} test
1	AR(2)	L	99	91
2	MA(2)	L	96	96
3	EXPAR	NL	34	37
4	BL	NL	96	100
5	SETAR	NL	100	100
6	SETAR	NL	100	100
7	NLMA	NL	14	21
8	NLMA	NL	96	95
9	BL	NL	98	89

* L: Linear; NL: Non-linear.

** CT: W.S. Chan and Tong (1986) results.

*** LR: Likelihood Ratio test results.

The poor performance of the likelihood ratio test with respect to model 7 (after a minor change) is perhaps due to the fact that $E(X_t|X_s \,.\, s < t)$ is linear in e_s , $s < t$. W.S. Chan and Tong (1986). Monte Carlo experimentation along the same lines as before with following models may give a better support in choosing the $3p$ degree of freedom for $p = 1$ and $p = 2$.

Model A: $X_t = e_t - 0.4e_{t-1} + 0.3e_{t-2} + 0.5e_{t-1}e_{t-2}$

Model B: $X_t = e_t - 0.4e_{t-1} + 0.3e_{t-2} + 0.5e_{t-1}^2$

Model C: $X_t = e_t - 0.4e_{t-1} + 0.3e_{t-2} + 0.5e_t e_{t-1}$

Model D: $X_t = e_t - 0.4e_{t-1} + 0.3e_{t-2} + 0.5e_t e_{t-2}$

The results with 100 replications in each case are given in Table 7.3.6 .

Table 7.3.6

Model	Frequency of correct decisions (5 per cent significance level)	
	CT test	LR test
A	95	89
B	97	100
C	14	29
D	12	22

7.4. A Comparative Study of the Likelihood Ratio Test and the Reverse CUSUM Test

To assess the performance of the likelihood ratio test by reference to the tabulated tail probability of the newly available asymptotic null distribution we simulate a number of SETAR processes which were studied by Petruccelli (1987 and 1988). In our comparative study we follow Petruccelli (1987) and assume that the true parameters p and d are known.

In this study, for those models which fall within the special cases (i) and (ii) described in §2, we refer the likelihood ratio statistic to the asymptotic 5% point of 7.84. An uncritical use of the standard likelihood ratio theory would refer the test statistic to the 5% of a χ^2_1 distribution (i.e. 3.84), and this would be wide off the mark. The correct value of 7.84 is closer to the 5% point of a χ^2_3 distribution! For those models outside the special cases (i) and (ii), we refer the test statistic to the empirically determined 5% point of 12.6; as we have said earlier, we await with keen interest results from probabilists working on higher dimensional Brownian bridges so that the empirical value may be replaced by an asymptotic value.

For each process, 100 replications each of length 50 and 100 replications each of length 100 were generated using the NAG routine G05DDF, which was initialized by NAG routine G05CCF. For each replication, start-up values were set to zero and the first 1499 observations discarded to remove transient effects. It is important to discard a sufficient number of observations. In our simulation studies we have found that discarding the first 500 observations, as suggested by Petruccelli (1987), is not always sufficient.

We have repeated all the models used by Petruccelli (1987) with the exception of two. He has included the two following models in his study:

$$X_t = \begin{cases} -1+0.5X_{t-1}+\varepsilon_t & \text{if } X_{t-1} \le 0 \\ 1+X_{t-1}+\varepsilon_t & \text{if } X_{t-1} > 0 \end{cases}, \tag{7.4.1}$$

and

$$X_t = \begin{cases} -1-0.5X_{t-1}+\varepsilon_t & \text{if } X_{t-1} \le 0 \\ 1+X_{t-1}+\varepsilon_t & \text{if } X_{t-1} > 0 \end{cases}. \tag{7.4.2}$$

These are not ergodic (K.S. Chan at. al 1985) and $|X_t|$ will tend to infinity in probability as $t \to \infty$. (See Fig. 7.4.1 for a partial realization of model (7.4.1).) It is not immediately obvious if existing tests, likelihood ratio approach or not, apply to the non-ergodic cases. We have therefore excluded them in our study.

In our comparative study, we have not been able to duplicate all of Petruccelli's (1987) results for his test; therefore we will give both his reported results (Petruccelli 1988) and our results together.

Note that in Moeanaddin and Tong (1988) the Petruccelli's (1987) results are slightly different from those of Petruccelli's (1988) which will be used here. Since the number of replications cannot be considered large some discrepancies are to be expected. However, in some cases the discrepancies are apparently considerable. In all the tables, the first column (marked RC^*) of power for each N is taken from Petruccelli (1988) and the second column (marked RC) is based on our running the programs † for the reverse CUSUM test.

Table 7.4.1 shows the results (all averaged over 100 replications each) for the 5% test and for N=50 and 100 on the AR(1) processes

$$X_t = b_1 X_{t-1} + e_t$$

for $b_1 = \pm 0.9, \pm 0.5, \pm 0.1$ and 0. It may be seen that the significance level of likelihood ratio test based on asymptotics is reasonable (See column under LR).

† Note: We are most grateful to Professor J. Petruccelli for providing us with his program.

Table 7.4.1

b_1	N=50			N=100		
	RC^*	RC	LR	RC^*	RC	LR
-0.9	0.07	0.06	0.04	0.09	0.06	0.06
-0.5	0.09	0.05	0.02	0.09	0.03	0.06
-0.1	0.08	0.03	0.08	0.05	0.04	0.06
0.0	0.06	0.03	0.04	0.03	0.03	0.06
0.1	0.06	0.03	0.06	0.09	0.01	0.02
0.5	0.03	0.01	0.09	0.04	0.03	0.03
0.9	0.02	0.06	0.09	0.05	0.03	0.09

Table 7.4.2 shows the power results for data simulated from models of the form (7.1.1) with d=p=1 ; l=2 ; r_1=0 , 0.2 ; sample size=50, 100. These models correspond to SETAR models with continuous and discontinuous autoregressive functions, with zero and non-zero thresholds.

Table 7.4.3 shows the power results of SETAR(3;1,1) with r_1=−1 , r_2=1.

Table 7.4.2

SETAR parameters				threshold	power					
					N=50			N=100		
$b_o^{(1)}$	$b_1^{(1)}$	$b_o^{(2)}$	$b_1^{(2)}$	r_1	RC^*	RC	LR	RC^*	RC	LR
0	0.9	0	-0.1	0	0.25	0.31	0.59	0.60	0.62	0.93
0	0.9	0	-0.77	0	0.48	0.47	0.79	0.79	0.75	0.99
0	0.5	0	-0.5	0	0.43	0.38	0.83	0.59	0.70	0.99
0	0.5	0	-0.1	0	0.62	0.62	1.00	0.95	0.91	1.00
0	0	0	-1.0	0	0.49	0.48	0.86	0.74	0.81	0.99
0	-0.5	0	-1.0	0	0.33	0.27	0.51	0.45	0.45	0.86
0	-1.0	0	-0.5	0	0.30	0.23	0.54	0.48	0.49	0.81
0	-1.0	0	0	0	0.52	0.46	0.88	0.81	0.77	0.99
0	-1.0	0	0.5	0	0.66	0.62	0.99	0.93	0.93	1.00
0	-0.5	0	0.5	0	0.43	0.31	0.81	0.66	0.73	0.98
0	-0.77	0	0.9	0	0.31	0.41	0.82	0.78	0.74	0.97
0	-0.1	0	0.9	0	0.24	0.25	0.56	0.55	0.61	0.90
1	0.5	-1	1	0	0.56	0.56	0.91	0.89	0.89	1.00
1	-0.5	-1	1	0	0.99	0.98	0.99	1.00	1.00	1.00
-1	0.5	1	-1	0	0.10	0.16	0.15	0.08	0.16	0.12
1	0.5	-1	-1	0	1.00	1.00	1.00	1.00	1.00	1.00

Note that the null hypothesis, H_o, states that $X_t = \alpha X_{t-1} + e_t$.

Table 7.4.2 (continued)

SETAR parameters				threshold	power					
					N=50			N=100		
$b_0^{(1)}$	$b_1^{(1)}$	$b_0^{(2)}$	$b_1^{(2)}$	r_1	RC^*	RC	LR	RC^*	RC	LR
-1	-0.5	1	-1	0	0.38	0.36	0.88	0.61	0.60	1.00
1	-0.5	-1	-1	0	0.51	0.48	0.82	0.84	0.78	1.00
1	0	1	-0.5	0	0.18	0.15	0.13	0.27	0.20	0.18
1	0	-1	-0.5	0	0.64	0.64	0.90	0.96	0.93	0.99
1	0.5	1	-1.5	0	1.00	1.00	0.99	1.00	1.00	1.00
1	0.5	-1	-1.5	0	1.00	1.00	1.00	1.00	1.00	1.00
0	0.6	0	0.3	0.2	0.07	0.07	0.11	0.16	0.07	0.23
0	0.6	0	-0.3	0.2	0.22	0.27	0.67	0.51	0.57	0.98
0.5	0.3	-1	0.3	0.2	0.42	0.37	0.56	0.75	0.62	0.93
-0.5	0.3	1	0.3	0.2	0.08	0.13	0.45	0.27	0.34	0.90

Table 7.4.3

SETAR parameters						threshold parameters		power					
								N=50			N=100		
$b_0^{(1)}$	$b_1^{(1)}$	$b_0^{(2)}$	$b_1^{(2)}$	$b_0^{(3)}$	$b_1^{(3)}$	r_1	r_2	RC^*	RC	LR	RC^*	RC	LR
0	0.5	0	-0.5	0	0.5	-1	1	0.15	0.11	0.45	0.32	0.26	0.78
-1	0	0	1	1	0	-1	1	0.16	0.22	0.12	0.32	0.27	0.18

After determining the empirical critical value for the likelihood ratio test of order 1 and 2, it is beneficial to compare the relative performance of the likelihood ratio test (LR), reverse CUSUM and Tsay's test on the models which have been used before in comparative studies. The results of Tsay's test have been reported by Petrucelli (1987), therefore we have not repeated them here. We applied the likelihood ratio test on the simulated data with $p=1$, $d=1$ and $p=2$,$d=1$,2. The simulated data is assumed to be non-linear if at least one of the different combination of p and d shows the non-linearity of the data.

Table 7.4.4 shows the results (all averaged over 100 replications) for the 5% test and for $N=50$ and $N=100$ on the AR(1) processes

$$X_t = b_1 X_{t-1} + e_t$$

for $b_1 = \pm 0.9$, ± 0.5 , ± 0.1 and 0.

Table 7.4.4

b_1	N=50			N=100		
	T test	*RC*	*LR*	T test	*RC*	*LR*
-0.9	0.02	0.06	0.05	0.06	0.06	0.09
-0.5	0.05	0.05	0.10	0.06	0.03	0.06
-0.1	0.05	0.03	0.09	0.03	0.04	0.06
0.0	0.06	0.03	0.10	0.04	0.03	0.04
0.1	0.02	0.03	0.10	0.06	0.01	0.09
0.5	0.05	0.01	0.09	0.04	0.03	0.05
0.9	0.07	0.06	0.08	0.03	0.03	0.06

Table 7.4.5 shows the power results for data simulated from models of the form 7.1.1 with $d=p=1$; $l=2$; $r_1=0$, 0.2; sample size = 50 and 100.

Table 7.4.5

SETAR parameters				threshold	power					
					N=50			N=100		
$b_o^{(1)}$	$b_1^{(1)}$	$b_o^{(2)}$	$b_1^{(2)}$	r_1	T test	*RC*	*LR*	T test	*RC*	*LR*
0	0.9	0	-0.1	0	0.11	0.31	0.36	0.20	0.62	0.60
0	0.9	0	-0.77	0	0.16	0.47	0.58	0.36	0.75	0.75
0	0.5	0	-0.5	0	0.17	0.38	0.39	0.45	0.70	0.65
0	0.5	0	-0.1	0	0.44	0.62	0.60	0.80	0.91	0.87
0	0	0	-1.0	0	0.35	0.48	0.47	0.58	0.81	0.65
0	-0.5	0	-1.0	0	0.18	0.27	0.29	0.26	0.45	0.39
0	-1.0	0	-0.5	0	0.18	0.23	0.29	0.32	0.49	0.30
0	-1.0	0	0	0	0.39	0.46	0.50	0.71	0.77	0.75
0	-1.0	0	0.5	0	0.61	0.62	0.68	0.90	0.93	0.87
0	-0.5	0	0.5	0	0.25	0.31	0.39	0.67	0.73	0.53
0	-0.77	0	0.9	0	0.42	0.41	0.56	0.84	0.74	0.73
0	-0.1	0	0.9	0	0.35	0.25	0.27	0.61	0.61	0.54
1	0.5	-1	1	0	0.31	0.56	0.91	0.50	0.89	1.00
1	-0.5	-1	1	0	0.97	0.98	0.99	1.00	1.00	1.00
-1	0.5	1	-1	0	0.02	0.16	0.15	0.11	0.16	0.12
1	0.5	-1	-1	0	1.00	1.00	1.00	1.00	1.00	1.00

Note that under H_0 , $X_t = \alpha_0 + \alpha_1 X_{t-1} + e_t$.

Table 7.4.5 (continued)

SETAR parameters				threshold	power					
					N=50			N=100		
$b_0^{(1)}$	$b_1^{(1)}$	$b_0^{(2)}$	$b_1^{(2)}$	r_1	T test	*RC*	LR	*Ttest*	*RC*	*LR*
-1	-0.5	1	-1	0	0.15	0.36	0.88	0.18	0.60	0.98
1	-0.5	-1	-1	0	0.52	0.48	0.92	0.69	0.78	1.00
1	0	1	-0.5	0	0.17	0.15	0.19	0.28	0.20	0.21
1	0	-1	-0.5	0	0.53	0.64	0.92	0.75	0.93	1.00
1	0.5	1	-1.5	0	0.97	1.00	1.00	1.00	1.00	1.00
1	0.5	-1	-1.5	0	1.00	1.00	1.00	1.00	1.00	1.00
0	0.6	0	0.3	0.2	0.12	0.07	0.13	0.21	0.07	0.07
0	0.6	0	-0.3	0.2	0.22	0.27	0.29	0.35	0.37	0.50
0.5	0.3	-1	0.3	0.2	0.10	0.37	0.70	0.30	0.62	0.93
-0.5	0.3	1	0.3	0.2	0.28	0.13	0.55	0.37	0.34	0.86

The run-time of the likelihood ratio test with reference to the asymptotic null distribution is negligible. For sample size of 50, the run-time varies from ½ to 1 second and for $N = 100$ it varies from 0.5 to 2.0 seconds. The run-time depends on the nature of the model. All in all, computing time involved in the application of the likelihood ratio test is no longer a problem. The performance of the test is quite encouraging as far as the study goes.

7.5. Power Plots

To assess the power performance of the likelihood ratio test (LR), we simulate two classes of SETAR models, SETAR(2;0,0) and SETAR(2;1,1). Specifically we simulate series of lengths 50, 100, 150, 200 and 250, and the power of each test is obtained from 100 replications of each model. We repeat the same experiments with the reverse CUSUM test with a view to comparison.

7.5.1. Piecewise Constant Models

Data are generated by SETAR(2;0,0) models of the form

$$X_t = \begin{cases} a+e_t & \text{if } X_{t-1} \le 0 \\ -a+e_t & \text{if } X_{t-1} > 0 \end{cases}, \qquad (7.5.1.1)$$

where $e_t \sim N(0,1)$ and $a \in [-3,3]$. The null hypothesis, H_o, states that $X_t = \alpha + e_t$, and the alternative hypothesis is a SETAR model of the general form

$$X_t = \begin{cases} b_1+e_t & \text{if } X_{t-1} \le r \\ b_2+e_2 & \text{if } X_{t-1} > r \end{cases}.$$

As described at the end of §2 the threshold parameter, r, is allowed to range over $r = X_{(i)}$, $i = \nu_1, \cdots, \nu_2$. The 5% asymptotic critical value is 7.84, which is adopted throughout our studying.

The power plots of the LR test and the reverse CUSUM test are given in Figs 7.5.1.1 and 7.5.1.2 respectively. It is quite clear that the reverse CUSUM test has low power for sample sizes 50 and 100. The improvement in power of the reverse CUSUM test seems quite marked when the sample size increases from 100 to 150. Overall, the LR test is more powerful in this case. The rather wavy power curves of the reverse CUSUM test also suggest the test's susceptibility to sampling fluctuations.

We repeat the experiments with the SETAR(2;0,0) model of the form

$$X_t = \begin{cases} 1+a+e_t & \text{if } X_{t-1} \le 0 \\ 1-a+e_t & \text{if } X_{t-1} > 0 \end{cases}, \qquad (7.5.1.2)$$

where $e_t \sim N(0,1)$ and $-1 \le a \le 1$. The power plots are given in Figs 7.5.1.3. and 7.5.1.4.

As can be seen in Fig 7.5.1.3, the performance of likelihood ratio test remains good for $a > 0$ but becomes poor for $a \le 0$ compared with the reverse CUSUM test. To investigate the situation further, we obtain the stationary marginal distribution function of the model ($a = -1$)

$$X_t = \begin{cases} e_t & \text{if } X_{t-1} \le 0 \\ 2+e_t & \text{if } X_{t-1} > 0 \end{cases}, \qquad (7.5.1.3)$$

by numerical integration, and then plot the density function of this model together with the normal density function, $N(\mu, \sigma^2)$, $\mu = E(X_t)$ and $\sigma^2 = Var(X_t)$. The result is given in Fig 7.5.1.5.

The similarity of the p.d.f's suggests that low power is not unexpected for the case $a \le 0$. The question still remain as to why the reverse CUSUM test performs almost as well with $a < 0$ as with $a > 0$. Since

the form of model (7.5.1.3) resembles an 'outlier' situation in time series, we suspect that the performance of the reverse CUSUM test may be connected with its sensitivity to 'outliers'. This suggests that the reverse CUSUM test may not be 'robust'. To investigate this aspect, we simulate 100 replications of the linear model

$$X_t = 2+e_t\,, \tag{7.5.1.4}$$

where $e_t \sim N(0,1)$. In each replication we replace $X(N/2)$ with some large value say k, $k=-20$, -15, -5, 10, 15, 20 (N is the sample size). The type I error probabilities of the tests are given in Table 7.5.1.1.

Table 7.5.1.1

N	k													
	-20		-15		-10		-5		10		15		20	
	RC	*LR*	*RC*	*LR*	*RC*	*LR*	*RC*	*LR*	*RC*	*LR*	*RC*	*LR*	*RC*	*LR*
50	0.75	0.12	0.56	0.08	0.37	0.08	0.13	0.02	0.08	0.06	0.51	0.14	0.25	0.08
100	0.44	0.03	0.34	0.02	0.19	0.02	0.06	0.05	0.05	0.02	0.24	0.02	0.15	0.02
150	0.27	0.01	0.18	0.01	0.11	0.01	0.06	0.04	0.07	0.02	0.14	0.01	0.09	0.01

RC = reverse CUSUM test

LR = likelihood ratio test

It is clear that the CUSUM test is more sensitive to outliers and tends to regard a linear series with an outlier as non-linear.

7.5.2. Piecewise Linear Models

We consider the SETAR(2;1,1)

$$X_t = \begin{cases} aX_{t-1}+e_t & \text{if } X_{t-1}\le 0 \\ -aX_{t-1}+e_t & \text{if } X_{t-1}>0 \end{cases}, \tag{7.5.2.1}$$

where $e_t \sim N(0,1)$ and $a\in[-0.9,0.9]$. Under the null hypothesis, H_o , the model is of the form

$X_t = \alpha X_{t-1} + e_t$; under the alternative, the model is of the form

$$X_t = \begin{cases} b_1 X_{t-1} + e_t & \text{if } X_{t-1} \le r \\ b_2 X_{t-1} + e_t & \text{if } X_{t-1} > r \end{cases}. \qquad (7.5.2.2)$$

Again, the asymptotic critical value is 7.84. The power plots of the tests are given in Figs 7.5.2.1 and 7.5.2.2.

Also, we consider the SETAR models

$$X_t = \begin{cases} (a+0.5)X_{t-1} + e_t & \text{if } X_{t-1} \le 0 \\ (0.5-a)X_{t-1} + e_t & \text{if } X_{t-1} > 0 \end{cases}, \qquad (7.5.2.3)$$

where $e_t \sim N(0,1)$ and $a \in [-0.5, 0.5]$. The power plots of the tests are given in Figs 7.5.2.3 and 7.5.2.4. The power of the reverse CUSUM test seems to drop abruptly as $|a|$ exceeds 0.45. Overall, the likelihood ratio test seems to perform better than the reverse CUSUM in this case.

In this chapter we have compared, via simulation studies, the performance of the likelihood ratio test with the reverse CUSUM test. The results of Tables 7.3.4 , 7.3.5 , 7.3.6 and 7.4.5 show that 6 and 9 degree of freedom as a critical 5% point for the likelihood ratio test with $p = 1$ and $p = 2$,-in presence of an intercept, is quite practical. Also, the results of Table 7.4.5 show that the performance of the likelihood ratio test in comparison to the CUSUM and Tsay's tests is quite remarkable.

The main strengths of the new implementation of the likelihood ratio test are (i) it does not depend on the optimal model selection (ii) the parameters threshold and delay need not to be estimated and (iii) it is fast in view of CPU time.

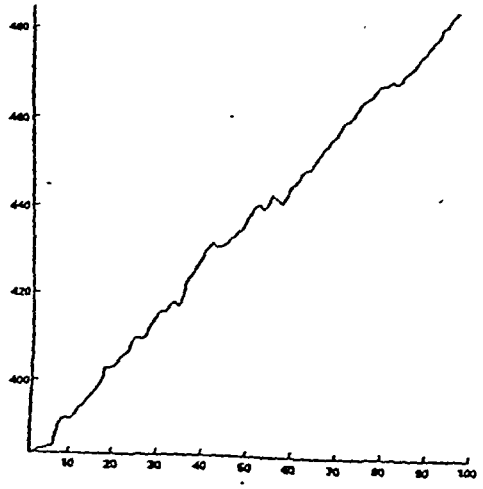


Fig. 7.4.1 : A realization of model (7.4.1).

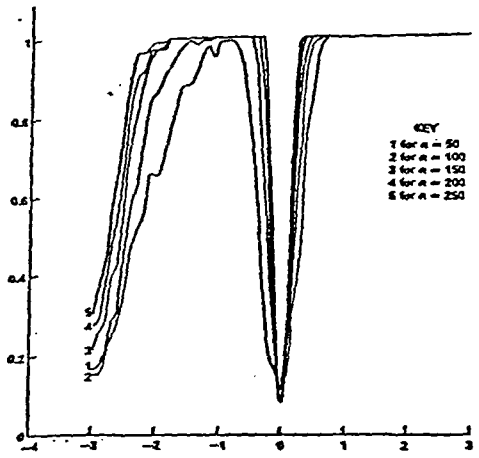


Fig. 7.5.1.1

: LR test power plot of model (7.5.1.1).

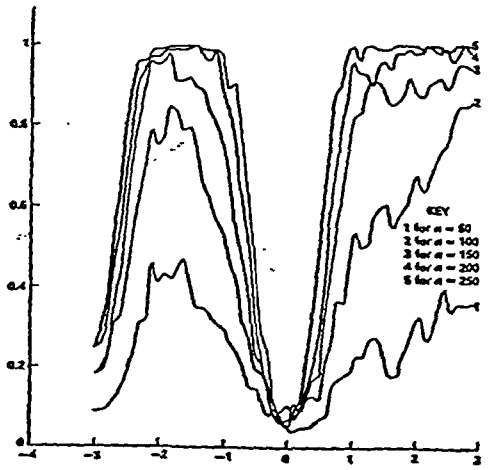


Fig. 7.5.1.2 :

: CUSUM test power plot of model (7.5.1.1).

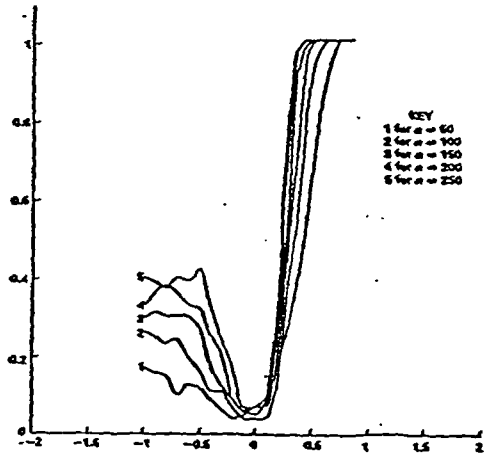


Fig. 7.5.1.3

LR test power plot of model (7.5.1.2).

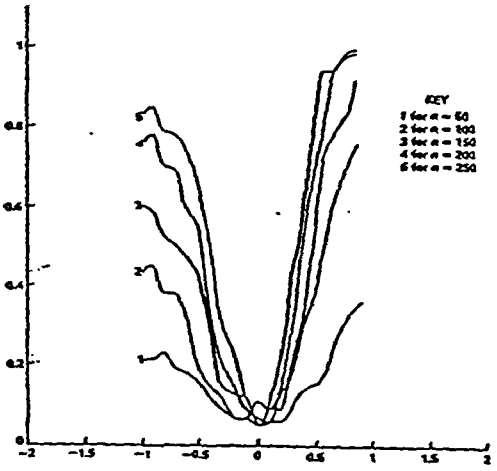


Fig. 7.5.1.4

CUSUM test power plot of model (7.5.1.2).

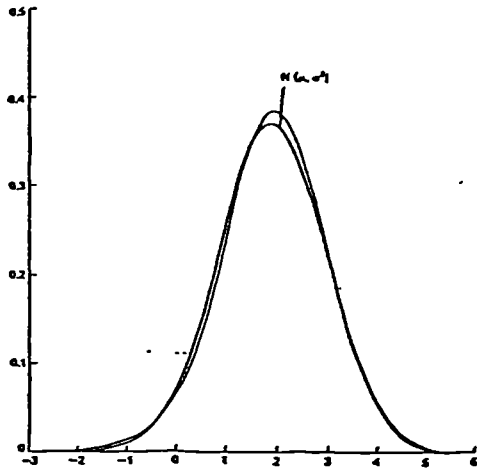


Fig. 7.5.1.5 : p.d.f of model (7.5.1.3).

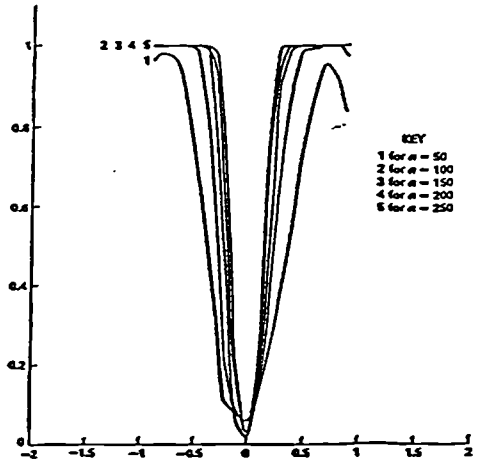


Fig. 7.5.2.1

: LR test power plot of model (7.5.2.1).

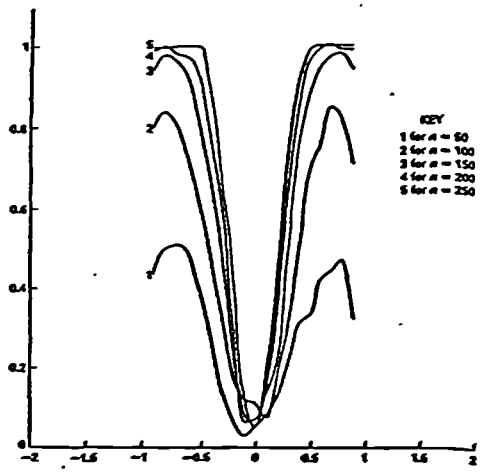


Fig. 7.5.2.2

CUSUM test power plot of model (7.5.2.1).

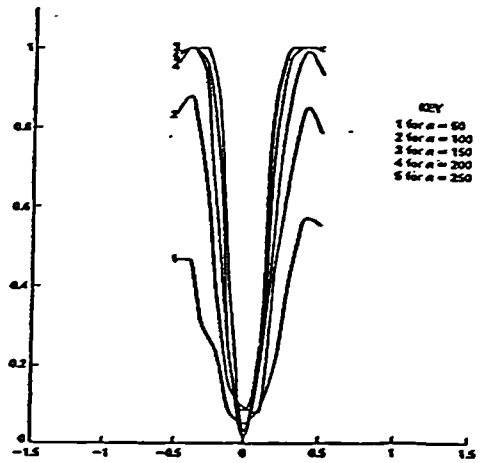


Fig. 7.5.2.3 : LR test power plot of model (7.5.2.3).

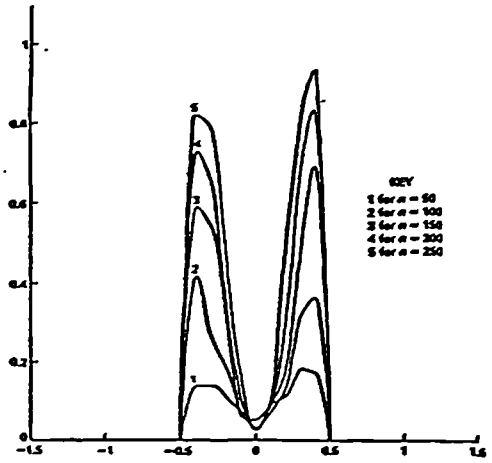


Fig. 7.5.2.4

Fig. 7.5.2.4 : CUSUM test power plot of model (7.5.2.3).

CHAPTER EIGHT

SETAR MODELLING OF BLOWFLY DATA

8.1. Some Background of A.J. Nicholson Data

In the 1950's, the Australian entomologist, A.J. Nicholson, conducted a series of experiments with blowflies, *Lucilia Cuprina*. His laboratory data have since become classic and simulated wide interest in population ecology. Of particular note are the cycles apparent in his population data.

The data we are going to analyse are abstracted from Brillinger et al (1980) and correspond to the bi-daily record of one of A.J. Nicholson's experiments extending over two years, in which a caged population of approximately 1000 blowflies was initiated with a reasonably balanced sex ratio. The caged blowflies were fed a limited amount (about 500 mg.) of ground liver daily as the only source of protein which is necessary for egg production. Experimental evidence suggests that egg production usually ceases when daily protein intake for the female fly drops below 0.14 mg. and levels out at 10 eggs per fly per day when protein supply is plentiful.

The first question is whether this data set is non-linear. The results of Table (7.3.4) in Chapter Seven shows the non-linearity of the first part (the first 182 observations) of the blowfly data. Also the results of three other different tests (i.e Subba Rao-Gabr-Hinich's, Keenan's and CUSUM) support the non-linearity of this data set. (See Tong 1987a and W.S. Chan & Tong 1986).

Since the marginal density function of all linear Gaussian processes are unimodal therefore, before following the Tong's SETAR fitting model procedure, it will be useful to consider a test for testing multi-modality of the marginal probability density function of a data set.

8.2. A Test for Multi-modality Based on Kernel Density Estimation

A test statistic for hypotheses concerning the number of modes in the density is obtained by con-

structing kernel density estimates of the data (Silverman, 1981). The kernel density estimate (Rosenblatt, 1956) for window width h based on univariate observations $X_1, X_2, ..., X_n$ is defined by

$$\hat{f}(t;h) = n^{-1}h^{-1}\sum_{i=1}^{n} K\{h^{-1}(t-X_i)\} ,$$

where K is a kernel function †, which we shall assume through out to be normal density function.

The window width h controls the amount by which the data are smoothed to obtain the kernel estimate. Suppose that we wish to test the null hypothesis that the density f underlying the data has k modes, against the alternative that f has more than k modes. Define the *h–critical* window width h_{crit} by

$$h_{crit} = inf\{h : \hat{f}(.,h) \text{ has at most } k \text{ modes}\} .$$

Large values of h_{crit} will reject the null hypothesis.

For any particular k-mode simple null-hypothesis it is easy to assess, by simulation, the significance of the value of the critical window width obtained from the data. Suppose the null hypothesis is that the true density is g and that the value of h_{crit} obtained from the data is h_o. Silverman (1981) has shown that

$$Pr_g(h_{crit} > h_o) = Pr\{\hat{f}(.;h_o) \text{ has more than } k \text{ modes} \mid \{X_1, \cdots, X_n\} \text{ is drawn from } g\} .$$

The hypothesis that the true density is at most k-modal is of course a compound hypothesis. To provide a conservative assessment of the significance of h_o, an appealing choice of the representative g_o from which to simulate is obtained by rescaling $\hat{f}(.;h_o)$, as constructed from the data, to have variance equal to the sample variance. It is easy to simulate from g_o. Efron (1979) pointed out that independent observations Y_i, from g_o, are given by

$$Y_i = (1+\frac{h_o^2}{\sigma^2})^{-\frac{1}{2}} (X_{I(i)}+h_o\varepsilon_i) ,$$

where $X_{I(i)}$ are sampled uniformly, with replacement, from the data $X_1, X_2, \cdots, X_n$, σ^2 is the sample variance of the data and ε_i is an independent sequence of standard normal random variables.

We applied this test to different sets of real and simulated data, which are typically serially dependent. It seems that the test works quite well .

---

† The kernel K is a symmetric function satisfying $\int K(t)\,dt = 1$, $\int t\,K(t)\,dt = 0$, $\int t^2 K(t)\,dt = k_2 \neq 0$.

Fig. 8.2.1 shows the histogram and critical window widths and significance levels for tests of the null hypothesis that the underlying density has at most k modes against the alternative that it has more than k modes of the monthly U.S. air passengers miles data . The data are taken from Cryer (1986, p. 270) .

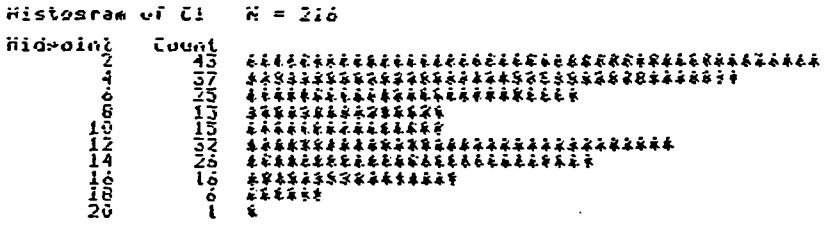


Fig. 8.2.1

Number of Modes	h_{crit}	p-value
1	3.239	0.01
2	0.634	0.92
3	0.544	0.84

The second example is $\log_{10}$ transfomation of lynx data. The data are taken from Tong (1983). The results are shown in Fig. 8.2.2 .

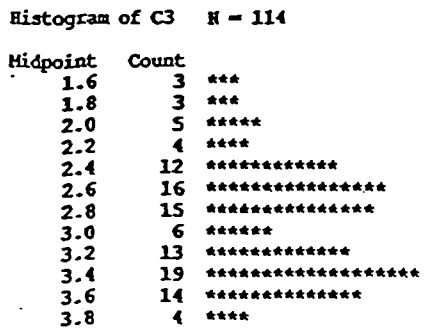


Fig. 8.2.2

Number of Modes	h_{crit}	p-value
1	0.297	0.03
2	0.145	0.45
3	0.087	0.72

In the third example we consider the Milk Production in Pound per Cow per Month of U.S. data . The data are taken from Cryer (1986, p. 269). The results are given in Fig. (8.2.3) .

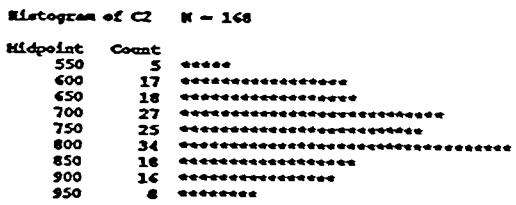


Fig. 8.2.3

Number of Modes	h_{crit}	p-value
1	25.572	0.75
2	21.264	0.53
3	19.253	0.65

In order to investigate the efficiency of the test by simulated data, we conduct two experiments.

Experiment I

$$X_t = \begin{cases} 1.5-0.9X_{t-1}+e_t & \text{if } X_{t-1} \le 0 \\ -0.4-0.6X_{t-1}+e_t & \text{if } X_{t-1} > 0 \end{cases},$$

and experiment II

$$X_t = \begin{cases} 1.5-0.7X_{t-1}+e_t & \text{if } X_{t-1} \le 0 \\ 1.0+0.8X_{t-1}+e_t & \text{if } X_t > 0 \end{cases}.$$

In each experiment, $e_t \sim N(0,1)$ and the sample size is equal to 100. The number of replication is 20.

The stationary marginal density function and the results of applying the test on the model I (biomodal) and model II (unimodal) are given by Figs. 8.2.4 and 8.2.5 respectively.

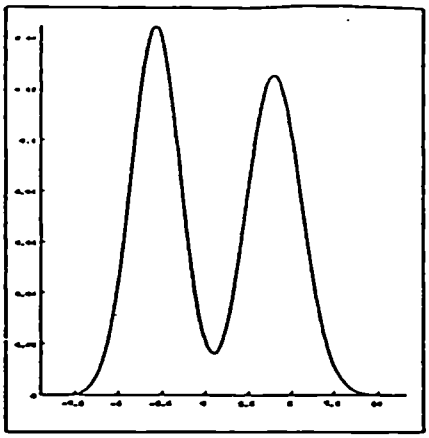


Fig. 8.2.4

Number of Modes	h_{crit}	p-value
1	2.90	0.00
2	0.59	0.58

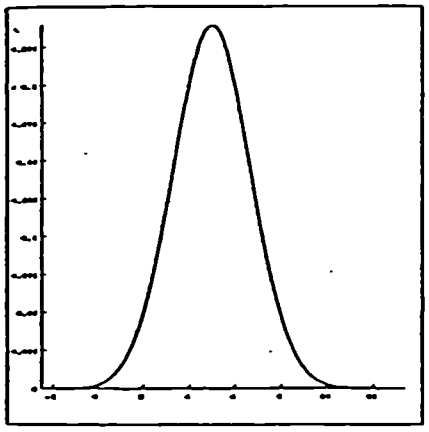


Fig. 8.2.5

Number of Modes	h_{crit}	p-value
1	0.51	0.57
2	0.38	0.49

8.3. Influential Data

Quite often, a simple residual check is not sufficient in suggesting improvement/inadequacy of the fitted model. It is always useful to look for the influential data.

A new approach, which is not based on residuals, has been explored by Hau and Tong (1984). Their emphasis were on the influential data.

Consider the linear Gaussian autoregressive model of order p

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t \,, \tag{8.2.1}$$

where e_t's are independent and identically distributed random variables and $e_t \sim N(0,\sigma^2)$.

Let

$$\mathbf{Z}_t^T = (Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}) \,,$$
$$\phi^T = (\phi_1, \phi_2, \cdots, \phi_p) \,,$$

then (8.2.1) can be rewritten as

$$Y_t = \mathbf{Z}_t^T \phi + e_t \,.$$

Suppose that n observations $Y_1, \cdots, Y_n$ are available, then we have the following normal equation

$$\underset{(n\times 1)}{\mathbf{Y}} = \underset{(n\times p)(p\times 1)}{\Gamma\ \phi} + \underset{(n\times 1)}{\mathbf{e}} \,,$$

where $\mathbf{Y} = (Y_1, \cdots, Y_n)^T$, $\mathbf{e} = (e_1, \cdots, e_n)^T$ and Γ is design matrix. Here

$$H = [h_{ij}] = \Gamma(\Gamma^T\Gamma)^{-1}\Gamma^T$$

is known as the *hat matrix*. It has been shown that the diagonal elements of the hat matrix i.e. $h_t = \mathbf{Z}_t^T(\Gamma^T\Gamma)^{-1}\mathbf{Z}_t$ contain some vital information about the influential data. (See Hau and Tong, 1984).

Consider the SETAR(2;k_1,k_2) model of the form

$$X_t = \begin{cases} a_o^{(1)} + \sum_{j=1}^{k_1} a_j^{(1)} X_{t-j} + e_t^{(1)} & \text{if } X_{t-d} \le r \\ a_o^{(2)} + \sum_{j=1}^{k_2} a_j^{(2)} X_{t-j} + e_t^{(2)} & \text{if } X_{t-d} > r \,, \end{cases} \tag{8.2.2}$$

where d and r are delay and threshold parameters respectively. For each $i = 1,2$, $e_t^{(i)}$'s are assumed to be i.i.d. and normally distributed with mean zero and variances σ_i^2, $i = 1,2$. Using the extended conventional least square procedure, since the model (8.2.2) consists of 2 piecewise linear models, we may have the fol-

lowing piecewise linear model formalism:

$$\begin{cases} \underset{(n_1\times 1)}{\mathbf{X}_1} = \underset{(n_1\times k_1)(k_1\times 1)}{\mathbf{A}_1\ \theta_1} + \underset{(n_1\times 1)}{\mathbf{e}_1} \\ \underset{(n_2\times 2)}{\mathbf{X}_2} = \underset{(n_2\times k_2)(k_2\times 1)}{\mathbf{A}_2\ \theta_2} + \underset{(n_2\times 1)}{\mathbf{e}_2} \end{cases},$$

where, for $i=1,2$,

$\mathbf{X}_i=(X_{j1}^{(1)},\cdots,X_{jn_i}^{(i)})^T$, $\mathbf{e}_i=(e_{j1}^{(i)},\cdots,e_{jn_i}^{(i)})^T$, $\theta_i=(a_o^{(i)},\cdots,a_{k_i}^{(i)})$ and

$$\mathbf{A}_i=\begin{bmatrix} \mathbf{Z}_1^{(i)T} \\ \mathbf{Z}_2^{(i)T} \\ \vdots \\ \mathbf{Z}_{n_i}^{(i)T} \end{bmatrix},$$

where $\mathbf{Z}_l^{(i)T}=(1\ X_{jl-1}^{(i)}\ X_{jl-2}^{(i)}\ \cdots\ X_{jl-k_i}^{(i)})$. Because of the piecewise linearity, we may define hat matrices $TARH_i$, $i=1,2$ as follows:

$$TARH_i=\mathbf{A}_i(\mathbf{A}_i^T\mathbf{A}_i)^{-1}\mathbf{A}_i^T,\ i=1,2.$$

For influential data detection in model (8.2.2), the diagonal elements $h_t^{(i)}$ of $TARH_i$ are examined. Large values of $h_t^{(i)}$'s correspond to the influential row vector $\mathbf{Z}_t^{(i)T}$'s of $\mathbf{A}_i$. In the linear AR case, a large value, h_{t_o} say, is due to the fact that some of the $X_{t_o-1},\cdots$, and $X_{t_o-k_i}$ are large. In the case of SETAR modelling, of course, $X_{t_o-1}^{(i)}$ is not necessarily, X_{t_o-1}. Therefore, unlike the linear AR case, even if $h_{t_o-1}^{(i)}$ is small and $h_{t_o}^{(i)}$ is large, we can not conclude that X_{t_o-1} is an influential data. To identify the influential data, it is suggested that we print out the design matrix $\mathbf{A}_i$ with the vector $\mathbf{h}_i$ augmented in the last column:

$$\underset{n_i\times(k_i+1)}{I_i}=(\mathbf{A}_i\,|\,\mathbf{h}_i),$$

where $\mathbf{h}_i^T=(h_1^{(i)},h_2^{(i)},\cdots,h_{n_i}^{(i)})$ and $h_j^{(i)}=\mathbf{Z}_j^{(i)T}(\mathbf{A}_i^T\mathbf{A}_i)^{-1}\mathbf{Z}_j^{(i)}$.

Examining the diagonal elements $h_t^{(i)}$ of $TARH_i$ and the original observations, can therefore identify those data which are the most influential.

8.4. Profile Likelihood

Inference in the presence of nuisance parameters is a widely encountered and difficult problem. Probably the simplest approach is to maximize out the nuisance parameters for fixed values of the parameters of

interest, and construct the so called *profile likelihood.* The profile likelihood is then treated as an ordinary likelihood function for estimation and inference about the parameters of interest.

We begin with a random N-sample $Y_1, Y_2, \cdots, Y_N$. The parameters vector θ can be partitioned as $\theta=(\psi, \lambda)$ where $\psi=(\psi_1, \cdots, \psi_r)$ is the parameter of interest and $\lambda=(\lambda_1, \cdots, \lambda_s)$ is the nuisance parameter. The log-likelihood will be denoted by $l(\theta)$.

Denote by $\hat{\theta}=(\hat{\psi}, \hat{\lambda})$ the overall maximum likelihood estimate. Let $\hat{\psi}_\lambda$ denote the maximum likelihood estimate of ψ for fixed λ and $\hat{\lambda}_\psi$ may be define in a similar way. The profile log-likelihood for ψ is defined by

$$l_p(\psi) = l(\psi, \hat{\lambda}_\psi).$$

To obtain the profile likelihood plot for a SETAR(2;p_1, p_2) model the vector of parameters

$$\theta = (r, d, a_o^{(1)}, \cdots, a_{p_1}^{(1)}, a_o^{(2)}, \cdots, a_{p_2}^{(2)})$$

can be partitioned as $\theta=(\psi, \lambda)$, where $\psi=(r, d)$ and λ the rest of parameters. For each $d \in \{1,2,...,m : m \ll N\}$, we can calculate the likelihood function or equivalently the residuals sum of squares, for each r, $r=Y_{(i)} \in [Q_1, Q_3]$, ($Y_{(i)}$'s being the ascending sorted data and Q_1, Q_3 the first and third quartial of data). For fixed d the family plots of profile likelihood vs r, may reveal some vital information about the location of threshold and the estimation of parameter delay.

In order to examine the application of profile likelihood in locating the threshold and estimating the delay, we conduct to the following simulation study. The model is

$$X_t = \begin{cases} 1+0.5X_{t-1}-0.6X_{t-2}+0.6X_{t-4}+0.3X_{t-5}+e_t & \text{if } X_{t-5} \le 0 \\ -2+0.5X_{t-1}-0.8X_{t-4}+0.5X_{t-5}+e_t & \text{if } X_{t-5} > 0, \end{cases} \tag{8.3.1}$$

where $e_t \sim N(0, 1)$ and the sample size is equal to 200. In this experiment we assume that the order of each piece is known and is equal to 5 . The profile likelihoods are plotted for $d=1,2,...,7$. Figs. (8.3.1) show the profile likelihood of simulated data with a good estimation of delay ($\hat{d}=5$) and a good interval estimation for threshold ($\hat{r} \in (-0.1, 0.1)$).

Our simulation studies suggest that the estimation of p_1 and p_2 is not crucial. Thus in practice, for each d and r, they can be considered as nuisance parameters and could be estimated by using AIC, BIC or any other criterion. In all our experiments with real data, the AIC criterion has been used for estimating the

order of each piece.

8.5. A Case Study

Now we will try to model the original (raw) blowfly data. Fig. (8.4.1) shows the time plot of the data, from which it is clear that the data exhibit population cycles and also shows a change in the dynamic system of this process (non-stationarity of this data set). Therefore, the data set is divided into two stationary sets, the first consisting of the first 200 observations and the other the rest of data.

Definition 8.4.1: A stationary time series $\{X_t\}$ is *time-reversible* if for every positive integer n, and every $t_1, t_2, ..., t_n \in \mathbf{Z}$, where $\mathbf{Z}$ denotes the set of integers, the vectors $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ and $(X_{-t_1}, X_{-t_2}, \cdots, X_{-t_n})$ have the same joint distribution.

A stationary time series which is not time-reversible is said to be *time-irreversible.*

Weiss (1975) has proved that ARMA processes with an autoregressive component are reversible if and only if they are Gaussian. The implications are far reaching, Stationary series which show evidence of irreversibility, cannot be modelled by Gaussian ARMA models, they need to be modelled by non-Gaussian ARMA models or some type of non-linear models. (See Lawrance, 1987)

In practice the reverse data plot may be used to detect time-irreversibility of a data set.

Comparing Fig. (8.4.2) (time plot of the first 200 observations) with Fig. (8,4,3) (reverse time plot of the data) reveals the time-irreversibility of the first part of blowfly data.

From the time plot of the data (Fig. 8.4.2), we may see that the mean length descent periods (5.7) tends to exceed that of ascent (3.8)periods. Table (8.4.1) shows the descent and ascent periods of blowfly data

Table 8.4.1

Ascent and Descent Period of blowfly data

Ascent period	1	1	4	1	2	2	6	2	12	6	
Descent period	9	1	13	1	12	3	9	2	9	9	
Ascent period	8	1	6	1	8	1	1	9	1	6	1
Descent period	9	1	8	2	8	1	1	9	1	11	1

Non-parametric sample estimates of $E(X_t \mid X_{t\pm j}=x)$, for $j=1,2,...,11$, reveal some vital information. Whilst $E(X_t \mid X_{t\pm j})$, $j=1,2$ are almost linear in x, for $j=3,...,11$ all are non-linear. There is a systematic change in $E(X_t \mid X_{t+j})$, $j=3,...,11$, which is different from that of $E(X_t \mid X_{t-j})$. The dissimilarity between $E(X_t \mid X_{t-j})$ and $E(X_t \mid X_{t+j})$ is further evidence of time-irreversibility of this data set. Also, it can be seen that the non-parametric regression for $j=8,...,11$ are all approximately the same. Thus it may be suggested that $\hat{d}\approx 8$ could be a crude estimate for the delay parameter. The results are given by Figs. 8.4.4a and 8.4.4b

A family of profile likelihoods of the first 200 observations of blowfly data is given by Fig. (8.4.5), which shows 8 could be a good estimate for parameter delay. The profile likelihoods for $d=13,...,18$, exhibit quite substantial fluctuation for different values of r, therefore $d \geq 13$ does not seem to be an adequate estimate for d.

The scatter diagrams of (X_t, X_{t+j}), $j=1,2,...,8$, consist of spirals which with j increasing, show a systematic change with an obvious hole in the center near the origin (Fig. 8.4.6), which tends to support the existence of a limit cycle in the data.

The marginal histogram (Fig. 8.4.7) of the data does not show any strong evidence of bimodality.

The test for bimodality rejects the unimodality of the data at the 34% point, and rejects bimodality at the 70% point. However, there is no strong evidence for bimodality of this data set. The marginal histogram shows clearly the skewness of data (skewness=0.77). The kurtosis is equal to -0.39 which together with skewness show the non-Gaussianity of the data .

The sample bivariate distribution of (X_t, X_{t+j}) , $j=1,...,10$ (Figs. 8.4.8), show particularly interesting non-Gaussianity of the data structure. Notice especially quite a regular change from unimodality ($j=1,...,6$) to bimodality ($j>6$) of sample bivariate distribution. Also, the "L" shape of sample bivariate distributions for $j>6$, suggests that a SETAR model with one threshold, could be a good preliminary model for this data set.

The power spectrum estimate of the data tends to suggest quite strongly the existence of higher harmonics to the fundamental period of approximately 36 to 40 days. (Fig. 8.7.2).

To fit a SETAR model, first we will describe a systematic procedure for obtaining estimates of d , r ; p_1 , $a_o^{(1)}$, $\cdots$, $a_{p_1}^{(1)}$, $\sigma_e^{(1)^2}$; $a_o^{(2)}$, $\cdots$, $a_{p_2}^{(2)}$, $\sigma_e^{(2)^2}$, where p_1 , p_2 , $\sigma_e^{(1)^2}$, $\sigma_e^{(2)^2}$ are the order and the variance of error term in the first and second piece respectively. (See Tong, 1983 p. 134).

STEP 1: Let d and r be fixed at d_o and r_o respectively. Let L denote the maximum order to be entertained for each of the two piecewise linear AR models. The choice of L is subjective and is usually dictated by the sample size.

STEP 2: Let d remain at d_o. Allow r to vary over $\{r_1, r_2, \cdots, r_s\}$ and minimize the AIC(d_o , r) over this set.

STEP 3: We search over $\{1,2, \cdots, \tau\}$ for the minimum AIC of d. Since the different choice of d changes the effective number of observations in estimating parameters and AIC, therefore we may use the normalized AIC by dividing AIC by $N-max\{d, L\}$. The minimum of normalized AIC is calculated by

$$\min_{d\in\{1,\ldots,\tau\}} \min_{r\in\{r_1,\ldots,r_s\}} \{NAIC(d,r)\} .$$

Now by using the above identification procedure, we may give a full analysis of the first 200 observations of the blowfly data. Table (8.4.2) details the results of the identification procedure to the first 200 observations with $\tau=10$ and $s=7$. The seven candidates for the threshold are obtained by using the

following relation, where S = standard deviation and $\bar{X}$ the data mean.

$$\bar{X}-(1-(2i-1)/7)S\ ,i=1,...,7$$

The entries under N_1 and N_2, the sample size for the two piecewise linear AR models, are quite important. Unduly small N_1 or N_2 generally leads to a poor model. The maximum possible autoregressive order,L, is set at 20. In the table, NAIC means normalized AIC, and σ_e^2 denotes the pooled mean square of the fitted residuals.

Table 8.4.2

d	r	N_1	N_2	p_1	p_2	NAIC	σ_e^2	limit cycle status	period
	795.6	44	136	2	7	13.4348	598684	limit point	
	1398.1	72	108	2	7	13.3848	594698	limit point	
	2000.6	90	90	2	7	13.2952	583229	limit point	
1	2603.1	104	76	3	7	13.2486	577076	limit point	
	3205.6	113	67	8	2	13.2086	616249	limit point	
	3808.1	130	50	7	2	13.2219	593581	limit point	
	4410.6	142	38	7	2	13.2511	544701	limit point	
	795.6	44	136	3	8	13.4450	591341	limit point	
	1398.1	71	109	3	7	13.3969	597570	limit point	
	2000.6	89	91	3	7	13.3302	591065	limit point	
2	2603.1	104	76	3	7	13.2992	586339	limit point	
	3205.6	113	67	8	7	13.3383	556922	3↑3↓	6
	3808.1	130	50	7	2	13.4141	613998	8↑7↓	15
	4410.6	142	38	7	1	13.3835	618855	limit point	
	795.6	44	136	4	8	13.4337	578277	limit point	
	1398.1	71	109	4	7	13.4274	591753	limit point	
	2000.6	88	92	4	16	13.3861	519495	non-stationary	
3	2603.1	103	77	9	7	13.4238	569966	limit point	
	3205.6	112	68	9	7	13.4631	587896	limit cycle with long period	
	3808.1	129	51	8	1	13.4158	601308	9↑8↓	17
	4410.6	142	38	8	1	13.4061	606612	limit point	
	795.6	44	136	5	16	13.4298	524050	limit point	
	1398.1	71	109	5	7	13.4363	587386	limit point	
	2000.6	107	93	5	16	13.4007	512413	non-stationary	
4	2603.1	102	78	5	2	13.4320	609885	9↑7↓	16
	3205.6	111	69	8	2	13.4351	592789	9↑8↓	17
	3808.1	128	52	8	1	13.4118	589666	limit cycle with long period	
	4410.6	141	39	8	2	13.4252	589267	limit point	

Table 8.4.2 (continue)

d	r	N_1	N_2	p_1	p_2	NAIC	σ_*^2	limit cycle status	period
	795.6	44	136	2	16	13.4046	537910	limit point	
	1398.1	71	109	6	7	13.4273	570446	limit point	
	2000.6	86	94	7	7	13.4102	561190	7 ↑ 10 ↓	17
5	2603.1	101	79	7	2	13.4202	600267	9 ↑ 8 ↓	17
	3205.6	110	70	7	2	13.4530	613308	10 ↑ 8 ↓	18
	3808.1	127	53	7	7	13.4778	599021	limit point	
	4410.6	140	40	7	1	13.4689	626151	limit point	
	795.6	44	136	2	7	13.3338	580233	limit point	
	1398.1	71	109	7	7	13.3309	548205	limit point	
	2000.6	85	95	7	2	13.3504	585035	9 ↑ 10 ↓	19
6	2603.1	100	80	7	1	13.3353	587518	limit cycle with long period	
	3205.6	109	71	7	1	13.3773	595389	limit cycle with long period	
	3808.1	126	54	7	2	13.1949	572825	9 ↑ 10 ↓	19
	4410.6	139	41	7	2	13.1843	590531	limit point	
	795.6	44	136	2	8	13.3120	559181	limit point	
	1398.1	71	109	2	7	13.1923	597020	8 ↑ 11 ↓	17
	2000.6	84	96	2	3	13.2563	623615	8 ↑ 11 ↓	17
	2603.1	99	81	2	3	12.7253	620900	9 ↑ 9 ↓	18
7	3205.6	108	72	7	2	12.8324	568193	limit cycle with long period	
	3808.1	125	55	7	2	12.8750	577809	limit cycle with long period	
	4410.6	138	42	7	2	12.9742	578411	limit cycle with long period	
	795.6	44	136	2	8	13.3359	571833	limit point	
	1398.1	61	109	2	7	12.9510	559186	11 ↑ 10 ↓ 11 ↑ 9 ↓	41
	2000.6	84	96	2	9	12.5976	575764	11 ↑ 9 ↓	20
8	2603.1	98	82	2	9	12.6053	615862	11 ↑ 9 ↓	20
	3205.6	107	73	10	2	12.7441	564266	8 ↑ 11 ↓ 9 ↑ 11 ↓	39
	3808.1	124	56	8	2	12.7618	568190	limit cycle with long period	
	4410.6	137	43	8	2	12.8941	571346	limit cycle with long period	

Table 8.4.2 (continue)

d	r	N_1	N_2	p_1	p_2	NAIC	σ_e^2	limit cycle status	period
	795.6	44	136	2	8	13.3238	588150	limit point	
	1398.1	71	109	2	6	12.5347	535619	11 ↑ 9 ↓	20
	2000.6	84	96	2	6	12.4712	588527	10 ↑ 11 ↓	21
9	2603.1	98	82	7	3	12.4207	552477	8 ↑ 13 ↓	21
	3205.6	107	73	10	3	12.5388	532957	limit cycle with long period	
	3808.1	123	57	10	2	12.7341	549085	limit point	
	4410.6	136	44	8	4	12.9009	581221	limit point	
	795.6	44	136	2	8	13.3718	623689	limit point	
	1398.1	71	109	7	8	13.2448	566405	limit point	
	2000.6	85	95	7	7	12.4382	513115	limit point	
10	2605.6	99	81	7	7	12.4656	532333	limit point	
	3205.6	108	72	7	7	12.58146	563039	limit point	
	3808.1	124	56	8	4	12.7368	565393	limit point	
	4410.6	136	44	7	2	12.8897	591513	limit point	

Models with no obvious limit cycles over the dynamic range of the data or with estimated order greater than 10 are discarded. After this discarding of unwanted models and by considering the NAIC, it seems that the best estimate for the delay parameter could be 7 or 8 or 9. Among the selected models, after searching over a fine grid tuning near the selected threshold the following model i.e. SETAR(2;2,9) is selected.

$$X_t = \begin{cases} \underset{(248.93)}{1369.46} + \underset{(0.12)}{0.90}X_{t-1} - \underset{(0.11)}{0.19}X_{t-2} + e_t^{(1)} & \text{if } X_{t-8} \le 1790 \\ \underset{(102.72)}{246.75} + \underset{(0.05)}{1.11}X_{t-1} - \underset{(0.06)}{0.23}X_{t-2} - \underset{(0.05)}{0.11}X_{t-3} - \underset{(0.05)}{0.03}X_{t-4} & \\ \quad + \underset{(0.04)}{0.04}X_{t-5} - \underset{(0.04)}{0.02}X_{t-6} - \underset{(0.04)}{0.03}X_{t-7} - \underset{(0.03)}{0.06}X_{t-8} & \\ \quad + \underset{(0.02)}{0.06}X_{t-9} + e_t^{(2)} & \text{if } X_{t-8} > 1790\,, \end{cases} \tag{8.4.1}$$

where $Var(e_t^{(1)}) = 1167684.0$, $Var(e_t^{(2)}) = 68844.8$, *pooled* $Var = 563322.4$, NAIC=12.5691.

On comparing the coefficients of the second piece with their respective standard deviations we may adopt the more parsimonious SETAR(2;2,2) model for this data set.

$$X_t = \begin{cases} \underset{(248.92)}{1369.46} + \underset{(0.12)}{0.90} X_{t-1} - \underset{(0.11)}{0.19} X_{t-2} + e_t^{(1)} & \text{if } X_{t-8} \le 1790 \\ \underset{(43.46)}{163.32} + \underset{(0.05)}{1.14} X_{t-1} - \underset{(0.04)}{0.35} X_{t-2} + e_t^{(2)} & \text{if } X_{t-8} > 1790 , \end{cases} \tag{8.4.2}$$

where $Var(e_t^{(1)}) = 1167684.0$, $Var(e_t^{(2)}) = 80851.8$ and *pooled Var* $= 567224.2$, the normalized AIC is equal to 12.5727 which is comparable with that of model (8.4.1).

8.6. Transformation

The scatter diagrams of (X_t , X_{t+j}) , $j = 1,...,8$ (Fig. 8.4.6), show frequent visits to the neighbourhood of the origin, which is a point of condensation. It has been suggested that the logarithmic transformation could remove this point of condensation and perhaps the fitted residuals of the selected model for $\log_{10}$ transformation of this data set will not be non-Gaussian. (See Tong, 1983 p. 210). Therefore we will use the $\log_{10}$ transformation for the first 200 observations, and after fitting a SETAR model, a full diagnostic check will be applied to the selected model.

Applying the same procedure which has been used for selecting model (8.4.1) to the $\log_{10}$ transformation of the blowfly data, the following SETAR(2;1,3) model is adopted.

$$X_t = \begin{cases} \underset{(0.23)}{2.66} + \underset{(0.07)}{0.27} X_{t-1} + e_t^{(1)} & \text{if } X_{t-8} \le 3.00 \\ \underset{(0.09)}{0.44} + \underset{(0.11)}{1.52} X_{t-1} - \underset{(0.20)}{0.33} X_{t-2} - \underset{(0.12)}{0.32} X_{t-3} + e_t^{(2)} & \text{if } X_{t-8} > 3.00 , \end{cases} \tag{8.4.3}$$

where $Var(e_t^{(1)}) = 0.0246$, $Var(e_t^{(2)}) = 0.0101$, *pooled Var* $= 0.0154$. The first 20 observations are discarded and the last 50 observations are set aside for one-step ahead prediction. The maximum lag for the first piece is 10 and for the second piece is 3.

8.7. Model Diagnostic Checking

Suppose that a TAR model has been identified. The question left is to examine the adequacy of the model. In other words, the *goodness of fit* of the fitted model need to be tested for.

According to the proposed requirements of non-linear time series models, we prefer a model which can capture the probabilistic structure of the data.

Some preliminary diagnostic checking on models (8.4.1) and (8.4.2) shows that these two models can not pass all the diagnostic checks successfully. Therefore, our discussion will centre around the SETAR(2;1,3) model fitted to the log transformation of blowfly data i.e. model (8.4.3).

Let $\{e_t^{(j)}\}$ denote the fitted residuals in the jth regime. It is assumed that $\{e_t^{(j)}\}$ is a zero mean Gaussian white noise sequence with $Var(e_t^{(j)})=\sigma_j^2$ and $cov(e_t^{(i)}, e_t^{(j)})=0$, $i \neq j$. The diagnostic checks of the fitted residuals need to be applied to the separate series of fitted residuals. It has been suggested that a normalization of fitted residuals by dividing $e_t^{(j)}$ by $S^{(j)}$, the root mean squares of the fitted residuals of the jth region, is satisfactory. (See Tong, 1983).

8.7.1. Test of Whiteness

The whiteness of the fitted (normalized) residuals $\{e_1,...,e_N\}$ can be tested by using the estimated autocorrelation function $\hat{\rho}_k$, where

$$\hat{\rho}_k = \frac{\hat{C}_k}{\hat{C}_o}, \quad \hat{C}_k = \frac{1}{N}\sum_{t=1}^{N-k}(e_t-\bar{e})(e_{t+k}-\bar{e}), \quad k=0,1,...,L,$$

$$\bar{e} = \frac{1}{N}\sum_{t=1}^{N} e_t .$$

Under the null hypothesis of whiteness, $\hat{\rho}_k$, $k=1,...,L$, are asymptotically independent and normally distributed with zero mean and variance N^{-1}. A simple test may be constructed by plotting the $\hat{\rho}_k$'s against k. We may expect that approximately 95% of $\hat{\rho}_k$'s should lie within the band $\pm 1.96 / \sqrt{N}$ if the e_t's are to be accepted as approximately white noise. (See, e.g. Tong, 1983 p. 157).

Fig. (8.7.1.1) shows the sample autocorrelation function of the normalized fitted residuals obtained from the SETAR(2;1,3), equation (8.4.3). It does not show any significant point outside the 95% confidence interval.

More formal tests based on $\hat{\rho}_k$'s have also been developed for the class of linear models. Ljung and Box (1978) have demonstrated that the statistic

$$Q = N(N+2)\sum_{i=1}^{M}\frac{\hat{\rho}_i}{(N-i)}, \quad M \ll N,$$

provides a closer small-sample approximation to $\chi^2_{(M-p-q)}$, where p & q are the orders of the AR and MA parts of a linear model.

The Ljung-Box statistic with $M=20$, is 24.61 which is not significant compared with the $\chi^2_{20}(5\%)$.

8.7.2. Test of Normality

Suppose that the e_t's are accepted as approximately white noise, we may then proceed with the Gaussian investigation. A histogram usually contains a fair amount of information, and provides a good graphical indicator. Fig. (8.7.2.1) shows the histogram of the normalized fitted residuals of model (8.4.3) which does not show a substantial departure from normality. Also, the normal probability plot of the normalized fitted residuals does not reveal any serious deviation from normality. (See Fig. 8.7.2.2). Sample coefficients of skewness and kurtosis are 0.52 and 1.93 respectively.

Using the fact that the population is normal if and only if the sample mean and sample variance are independently distributed, Lin and Mudholkar (1980) proposed the following test for normality against asymmetric alternatives.

If $e_1, e_2, \cdots, e_N$ are independent and identically normally distributed, then the test statistic $2^{-1}(N/3)^{\frac{1}{2}} \ln\{(1+R)/(1-R)\}$, where

$$R = \frac{\sum_{i=1}^{N}(e_i-\bar{e})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{N}(e_i-\bar{e})^2 \sum_{i=1}^{N}(Y_i-\bar{Y})^2}},$$

$$\text{and } Y_i = \left\{\frac{1}{N}\left[\sum_{j=1}^{N} e_j^2 - \frac{1}{N-1}\left(\sum_{j=1}^{N} e_j\right)^2\right]\right\}^{\frac{1}{3}}, \quad i=1,\ldots,N,$$

is asymptotically normally distributed with mean zero and unit variance. The Lin-Mudholkar (L-M) statistic is -1.74 which is slightly significant at the nominal 5% point of N(0,1). The autocorrelation function plot of the square of the normalized fitted residuals (Fig. 8.7.2.3) does not show any significant results at the 5% point.

McLeod and Li (1983) have proposed a test for testing statistical independence based on the squares of fitted residuals, autocorrelations of a linear model fitted to the data.

The autocorrelation function of e_t^2 is estimated by

$$\hat{\rho}_k = \sum_{t=k+1}^{N} (e_t^2 - \hat{\sigma}^2)(e_{t-k}^2 - \hat{\sigma}^2) \Big/ \sum_{t=1}^{N} (e_t^2 - \hat{\sigma}^2)^2 ,$$

where $\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^{N} e_t^2$.

A significance test is provided the portmanteau statistic

$$Q = N(N+2) \sum_{i=1}^{M} \frac{\hat{\rho}_i^2}{N-i} \quad , M \ll N ,$$

which is asymptotically $\chi^2_{(M)}$ if the e_t are independent.

The Mcleod and Li statistic is 13.69 which is not significant at the nominal 5% point of $\chi^2_{(20)}$.

In order to examine the mean, variance and higher moments of model (8.4.3), a sample of 10000 observations, using a simulation technique, is generated. Mean, variance, third and fourth sample moments of the data and simulated data are given in Table (8.7.1). It seems that most of the statistics of the model are quite close to those of the data

Table 8.7.1

Stat.	$E(X)$	$E(X^2)$	$E(X^3)$	$E(X^4)$	MIN	MAX	Q_1	Q_2
DATA	3.21	10.55	35.31	120.02	1.78	3.95	2.90	3.60
SIMULATION	3.22	10.66	36.33	125.30	0.72	5.16	2.87	3.62

Fig. (8.7.1) shows the sample autocovariance function of the blowfly data and the 10000 simulated observations for model (8.4.3). The agreement of the autocovariance function of simulated data with that of the data is quite good.

Using the Tukey window, the smoothed spectral density functions of the data and 10000 simulated observations are estimated. Fig. (8.7.2) shows a good agreement between the spectral density function for the data and simulated observations.

The sample probability distribution of the blowfly data may be used to check the adequacy of the fitted model. It is expected that a good fitted model, will generate a similar probabilistic structure as that of

the data. We will use the graphical methods based on marginal distributions, bivariate distributions and regression functions to check the adequacy of model (8.4.3) for the first 200 log transformation of blowfly data.

Using the Monte Carlo method, we obtained a simulated data size of 10000 observations from SETAR(2;1,3), equation (8.4.3). The sample histograms of data and simulated data are given by Figs. (8.7.3) and (8.7.4) respectively. The marginal distribution of the SETAR model and that of data are quite similar.

Figs. (8.7.5) show a remarkable agreement between the bivariate frequency distributions of simulated observations and those of the data.

Non-parametric estimates of $E(X_t|X_{t+j})$ and $Var(X_t|X_{t+j})$ (Figs. 8.7.6), show that the lag regression and the conditional variance of the fitted model and those of the data have a good agreement.

The time plot of a realization of SETAR(2;1,3) model (Fig. 8.7.7), shows a reasonable agreement with the time plot of the data.

Fig. (8.7.8) and Fig. (8.7.9) show the one-step ahead predictions and the systematic part of the model (8.4.3) with the white noise suppressed and with the latest few observations as initial values of the recursion.

Although the model (8.4.3) could pass all (except the test for normality) diagnostic checks, the diagonal elements of hat matrix may reveal a very interesting feature of this data set. Figs. (8.7.10) show the diagonal elements of the hat matrixes associated with the piecewise linear autoregressive model. From these, it is clear that the 'influential pattern' remains cyclical with a period of about 20 units of time, which is about the same as that of the blowfly population cycle. This suggests the need of at least a bivariate time series, incorporating the egg count, for example. It is feasible that the interaction between the fly population and the egg population will give a better understanding of the cycle generating mechanism.

Tsay (1988) has pointed out that for all non-linear models fitted to the blowfly data which have been considered by him, large residuals persistently appear at the same time points. Many of the large residuals correspond to the big jumps in blowfly population caused by an unusual amount of emergence. Since emer-

gence depends on the number of eggs laid, therefore the egg series can be incorporated. Tsay (1988) by incorporating the egg series has managed to reduce the variance of fitted residuals 57.2%. However, the residual analysis shows that the resulting residual is not Gaussian, with some large values. In order to assess the effect of outlying (influential) observations, he has used some dummy (binary) variables. Then by using four dummy variables, the skewness and kurtosis of fitted residuals are reduced substantially, but there is some evidence that the fitted residuals are non-Gaussian.

The general message is that we should always look for influential data, since influential data pattern may suggest a direction of modification.

Finally the following AR(3) model is fitted to the second part ($201 \le t \le 361$) of $\log_{10}$ transformation blowfly data.

$$X_t = \underset{(0.15)}{0.64} + \underset{(0.08)}{1.03} X_{t-1} - \underset{(0.11)}{0.09} X_{t-2} - \underset{(0.08)}{0.12} X_{t-3} + e_t ,$$

where $Var(e_t) = 0.0115$. The maximum lag is equal to 10 and the AIC criterion is employed for model selection.

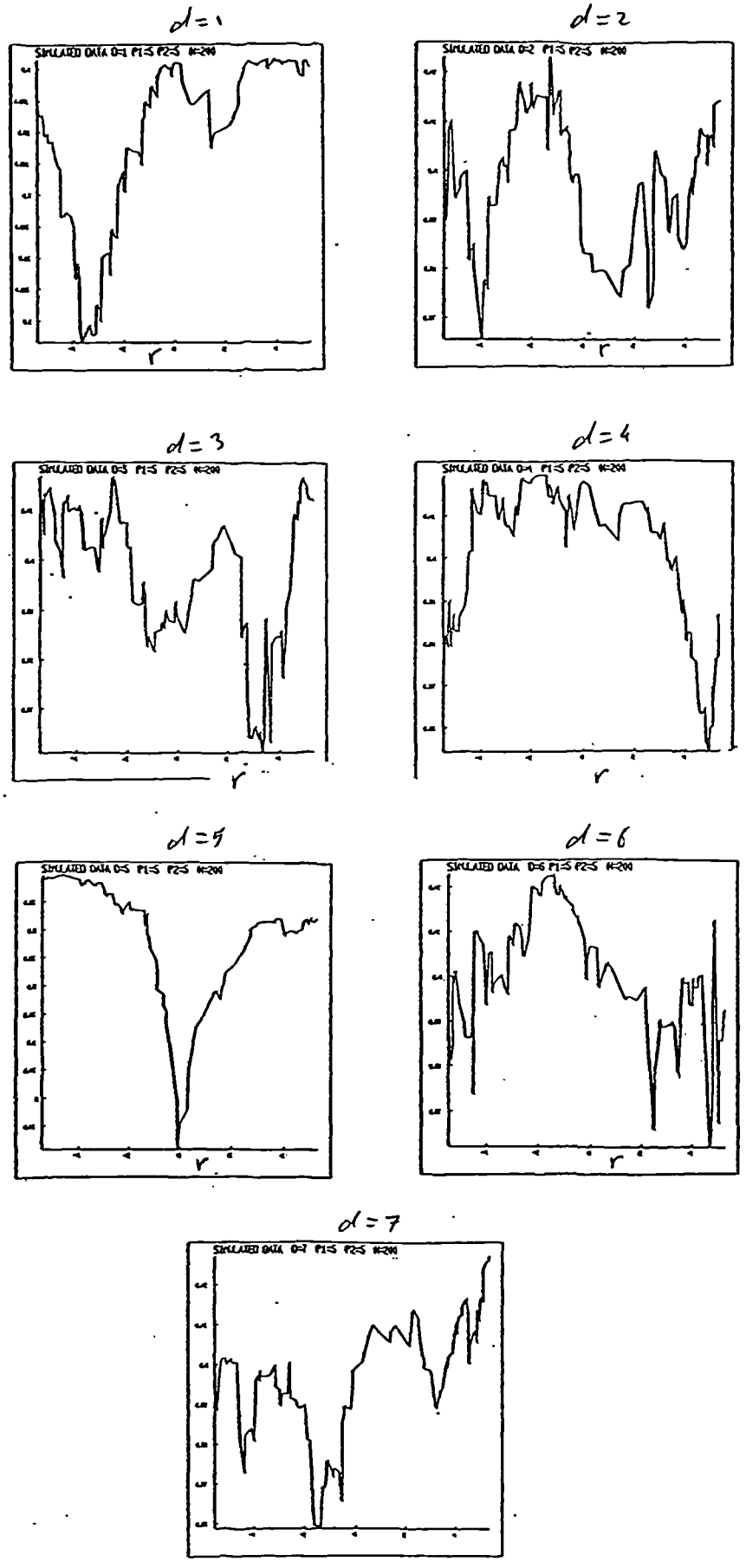


Fig. 8.3.1 : Profile likelihood functions of model (8.3.1).

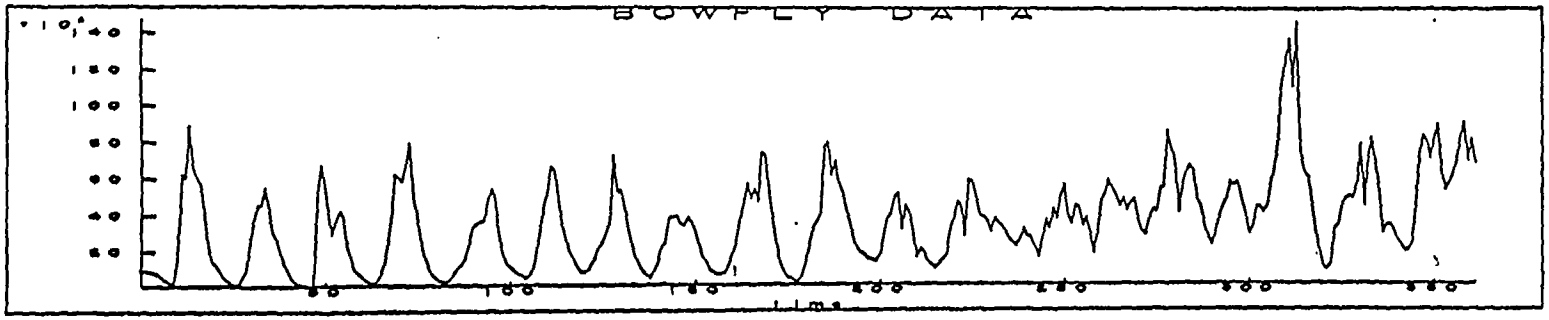


Fig. 8.4.1 : Time plot of data.

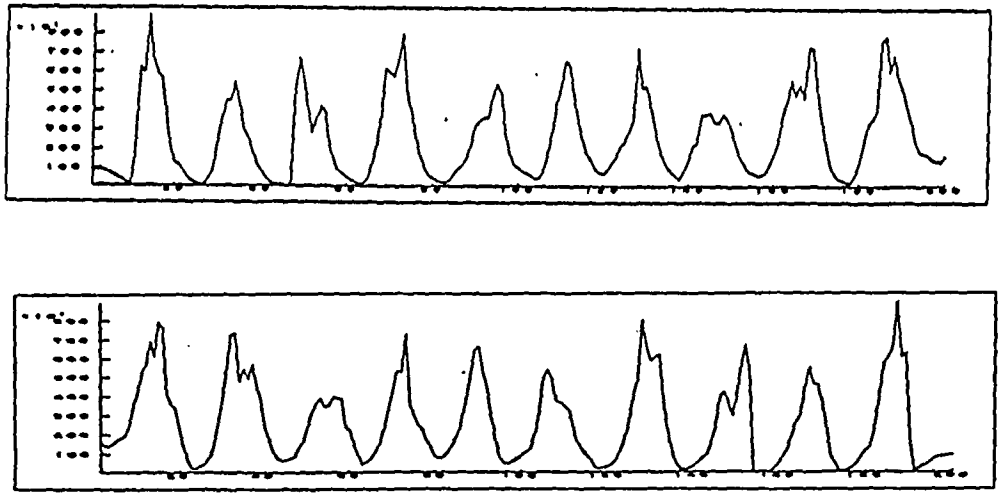


Fig. 8.4.2-8.4.3 : Time and reverse time plot of first 200 observations.

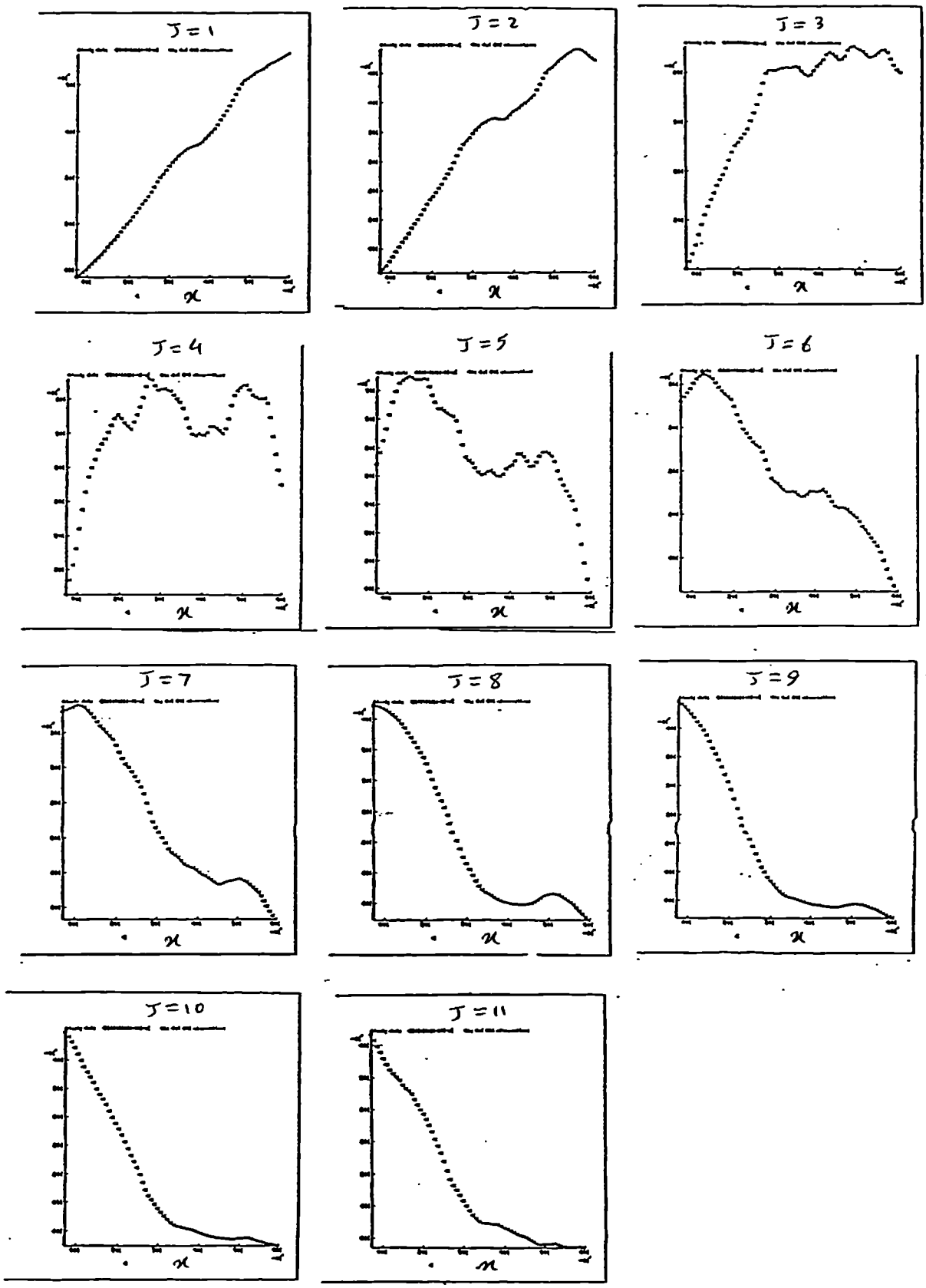


Fig. 8.4.4a : Non-parametric estimation of $E(X_t \mid X_{t-j}=x)$, j 1, . . . , 11 (data)

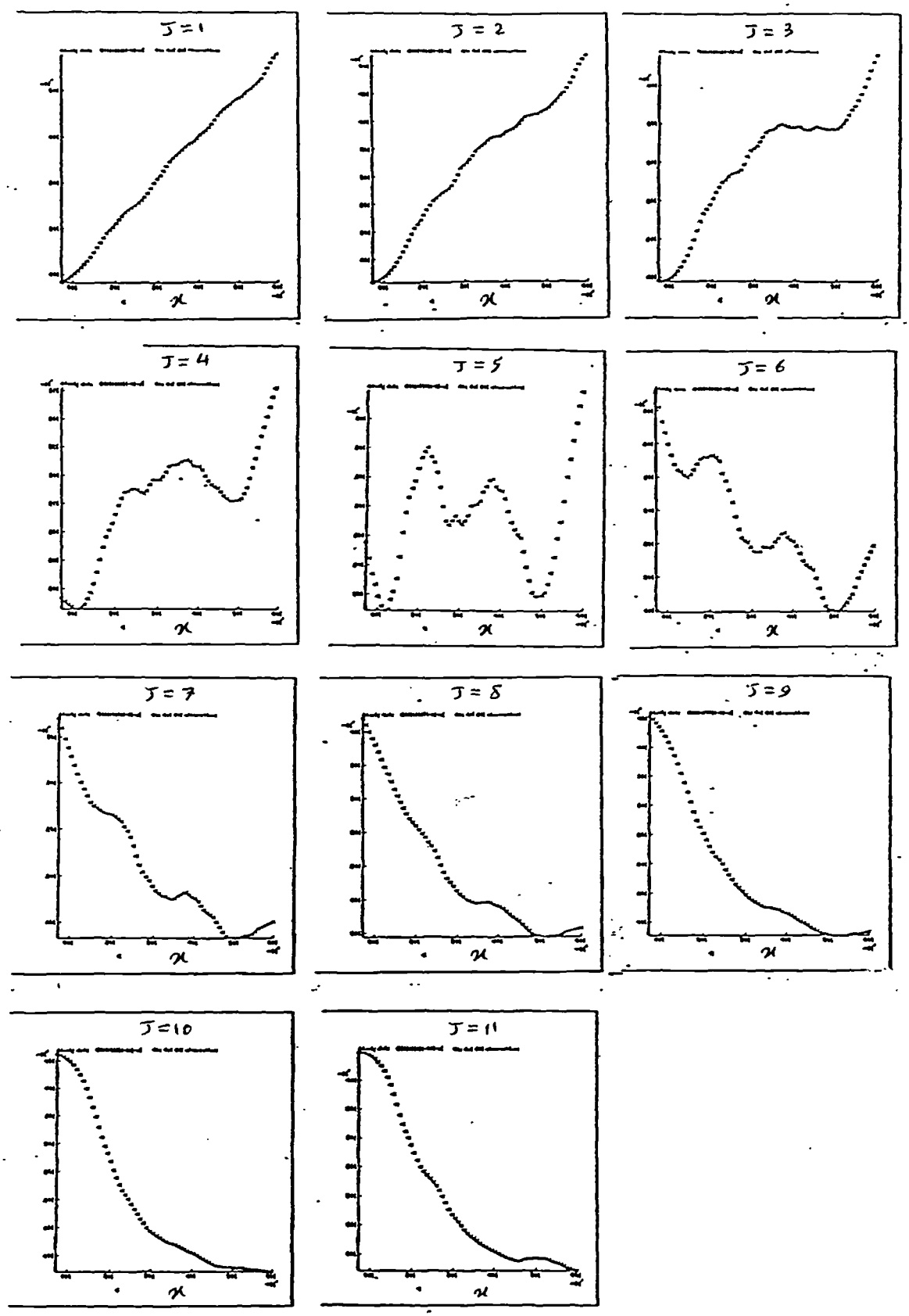


Fig. 8.4.4b : Non-parametric estimation of $E(X_t \mid X_{t+j}=x)$, j 1, . . . , 11 (data)

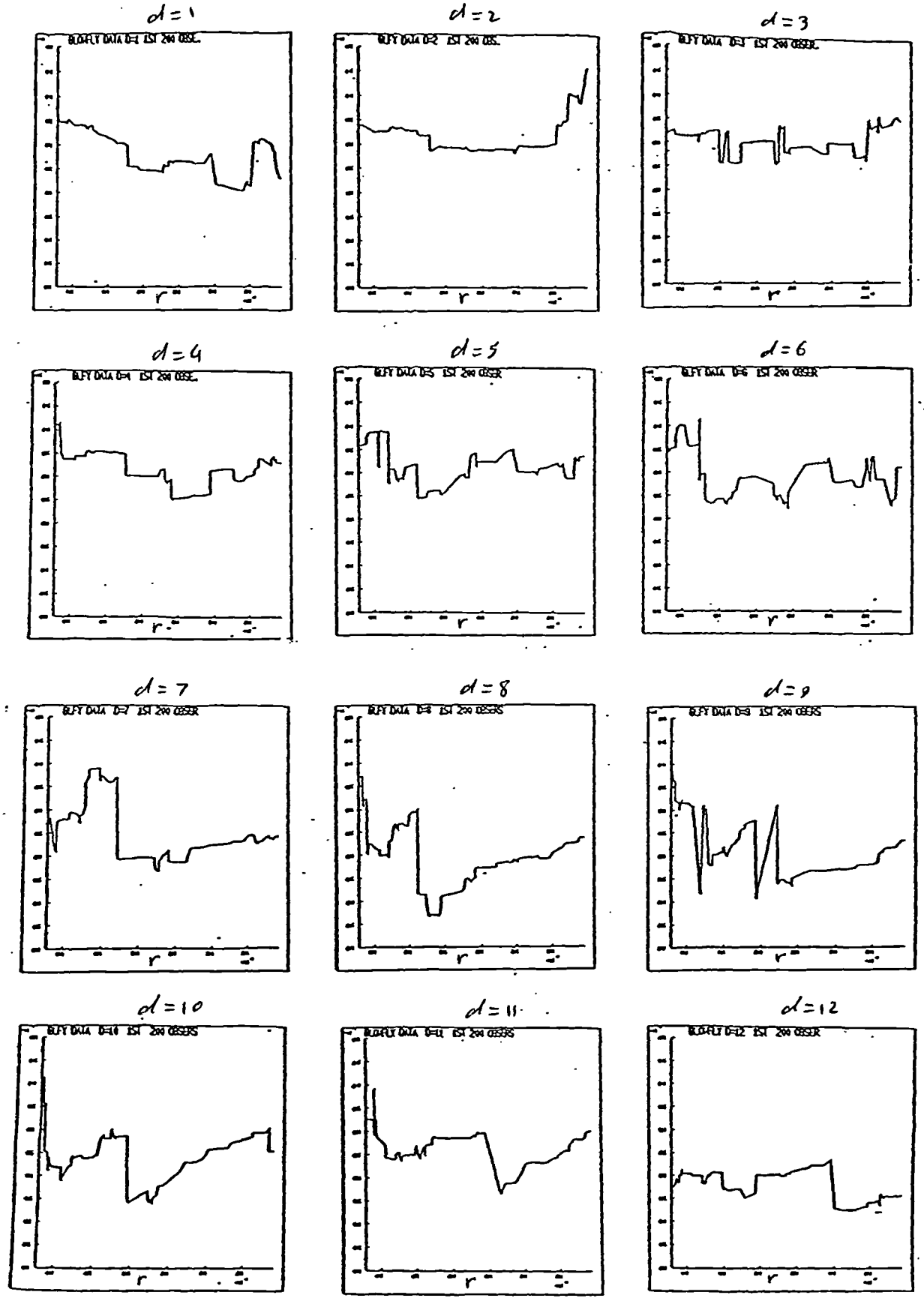


Fig. 8.4.5 : Profile likelihood functions of data for d=1,...,18 .

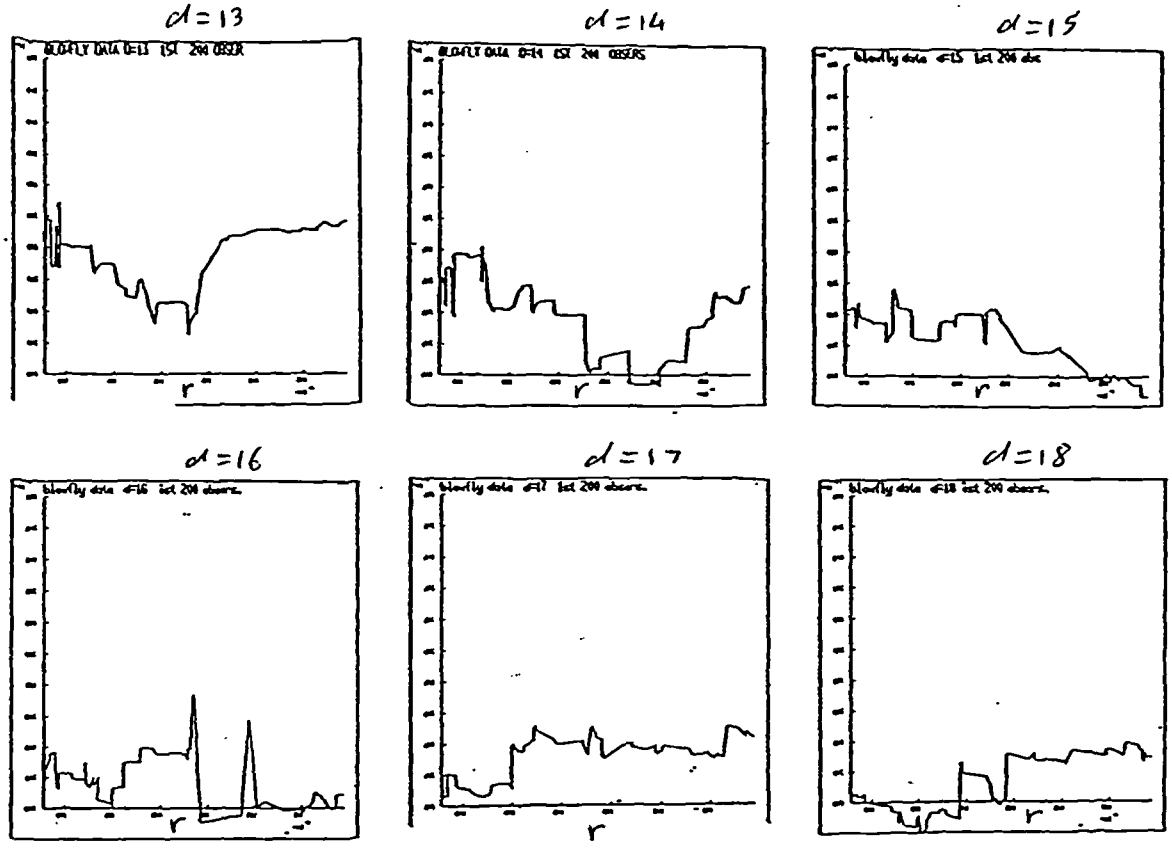


Fig. 8.4.5

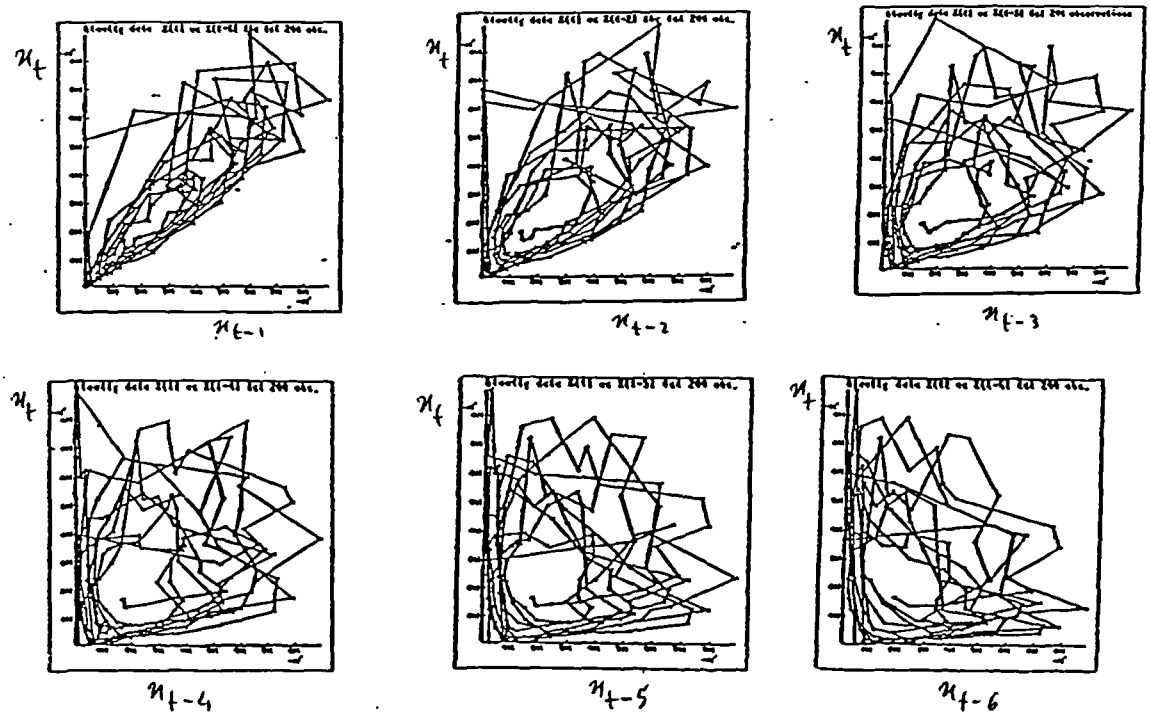


Fig. 8.4.6 : Scatter diagrams of X_t vs X_{t-j} for $j=1,\ldots,8$.

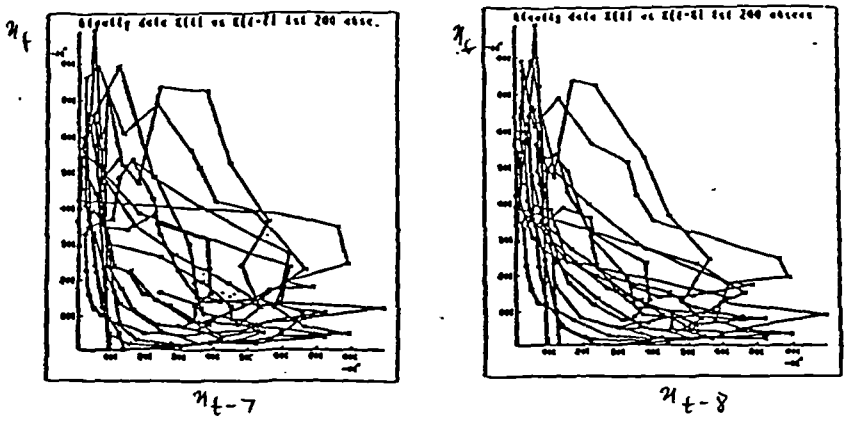


Fig. 8.4.6

Midpoint	Count	
0	27	**********
1000	58	**********
2000	32	**********
3000	16	******
4000	25	**********
5000	17	*******
6000	11	*****
7000	3	**
8000	3	**
9000	1	*

Fig. 8.4.7 : Histogram of data.

Fig. 8.4.8 : Bivariate histograms of data.

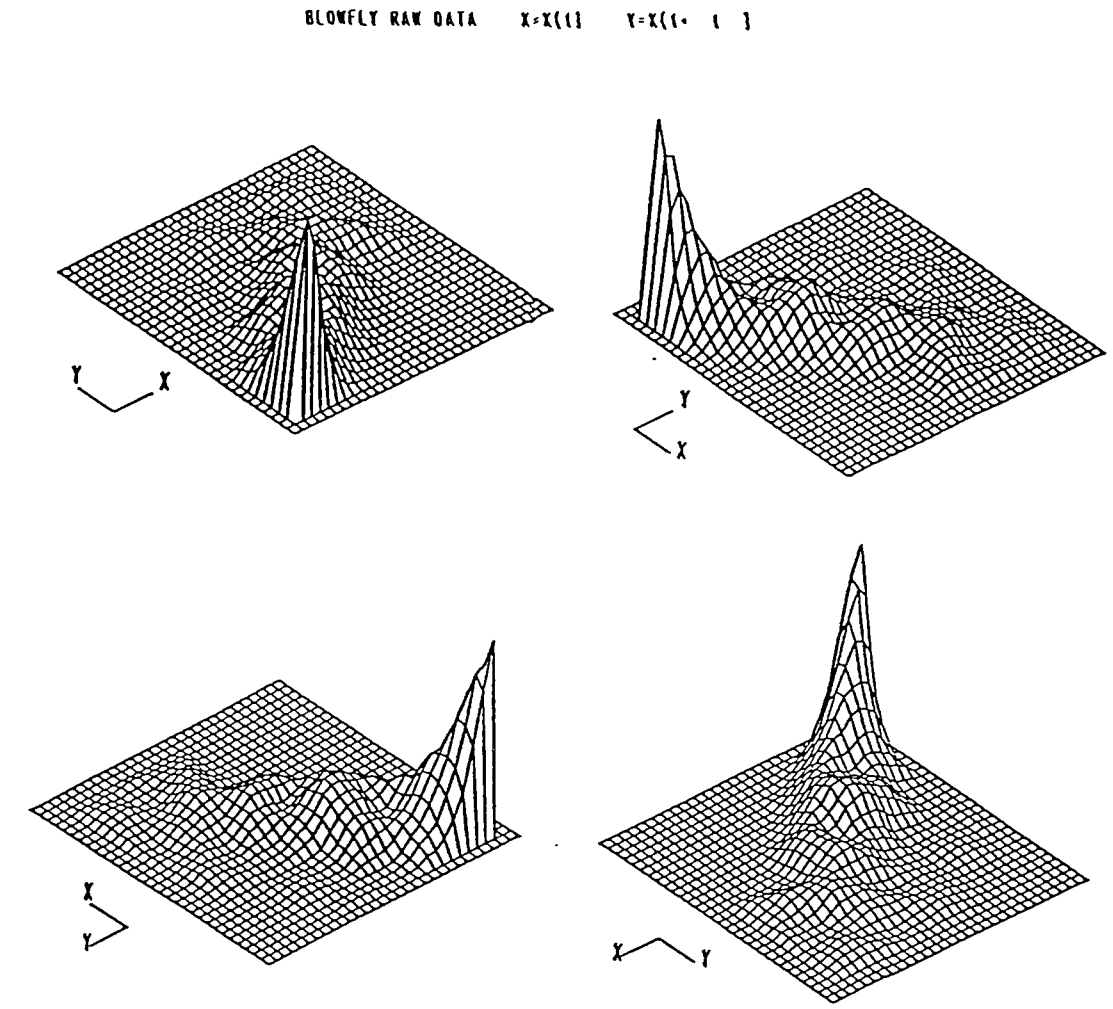
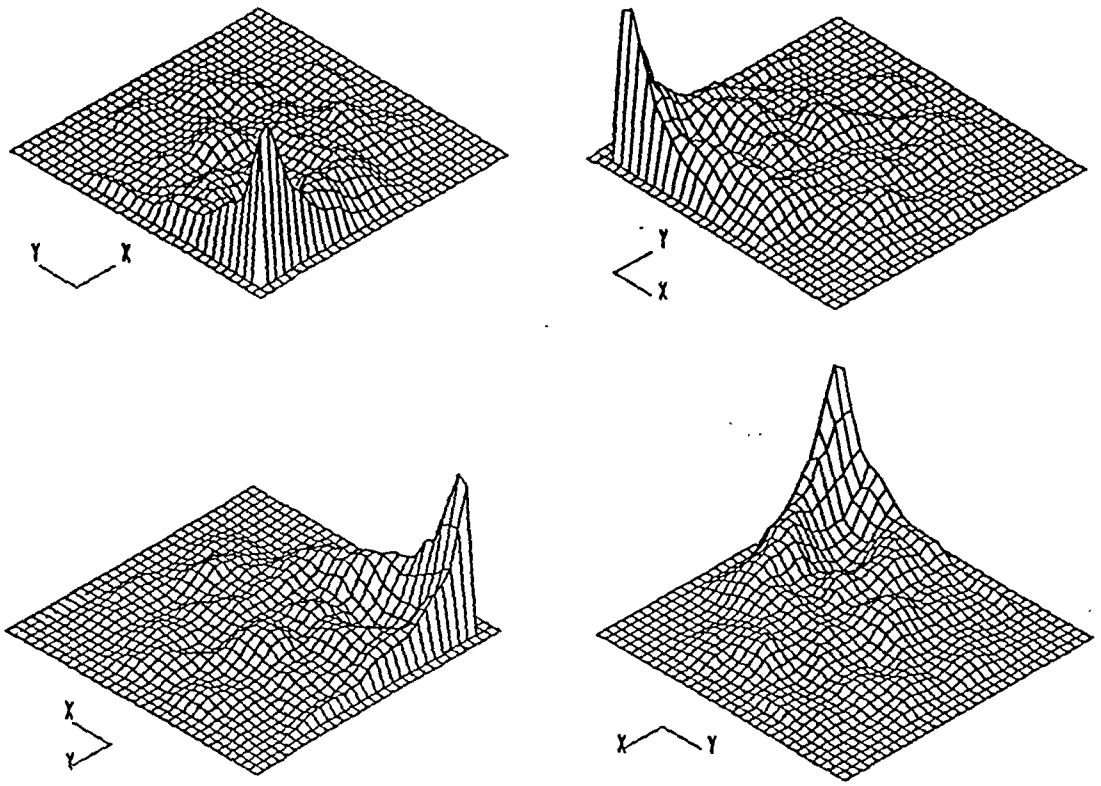
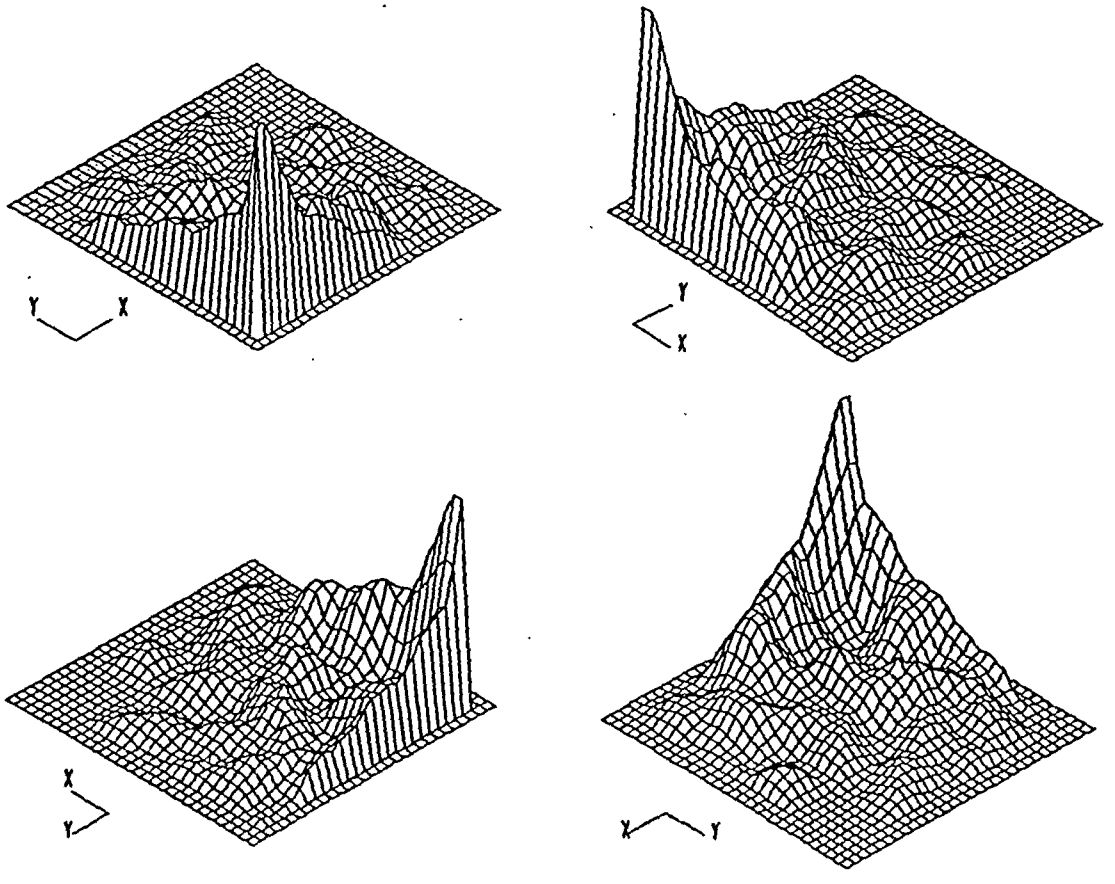


Fig. 8.4.8

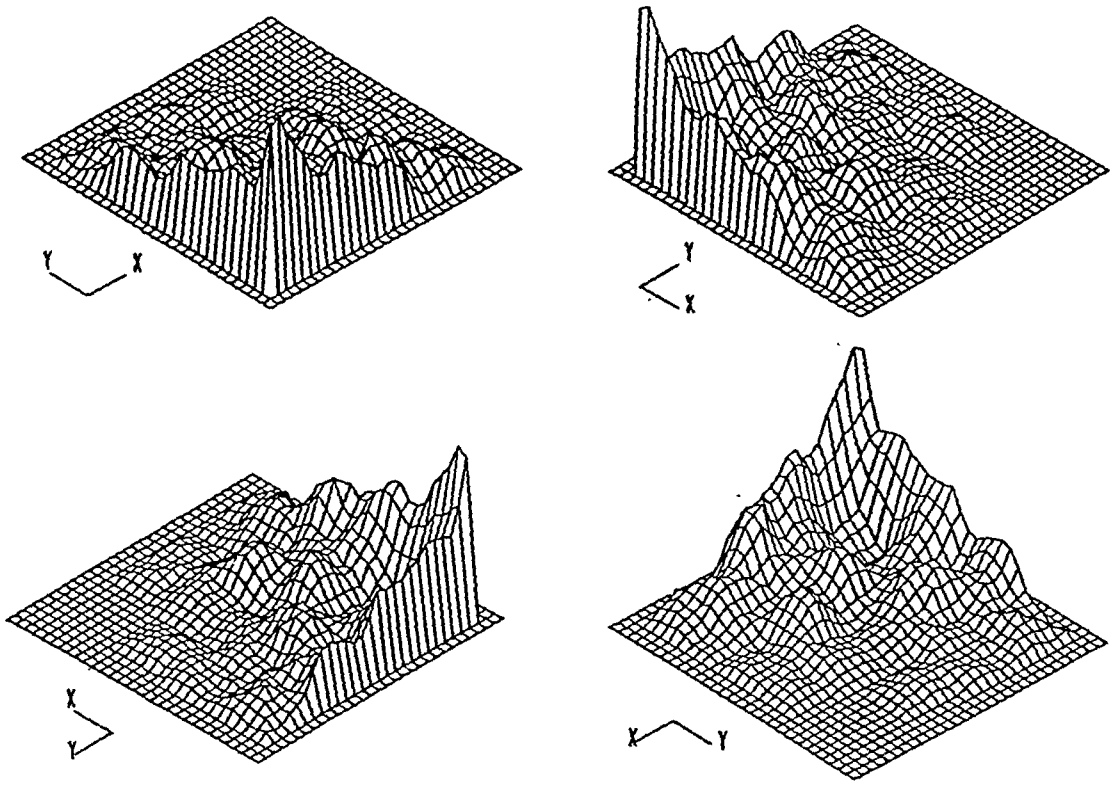
BLOWFLY RAW DATA X=X(1) Y=X(1, 3)



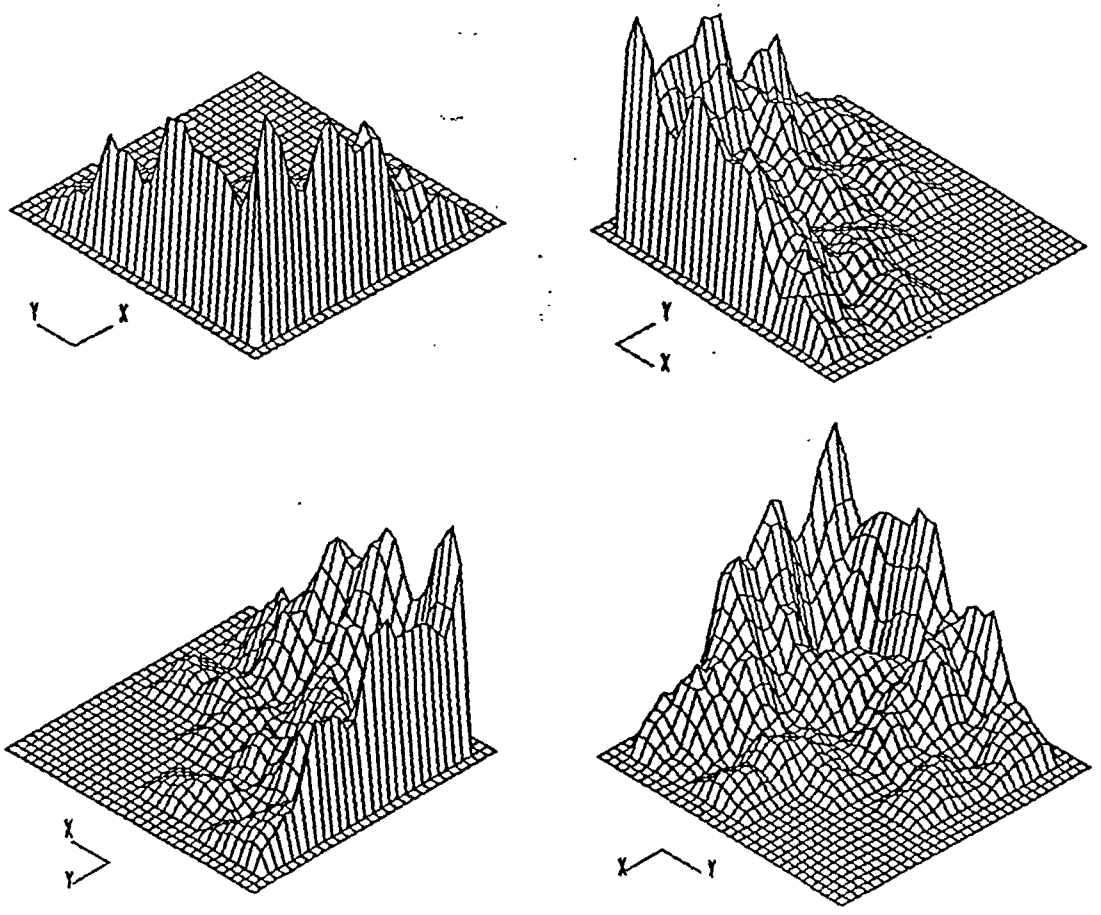
BLOWFLY RAW DATA X=X(1) Y=X(1, 4)



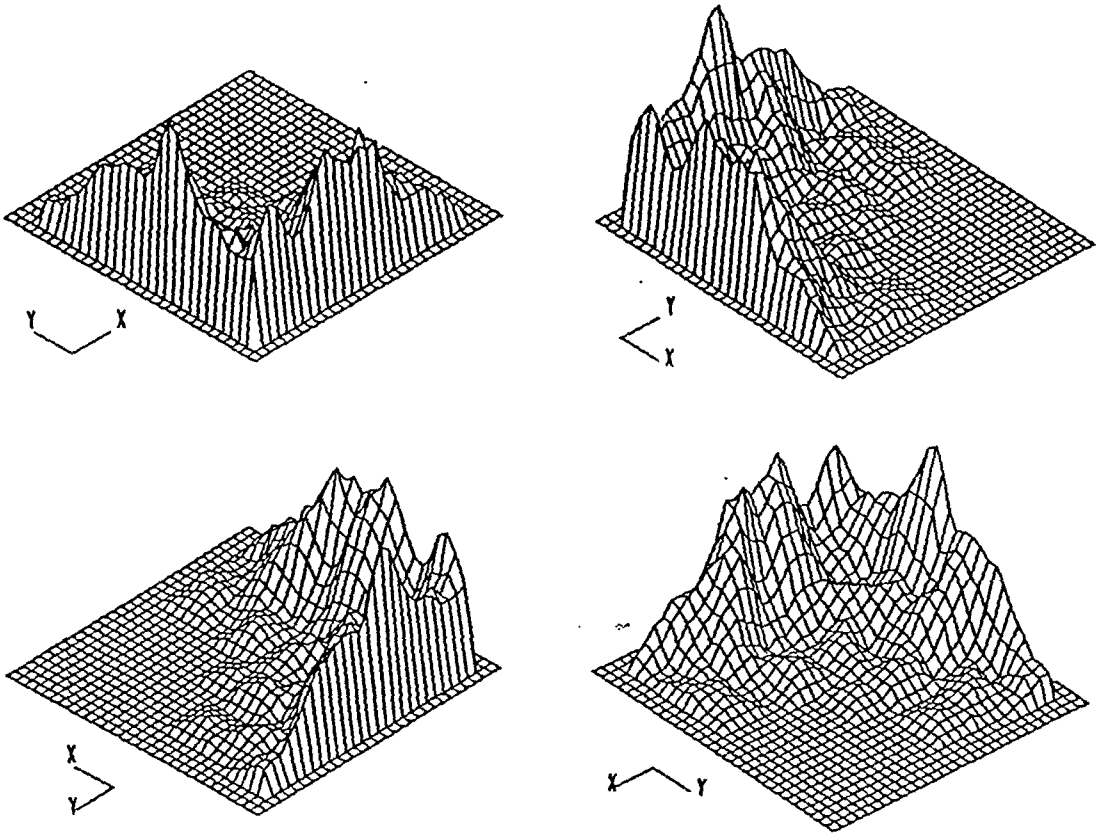
BLOWFLY RAW DATA X=X(I) Y=X(I+ 5)



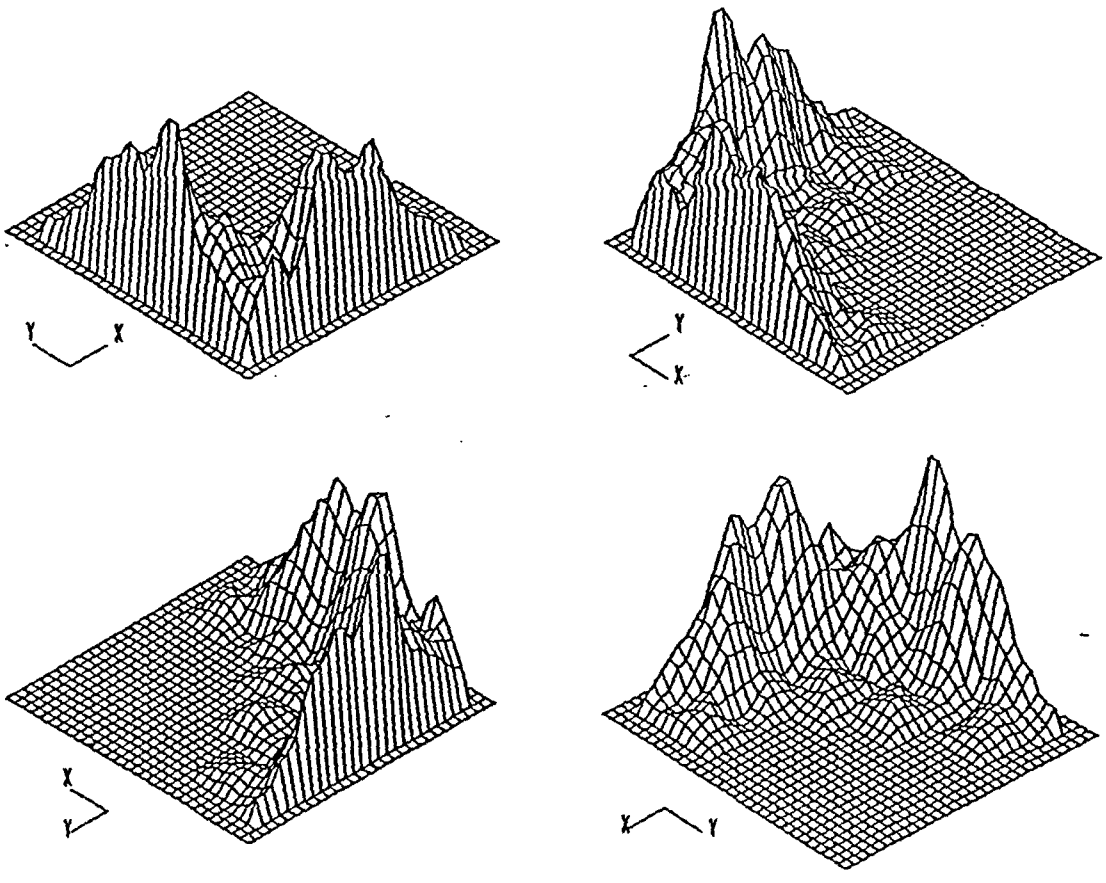
BLOWFLY RAW DATA X=X(I) Y=X(I+ 6)



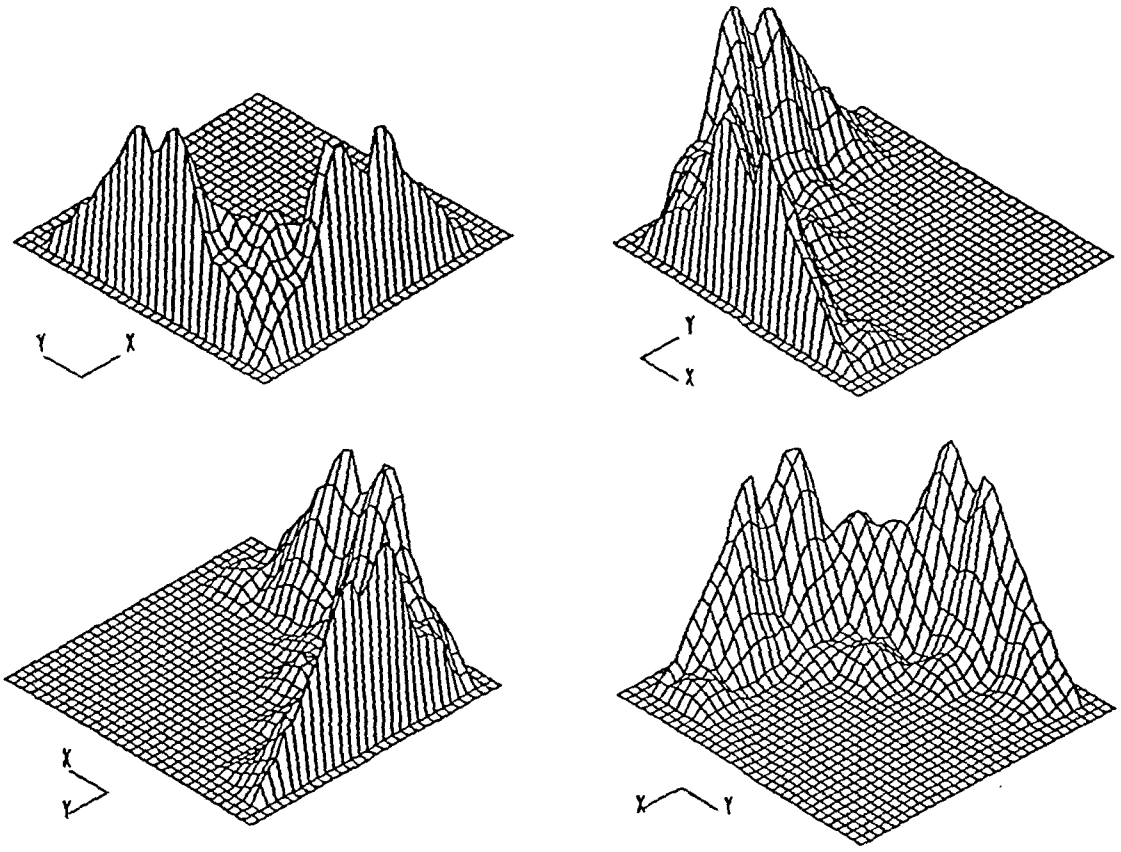
BLOWFLY RAW DATA X=X(1) Y=X(1, 2)



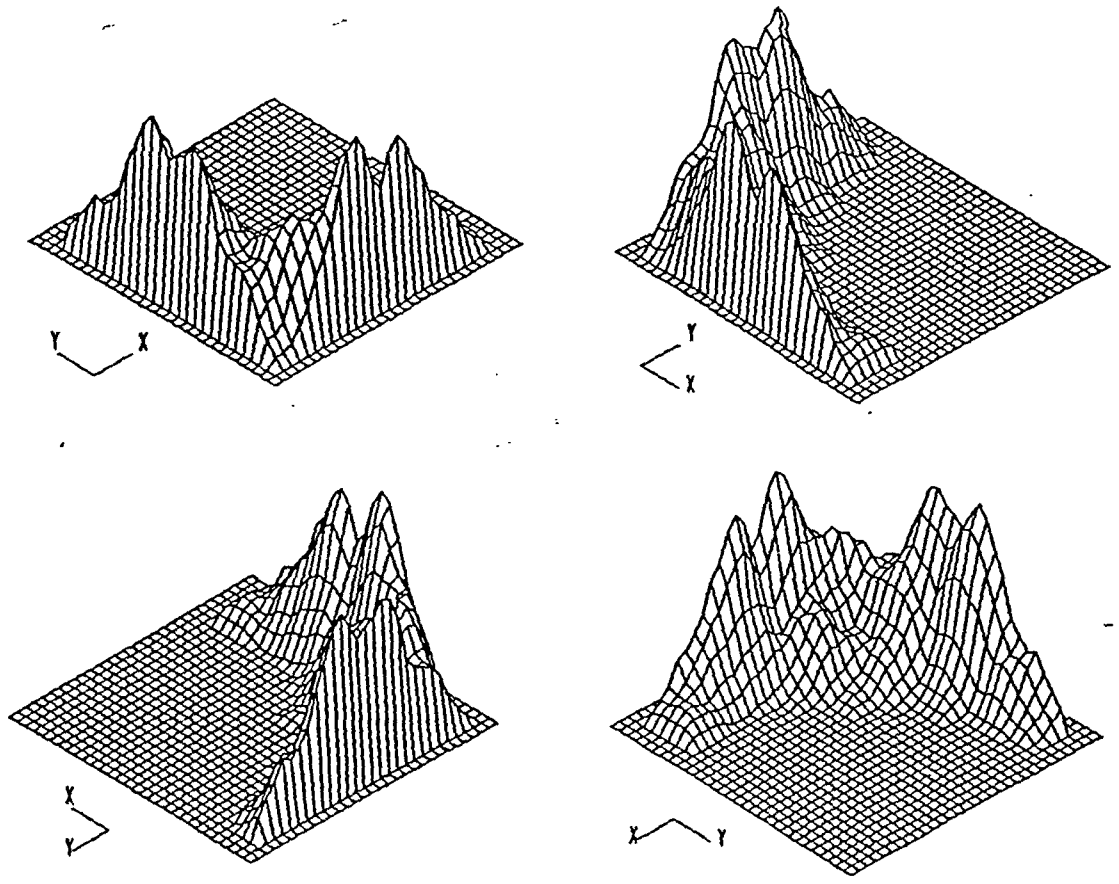
BLOWFLY RAW DATA X=X(1) Y=X(1, 8)



BLOWFLY RAW DATA X=X(1) Y=X(1, 9)



BLOWFLY RAW DATA X=X(1) Y=X(1, 10)



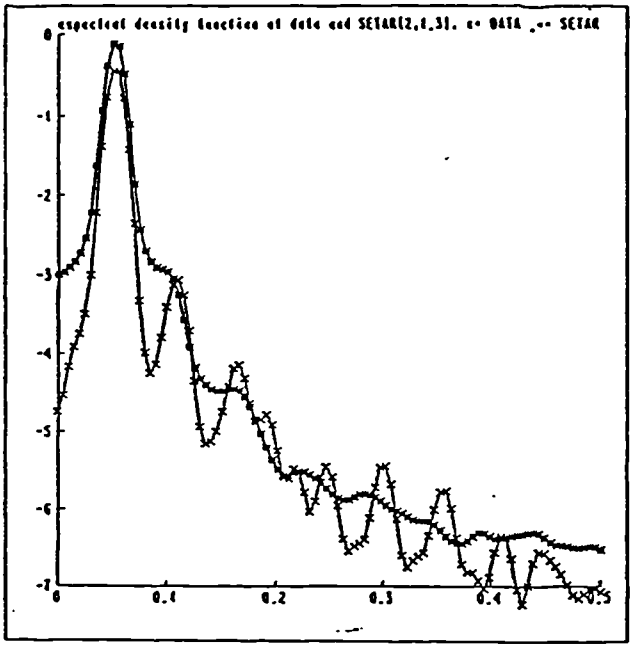


Fig. 8.7.2 : Spectral density functions of data and fitted model.

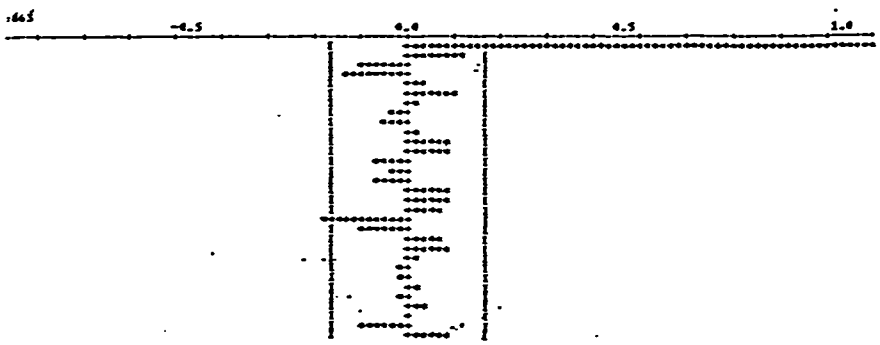


Fig. 8.7.1.1 : ACF of fitted residuals.

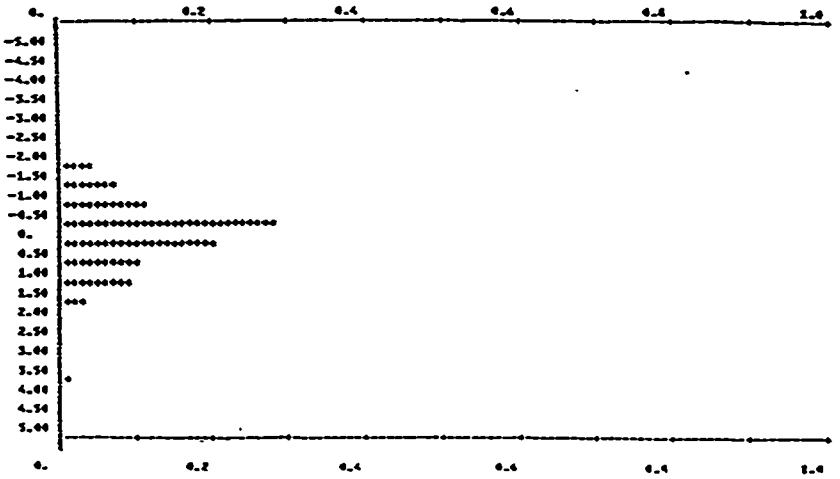


Fig. 8.7.2.1 : Histogram of fitted residuals.

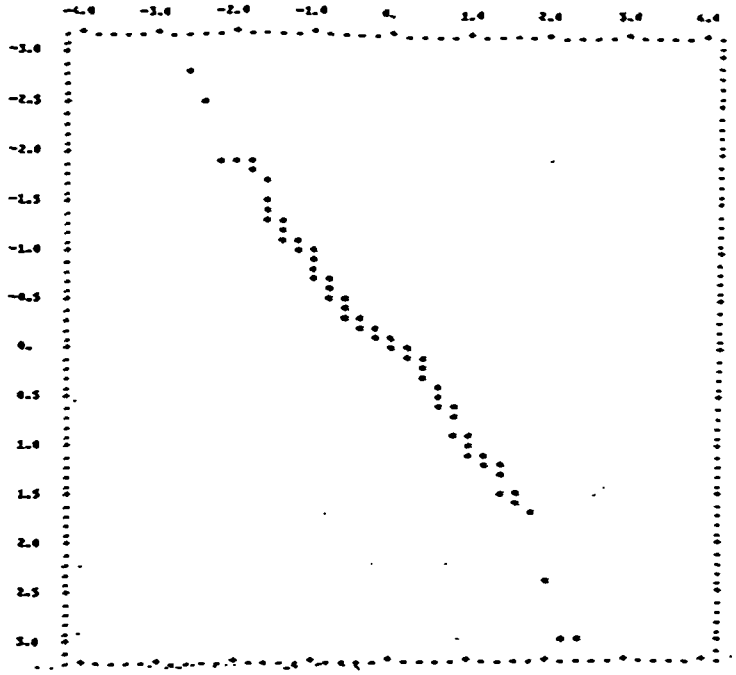


Fig. 8.7.2.2 : Normal plot of fitted residuals.

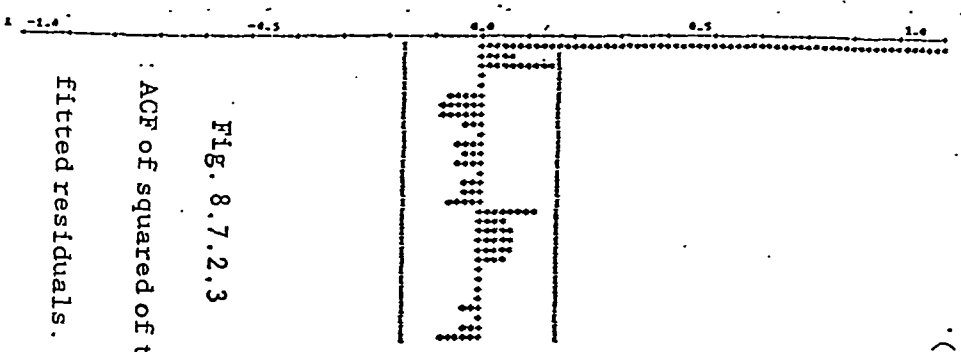


Fig. 8.7.2.3 : ACF of squared of the fitted residuals.

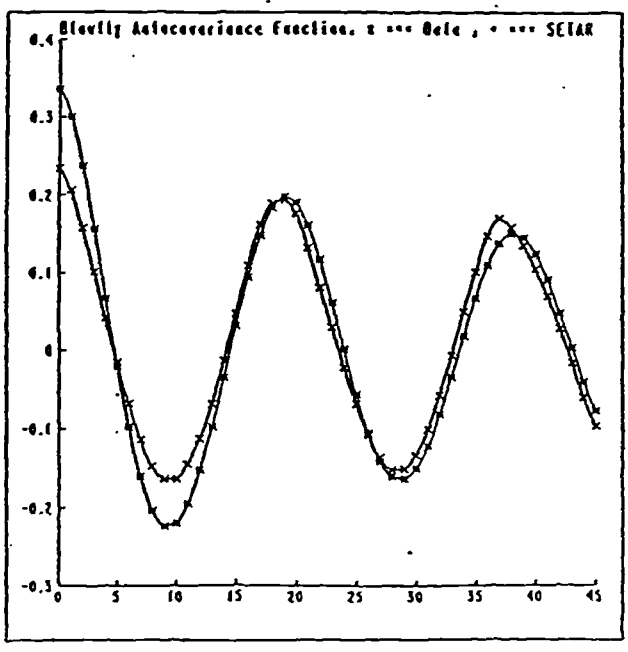


Fig. 8.7.1 ACF of data and simulated data form model (8.4.3).

Midpoint	Count	
0.8	0	
1.2	0	
1.6	1	*
2.0	6	***
2.4	15	********
2.8	43	**********************
3.2	52	***************************
3.6	70	************************************
4.0	13	*******
4.4	0	
4.8	0	
5.2	0	

Fig. 8.7.3

Midpoint	Count	
0.8	3	*
1.2	14	*
1.6	82	**
2.0	375	*****
2.4	1045	**************
2.8	1558	**********************
3.2	2218	********************************
3.6	3771	**
4.0	865	**********
4.4	61	*
4.8	6	*
5.2	2	*

Fig. 8.7.4

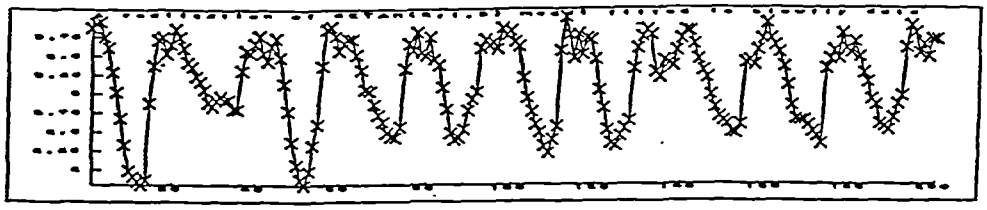


Fig. 8.7.7

Fig. 8.7.3 : Histogram of the data.

Fig. 8.7.4 : Histogram of simulated data.

Fig. 8.7.7 : A realization of model (8.4.3).

Fig. 8.7.5 : Bivariate histograms of data and simulated data.

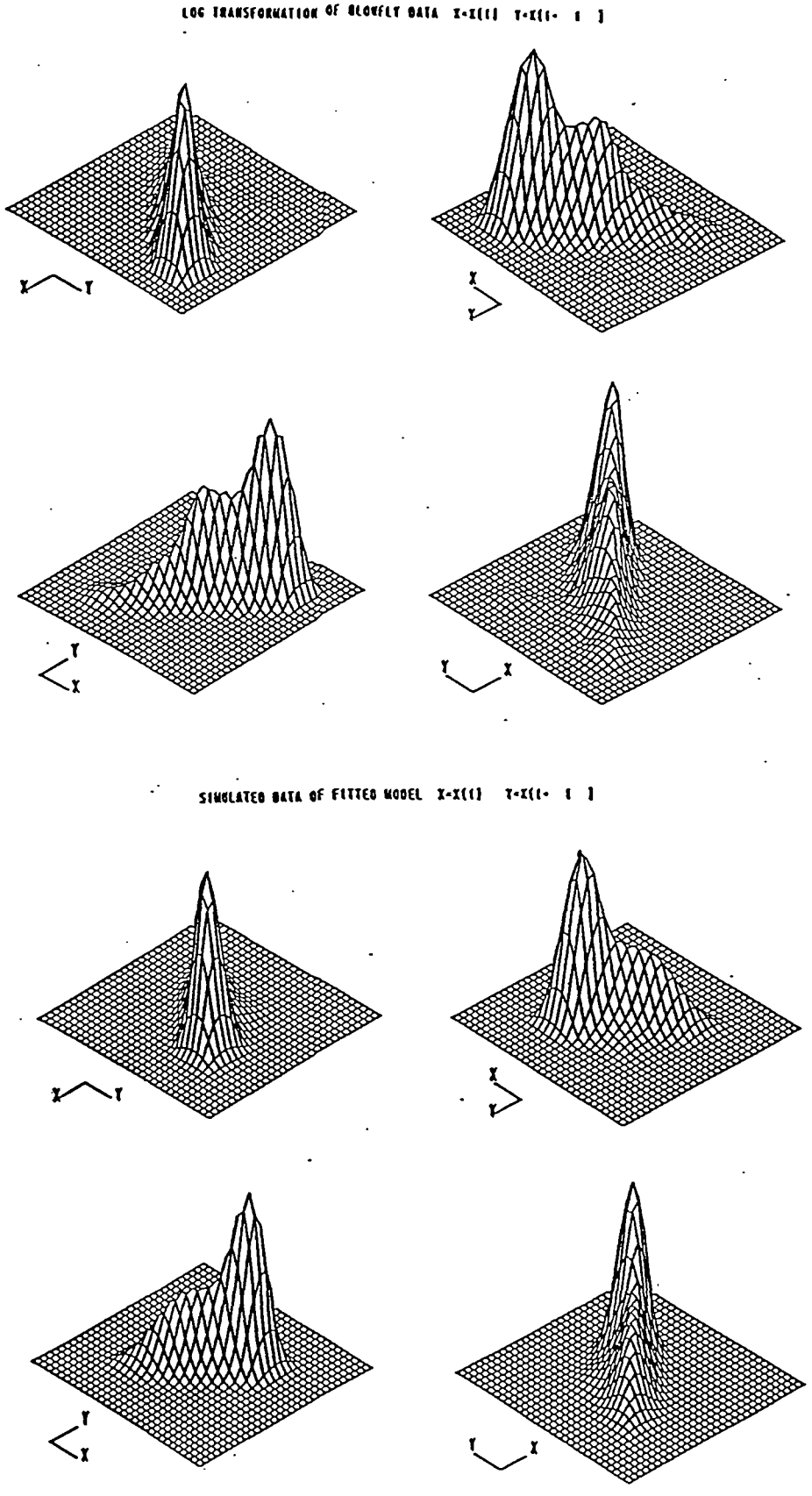
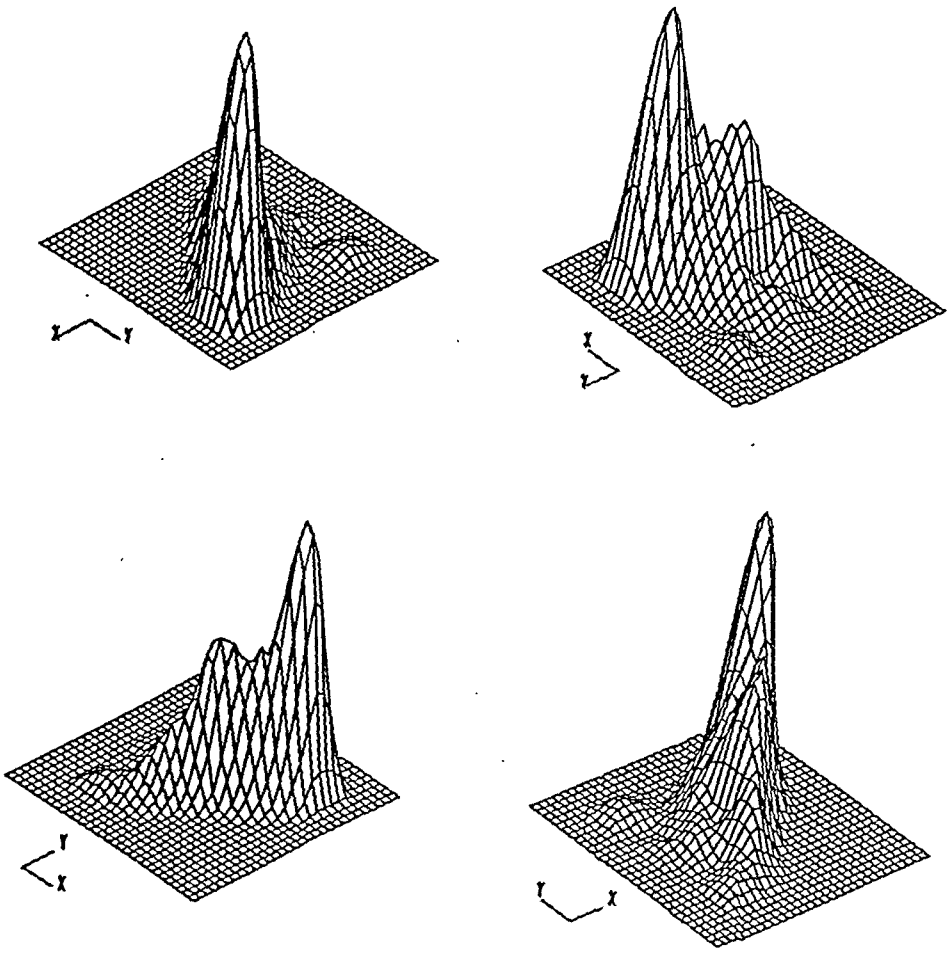
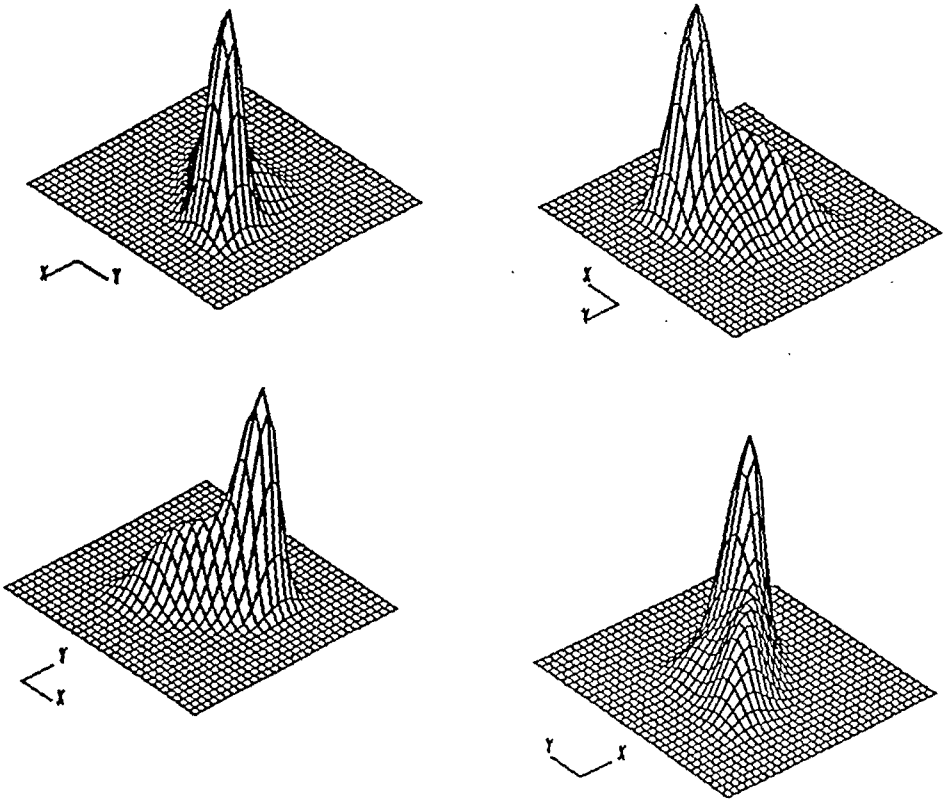


Fig. 8.7.5

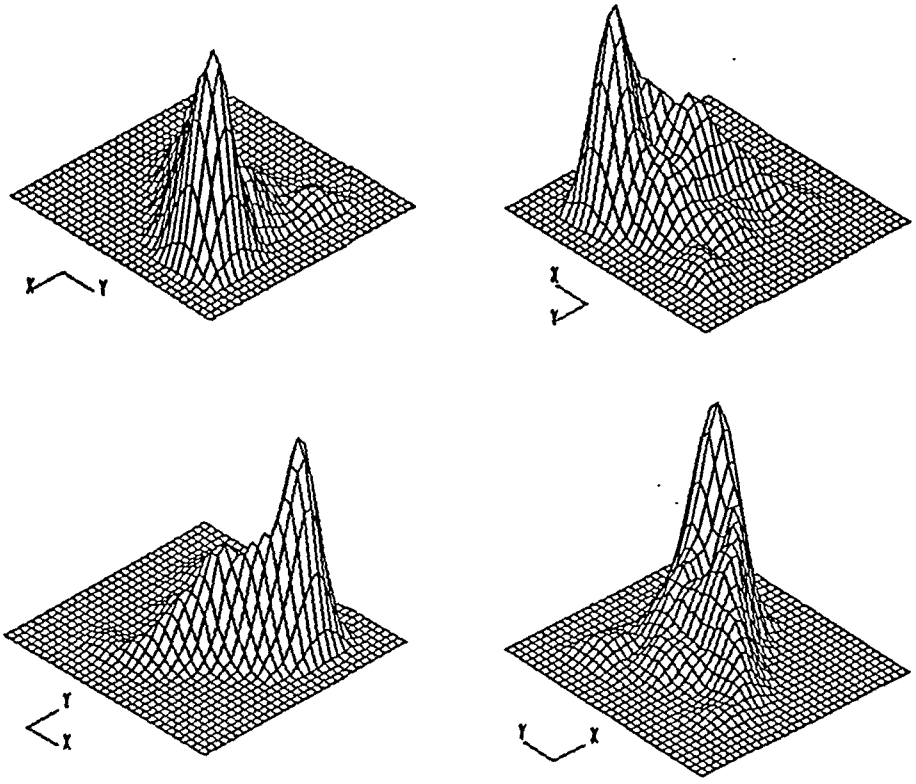
LOG TRANSFORMATION OF BLOWFLY DATA X=X[1] Y=X[1+ 2]



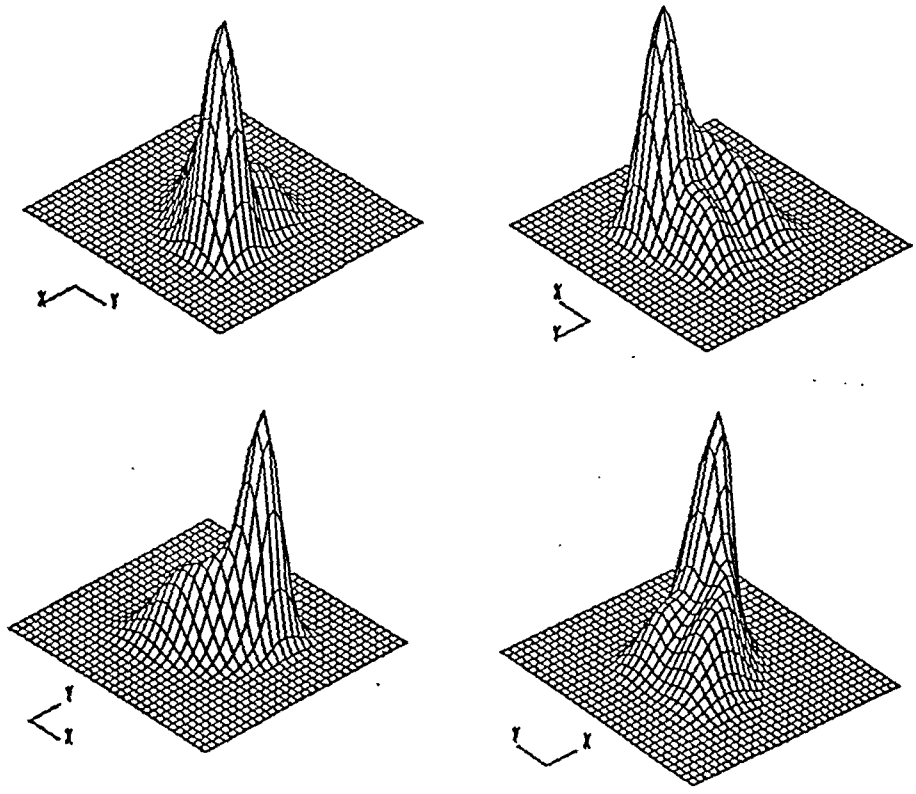
SIMULATED DATA OF FITTED MODEL X=X[1] Y=X[1+ 2]



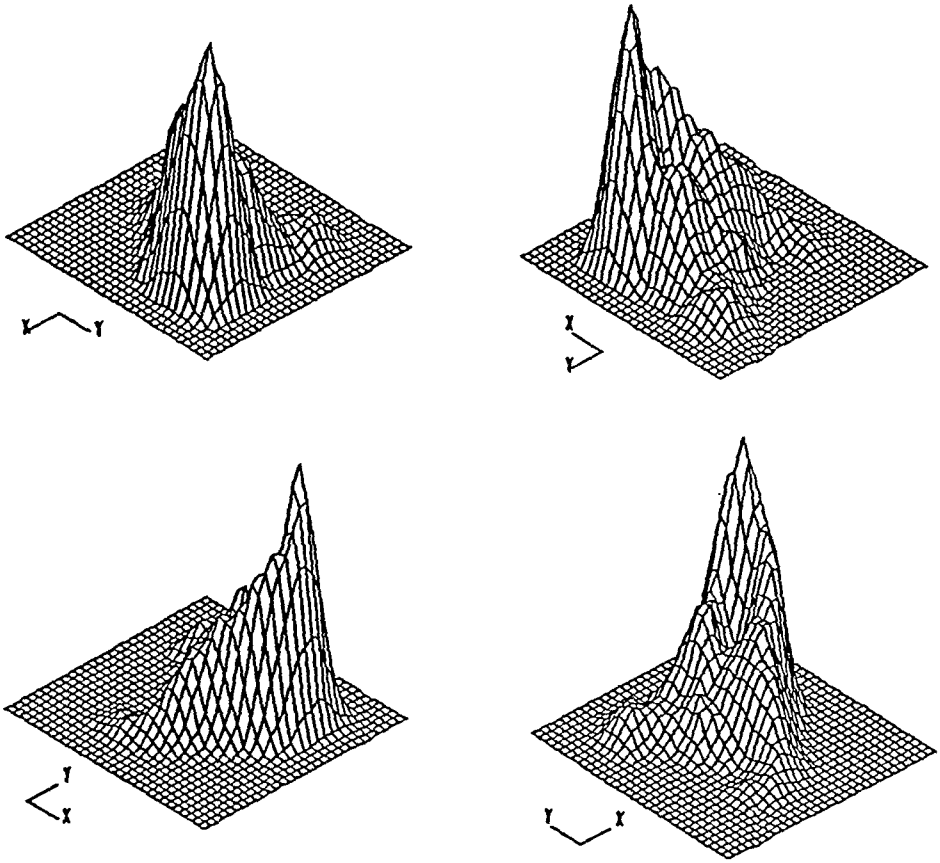
LOG TRANSFORMATION OF BLOWFLY DATA X=X[1] Y=X[1: 3]



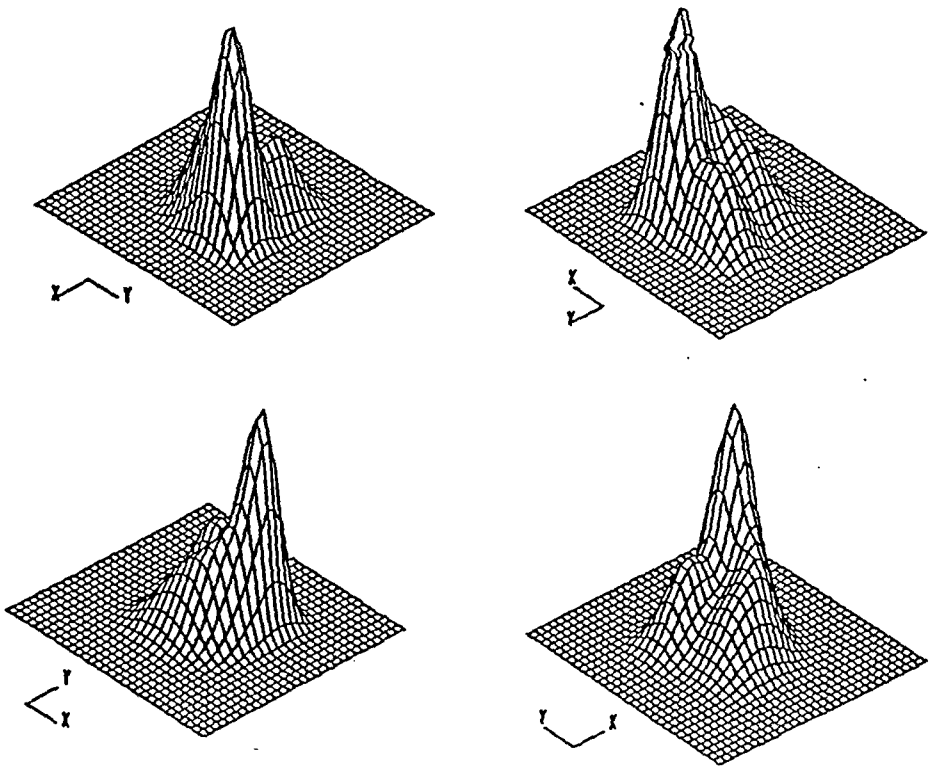
SIMULATED DATA OF FITTED MODEL X=X[1] Y=X[1: 3]



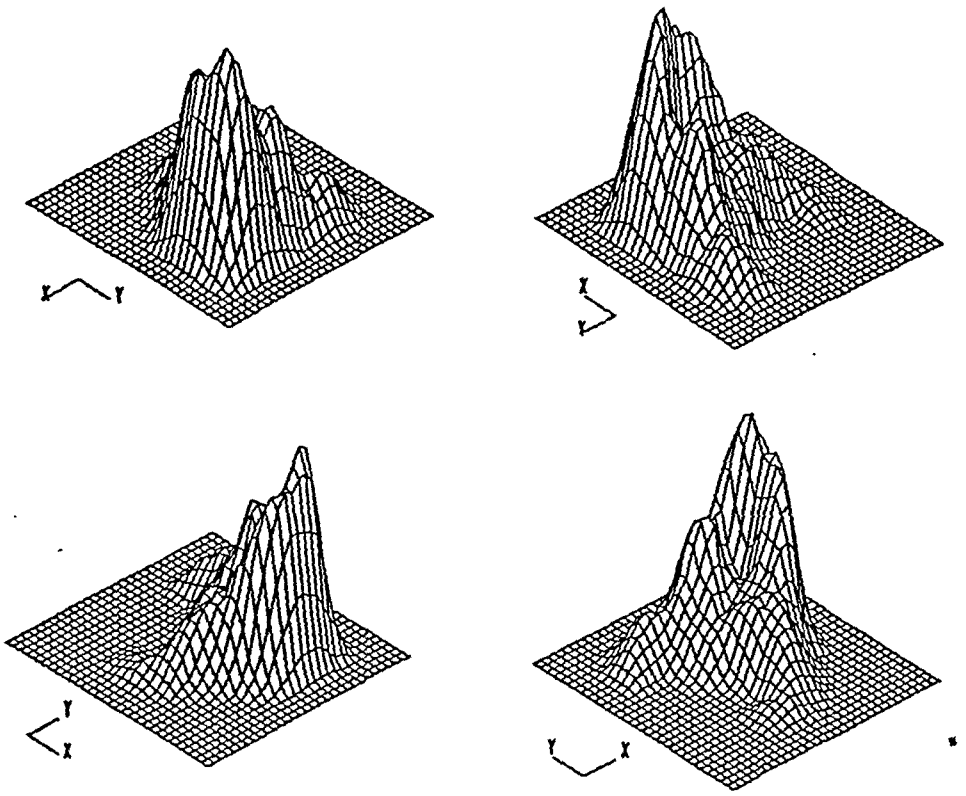
LOG TRANSFORMATION OF BLOWFLY DATA X=X(I) Y=X(I, 4)



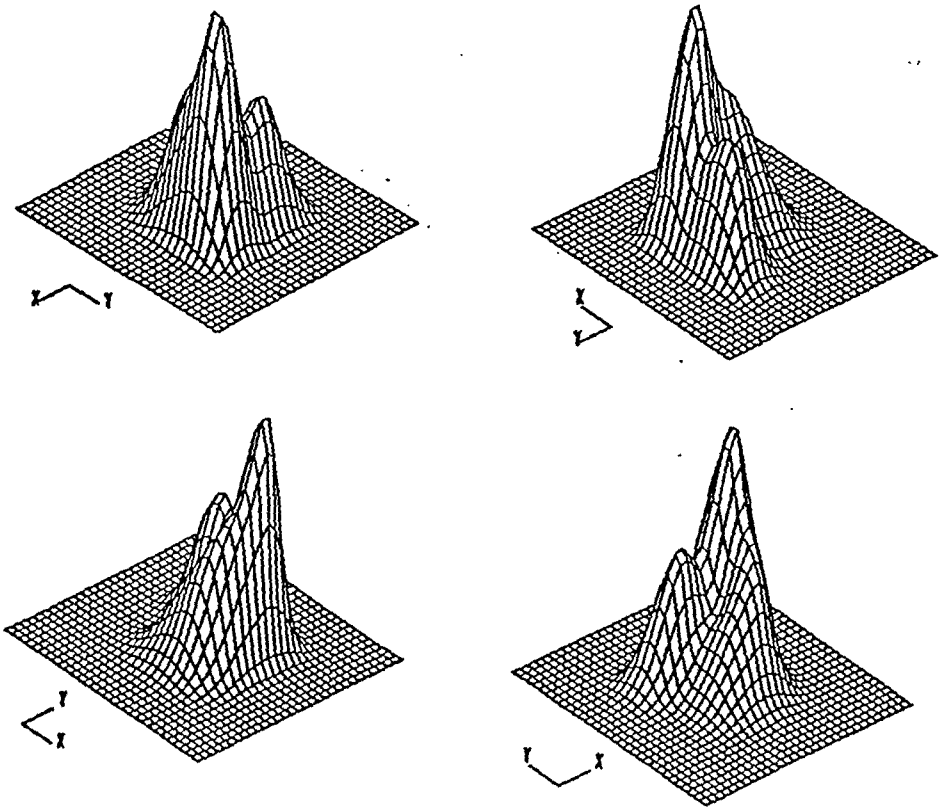
SIMULATED DATA OF FITTED MODEL X=X(I) Y=X(I, 4)



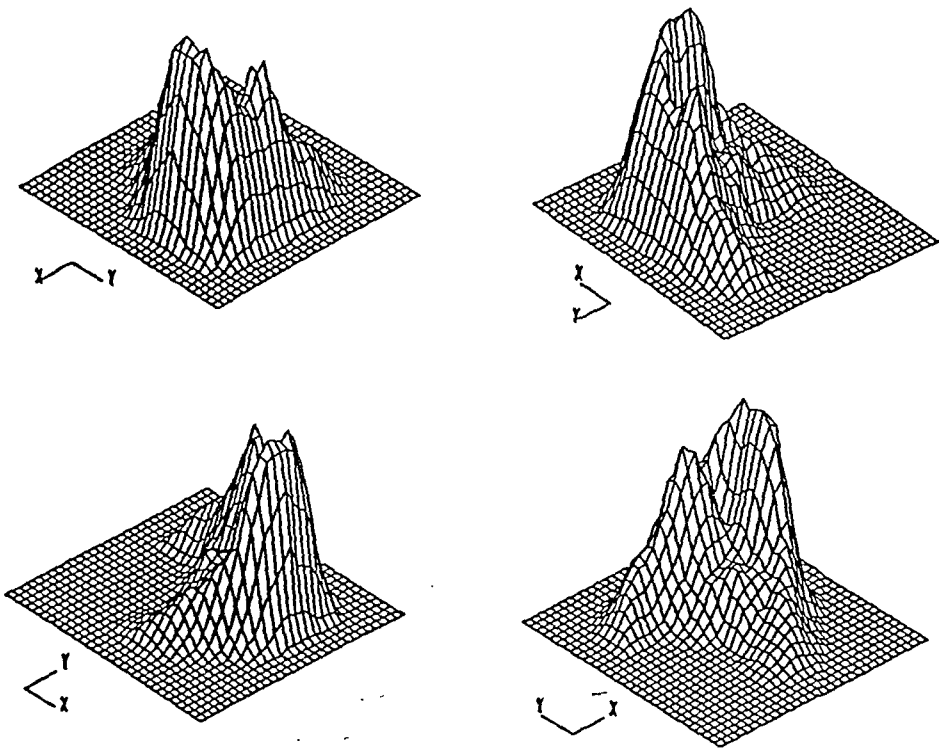
LOG TRANSFORMATION OF BLOWFLY DATA X←X[I] Y←X[I+ 5]



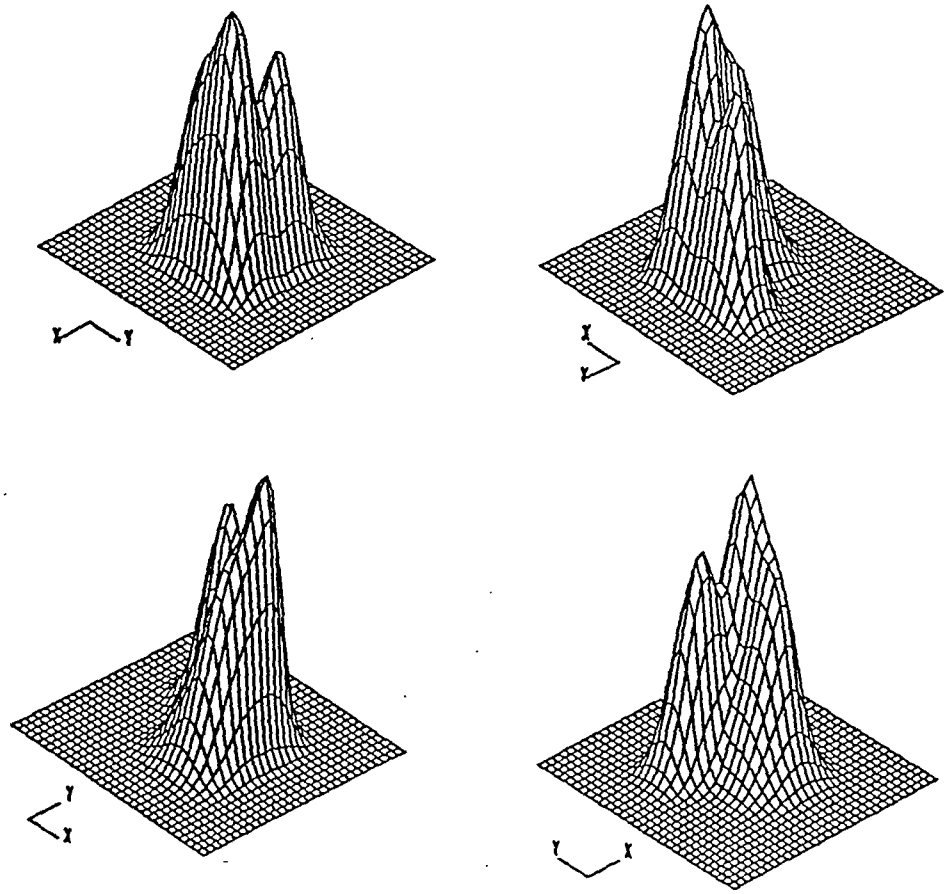
SIMULATED DATA OF FITTED MODEL X←X[I] Y←X[I+ 5]



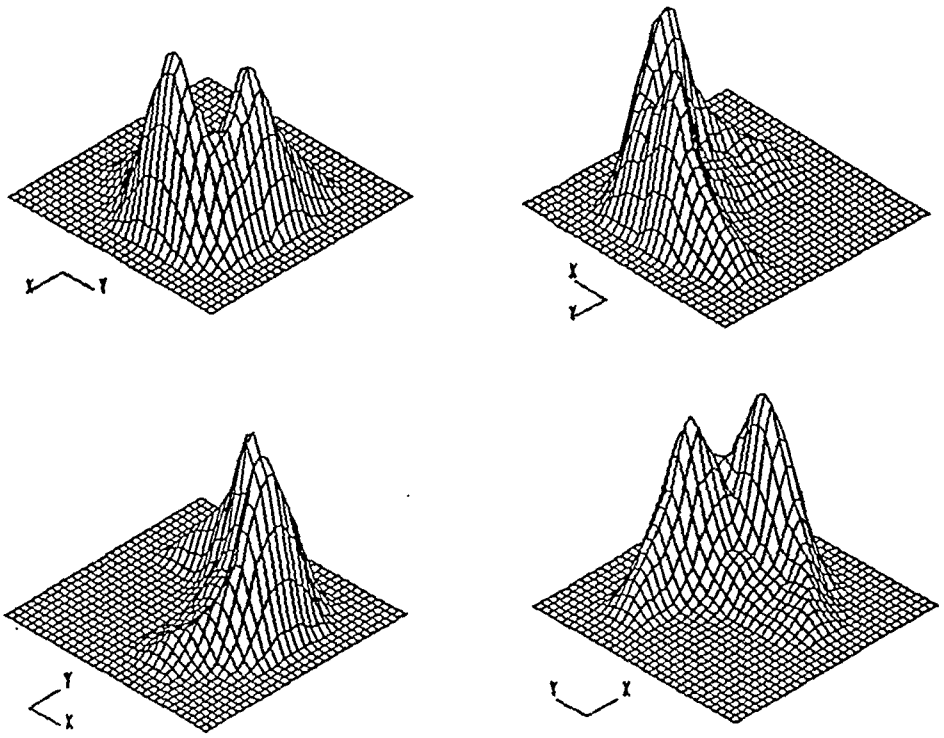
LOG TRANSFORMATION OF BLOWFLY DATA X=X(1) Y=X(1+ 6)



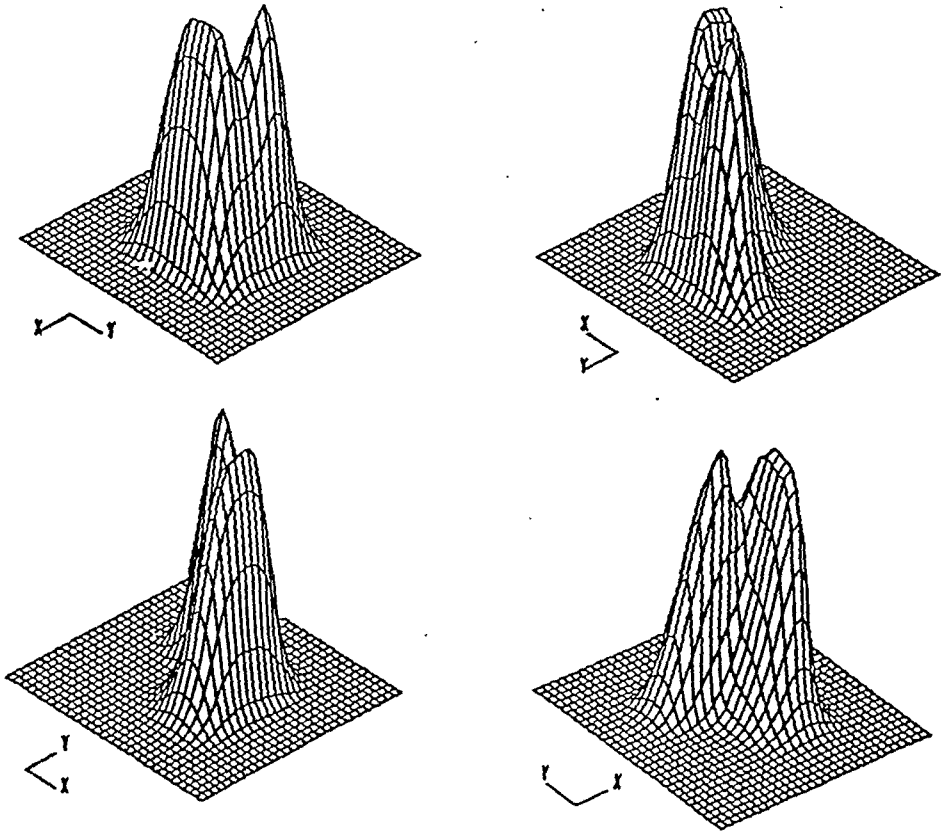
SIMULATED DATA OF FITTED MODEL X=X(1) Y=X(1+ 6)



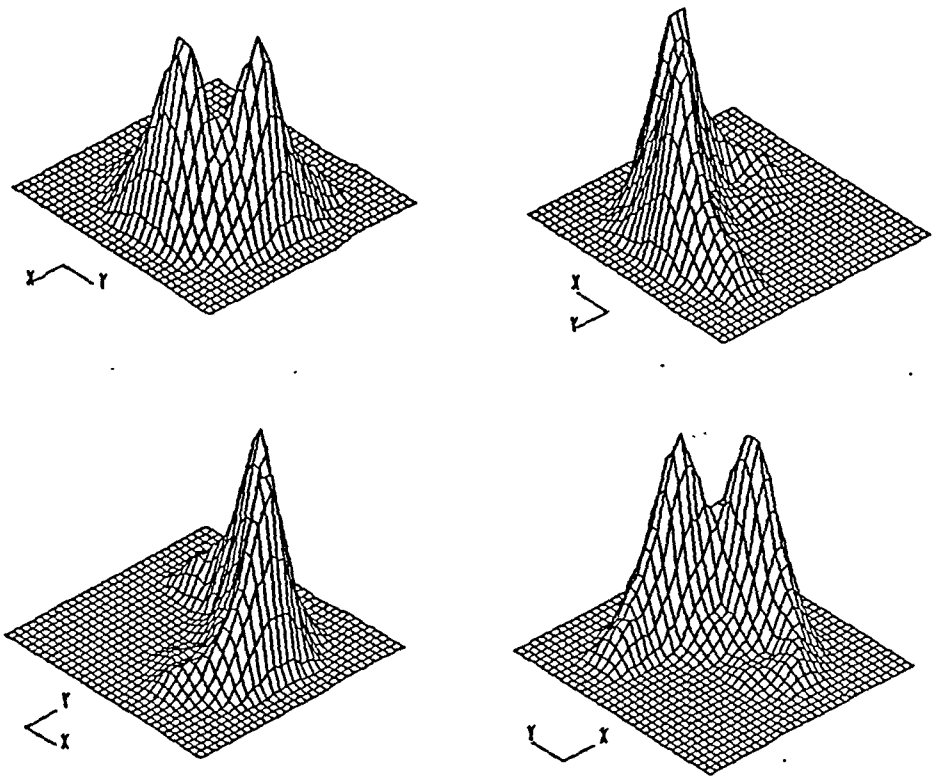
LOG TRANSFORMATION OF BLOWFLY DATA X=X(1) Y=X(1+ 7)



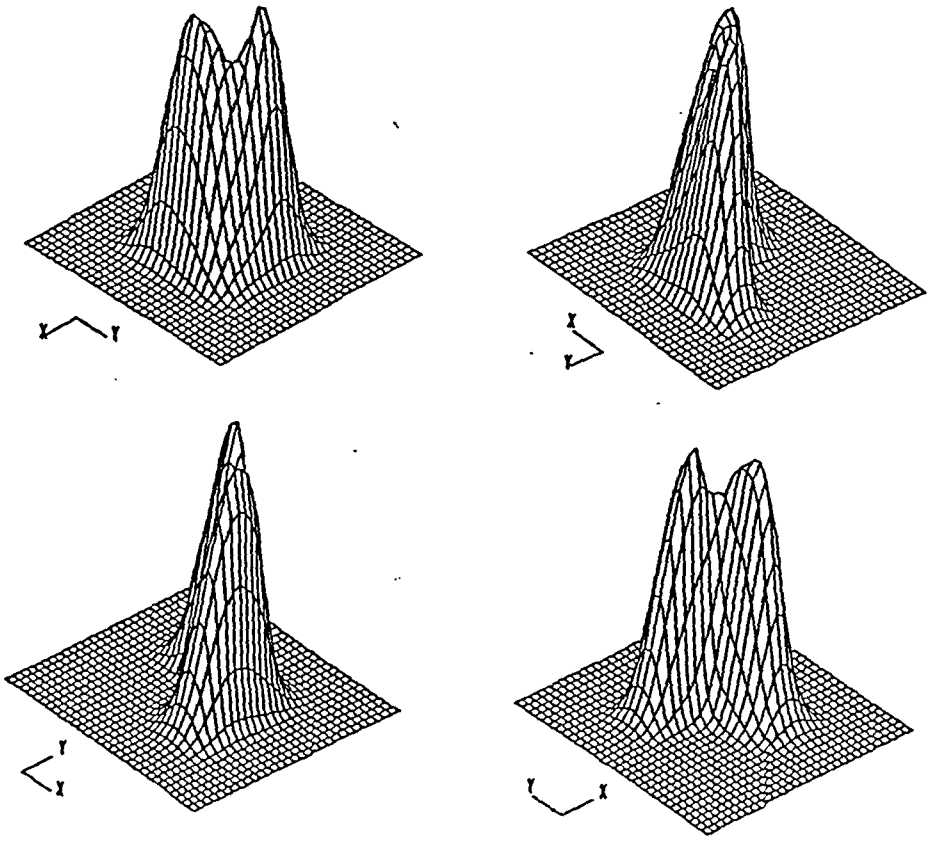
SIMULATED DATA OF FITTED MODEL X=X(1) Y=X(1+ 7)



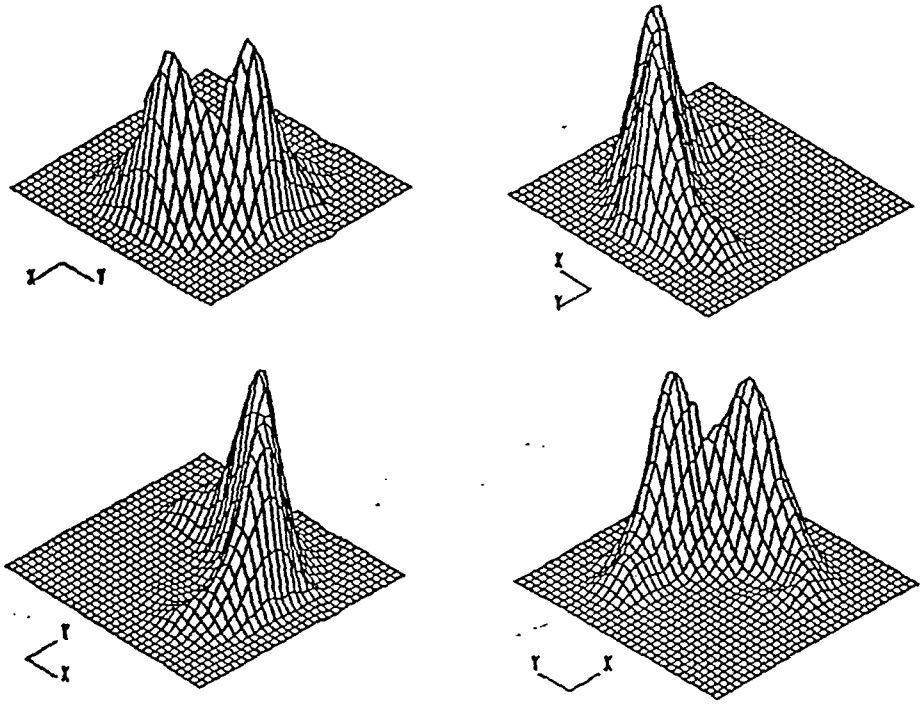
LOG TRANSFORMATION OF BLOWFLY DATA X=X(I) Y=X(I+ 6)



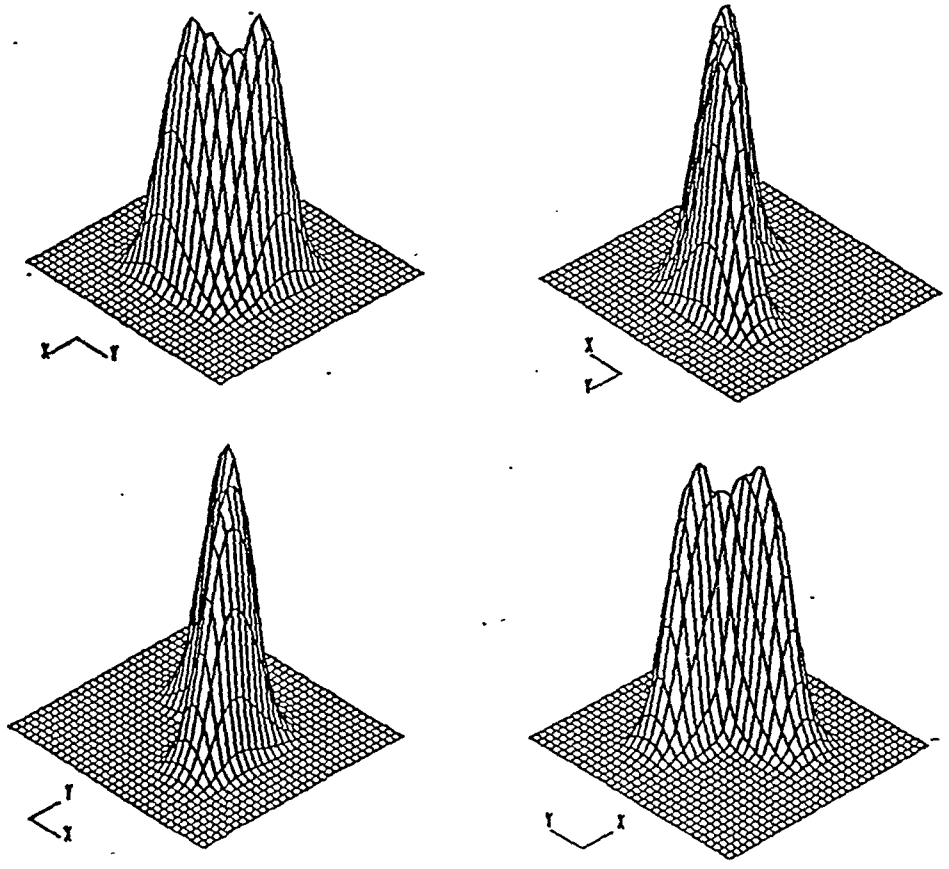
SIMULATED DATA OF FITTED MODEL X=X(I) Y=X(I+ 6)



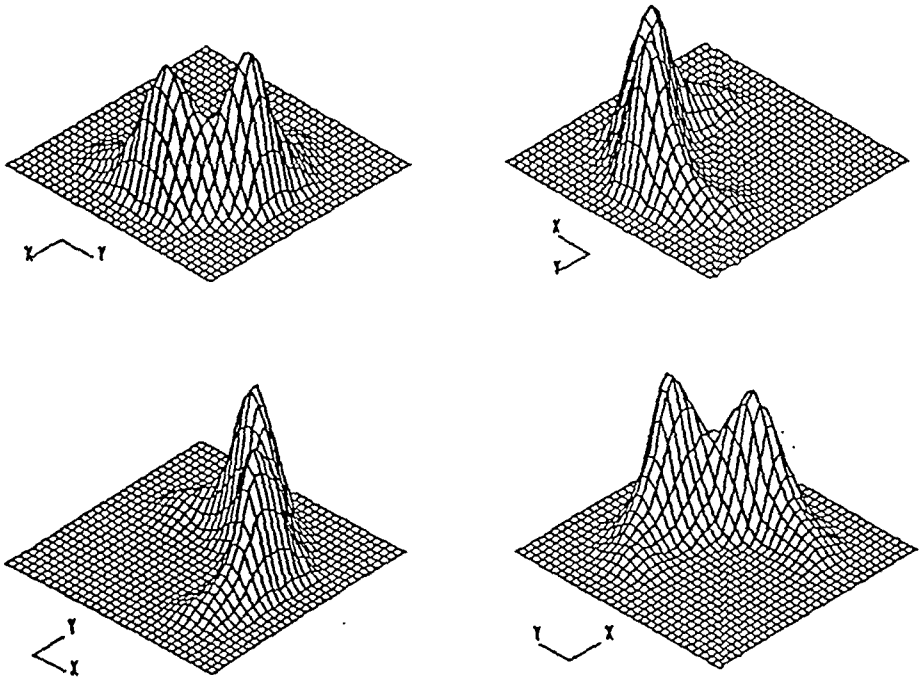
LOG TRANSFORMATION OF BLOWFLY DATA X=I[1] Y=I[1: 9]



SIMULATED DATA OF FITTED MODEL X=I[1] Y=I[1: 9]



LOG TRANSFORMATION OF BLOWFLY DATA X=X(1) Y=X(1+ 10)



SIMULATED DATA OF FITTED MODEL X=X(1) Y=X(1+ 10)

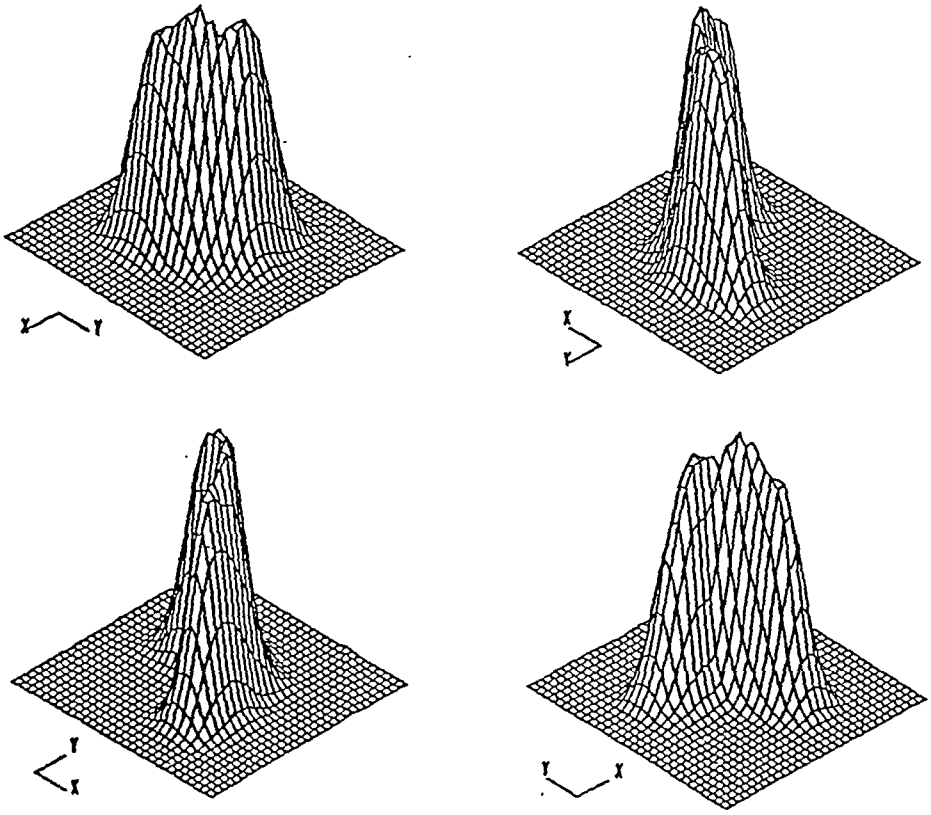


Fig. 8.7.5

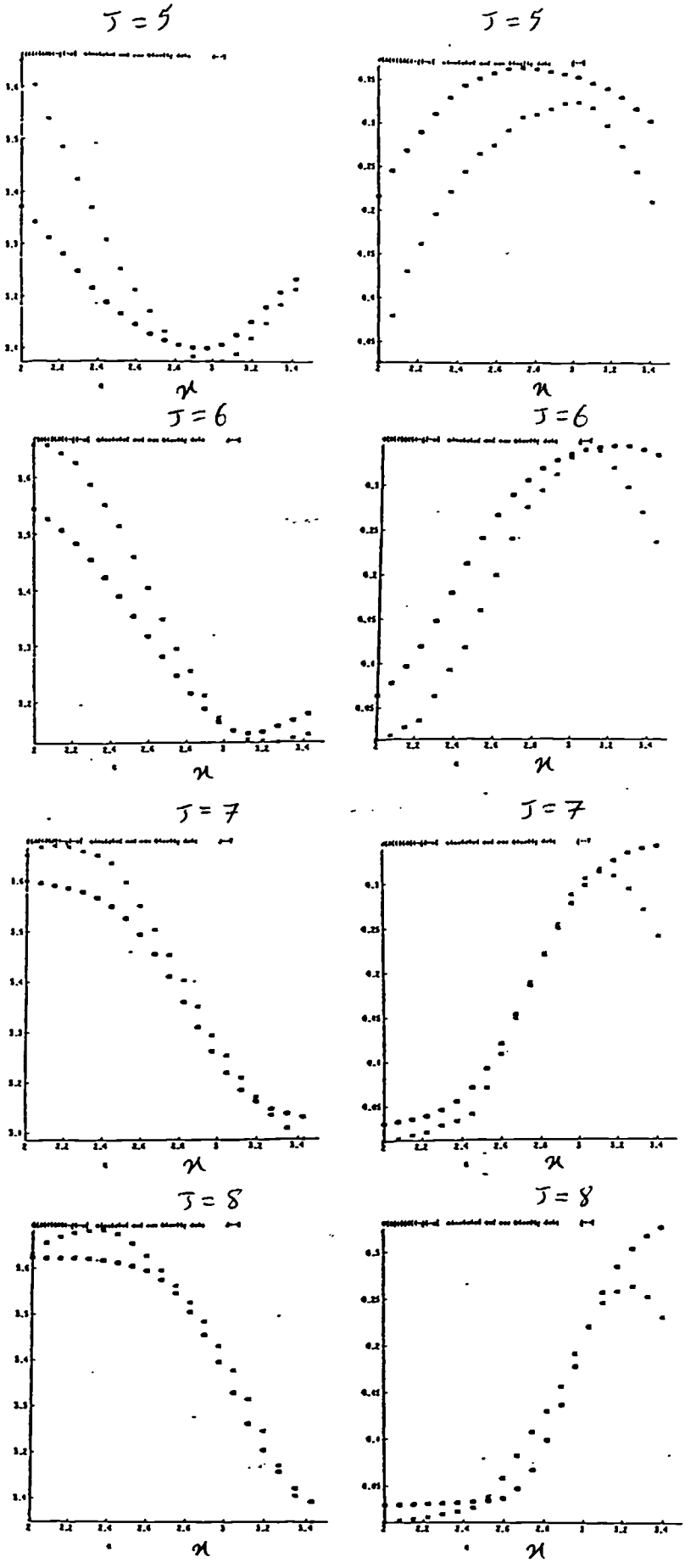


Fig. 8.7.6 : Non-parametric estimation of $E(X_t \mid X_{t-j}=x)$ and $V(X_t \mid X_{t-j}=x)$, $j=1,\ldots,8$ of the data and simulated data.

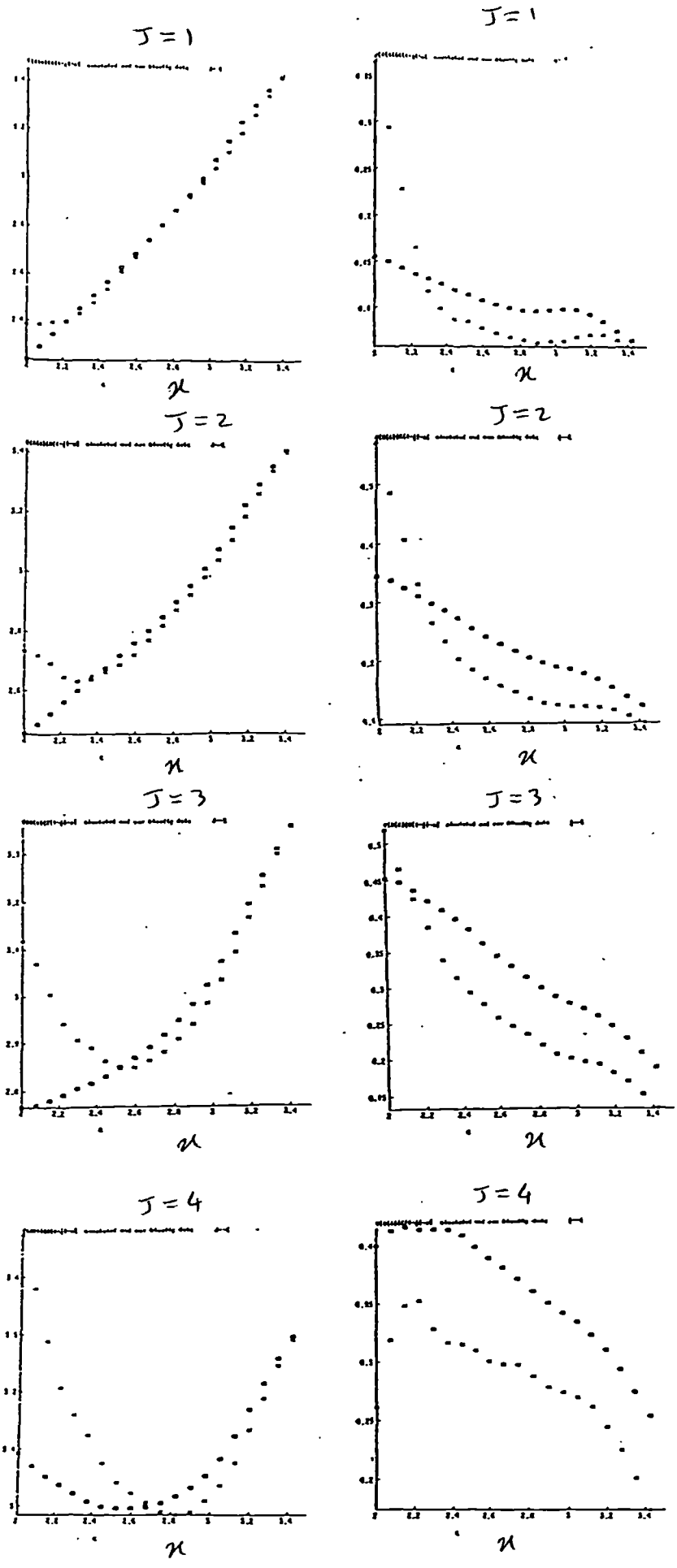


Fig. 8.7.6 : Non-parametric estimation of $E(X_t \mid X_{t-j}=x)$ and $V(X_t \mid X_{t-j}=x)$, $j=1, \ldots, 8$ of the data and simulated data.

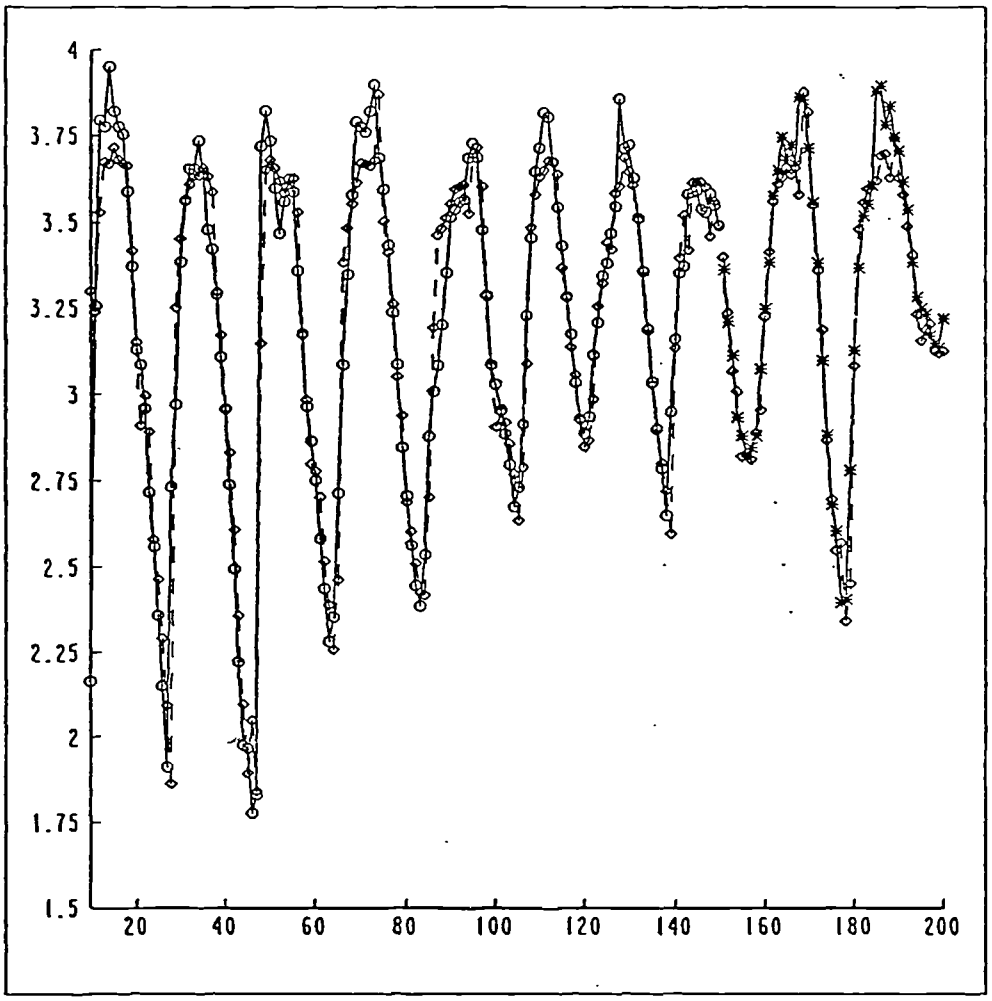


Fig. 8.7.8 : The data (□), fitted values (o) and the (*) are the one-step ahead prediction of model (8.4.3).

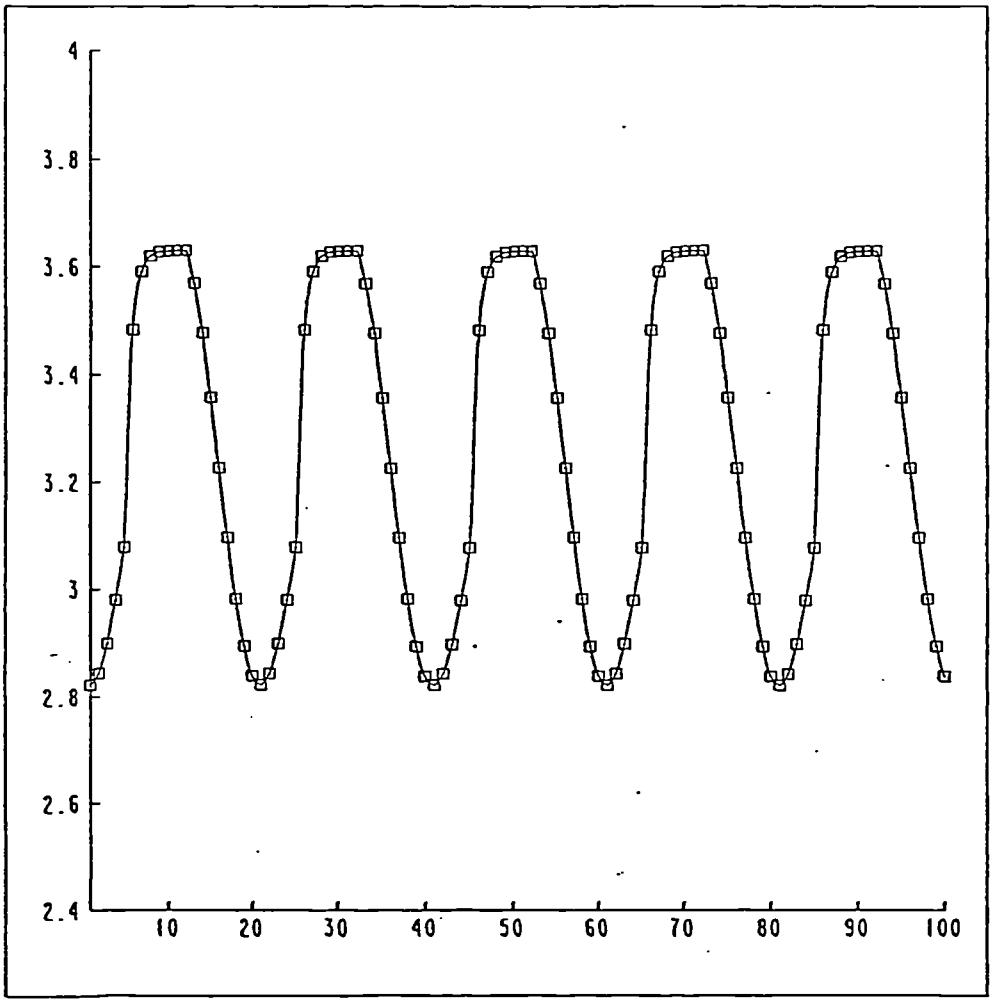
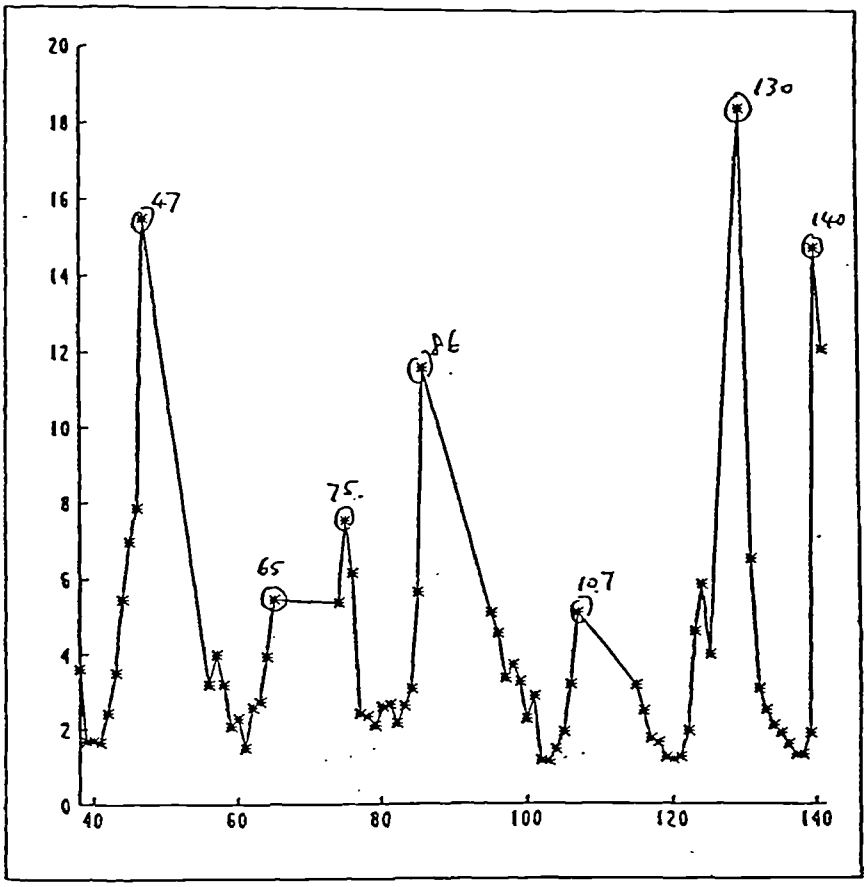


Fig. 8.7.9 : Systematic part of model (8.4.3).



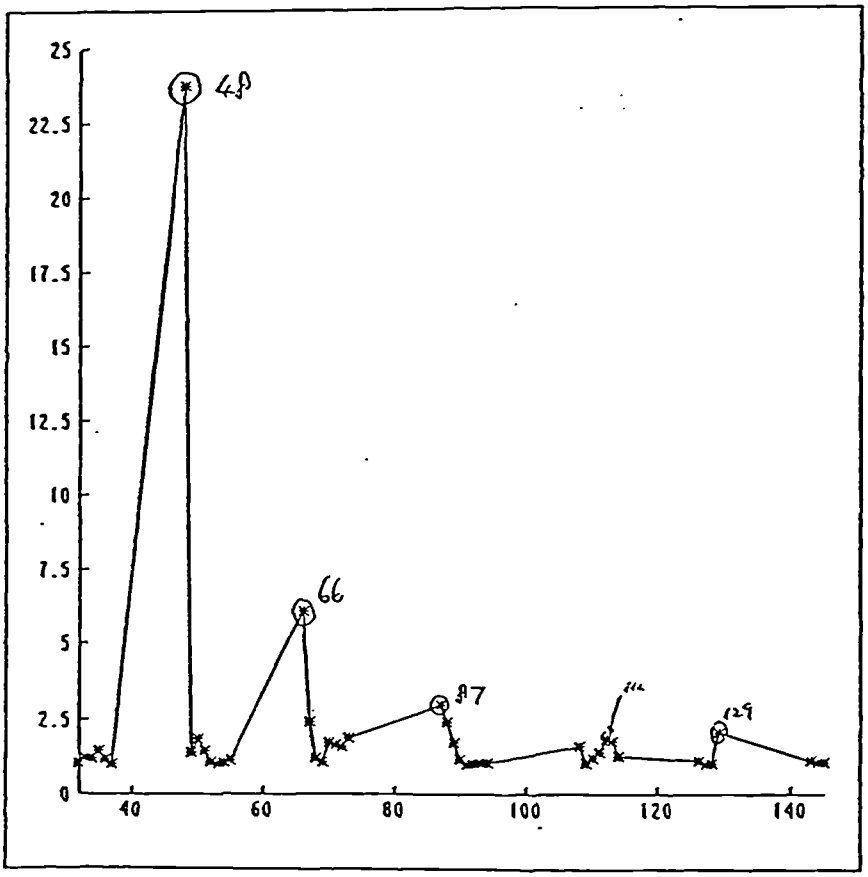


Fig. 8.7.10 : The elements of hat matrixes for first and second regimes.

CHAPTER NINE

MULTI-STEP NON-LINEAR LEAST SQUARES PREDICTION

9.1. Introduction

Let $\{X_t\}$ be some discrete time stationary stochastic process, which for the time being will be assumed to be stationary. Suppose that we are at time t ($\equiv now$) and we wish to forecast † m units ahead to time $t+m$ ($m \equiv hence$), so that attention is directed to the random variable X_{t+m}. Any forecasting procedure will have to be based on some information set, e.g. $B_t^{(X)}$ the σ-algebra generated by $\{X_s\ ;\ s \le t\}$.

Since the variable to be forecast X_{t+m} is a random variable, it can be fully characterized only in terms of a probability density function or some equivalent function. However, since the information set is to be utilized we need to use a *conditional* density function, i.e.

$$Prob(x < X_{t+m} \le x+dx \,|\, \text{Information set}) = g_c(x)dx$$

where the subscript c denotes "conditional". If $g_c(x)$ were available, then all other properties of X_{t+m}, such as the conditional mean $E_c(X_{t+m})$, could be immediately determined. However, in practice, it is generally rather too ambitious to hope to be able to characterize fully X_{t+m}, and so we attempt the less ambitious task of finding some confidence band for X_{t+m} or some single value, called a *point forecast*, that in some way "best" represents the random variable X_{t+m}.

To obtain any kind of best value for a point forecast, one requires a criterion against which various alternatives can be judged. An intellectually satisfying way to proceed is to introduce the idea of a *cost function*. Since forecast errors are virtually certain to occur in connection with a random process, suppose that we can cost the effect of an error of size e to be $C(e)$, with $C(0)=0$. One particular cost function

† The words forecast and prediction will be used interchangeably.

gives a rather tidy and easily used solution, and that is the function

$$C(e) = ae^2$$

where a is some positive constant.

Suppose we wish to form an optimal point forecast $f_{t,m}$ (the point forecast of X_{t+m} based on the information set at time t) of X_{t+m} using the information set and a cost function $C(e)$.

The cost function will be assumed to have the properties that $C(0) = 0$, $C(e)$ is monotonic nondecreasing for $e > 0$ and monotonic nondecreasing for $e < 0$. Let $g_{c,m}$ be the conditional probability density function of X_{t+m} given X_s , $s \le t$, then the required optimal forecast $f_{t,m}$, which will be a function of only X_s , $s \le t$, will be found by choosing $f_{t,m}$ so that the expected cost is a minimum, i.e. minimizing

$$J = \int_{-\infty}^{\infty} C(x - f_{t,m}) g_{c,m}(x)\, dx \, .$$

In the case when $C(e) = ae^2$ with $a > 0$, J becomes

$$J = \int_{-\infty}^{\infty} a(x - f_{t,m})^2 g_{c,m}(x)\, dx \, .$$

Define $\tilde{X}_t(m)$ to be the conditional mean of X_{t+m} given X_s , $s \le t$, so that $\tilde{X}_t(m) = \int_{-\infty}^{\infty} x g_{c,m}(x) dx$, then it can be shown that

$$J = a(\tilde{X}_t(m) - f_{t,m})^2 + a\int_{-\infty}^{\infty} (x - \tilde{X}_t(m))^2 g_{c,m}(x)\, dx \, .$$

So J is minimized by taking (Granger, C.W.J. and Newbold, 1986)

$$f_{t,m} = E_c\{X_{t+m} \mid X_s \, , \, s \le t\}.$$

In the Gaussian situation, where every finite subset of variables is normally distributed, this leads to the important result that the optimal least-squares prediction of X_{t+m} is a *linear* sum of the terms in $\{X_s \, , \, s \le t\}$. It should also be noted that in this case $g_{c,m}(x)$ is a normal distribution with mean $\tilde{X}_t(m)$ and variance that does not depend on $\{X_s \, , \, s \le t\}$.

Let $\{X_t : t = 0, \pm 1, \pm 2, \cdots\}$ be a second order stationary time series with zero mean and finite variance which admits a strict linear representation,

$$X_t = \sum_{j=0}^{\infty} a_j e_{t-j} \, , \qquad (9.1.1)$$

where the e_t's are independent and identically distributed random variables with

$$E(e_t)=0\,,\ Var(e_t)=\sigma_e^2<\infty\,,\ \forall t\,.$$

Thus

$$E(X_t)=0\ \text{ and }\ Var(X_t)=\sigma_e^2\sum_{j=0}^{\infty}a_j^2<\infty\ .$$

Moreover, the linear representation is assumed to be invertible in the usual sense.

The problem is to find a linear forecast for X_{t+m} of the form

$$f_{t,m}=\sum_{j=0}^{\infty}w_{j,m}X_{t-j}$$

using a least-squares criterion. Thus, the $w_{j,m}$ need to be chosen so that

$$J=E\left\{(X_{t+m}-\sum_{j=0}^{\infty}w_{j,m}X_{t-j})^2\right\}\,,$$

is minimized. Since $f_{t,m}$ is a linear function of X_t, therefore it will also be a linear function of current and past e_t. Thus the forecast $f_{t,m}=\sum_{j=0}^{\infty}\phi_{j,m}e_{t-j}$, is expressed in terms of the white noise e_t.

By letting t take the value $t+m$, it is seen that J may be expressed as

$$J=E\left\{\left(\sum_{j=0}^{m-1}a_je_{t+m-j}+\sum_{j=0}^{\infty}(a_{j+m}-\phi_{j,m})e_{t-j}\right)^2\right\}\,,$$

which gives

$$J=\sigma_e^2\sum_{j=0}^{m-1}a_j^2+\sigma_e^2\sum_{j=0}^{\infty}(a_{j+m}-\phi_{j,m})^2\,.$$

It is immediately clear that J is minimized by taking

$$\phi_{j,m}=a_{j+m}\,,$$

so that the optimal m-step ahead forecast may then be written as

$$f_{t,m}=\sum_{j=0}^{\infty}a_{j+m}e_{t-j}\,,$$

and the m-step ahead error will be

$$e_{t,m}=X_{t+m}-f_{t,m}=\sum_{j=0}^{m-1}a_je_{t+m-j}\,.$$

Simply it follows that the errors have mean zero and variances given by

$$Var(e_{t,m})=E(e_{t,m}^2)=\sigma_e^2\sum_{j=0}^{m-1}a_j^2\,,$$

which

$$\lim_{m \to \infty} Var(e_{t,m}) = Var(X_t) = \sigma_e^2 \sum_{j=0}^{\infty} a_j^2 < \infty .$$

The sequence of error variances is seen to be monotonically nondecreasing. It thus follows that, when using optimal linear forecasts, the further ahead one forecasts the worse one does, on average.

9.2. Non-monotonicity

As we have shown that for linear forecasting the conditional variance of $e_{t,m}$ given X_s , $s \le t$, is a monotonic nondecreasing function of m. Results of this type seem to have generated the popular belief that the further ahead we forecast the less "reliable" is the forecast. However, as first pointed out in our abstract in the January 1987 issue of the IMS *Bulletin* (Tong and Moeanaddin, 1987a) and in a concurrent technical report (Tong and Moeanaddin, 1987b) and in Tong and Moeanaddin (1988), *the belief could be misleading* with respect to *non-linear* least-squares predictors.

To demonstrate the point, let us consider first a simple strictly stationary bilinear time series model of the type which has been studied extensively by Quinn (1982),

$$X_t = \alpha e_{t-1} X_{t-2} + e_t \ , \ t = 0 \ , \pm 1 \ , \ \cdots \ , \tag{9.2.1}$$

where $\{e_t\}$ is a sequence of independent and identically distributed random variables with $E(e_t) = 0$, and $Var(e_t) = \sigma_e^2 < \infty$, and $\sigma_e^2 \alpha^2 < \frac{1}{2}$. Here, e_t is independent of X_s , $s < t$. Clearly $0 < Var(X_t) < \infty$, all t, and let $\hat{X}_t(m)$ denote the conditional expectation and therefore the least-squares predictor of X_{t+m} given X_s , $s \le t$. To evaluate $Var(X_{t+m} - \hat{X}_t(m))$ conditional on X_s , $s \le t$, we only need to evaluate $Var(X_{t+m})$ conditional on X_s , $s \le t$. Let this be denoted by $\hat{\sigma}_t^2(m)$. Simple calculation gives

$$\begin{aligned}
\hat{\sigma}_t^2(1) &= Var(X_{t+1} \,|\, X_s \ , \ s \le t) \\
&= Var(\alpha e_t X_{t-1} + e_{t+1} \,|\, X_s \ , \ s \le t) \\
&= Var(\alpha e_t x_{t-1} + e_{t+1} \,|\, X_s \ , \ s \le t) \\
&= Var(e_{t+1}) \\
&= \sigma_e^2 \ . \\
\hat{\sigma}_t^2(2) &= Var(X_{t+2} \,|\, X_s \ , \ s \le t) \\
&= Var(\alpha e_{t+1} X_t + e_{t+2} \,|\, X_s \ , \ s \le t) \\
&= Var(e_{t+2}) + \alpha^2 x_t^2 Var(e_{t+1})
\end{aligned}$$

$$= \sigma_e^2 + \alpha^2 \sigma_e^2 x_t^2 .$$

$$\hat{\sigma}_t^2(3) = Var(X_{t+3} \mid X_s , s \le t)$$

$$= Var(\alpha^2 e_{t+2} X_{t+1} + \alpha e_{t+1} e_{t+2} + e_{t+3} \mid X_s , s \le t)$$

$$= \sigma_e^2 + \alpha^2 \sigma_e^4 + \alpha^4 \sigma_e^2 e_t^2 x_{t-1}^2$$

etc.

[Notice that $\alpha^2 Var(e_t) < \frac{1}{2}$ ensures that e_t is measurable with respect to the σ-field generated by $\{X_t, X_{t-1}, \cdots\}$.] It is therefore clear that, as expected,

$$\hat{\sigma}_t^2(1) \ge \hat{\sigma}_t^2(2) .$$

However,

$$\hat{\sigma}_t^2(2) \ge \hat{\sigma}_t^2(3) \text{ if and only if } x_t^2 \ge \sigma_e^2 + \alpha^2 e_t^2 x_{t-1}^2 . \qquad (9.2.2)$$

Inequality (9.2.2) is an event with non-zero probability and has the interpretation that 3-step ahead prediction need not be less `reliable' than 2-step ahead prediction. In hindsight, inequality (9.2.2) simply reflects the sometimes neglected fact that `prediction accuracy' is also influenced by where we are currently (i.e. at time t) . This last observation is particularly relevant when we are dealing with cyclical data, to which we turn our attention next. We shall also demonstrate that the nonmonotonicity just described is not pathological.

9.3. Non-linear Models

Almost all of the specific non-linear models give rise to fairly straightforward one-step ahead least-squares predictions. For more than *one-step ahead* least-squares predictions, the situation becomes radically different.

9.3.1. Bilinear Models

As far as bilinear models are concerned, short-to-medium-term (say up to five-step ahead) forecast is manageable without straining the algebra too much and presumably with due regard to rounding errors, recalling the multiplicative nature of the models. It should be noted that linear / non-linear models (e.g. subset bilinear model) are used more often than full model in the time series literature. However, the results of Copas (1983) on this kind of practice suggest its potential complication. Subba Rao and Gabr (1984) have

given the method of calculating the optimal m-step ahead ($E(X_{t+m}|X_t,X_{t-1},\cdots)=\hat{X}_t(m)$) prediction of the general bilinear model of the form (3.2.1).

In this section, very briefly, we will consider the multi-step ahead least-squares prediction of some sub-classes of bilinear model.

9.3.1.1. Superdiagonal Models

The first class to analyze are the super diagonal completely bilinear models, the general form of which is

$$X_t=\sum_{\substack{k=1\\k<l}}^{Q}\sum_{l=2}^{P}\beta_{kl}e_{t-k}X_{t-l}+e_t\,,$$

where e_t is strict white noise, and a special case is

$$X_t=\beta e_{t-k}X_{t-l}+e_t\,,\ k<l\,, \tag{9.3.1.1.1}$$

which will be used here as a particular example. Assuming $e_t\sim N(0,\sigma_e^2)$, it follows from (9.3.1.1.1) that $E(X_t)=0$ and $Var(X_t)=\dfrac{\sigma_e^2}{1-\beta^2\sigma_e^2}$ provided $\beta^2\sigma_e^2<1$. Suppose that the model (9.3.1.1.1) is invertible, therefore knowledge of $\{X_1,\cdots,X_N\}$ implies the knowledge of $\{e_1,\cdots,e_N\}$.

Now

$$X_{t+m}=\beta e_{t+m-k}X_{t+m-l}\,,\ k,l\geq 1\,.$$

Then the m-step ahead prediction of this model is given by term

$$\hat{X}_t(m)=\begin{cases}\beta\varepsilon_{t+m-k}x_{t+m-l} & \text{if } m\leq k\\ 0 & \text{if } m>k\,,\end{cases}$$

where x_t and ε_t are the observed values of the random variables X_t and e_t. Let $V_t(.)$ denote the $Var(.|X_s\,,\ s\leq t)$, then

$$V_t(\hat{e}_t(m))=V_t(X_{t+m})=\begin{cases}\sigma_e^2 & \text{if } m\leq l\\ \beta^2\sigma_e^2x_{t+m-l}^2+\sigma_e^2 & \text{if } k<m\leq l\,.\end{cases}$$

For $m>l$ it can be shown that

$$V_t(X_{t+m})=\beta^2\sigma_e^2V_t(X_{t+m-l})+\sigma_e^2\,,$$

and as $m\to\infty$, we have

$$V_t(X_{t+m})=\sigma_e^2+\beta^2\sigma_e^4+\beta^4\sigma_e^6+\cdots$$

$$= \sigma_e^2+\beta^2\sigma_e^4(1+\beta^2\sigma_e^2+\beta^4\sigma_e^4+\cdots)$$
$$= \sigma_e^2+\beta^2\sigma_e^4\frac{1}{1-\beta^2\sigma_e^2}$$
$$= \frac{\sigma_e^2}{1-\beta^2\sigma_e^2}$$
$$= Var(X_t) .$$

9.3.1.2. Diagonal Models

The second group of models are diagonal models. The following special model will be considered here.

$$X_t = \beta X_{t-k}e_{t-k}+e_t , \qquad (9.3.1.2.1)$$

where it is assumed that the model is invertible and $e_t \sim N(0, \sigma_e^2)$. From equation (9.3.1.2.1) it follows that $E(X_t) = \beta\sigma_e^2$ and

$$Var(X_t) = \frac{1+\beta^2\sigma_e^2+\beta^4\sigma_e^4}{1-\beta^2\sigma_e^2} , \; 1-\beta^2\sigma_e^2 > 0 ,$$

which is a necessary and sufficient condition for second order stationarity. Now

$$X_{t+m} = \beta X_{t+m-k}e_{t+m-k}+e_{t+m} ,$$
$$\hat{X}_t(m) = E(X_{t+m} \mid X_t, X_{t-1}, \cdots)$$
$$= \begin{cases} \beta x_{t+m-k}\varepsilon_{t+m-k} & \text{if } m \le k \\ \beta\sigma_e^2 & \text{if } m > k , \end{cases}$$

and

$$V_t(X_{t+m}) = Var(X_{t+m} \mid X_t, X_{t-1}, \ldots)$$
$$= \beta^2 V_t(X_{t+m-k}e_{t+m-k})+\sigma_e^2$$
$$= \sigma_e^2 \quad \text{if } m \le k .$$

For $m > k$ we have

$$V_t(X_{t+m}) = \beta^2 V_t(X_{t+m-k}e_{t+m-k})+\sigma_e^2$$
$$= \beta^2\left[E_t(X_{t+m-k}^2 e_{t+m-k}^2)-\sigma_e^4\right]+\sigma_e^2$$
$$= \beta^4\sigma_e^2 E_t(X_{t+m-2k}e_{t+m-2k})+3\beta^2\sigma_e^4-\beta^2\sigma_e^4+\sigma_e^2$$
$$= \beta^6\sigma_e^6 E_t(X_{t+m-3k}^2 e_{t+m-3k}^2)+3\beta^4\sigma_e^6-\beta^2\sigma_e^4+\sigma_e^2$$
$$\vdots$$
$$= 3\beta^2\sigma_e^4(1+\beta^2\sigma_e^2+\beta^4\sigma_e^4+\cdots)-\beta^2\sigma_e^4+\sigma_e^2$$
$$= 3\beta^2\sigma_e^4\frac{1}{1-\beta^2\sigma_e^2}-\beta^2\sigma_e^2+\sigma_e^2$$

$$= \frac{1+\beta^2\sigma_e^2+\beta^4\sigma_e^4}{1-\beta^2\sigma_e^2}$$

$$= Var(X_t) \ \ as \ m \to \infty .$$

9.3.1.3. Subdiagonal Models

The third group of models which will be considered here are those of the form

$$X_t = \beta X_{t-k} e_{t-k-j} + e_t \ , \ k,j \geq 1 \ , \qquad (9.3.1.3.1)$$

where it is assumed that the models are invertible and $e_t \sim N(0,\sigma_e^2)$. From (9.3.1.3.1) it can be shown that

$$E(X_t) = 0 \ \text{ and } \ Var(X_t) = \frac{\sigma_e^2}{1-\beta^2\sigma_e^2} \ , \text{ with } \beta^2\sigma_e^2 < 1 \ .$$

Now we can show that

$$\hat{X}(m) = E_t(X_{t+m}) = E(X_{t+m} \mid X_t, X_{t-1}, \cdots)$$
$$= \begin{cases} \beta x_{t+m-k}\varepsilon_{t+m-k-j} & \text{if } m \leq k+j \\ \beta\varepsilon_{t+m-k-j}\hat{X}(m-k) & \text{if } k+j < m \leq k \\ 0 & \text{if } m > k \ , \end{cases}$$

and

$$V_t(X_{t+m}) = Var(X_{t+m} \mid X_t, X_{t-1}, \cdots)$$
$$= V_t(\beta X_{t+m-k} e_{t+m-k-j} + e_{t+m})$$
$$= \sigma_e^2 \ \text{ if } m \leq k$$
$$= \beta^2\varepsilon_{t+m-k-j}^2 V_t(X_{t+m-k}) + \sigma_e^2 \ \text{ if } k < m \leq k+j \ ,$$

and for $m > k+j$ we have

$$V_t(X_{t+m}) = \beta^2 V_t(X_{t+m-k} e_{t+m-k-j}) + \sigma_e^2$$
$$= \beta^2 E_t(X_{t+m-k}^2 e_{t+m-k-j}^2) + \sigma_e^2$$
$$= \beta^4 E_t(X_{t+m-2k}^2 e_{t+m-2k-j}^2) + \beta^2\sigma_e^4 + \sigma_e^2$$
$$\vdots$$
$$= \beta^{2n} E_t(X_{t+m-nk}^2 e_{t+m-nk-j}^2 \cdots) + R$$

where

$$R = \beta^{2(n-1)}\sigma_e^{2n} + \cdots + \beta^2\sigma_e^4 + \sigma_e^2 = \left(\frac{1-\beta^{2n}\sigma_e^{2n}}{1-\beta^2\sigma_e^2}\right)\sigma_e^2 \ .$$

Therefore, as $m \to \infty$,

$$V_t(X_{t+m}) \to \frac{\sigma_e^2}{1-\beta^2\sigma_e^2} = Var(X_t) \ .$$

9.3.2. Non-linear Autoregression Models

The class of non-linear autoregression (NLAR) models for discrete time series has been studied extensively for the past ten years. It has, however, apparently been thought that optimal least-squares forecasts for more than one-step ahead are difficult to obtain. In this section we will show that a simple recurrence relation exists from which these forecasts can be calculated.

Without loss of generality, we restrict our discussion to the first order case:

$$X_t = \lambda(X_{t-1}) + e_t \ , \ t = 0 \ , \pm 1 \ , \pm 2 \ , \ldots \ , \tag{9.3.2.1}$$

where $\{e_t\}$ is a sequence of independent and identically distributed random variables with $E(e_t) = 0$ and $Var(e_t) = \sigma_e^2 < \infty$, all t. Since we have a Markov chain over R, we may recall the *Chapman-Kolmogorov relation*

$$f(x_{t+m} \,|\, x_t) = \int_{-\infty}^{\infty} f(x_{t+m} \,|\, x_{t+1}) \, f(x_{t+1} \,|\, x_t) \, dx_{t+1} \ , \tag{9.3.2.2}$$

where $f(x_s \,|\, x_t)$ denotes the conditional probability density function of X_s given $X_t = x_t$ (assumed to exist). Suppose that model (9.3.2.1) is strictly stationary. (Sufficient conditions for strict stationarity are available; see e.g. Tong 1987a.) Let g denote the probability density function of e_t. Let k denote a well-behaved (i.e. Baire) function of X_t and suppose $E(|k(X_t)|) < \infty$. Let $K_m(X_t)$ denote the conditional expectation $E(k(X_{t+m}) \,|\, X_t)$. Equation (9.3.2.2) gives immediately

$$K_m(x_t) = \int_{-\infty}^{\infty} K_{m-1}(x_{t+1}) \, f(x_{t+1} \,|\, x_t) \, dx_{t+1} \ , \tag{9.3.2.3}$$

or

$$K_m(x) = \int_{-\infty}^{\infty} K_{m-1}(y) \, g(y - \lambda(x)) \, dy \ , \tag{9.3.2.4}$$

where $K_1(X_t) = E(k(X_{t+1} \,|\, X_t = x_t) = \lambda(x_t)$. Equation (9.3.2.4) gives, in particular, recursive formulae for conditional expectations and conditional variances. These are the type recently discussed by, e.g Al-Qassem and Lane (1987) and Pemberton (1987) as alternative to the methods developed by Jones (1976). Apparently, without realizing it, Al-Qassem & Lane (1987) and Pemberton (1987) have actually implicitly used the Chapman-Kolmogorov relation, except that they have used it in two different ways. Except for special cases of λ (e.g. the linear case), the integral in equation (9.3.2.4) does not readily admit analytic solution and numerical integration is commonly the only solution. Needless to say, if numerical integration

is employed, care must be taken to handle accumulation of rounding errors. An alternative method which is especially useful for higher order autoregression models is the Monte Carlo method. Here, a sufficiently long record of data is simulated in accordance with the model (i.e. equation (9.3.2.1) or its higher order generalization and the given values $X_t, X_{t-1}, \cdots$, as the starting point) and the sample estimate of $K_m(x)$ is taken as an approximation of $K_m(x)$. Details of comparison will be given later. Suffice it to say that both methods give reasonable approximations.

It is clear that $K_\infty(x)\equiv$constant is one solution of (9.3.2.4) for the case $m=\infty$. In fact, this is the only solution within our context, which can be shown by appealing to results in the martingale theory (Russel Gerrard, private communication). Thus, in our case, the remote past plays no material role in prediction.

For the general non-linear autoregressive model of order k (*NLAR* (k)), of the form

$$X_t = \lambda(\mathbf{X}_{t-1}) + e_t ,$$

where $\mathbf{X}_t = (X_t, X_{t-1}, \cdots, X_{t-k+1})^T \in R^k$ and e_t a sequence of zero mean, independent and identically distributed random variables, and $\lambda : R^k \to R$, the following recursive integral equation (using the Chapman-Kolmogorov relation) can be obtained

$$K_m(\mathbf{x}) = \int_{-\infty}^{\infty} K_{m-1}(\mathbf{x}_{k+1})\, g(x_{k+1} - \lambda(\mathbf{x}))\, dx_{k+1} , \qquad (9.3.2.5)$$

where $K_1(\mathbf{x}) = \lambda(\mathbf{x})$ and $\mathbf{x} = (x_k, x_{k-1}, \cdots, x_1)^T$.

Note that for *NLAR* (k), the prediction functions are of course functions of k variables and that calculation of $K_m(\mathbf{x})$ requires knowledge of $K_{m-1}(\mathbf{x}_{k+1})$ for a range of $\mathbf{x}_{k+1} \in \mathbf{R}^k$ and not, as it may appear from (9.3.2.5), only for a range of the first component of $\mathbf{x}_{k+1}$. Thus the problem is one of extensive numerical evaluation over $\mathbf{R}^k$.

Clearly for a threshold autoregression, SETAR(2;1,1), of the form

$$X_t = \lambda(X_{t-1}) + e_t ,$$

where $\lambda(x) = a_o^{(i)} + a_1^{(i)} x$, $x \in (r_{i-1}, r_i]$, $i = 1,2,3,...,l$ with $-\infty = r_o < r_1 < \cdots < r_l = \infty$, and $(a_o^{(i)}, a_1^{(i)})$ are the l pairs of constants coefficients, equation (9.3.2.4) simplifies to

$$K_m(x) = \sum_{j=1}^{l} \int_{r_{j-1}}^{r_j} K_{m-1}(y) g(y - \lambda(x)) dy , \qquad (9.3.2.6)$$

with boundary condition

$$K_1(x) = \lambda(x) .$$

In practice, we may employ some Gauss-type formula to generate an appropriate set of points $x_1, x_2, \cdots, x_n$ with corresponding weights $w_1, w_2, \cdots, w_n$ and by using the recursive formula (9.3.2.6) a sequence of prediction values can be obtained. In our experiments, NAG routine D01BBF is employed to generate an appropriate set of points and corresponding weights.

As a check on the method of numerical integration and Monte Carlo method, they have been applied to obtain the conditional variance of m-step ahead prediction error of the linear AR(2) process

$$X_t = 1.7X_{t-1} - 0.8X_{t-2} + e_t , \qquad (9.3.2.7)$$

where $e_t \sim N(0,1)$. The results with 1000 replications in Monte Carlo method, are given by Table (9.3.2.1).

Table (9.3.2.1)

m	theoretical	numerical	Monte Carlo
1	1.00	1.00	1.00
2	3.89	3.89	3.89
3	7.26	7.26	7.22
4	12.07	12.07	12.14
5	16.29	16.29	16.47
6	19.33	19.33	19.23
7	21.05	21.05	21.17
8	21.76	21.76	21.50

9.3.3. Two Experiments

To investigate the influence of the *skeleton* i.e. the equation $x_t = \lambda(x_{t-1})$ without the noise (i.e. *flesh*), on the multi-step prediction of (9.3.2.1), we conduct two experiments. (The term skeleton was first introduced in the present context by Tong, 1987a). It carried a different meaning from the conventional definition in the theory of Markov Processes. However, our usage is closer to the conventional meaning of the word in everyday English).

On using the numerical integration technique with accuracy to more than two decimal places, Table (9.3.3.1) gives the results of the two experiments. In experiment I, $x_o = 4.0435$ and

$$\lambda(x) = \begin{cases} 1.5-0.9x & \text{if } x \le 0 \\ -0.4-0.6x & \text{if } x > 0 \end{cases}, \tag{9.3.3.1}$$

and in experiment II, $x_o = 5.0$ and

$$\lambda(x) = \begin{cases} 1.5-0.7x & \text{if } x \le 0 \\ 1.0+0.8x & \text{if } x > 0 \end{cases}. \tag{9.3.3.2}$$

In each experiment, $e_t \sim N(0, \sigma^2)$.

Table (9.3.3.1)

Experiment I

	σ=0.4		σ=1.0		σ=2.0	
m	$\hat{X}_o(m)$	$\hat{\sigma}(m)$	$\hat{X}_o(m)$	$\hat{\sigma}(m)$	$\hat{X}_o(m)$	$\hat{\sigma}(m)$
1	-2.8262	0.4000	-2.8261	1.0000	-2.8261	2.0000
2	4.0435	0.5381	4.0392	1.3571	3.9151	2.8761
3	-2.8261	0.5141	-2.8177	1.3076	-2.5339	2.9890
4	4.0435	0.6116	4.0092	1.6107	3.5602	3.6110
5	-2.8261	0.5428	-2.7736	1.4859	-2.1061	3.5869
6	4.0435	0.6314	3.9524	1.7772	3.1537	4.0278
7	-2.8261	0.5509	-2.7073	1.6497	-1.6943	3.9603
8	4.0435	0.6371	3.8805	1.9228	2.7753	4.2814
9	-2.8261	0.5533	-2.6309	1.8038	-1.3321	4.2007
10	4.0435	0.6387	3.8022	2.0557	2.4444	4.4415
11	-2.8261	0.5540	-2.5508	1.9454	-1.0225	4.3601
12	4.0435	0.6392	3.7217	2.1769	2.1615	4.5453
13	-2.8261	0.5542	-2.4701	2.0741	-0.7604	4.4685
14	4.0435	0.6393	3.6414	2.2871	1.9217	4.6137
15	-2.8261	0.5542	-2.3902	2.1905	-0.5394	4.5438

Table (9.3.3.1)

Experiment II

	$\sigma = 0.4$		$\sigma = 1.0$		$\sigma = 2.0$	
m	$\hat{X}_o(m)$	$\hat{\sigma}(m)$	$\hat{X}_o(m)$	$\hat{\sigma}(m)$	$\hat{X}_o(m)$	$\hat{\sigma}(m)$
1	5.0000	0.4000	5.0000	1.0000	5.0000	2.0000
2	5.0000	0.5123	5.0000	1.2806	5.0091	2.5477
3	5.0000	0.5727	5.0000	1.4315	5.0524	2.7939
4	5.0000	0.6082	5.0003	1.5198	5.0159	2.9150
5	5.0000	0.6299	5.0007	1.5731	5.1551	2.9802
6	5.0000	0.6433	5.0012	1.6058	5.1966	3.0177
7	5.0000	0.6518	5.0018	1.6260	5.2300	3.0401
8	5.0000	0.6572	5.0024	1.6387	5.2565	3.0542
9	5.0000	0.6606	5.0029	1.6466	5.2772	3.0633
10	5.0000	0.6628	5.0033	1.6515	5.2934	3.0693
11	5.0000	0.6642	5.0037	1.6546	5.3058	3.0734
12	5.0000	0.6651	5.0040	1.6566	5.3154	3.0763
13	5.0000	0.6657	5.0042	1.6578	5.3228	3.0784
14	5.0000	0.6660	5.0044	1.6586	5.3285	3.0799
15	5.0000	0.6663	5.0046	1.6591	5.3329	3.0810

[Note: $\hat{\sigma}(m) = \sqrt{Var(X_m | X_o)}$].

Some comments are in order:

(i) Let $\lambda^m(x)$ denote $\lambda(\lambda(\cdots(\lambda(x))\cdots)$, the m-fold application of λ. With decreasing signal-to-noise ratio, the difference between $\hat{X}_o(m)$ and $\lambda^m(x_o)$ increases.

(ii) For model (9.3.3.2) which admits only a periodic attractor of period one (i.e. $\lambda^m(x) \to 5.0$ as $m \to \infty$, $\forall x$), $\hat{\sigma}(m)$ is a monotonic increasing function of m for all three choices of σ. The limit is clearly $Var(X_t)$.

(iii) In contrast, for model (9.3.3.1), which admits a periodic attractor of period two at $C = \{-2.8261, 4.0435\}$, (i.e. $\lambda^m(x) \to C$ as $m \to \infty$, $\forall x$), we observe that for $\sigma = 0.4$ and 1.0,

$$\hat{\sigma}_t(1) < \hat{\sigma}_t(3) < \hat{\sigma}_t(5) < \cdots \leq Var(X_t)$$

and

$$\hat{\sigma}_t(2) < \hat{\sigma}_t(4) < \hat{\sigma}_t(6) < \cdots \leq Var(X_t) .$$

However, $\hat{\sigma}_t(2m) \nless \hat{\sigma}_t(2m+1)$, $m = 1,2,\ldots$. It is clear that it would be interesting and important to establish the precise role of periodic attractors in m-step ahead prediction. We leave this challenging problem to future research.

(iv) Gerrard (private communications, 1987) has given heuristic explanation of points (ii) and (iii), which runs in the same spirit as Tong (1983), pp. 107-114 and p. 192). If λ has a limit cycle of period r say, i.e.

$$\lambda^m(x) \to \{c_1, c_2, \cdots, c_r\} , \text{ as } m \to \infty ,$$

where

$$\lambda : c_i \to c_{i+1} , \; i = 1,2,\ldots,r-1 , \; \lambda : c_r \to c_1 ,$$

then the stationary density is of the form

$$\{g_1(x-c_1) + g_2(x-c_2) + \cdots + g_r(x-c_r)\}/r ,$$

where g_i is a density function with variance σ_i^2, tending to a δ-function as the noise variance tends to zero. Therefore, unless σ_i's are all equal (which may be shown to be the exception rather than the rule) monotonicity of $\hat{\sigma}_t(m)$ in m will not be obtained.

9.3.4. Prediction from SETAR Models

Tong and Wu (1981) have conducted a fairly extensive study on the multi-step ahead forecast performance of SETAR models for sunspot data. Generally in the case of SETAR model, the prediction problem becomes very easy for at most up to d-step ahead (d being the delay parameter). The main point is that we

know which branch in the SETAR model the future prediction data belongs to. In this case similar to the linear autoregressive model we can easily obtain the point forecast and the conditional variance of predictors. Obviously in more than d-step ahead prediction it is difficult to choose a reasonable linear autoregressive branch instantaneously in the SETAR model to predict the new future value. In this case we have to use the formula (9.3.2.5) to calculate $\hat{X}_t(d+1)$, $\hat{\sigma}_t(d+1)$,... and so on.

To show the performance of Monte Carlo method in approximation of m-step ahead forecasting of non-linear models, the following example is in order.

To the Wolf's annual sunspot numbers for the period (1700-1979) after an instantaneous square root transformation, Ghadder and Tong (1981) fitted a threshold autoregressive (SETAR(2;9,2)) model. Let X_t denote the sunspot number for the year $1699+t\,(t=1,2,...,280)$, and let $Y_t=2[(X_t+1)^{1/2}-1]$. Then the SETAR(2;9,2) model for $\{Y_t\}$ with $d=8$ is given by

$$Y_t=\begin{cases}1.89+0.86Y_{t-1}+0.08Y_{t-3}+0.16Y_{t-4}-0.21Y_{t-5} \\ \quad -0.00Y_{t-6}+0.19Y_{t-7}-0.28Y_{t-8}+0.20Y_{t-9}+e_t^{(1)} & \text{if } Y_{t-8}\le 11.9284 \\ 4.53+1.41Y_{t-1}-0.78Y_{t-2}+e_t^{(2)} & \text{if } Y_{t-8}>11.9284\,,\end{cases} \tag{9.3.4.1}$$

where $Var(e_t^{(1)})=1.946$, $Var(e_t^{(2)})=6.302$ and pooled $Var=3.734$. Using the above model, we can obtain the exact multi-step ahead forecasts of the transformed series i.e. $\{Y_t\}$, for eight years. Having the exact point forecast and the exact conditional variance of the predictor of Y_{t+m}, and under normality assumption of $e_t^{(1)}$ and $e_t^{(2)}$, we may use the following relations and calculate the exact conditional mean and conditional variance of X_{t+m}, $m=1,2,...,8$.

$Y_{t+m}=2[(X_{t+m}+1)^{1/2}-1]$ is equivalent to $X_{t+m}=\dfrac{Y_{t+m}^2}{4}+Y_{t+m}$. The optimal m-step ahead point forecast of series X_t is given by

$$\begin{aligned}E_t(X_{t+m}) &= E(X_{t+m}\,|\,X_t,X_{t-1},\cdots) \\ &= \frac{1}{4}E_t(Y_{t+m}^2)+E_t(Y_{t+m}) \\ &= \frac{1}{4}[V_t(Y_{t+m})+\hat{Y}_t^2(m)]+\hat{Y}_t(m)\,.\end{aligned}$$

Similarly

$$V_t(X_{t+m}) = E_t\left[\left(\frac{Y_{t+m}^2}{4}+Y_{t+m}\right)^2\right]-E_t^2(X_{t+m})$$

$$= \frac{1}{16}E_t(Y_{t+m}^4)+E_t(Y_{t+m}^2)+\frac{1}{2}E_t(Y_{t+m}^3)-\hat{X}_t^2(m)$$

$$= \frac{1}{16}\left[3\hat{\sigma}_t^4(m)+\hat{Y}_t^4(m)+6\hat{Y}_t^2(m)\sigma_t^2(m)\right]$$

$$+\hat{\sigma}_t^2(m)+\hat{Y}_t^2(m)+\frac{1}{2}\left[3\hat{Y}_t(m)\sigma_t^2(m)+\hat{Y}_t^3(m)\right]-\hat{X}_t^2(m),$$

where $\sigma_t^2(m)=V_t(Y_{t+m})=Var(Y_{t+m}\,|\,Y_t,Y_{t-1},\cdots)$.

In our Monte Carlo study, we simply record the simple back-transformation of the simulated data, and then we use the sample mean and sample variance of 1000 simulated series as an approximation of $\hat{X}_t(m)$ and $\hat{\sigma}_t^2(m)$. The results are given in Table (9.3.4.1).

Table (9.3.4.1)

	Theoretical (exact)		Monte Carlo	
m	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$
1	162.5	32.02	162.6	32.05
2	144.0	30.86	144.3	30.89
3	102.8	27.99	103.1	28.07
4	65.8	19.89	66.0	19.84
5	35.1	14.36	35.3	14.37
6	19.5	10.27	19.5	10.29
7	19.5	14.62	19.5	14.76
8	31.3	15.96	31.1	25.96

It can be seen that the performance of the Monte Carlo method is quite remarkable. Our experiments show that similar accuracy can be achieved for the bilinear model.

To compare the performance of the multi-step ahead prediction of SETAR and bilinear models on

real data, we choose the following SETAR(2;3,11) and the subset bilinear model which is fitted to the same period (1700-1920) of the sunspot numbers. The SETAR(2;3,11) which is reported by Tong (1983) is given by

$$X_t = \begin{cases} 11.97+1.71X_{t-1}-1.26X_{t-2}+0.43X_{t-3}+e_t^{(1)} & \text{if } X_{t-3} \le 36.6 \\ 7.84+0.73X_{t-1}-0.04X_{t-2}-0.20X_{t-3}+0.16X_{t-4} & \\ \quad -0.23X_{t-5}+0.02X_{t-6}+0.15X_{t-7}-0.24X_{t-8} & \\ \quad +0.31X_{t-9}-0.37X_{t-10}+0.38X_{t-11}+e_t^{(2)} & \text{if } X_{t-3} > 36.6\,, \end{cases} \tag{9.3.4.2}$$

where $Var(e_t^{(1)})=254.64$, $Var(e_t^{(2)})=66.80$ and pooled $Var=153.7$. The subset bilinear model which is reported by Gabr and Subba Rao (1981) is given by

$$\begin{aligned} X_t = {} & 6.8860+1.5012X_{t-1}-0.7670X_{t-2}+0.1152X_{t-9} \\ & -0.01458X_{t-2}e_{t-1}+0.006312X_{t-8}e_{t-1}-0.007152X_{t-1}e_{t-3} \\ & +0.006047X_{t-4}e_{t-3}+0.003619X_{t-1}e_{t-6}+0.004334X_{t-2}e_{t-4} \\ & +0.001782X_{t-3}e_{t-2}+e_t\,, \end{aligned} \tag{9.3.4.3}$$

where $Var(e_t) = 124.33$.

The point forecast together with the error bounds $(\hat{X}_t(m) \pm \hat{\sigma}_t(m))$ of models (9.3.4.2) and (9.3.4.3) are given by Figs 9.3.4.1 and 9.3.4.2 respectively. The non-monotonicity of $\hat{\sigma}_t(m)$ for both models are quite clear.

For comparison, we also considered the linear autoregresive AR(9) model which is fitted by Gabr and Subba Rao (1981) to the mean deleted sunspot data with the same period (1700-1920). Their fitted linear autoregression model is given by

$$\begin{aligned} X_t = {} & 1.2163X_{t-1}-0.4670X_{t-2}-0.1416X_{t-3}+0.1691X_{t-4} \\ & -0.1473X_{t-5}+0.0543X_{t-6}-0.0534X_{t-7}+0.0667X_{t-8} \\ & +0.1129X_{t-9}+e_t\,, \end{aligned} \tag{9.3.4.4}$$

where $Var(e_t) = 199.27$. Fig. (9.3.4.3) shows the point forecasts and the error bounds of the linear model predictions.

Some comments are now in order

(i) Over the short period of (1921-1935)(especially for the first five years) the performance of SETAR model prediction is better than the other models.

(ii) Associated with the SETAR model, for the non-linear predictors of peak value those are greater conditional variances.

(iii) Notice that the vertical lines at the troughs in Figs. 9.3.4.3 and 9.3.4.2 have a greater tendency (especially in Fig. 9.3.4.2) of extending below zero into the negative regime, sometimes quite substantially.

Note that the values of $\hat{X}_t(m)$ and $\hat{\sigma}_t(m)$ in the bilinear case for $m > 6$ and in the SETAR case for $m > 3 = d$ are obtained by Monte Carlo method with 1000 replications.

9.3.5. Bias Correction of Forecasting Transformed Series

Suppose that a time series analyst has derived a model for a series $Y_t = f(X_t)$, an instantaneous transformation of time series X_t, where $f(.)$ is some well behaved function and thus can forecast the series Y_t. How should forecasts of X_t be formed ? It is generally accepted that the simple back transformation of Y_t as a forecast of X_t is biased, and a bias correction method is needed. This problem has been studied by Granger and Newbold (1976), but it does not seem to be possible to apply their method to the non-linear models.

The bias due to the square root transformation or more generally any smooth one to one instantaneous transformation, may be corrected to a large extent by appealing to the usual δ-technique. Specifically, let f be a smooth one to one map from x to y. Let g denote the smooth inverse map. Suppose that a model is fitted to the transformed series $\{Y_t\}$ and we wish to forecast the original series $\{X_t\}$. Let $B_t^{(X)}$ denote the σ-algebra generated by $\{X_s, s \le t\}$. $B_t^{(Y)}$ may be defined in a similar way. Now

$$\begin{aligned} E[X_{t+m} \mid B_t^{(X)}] &= E[g(Y_{t+m}) \mid B_t^{(Y)}] \\ &\cong g(\hat{Y}_t(m)) + E[(Y_{t+m} - \hat{Y}_t(m)) g'(\hat{Y}_t(m)) \mid B_t^{(Y)}] \\ &\quad + \frac{1}{2} E[(Y_{t+m} - \hat{Y}_t(m))^2 g''(\hat{Y}_t(m)) \mid B_t^{(Y)}] \\ &= g(\hat{Y}_t(m)) + \frac{1}{2} g''(\hat{Y}_t(m))\, Var[Y_{t+m} \mid B_t^{(Y)}] . \end{aligned}$$

In other words, we only need to adjust the naive forecast $g(\hat{Y}_t(m))$ by the correction term specified above. Note that the conditional variance of Y's is involved in the correction term. A similar argument yields

$$Var[X_{t+m} \mid B_t^{(X)}] \cong [g'(\hat{Y}_t(m))]^2 \, Var[Y_{t+m} \mid B_t^{(Y)}] .$$

Let $y=2[(x+1)^{1/2}-1]=f(x)$. Thus $g(y)=f^{-1}(y)=(1+\frac{1}{2}y)^2-1$ and $g'(y)=1+\frac{1}{2}y$ and $g''(y)=\frac{1}{2}$.

Table (9.3.5.1) shows the results of applying the the above bias correction method to the sunspot predicted numbers of SETAR model (9.3.4.2) under Box-Cox transformation with $\lambda=0.5$.

Table (9.3.5.1)

lead time	naive		theoretical		bias corrected		Monte Carlo	
m	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$	$\hat{X}_t(m)$	$\hat{\sigma}_t(m)$
1	160.9	-	162.5	32.02	162.5	31.10	162.6	32.05
2	142.4	-	144.0	30.86	144.0	30.95	144.3	30.89
3	100.9	-	102.8	27.99	102.8	28.12	103.1	28.07
4	64.3	-	65.8	19.89	65.8	19.99	66.0	19.84
5	33.7	-	35.1	14.36	35.1	14.50	35.3	14.37
6	18.1	-	19.5	10.27	19.5	10.43	19.5	10.29
7	16.7	-	19.5	14.62	19.4	15.11	19.5	14.76
8	25.6	-	31.3	25.96	31.1	27.08	31.1	25.96

The performance of the bias correction method and the Monte Carlo procedures are remarkable. Note that variance is a non-linear operator, therefore using the simple minded back transformation of $V_t(Y_{t+m})$ as a naive conditional variance of X_{t+m} is not valid.

9.3.6. A Comparative Study of Linear, Bilinear and SETAR models in Predicting sunspot numbers

To assess the performance of some non-linear models by reference to the Monte Carlo and bias correction method, we considered a number of SETAR models which have been fitted to raw and transformed sunspot numbers and the bilinear model (9.3.4.3). The linear model (9.3.4.4) is also considered with a view to comparison.

Let $\hat{\sigma}^2(m)$ denote the sample mean square error of m-step ahead predictions given by

$$\hat{\sigma}^2(m) = \frac{1}{N-t} \sum_{t_*=t-m}^{N-m} (X_{t_*+m} - \hat{X}_{t_*}(m))^2 \ , \ m = 1,2,\ldots , \qquad (9.3.6.1)$$

where N is the number of data (sample size) and t is the starting point of forecasting.

Smaller sample mean square error of prediction, means a better performance of prediction.

Tong (1983) has reported three different SETAR models with the Serial numbers AS7133, BS8103 and CS8103 (see Tong, 1983 p. 253) which are fitted to the different period of raw and transformed sunspot numbers.

Tong and Lim (1980, p. 261) have fitted a threshold autoregressive model to the annual Wolf's sunspot numbers of (1700-1920), using the data bank available at the University of Manchester Institute of Science and Technology at the time, which contained the data only for the period (1700-1955). We have listed in Table (9.3.6.1) and (9.3.6.2) the prediction results of the SETAR models and those of linear, subset linear and subset bilinear models fitted by Gabr and Subba Rao (1981) using Akaike's Information Criterion. The fitting period of the linear (equation 9.3.4.4), bilinear (equation 9.3.4.3) and SETAR(2;3,11) (equation 9.3.4.2) and subset linear model (equation 9.3.6.2) and the prediction period for all models are identical. These are (1700-1920, fitting period) and (1921-1955 and 1956-1979, prediction periods). The prediction period of (1921-1955 and 1956-1979) represent roughly three and two sunspot cycles and the latter period (1956-1979) is so chosen to retain some realism in the comparison (risky prediction) since these data were unknown to Tong & Lim and Gabr & Subba Rao at the time of fitting their models. Another motivation is to investigate the robustness of various forecasts to unusually influential observations like 1956. The period (1980-1987) is reserved for further experimentations.

Tables (9.3.6.1) and (9.3.6.2) show the results of the performance of these models. Note that the subset linear model which is reported by Gabr & Subba Rao (1981) is given by

$$X_t = 1.2496X_{t-1} - 0.5510X_{t-2} + 0.1450X_{t-9} + e_t \ , \qquad (9.3.6.2)$$

where $Var(e_t) = 203.21$.

Table (9.3.6.1)

period 1921-1955

sample predic. MSE	full AR(9)	subset AR(9)	SETAR(2;3,11)	bilinear	SETAR(AS7133)	SETAR(BS103		SETAR(CS8103)	
						naive	bias corrected	naive	bias corrected
$\hat{\sigma}_e^2$	199.27	203.21	153.7	124.33	152.31	3.69		3.72	
$\hat{\sigma}^2(1)$	190.87	216.09	153.88	123.79	161.90	120.63	115.01	119.77	115.05
$\hat{\sigma}^2(2)$	414.75	429.97	388.35	337.60	362.55	310.31	271.16	303.94	267.13
$\hat{\sigma}^2(3)$	652.07	676.22	672.67	569.69	593.24	563.28	475.36	549.84	467.49
$\hat{\sigma}^2(4)$	797.32	650.09	641.16	621.29	650.09	641.16	530.65	621.29	516.81
$\hat{\sigma}^2(5)$	770.84	775.86	835.31	718.43	613.15	714.76	589.17	691.14	570.91
$\hat{\sigma}^2(6)$	786.40	797.50	900.67	732.42	584.76	758.01	619.93	732.06	603.04
$\hat{\sigma}^2(7)$	789.01	814.67	993.83	781.65	508.12	780.22	629.97	751.31	610.89
$\hat{\sigma}^2(8)$	827.79	860.76	1083.60	833.18	531.76	828.22	663.03	796.58	641.63
$\hat{\sigma}^2(9)$	862.06	899.37	1111.91	900.61	518.80	873.57	694.32	836.52	666.58
$\hat{\sigma}^2(10)$	895.58	936.34	1124.56	961.93	520.93	930.07	734.18	886.15	702.41
$\hat{\sigma}^2(11)$	982.82	1040.50	1173.35	1013.80	562.99	988.52	775.36	941.02	740.29
$\hat{\sigma}^2(12)$	1168.52	1247.49	1304.26	1139.21	650.51	1071.68	841.27	1022.56	800.83

Table (9.3.6.2)

period 1956-1979

sample predic. MSE	full AR(9)	subset AR(9)	SETAR(2;3,11)	bilinear	SETAR(AS7133)	SETAR(BS103		SETAR(CS8103)	
						naive	bias corrected	naive	bias corrected
$\hat{\sigma}_e^2$	199.27	203.21	153.7	124.33	152.31	3.69		3.72	
$\hat{\sigma}^2(1)$	525.70	493.85	428.10	21122.90	504.26	438.03	412.79	433.44	414.02
$\hat{\sigma}^2(2)$	1049.10	1004.18	1006.43	14651.21	1051.20	1133.09	993.90	1102.66	982.22
$\hat{\sigma}^2(3)$	970.11	905.91	946.50	592805.76	1099.28	1370.27	1128.74	1320.17	1098.70
$\hat{\sigma}^2(4)$	747.12	712.30	900.77	174075.50	939.11	1264.03	1025.79	1264.03	988.60
$\hat{\sigma}^2(5)$	618.33	577.36	773.59	126232.63	896.72	1111.87	901.53	1079.01	875.21
$\hat{\sigma}^2(6)$	603.61	560.57	866.21	152870.92	353.65	906.38	720.99	881.49	698.64
$\hat{\sigma}^2(7)$	640.95	611.79	908.09	174587.72	439.73	944.96	757.455	916.60	653.55
$\hat{\sigma}^2(8)$	565.08	590.24	880.36	140566.50	575.73	855.07	674.51	812.72	634.83
$\hat{\sigma}^2(9)$	619.68	652.30	929.24	89642.94	563.97	865.64	702.72	830.47	668.78
$\hat{\sigma}^2(10)$	660.25	687.23	955.55	59455.70	716.90	816.79	652.48	777.61	620.42
$\hat{\sigma}^2(11)$	684.67	676.42	1058.96	71424.98	702.48	854.33	685.21	816.57	645.38
$\hat{\sigma}^2(12)$	808.98	758.12	1275.26	151970.92	692.51	989.16	786.30	937.34	736.99

Considering the results of Table (9.3.6.1), it is clear that the short term forecasting of SETAR(2;3,11) and bilinear model are better than linear models. The SETAR(CS8103) has the best performance for all m, and SETAR(AS7133) has a quite good performance over the short term predictions. Although the bilinear model has a better performance for the first period than the SETAR(2;3,11) and SETAR(BS103), the performance of SETAR(AS7133) after five steps is better than bilinear model. Among those models with the same fitting period, the performance of non-linear models over short term prediction are better than linear

models.

A very poor performance of the bilinear model in predicting sunspot numbers, over the second period (i.e. 1956-1979) may be due to non-invertibility of this model (See chapter four) and the effect of the influential observation at 1956. In the former case it seems that the SETAR and linear models except in the 2-step ahead prediction are not seriously affected by this influential observation. For the bilinear model, the affect of the 1956 observed value has carried on for quite a long time, which seems to be due to the product terms of e and X.

The prediction results of SETAR models after correcting the bias seems to give a good support for our bias correction method.

Despite the encouraging performance demonstrated in Table (9.3.6.2), as far as medium to long range forecasting is concerned our limited experience suggests that a linear multi-step ahead forecast often provides a *robust* bench-mark especially in dealing with a fairly complex situation. Now, it is well known that the Wolf's annual sunspot numbers represent one such complex situation. (See, e.g. Tong 1983, p. 230). For example, the rise from 38.0 in 1955 to 141.7 in 1956 is unusually steep.

9.3.7. Combination of Forecasts

From Table (9.3.6.2) it is clear that the non-linear forecast of SETAR(2;3,11) out performs the linear forecast over the short range, the reverse is true over the longer range. Closer investigation reveals that whilst the non-linear (SETAR) forecast does a better job than a linear forecast over the `troughs', the reverse is true over the `peaks'. See the results of Table (9.3.7.1).

Table (9.3.7.1)

period (1921-1955)

	Full AR(9)			SETAR			Combination
$\hat{\sigma}^2(m)$	below the mean	above the mean	over all	below the mean	above the mean	over all	forecast
$\hat{\sigma}^2_*$			199.270			153.710	
$\hat{\sigma}^2(1)$	319.314	649.528	525.698	101.848	693.617	422.389	422.269
$\hat{\sigma}^2(2)$	524.930	1386.058	1049.095	135.520	1579.983	951.956	873.047
$\hat{\sigma}^2(3)$	472.676	1314.480	970.106	131.457	1516.083	886.707	766.840
$\hat{\sigma}^2(4)$	420.746	991.899	747.119	107.278	1549.852	862.912	560.380
$\hat{\sigma}^2(5)$	408.146	790.305	618.333	290.332	1287.316	738.974	460.611
$\hat{\sigma}^2(6)$	362.916	778.666	603.613	87.201	1619.232	812.900	443.408
$\hat{\sigma}^2(7)$	340.204	881.548	640.951	112.816	1553.409	833.112	490.115
$\hat{\sigma}^2(8)$	278.068	820.207	565.083	41.246	1701.446	822.516	407.398
$\hat{\sigma}^2(9)$	303.954	865.245	619.680	49.513	1477.544	852.780	462.638
$\hat{\sigma}^2(10)$	349.445	867.454	660.250	23.627	1469.052	890.882	488.420
$\hat{\sigma}^2(11)$	433.262	873.236	684.675	213.992	1412.040	984.166	580.982
$\hat{\sigma}^2(12)$	709.895	870.906	808.979	626.640	1652.273	1178.904	620.539

This therefore suggests that it might be profitable to combine the two forecasts so as to exploit their complementary preformances. The final column of Table (9.3.7.1) gives the results when the non-linear forecasts (NL) and the linear forecasts (L) are combined in the following manner. (Let M stand for the historical mean over the fitting period 1700-1920):

If both NL and L $\leq$ M, then adopt NL.

If both NL and L > M, then adopt L.

If NL ≤ M, and L > M, then adopt $\frac{1}{2}$(NL+L).

IF NL > M and L ≤ M, then adopt $\frac{1}{2}$(L+NL).

Our combination rule implies a probability distribution of combination weights of forecasts which is 'self-excited' by the forecasts themselves (c.f. Granger and Newbold, 1986). Our approach seems to be closer to an empirical Bayes approach in spirit. Accordingly, we should emphasis the pragmatic (empirical) nature of our results in this respect at present.

Finally, we return to the multi-step ahead prediction of Tong-lim's threshold autoregressive model and Subba Rao-Gabr's linear autoregressive model for the period of 1980-1987. The point forecasts are shown in Figs (9.3.7.1) and (9.3.7.2) and Table (9.3.7.2).

Table (9.3.7.2)

year	observations	linear forecast	prediction error	SETAR forecast	prediction error	combination forecast	prediction error
1980	154.7	162.8	-8.1	174.7	-20.0	162.8	-8.1
1981	140.5	133.5	7.0	147.5	-7.0	133.5	7.0
1982	115.9	91.5	24.4	92.4	23.5	91.5	23.5
1983	66.6	53.3	13.3	62.9	3.7	53.3	13.3
1984	45.9	23.1	22.8	24.1	21.8	24.1	21.8
1985	17.9	4.6	13.3	12.6	5.3	12.6	5.3
1986	13.4	2.0	11.4	0.2	13.2	0.2	13.2
1987	29.2	21.1	8.1	10.8	18.4	10.8	18.4
MSE			222.4		253.9		237.9

[Note: The base year is 1979]

It seems that the adverse effect of the 1956 datum on the threshold prediction performance has subsided. With a mean-squared error (MSE) of 253.9, the non-linear forecast compares nearly well with the performance of the linear forecast(MSE at 222.4). The combination forecast as prescribed earlier again gives an improvement over the non-linear prediction with a lower MSE at 237.9. See Fig (9.3.7.3) and Table (9.3.7.2).

It could be argued that we should adjust our forecasts/models in the light of the rather influential datum of 1956. For example, the parameters of the models could be more robustly fitted by modifying the least-squares method. Also, in the bilinear case it is very important to check the invertibility of the fitted model especially if forecasting is the major concern. However, is the adjustment really relevant prior to 1956? It would nevertheless seem worthwhile to explore the possibilities of a combination of forecasts in the manner that we have experimented with as a means of reducing the adverse effect on prediction of unusually influential prospective observations.

9.3.8. Risky Prediction of Wolf's Annual Sunspot Numbers

Ghaddar and Tong (1981) have fitted a threshold autoregressive,SETAR(2;9,3) model to the square root transformation of Wolf's annual sunspot numbers, on this basis they derived risky predictions for the period (1980-1987) and their results were in print in 1981. The model is given by equation(9.3.4.1). However, at that stage of the development, $\hat{\sigma}(m)$'s were not available for their point forecasts. The situation may now be remedied as shown in Fig. (9.3.8.1). Shown in Fig. (9.3.8.2) ares the risky prediction of sunspot numbers after correcting the bias.

The linear autoregressive model for the original i.e. untransformed, sunspot numbers, which has been selected by Akaike's information criterion, is given by

$$\begin{aligned} X_t = {} & 6.9627+1.2064X_{t-1}-0.4507X_{t-2}-0.1747X_{t-3}+0.1974X_{t-4} \\ & -0.1366X_{t-5}+0.0268X_{t-6}+0.0128X_{t-7}-0.0312X_{t-8} \\ & +0.2123X_{t-9}+e_t\ , \end{aligned} \tag{9.3.8.1}$$

where $Var\,(e_t) = 221.2366$. Fig. (9.3.8.3) shows the risky prediction of linear autoregressive model.

Some comments are now in order.

(i) Over the period (1980-1987), it would appear that the non-linear prediction has consistently out-performed the linear prediction by a noticeable margin in respect of $\hat{X}_t(m)-X_{t+m}$. However, associated with the non-linear predictors of peak values are greater conditional variances.

(ii) Notice that the vertical lines at the troughs in Fig. (9.3.8.3) have the tendency of extending below zero into the negative regime, sometimes quite substantially. In contrast, the vertical lines in Fig. (9.3.8.2) have all managed to steer clear of the negative regime. It may also be remarked that the 'coefficient of variation', $\hat{\sigma}(m)/\hat{X}_t(m)$, fluctuates quite violently over m in the linear case but much less so in the non-linear case. The ranges are 0.09 ~ 4.92 and 0.19 ~ 1.01 respectively.

(cf. approximately 0.8 for the coefficient of variation of the data).

(iii) The non-linear forecasts for (1988-1993) seem to be much vaguer than their linear counterparts. Both approaches appear to forecast that (1988-1993) will correspond to an 'upward' swing of a sunspot cycle.

For completeness, we fit a linear autoregressive model of order 9 to the first 280 square root transformed sunspot numbers. The model which has been selected by Akaike's information criterion is given by

$$\begin{aligned} X_t = {} & 1.6019+1.2212X_{t-1}-0.4896X_{t-2}-0.1579X_{t-3} \\ & +0.2746X_{t-4}-0.2489X_{t-5}+0.0257X_{t-6} \\ & +0.1593X_{t-7}-0.2222X_{t-8}+0.2980X_{t-9}+e_t , \end{aligned} \qquad (9.3.8.2)$$

where $Var(e_t) = 4.0560$. To avoid bias due to square root transformation, we have calculated the raw risky prediction of linear and SETAR models (equations 9.3.8.2 and 9.3.4.1) for the period (1980-1987). The results are given in Table 9.3.8.3 and Fig. 9.3.8.4. Again the performance of SETAR model in predicting the sunspot numbers is quite clear.

Table 9.3.8.3

year	observations	linear forecast	prediction error	SETAR forecast	prediction error
1980	22.96	22.19	0.76	23.38	-0.42
1981	21.79	20.22	1.57	21.95	-0.16
1982	19.62	16.16	3.46	18.19	1.43
1983	14.44	12.52	1.92	14.16	0.28
1984	11.70	8.77	2.93	9.78	1.92
1985	6.69	6.54	0.15	6.74	-0.05
1986	5.59	5.71	-0.12	6.41	-0.82
1987	8.99	7.60	1.39	8.32	0.67
MSE			3.66		0.89

9.3.9. Linear Non-Gaussian Autoregressive Model

Li and McLeod (1988), have fitted a linear non-Gaussian autoregressive model of order 2, for the period (1700-1955), to the sunspot data. Their model is given by

$$X_t = 1.6759X_{t-1} - 0.7840X_{t-2} + a_t\,, \qquad (9.3.9.1)$$

where a_t is log-normal distributed with estimated mean 13.88 and variance of 153.39 . Although the residual variance of 153.39 is comparable to that of the threshold autoregressive model (equation (9.3.4.2)), and it is better then the Gaussian linear autoregressive model (equation (9.10.3.4.4)), they have claimed that the non-Gaussian linear model (equation (9.3.9.1)), is certainly more parsimonious.

It may be interesting to see how comparable is the mean of model (9.3.9.1) with the historical mean of data. Under the stationarity assumption it is easy to show that the mean of model (9.3.9.1) is equal to 128.4, which is quite different the historical mean of data which is close to 43.

To see the performance of this non-Gaussian linear model in predicting sunspot data, we calculate the risky prediction of model (9.3.9.1) over the period (1956-1979), i.e. 24 points. For comparison, we have

also included the results based on the linear Gaussian autoregressive model of order 2, which is fitted (based on the AIC criterion) to the same period (i.e. 1700-1955) of sunspot numbers. The model is given by

$$X_t = 14.0531+1.3605X_{t-1}-0.6706X_{t-2}+e_t\,, \qquad (9.3.9.2)$$

where $Var(e_t)=232.41$. The sample mean square error of m-step ahead prediction of these two linear models are given by Table (9.3.9.1).

Table (9.3.9.1)

Lead time	Linear Gaussian	Linear Non-Gaussian
1	656.38	710.12
2	1689.70	2345.89
3	2054.01	4445.43
4	1981.48	6759.83
5	1692.54	8830.82
6	1476.65	10193.57
7	1491.94	10303.46
8	1356.90	9988.48
9	1495.60	9244.70
10	1530.11	8412.47
11	1603.09	7964.47
12	1844.03	7951.79

On looking at the two linear models (Gaussian and non-Gaussian) more closely, it seems that the constant term of the Gaussian linear model(i.e. 14.05) is comparable with the mean of the error term in the non-Gaussian case (i.e. 13.88). Also the coefficients (i.e. 1.36 and -0.67) of the Gaussian model are comparable with those of the non-Gaussian one (i.e. 1.6759 and -0.7840) .

Based on Table (9.3.9.1), we can say that if prediction is a major concern, then the Gaussian linear autoregressive model did a better job than the non-Gaussian linear autoregressive model.

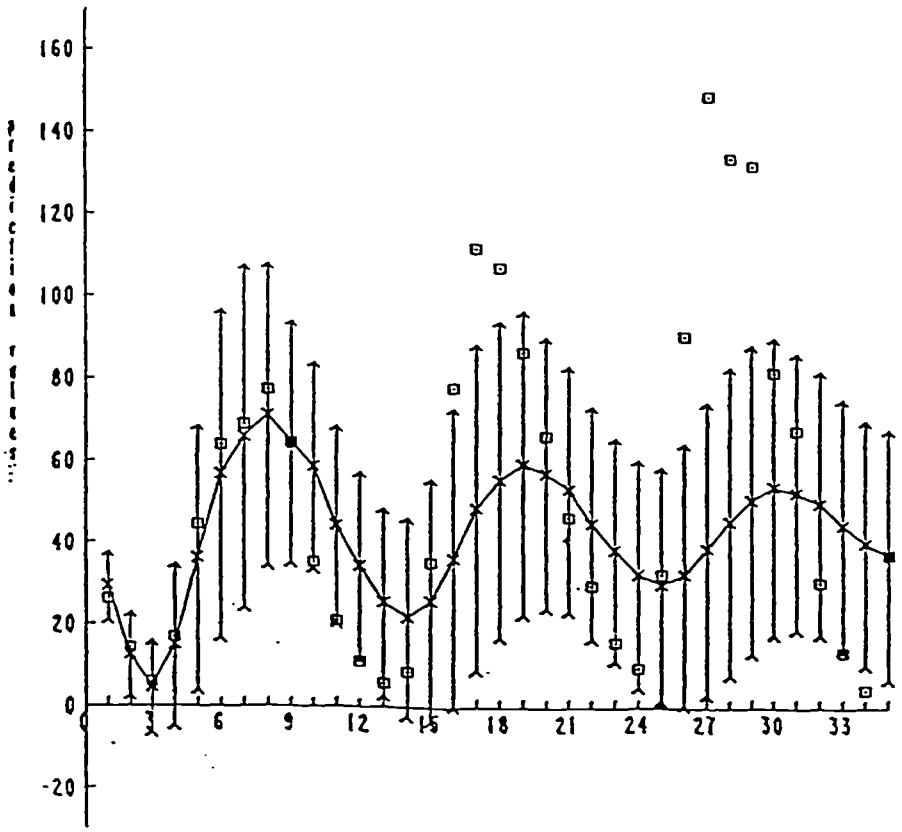


Fig. 9.3.4.1 : □ observed values, × predicted values of model (9.3.4.2). Lead time t is equivalent to year 1920+t.

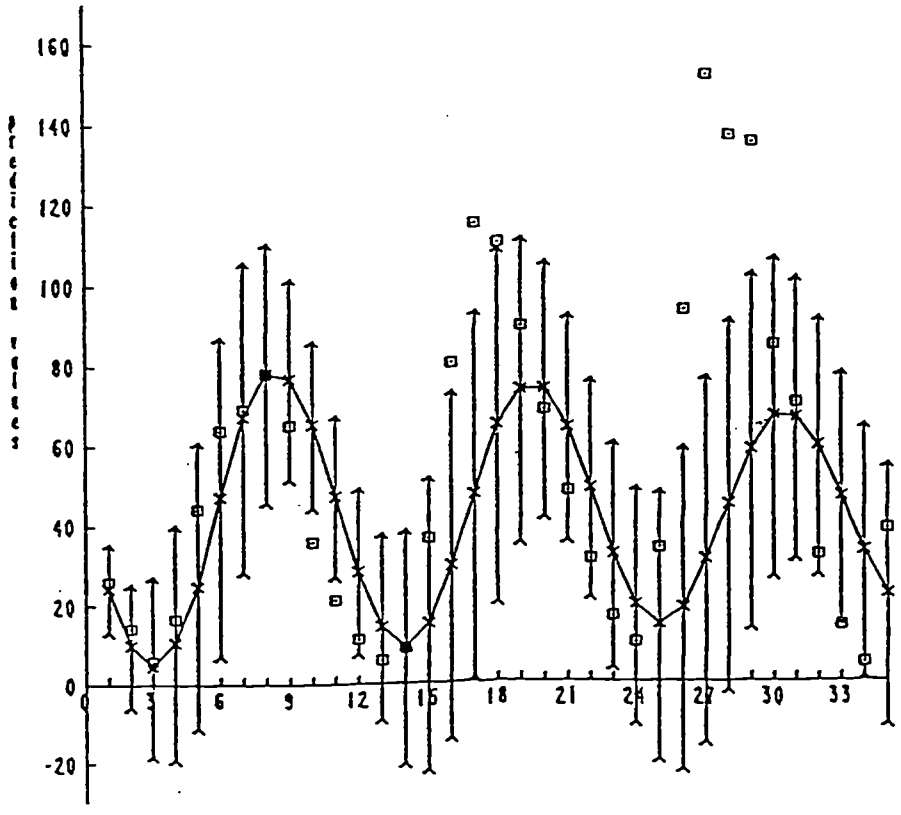


Fig. 9.3.4.2 : □ observed values, × predicted values of model (9.3.4.3). Lead time t is equivalent to year 1920+t.

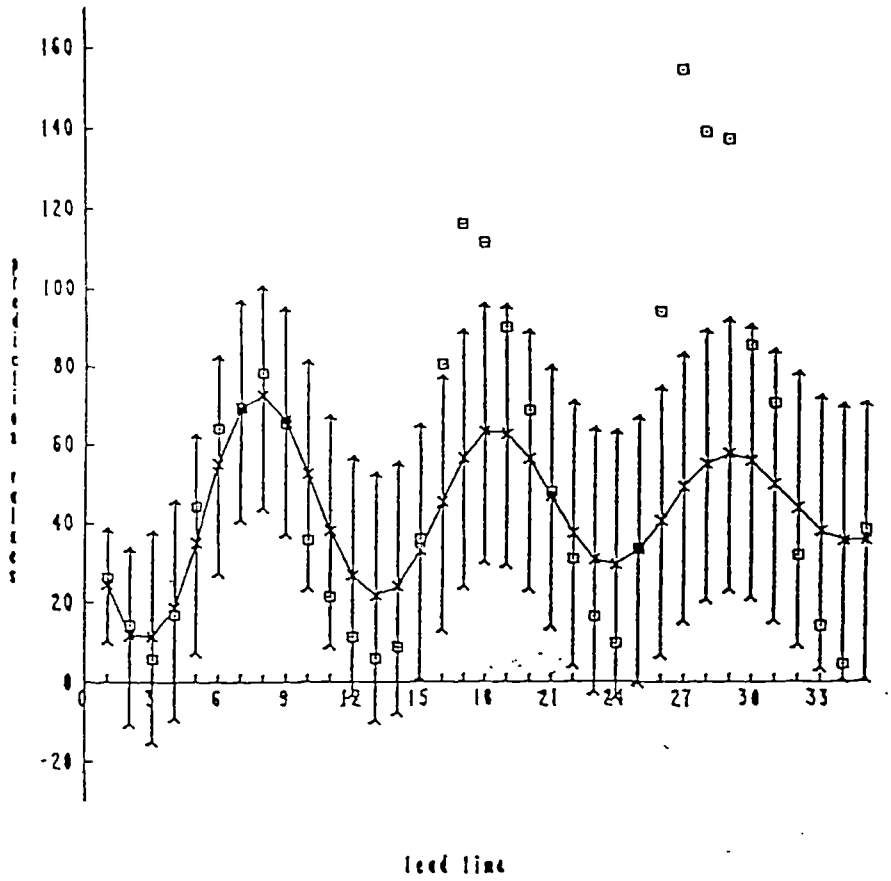


Fig. 9.3.4.3 : □ observed values, × predicted values of model (9.3.4.4). Lead time t is equivalent to year 1920+t.

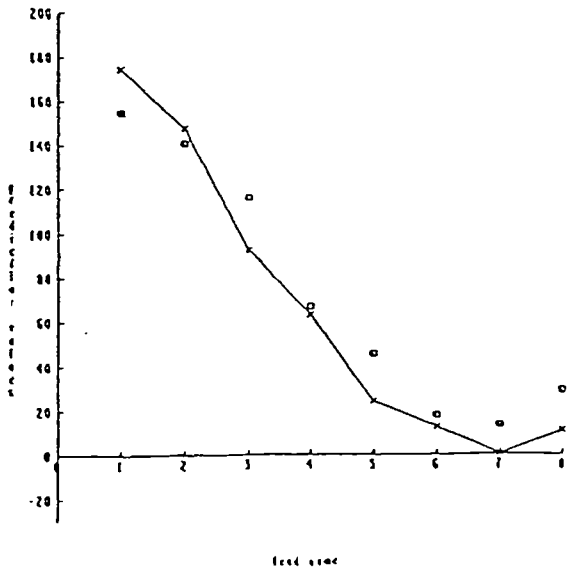


Fig. 9.3.7.1

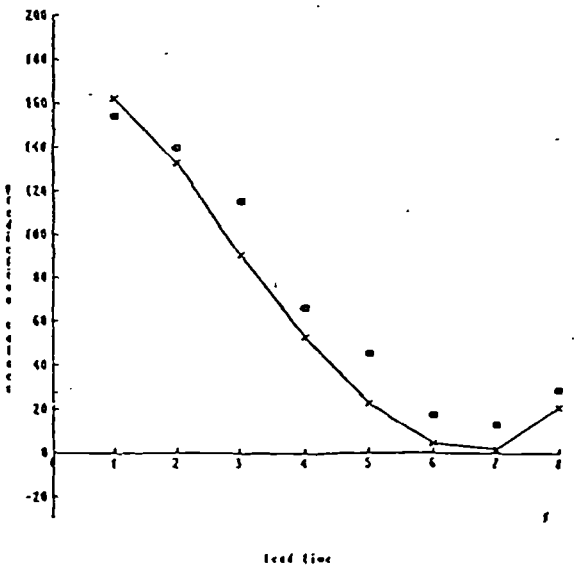


Fig. 9.3.7.2

Fig. 9.3.7.1 : Risky prediction of model (9.3.4.2). Lead time t is equivalent to year 1979+t.

Fig. 9.3.7.2 : Risky prediction of model (9.3.4.4). Lead time t is equivalent to year 1979+t.

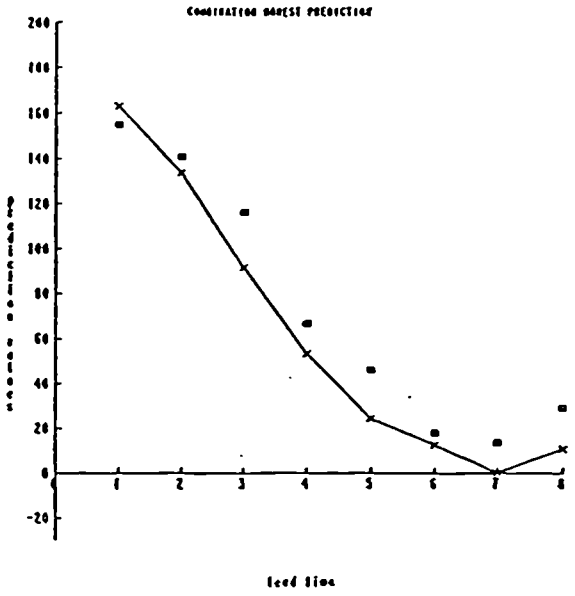


Fig. 9.3.7.3

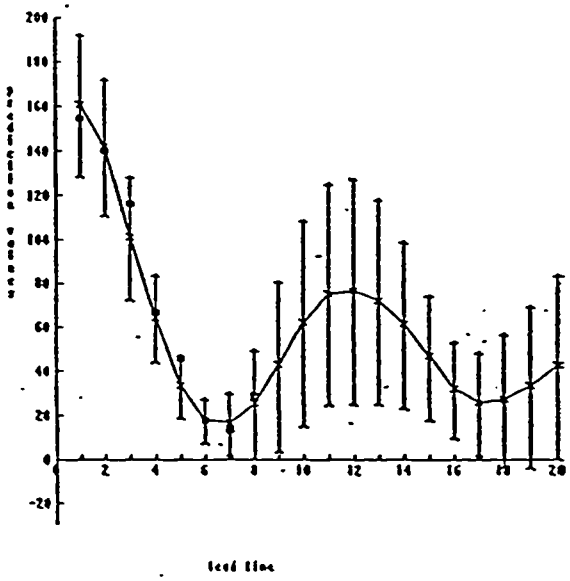


Fig. 9.3.8.1

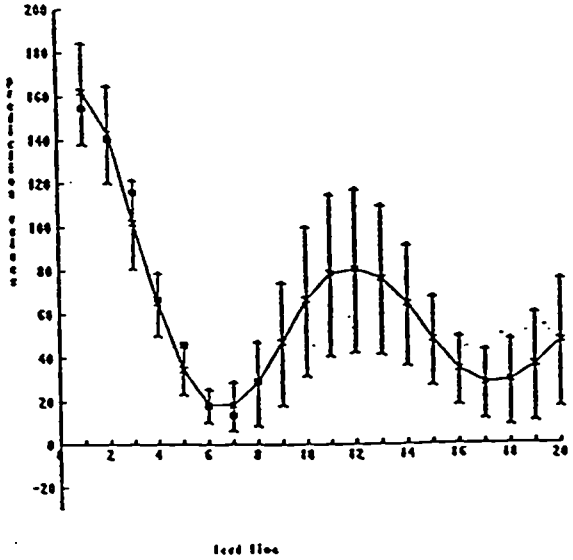


Fig. 9.3.8.2

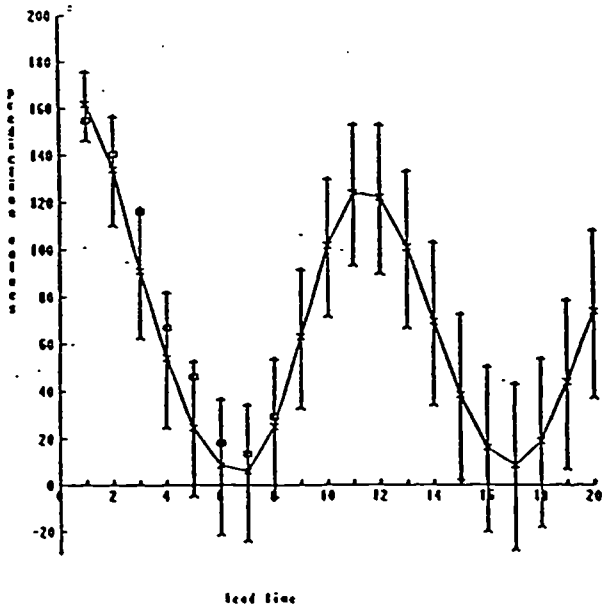


Fig. 9.3.8.3

Fig. 9.3.7.3 : Combination of forecasts.

Fig. 9.3.8.1 : Risky prediction of model (9.3.4.1).

Fig. 9.3.8.2 : Biased corrected of risky prediction of model (9.3.4.1).

Fig. 9.3.8.3 : Risky prediction of linear model (9.3.8.1).

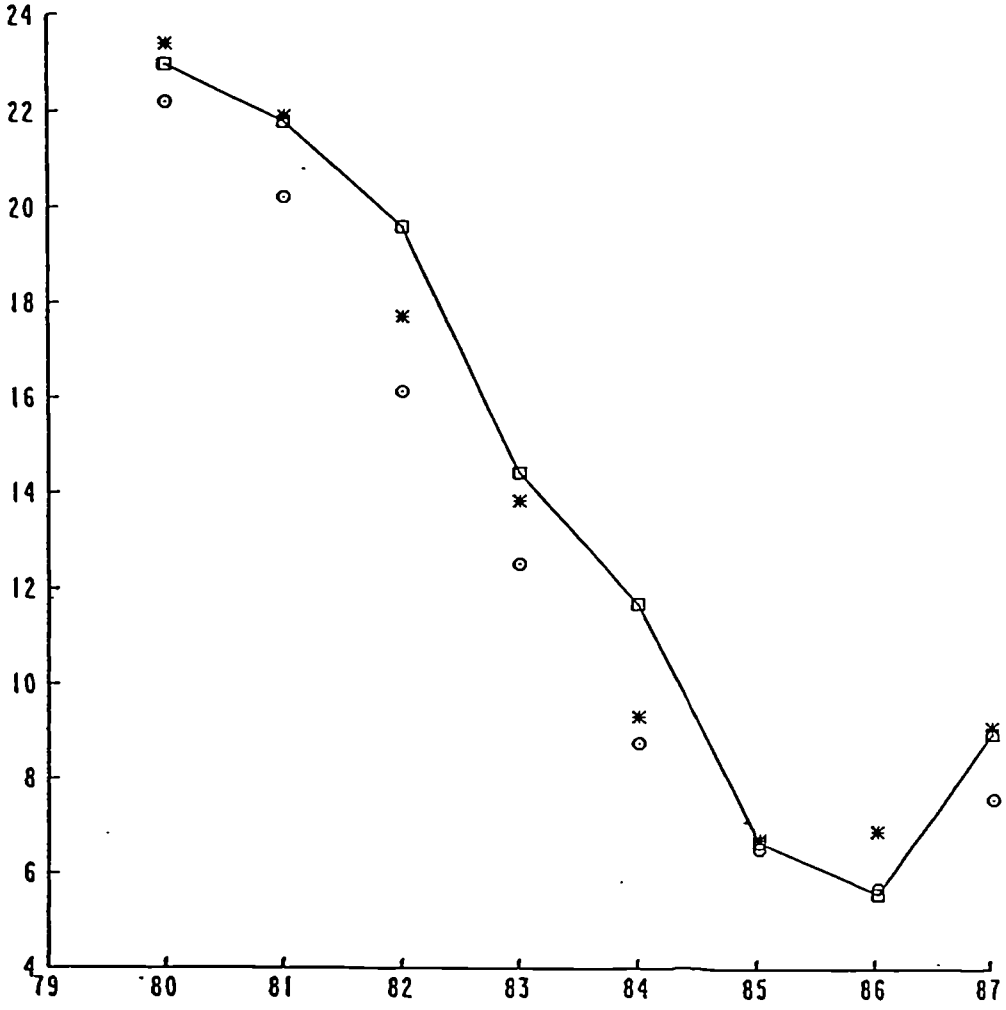


Fig. 9.3.8.4 □ observed values

* SETAR prediction

o Linear prediction

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification, *IEEE Vol. AC-19*, 716-723.

Al-Qassem, M.S. and Lane, J.A. (1987). Forecasting exponential autoregressive models of order 1. *Technical Report.* Dept. of Maths, Univ. College of Wales, Aberystwyth, Wales.

Andel, J., Netuka, I., Svara, K. (1984). On threshold autoregressive processes. *Kybernetika*, **20**, 89-106.

Andel, J. and Barton, T. (1986). A note on threshold AR(1) Model with Cauchy innovations. *J. Time Series Anal.*, **7**, 1-5.

Anderson, T.W. (1986). Why do noninvertible estimated moving averages occur ? *J. Time Series Anal., Vol.* **7**, No. 4. pp. 235-254.

Basawa, I.V. and Scott, D.J. (1980). Remarks on an asymptotic optimality Criterion for estimation in Stochastic Processes. *Sankhya: The Indian Journal of Statistics. Vol.* **42**, Series A, Pts. 3 and 4, pp. 262-271.

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

Brillinger, D.R., Guckenheimer, J. Guttorp, P.E. and Oster, G. (1980). Empirical modelling of population time series data: the case of age and density dependent vital rates. *Lectures on Mathematics in the Life Sciences*, (Amer. Math. Soc.), Vol. **13**, 65-90.

Chan, K.S. (1987). Strong Consistency of the Least Squares Estimators of a Threshold Autoregressive Model. *Technical Report.* City Polytechnic of Hong Kong and Univ. of Chicago U.S.A.

Chan, K.S. (1988). Consistency and Limiting Distribution of the Least Squares Estimator of a threshold Autoregressive Model. *Technical Report.* Dept. of Statistics, Univ. of Chicago, U.S.A.

Chan, W.S. and Tong, H. (1986). On test for non-linearity in time series analysis. *Journal of forecasting*, **5**, pp. 217-228.

Chan, K.S. and Tong, H. (1984). A note on sub-system stability and system stability. *J. Eng. Maths.* (China), **1**, 43-51.

Chan, K.S. and Tong, H. (1985). On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. Appl. Prob.*, **17**, 666-678.

Chan, K.S. and Tong, H. (1986a). A survey of the statistical analysis of univariate threshold autoregressive models. To appear in *Advance in Statistical Analysis and Statistical Computing: Theory and Application*, Ed. R.S. Mariano, JAI Press Inc., U.S.A.

Chan, K.S. and Tong, H. (1986b). A note on certain integral equations associated with non-linear time series analysis. *Probab. Th. Rel. Fields*, **73**, 153-158.

Chan, K.S., Moeanaddin, R. and Tong, H. (1988). Some Difficulties of Non-linear Time Series Modelling. *Technical Report*. Mathematical Institute, Univ. of Kent Canterbury, U.K.

Chan, K.S., Petruccelli, J.D., Tong, H. and Woolford, S.W. (1985). A multiple-threshold AR(1) model. *J. Appl. Prob.*, **22**, 267-279.

Chan, K.S. and Tong, H. (1988). On Likelihood Ratio Test for Threshold Autoregression. *Technical Report*. Mathematical Institute, Univ. of Kent Canterbury, U.K.

Copas, J.B. (1983). Regression, Prediction and Shrinkage. *J. Roy. Statist. Soc.* B, **45**, No. 3, pp. 311-354.

Cryer, J.D. (1986). *Time Series Analysis* , Duxbury Press, Boston.

Cryer, J.D. and Ledolter, J. (1981). Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika*, **68, 3**, pp. 691-694.

Draper, N.R. and Smith, H. (1981). *Applied Regression Analysis* 2nd ed. New York: J. Wiley.

Efron, B. (1979) Bootstrap method: another look at the jackknife. *Ann. Statist.*, **7**, 1-26.

Gabr, M.M. and Subba Rao, T. (1981). The estimation and prediction of subset bilinear time series models with applications. *J. Time Series Anal.*, **2**, 153-171.

Ghaddar, D.K. and Tong, H. (1981). Data transformation and self-exciting threshold autoregression. *J. Roy. Statis. Soc.*, C, **30**, 238-248.

Granger, C.W.J. and Andersen, A.P. (1978). *An Introduction to Bilinear Time Series Models*. Vandenhoeck & Rupercht.

Granger, C.W.J. and Newbold, P. (1986). *Forecasting Economic Time Series*. 2nd ed. Academic Press.

Haggan, V. and Ozaki, T. (1981). Modelling non-linear vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189-196.

Hau, M.C. and Tong, H. (1984). Outlier detection in autoregressive time series modelling. *Technical Report. No. 15*, Dept. of statistics, Chinese Univ. of Hong Kong, Hong Kong.

Hida, T. (1980). *Brownian Motion*. Springer-Verlag, New York, Heidelberg Berlin.

Hôgnàs, G. (1986). Comparison of some non-linear autoregressive processes. *J. Time Series Anal.*, **7**, pp. 205-212.

Jones, D.A. (1976). Non-linear autoregressive processes. Unpublished Ph.D. Thesis, Univ. of London, U.K.

Jones, D.A. (1978). Non-linear autoregressive processes. *Proc. R. Soc. London*, A, **360**, 71-95.

Kang, K.M. (1975). A Comparison of Estimators for Moving Average Processes. Unpublished Australian Bureau of Statistics.

Klimko, L.A. and Nelson, P.I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.*, **6**, 629-642.

Lasalle, J.P. (1976). *The Stability of Dynamical Systems*. Society for Industrial and Applied Mathematics.

Lawrance, A.J. (1987). Directionality and Reversibility in Time Series. *Technical Report*. Dept. of Statistics, Univ. of Birmingham, Birmingham, U.K

Lawrance, A.J. and Lewis, P.A.W. (1977). An exponential moving average sequence and point process, EMA(1). *J. Appl. Prob.*, **14**, 98-113.

Lawrance, A.J. and Lewis, P.A.W. (1980). The Exponential Autoregressive-Moving Average EARMA(p,q) process. *J. Roy. Statist. Soc., B*, **42**, No. 2, pp. 150-161.

Li, W.K. and McLeod, A.I. (1988). ARMA Modelling with Non-Gaussian Innovations. *J. Time Series Anal., Vol.*, **9**, No. 2, pp. 155-168.

Lim, K.S. (1987). A Comparative study of various univariate time series models for Canadian lynx data. *J. Time Series Anal.*, **8**, pp. 161-176.

Lin,C.C. and Mudholker, G.S. (1980). A simple test for normality against asymmetric alternatives. *Biometrika*, **67**, 455-461.

Liu, S.I. (1985). Theory of Bilinear Time Series Models, *Comm. Statist. Theor. Math.*, **14** (10), 2549-2561.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, **65**, 297-303.

McKenzie, E. (1984) Autoregressive-Moving Average Process with Negative Binomial and Geometric Marginal Distribution. *Technical Report*, Univ. of Strathclyde.

McLeod, A.I. and Li, W.K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Anal.*, **4**, 269-273.

Moeanaddin, R. and Tong, H. (1988). A comparison of likelihood ratio test and CUSUM test for threshold autoregression. *The Statistician*, **37**, 213-225.

Moeanaddin, R. and Tong, H. (1989). Numerical Evaluation of Distributions in Non-linear Autoregression. *Technical Report*, Mathematical Institute, Univ. of Kent at Canterbury, U.K.

Mohler, R.R. (1973). *Bilinear Control Processes*. Academic Press, New York and London.

Moran, P.A.P. (1984). Some experiments on the prediction of sunspot numbers. *J. Roy. Statist. Soc.*, B, **116**, 112-117.

Nelson, J.Z. and Van Ness, J.W. (1973). Choosing a Nonlinear Predictor. *Technometrics*, **15**, 219-231.

Neudecker, H. (1969). Some theorem on matrix differentiation with special references to Kronecker matrix products. *J. Amer. Statist. Assoc.*, **69**, 153-159.

Ozaki, T. (1982). The statistical analysis of perturbed limit cycle processes using nonlinear time series models. *J. Time Series Anal.*, **3**, 29-41.

Ozaki, T. and Oda, H. (1987). Non-linear time series model identification by Akaikae's Information Criterion. In *Information and system*, ed. B. Dabuisson, Pergamon Press, 83-91.

Pemberton, J. (1985). Contributions to the theory of non-linear time series models. Unpublished Ph.D. Thesis, Univ. of Manchester, U.K.

Pemberton, J. (1987). Exact least squares multi-step prediction from nonlinear autoregressive models. *J. Time Series Anal.*, **8**, pp. 443-448.

Pemberton, J. and Tong, H. (1981). A note on the distribution of non-linear autoregressive stochastic models. *J. Time Series Anal.*, **2**, 49-42.

Petruccelli, J.D. and Woolford, S.W. (1984). A threshold AR(1) model. *J. Appl. Prob,*. **21**, 270-286.

Petruccelli, J.D. and Davies, N. (1986). A Portmanteau test for self-exciting threshold autoregressive-type nonlinearity in time series. *Biometrika*, **73**, pp. 687-694.

Petruccelli, J.D (1987). On test for SETAR-type nonlinearity in time series. *Technical Report*, Worcester Polytechnic Institute, Worcester, MA, Worcester Polytechnic Institute.

Petruccelli, J.D (1988). On test for SETAR-type nonlinearity in time series. *Technical Report*, Worcester Polytechnic Institute, Worcester, MA, Worcester Polytechnic Institute.

Pham-Dinh Tuan and Lanh, Tat Tran (1981). On the first order bilinear time series model. *J. Appl. Prob.*, **18**, 617-627.

Priestley, M.B. (1980). State Dependent Models: A General Approach to Non-linear Time Series Analysis. *J. Time Series Anal.*, **1**, 47-71.

Priestley, M.B. (1988). *Non-linear and Non-stationary Time Series Analysis*. Academic Press, London.

Quinn, B.G. (1982). A note on the existence of strictly stationary solutions to bilinear equations. *J. Time Series Anal.*, **3**, 249-252.

Raftery, A.E. (1982). Generalized Non-Normal Time Series Models. *Time Series Analysis, Theory and Practice.*, **1**, Anderson ed. North Holland.

Robinson, P.M. (1983). Nonparametric Estimators for Time Series. *J. Time Series Anal., Vol.* **4**, No. 3. 185-207.

Rosenblatt, M. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Statist.*, **27**, 832-837.

Silverman, B.W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc.*, B, **43**, 97-99.

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, New York.

Subba Rao, T. (1981). On the theory of bilinear models. *J. Roy. Statist. Soc. Ser. B,* **43,** 244-255.

Subba Rao, T. and Gabr, M.M. (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models.* Springer-verlag, New York Berlin Heidelberg Tokyo.

Tjostheim, D. (1986). Estimation in Nonlinear Time Series Models. *Stochastic Processes and their Applications,* **21,** 251-273.

Tong, H. (1977a). Discussion of a paper by A.J. Lawrance and N.T. Kottegoda. *J. Roy. Statist. Soc.,* A, **140,** 34-35.

Tong, H. (1977b). Some comments on the Canadian lynx data - with discussion. *J. Roy. Statist. Soc.,* A, **140,** 432-435, 448-468.

Tong, H. (1978) On threshold model . *Pattern Recognition and Signal Processing* (ed. C. H. Chen), The Netherlands: Sijthoff and Noordhoof.

Tong, H. (1981). A note on a Markov bilinear stochastic process in discrete time. *J. Time Series Anal.,* **2,** 279-284.

Tong, H. (1983). *Threshold Models in Non-linear Time Series Analysis.* Springer Verlage, New York.

Tong, H. (1987a). Non-linear Time Series Models of Regularly Sampled Data: a review, in: Y.U. PROHOROV & V.V. SASANOV (Eds) *First World Congress of the Bernoulli Society,* 2, pp. 355-367, September 1986, Tashkend, USSR. (Holland, VNU Science Press).

Tong, H. (1987b). Threshold, stability, non-linear forecasting and irregularly sampled data. Invited paper to be included in: *Statistical Analysis and Forecasting Economic Structural Chang.* IIASA, ed. P. Hackl (1987).

Tong, H. and Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc.,* **42,** 245-292.

Tong, H. and Moeanaddin, R. (1987a). On the non-monotonicity of the mean-squared error of multi-step-ahead prediction. Abstract no. 87t-24, IMS *Bulletin,* 16, no. 1, Issue no. 90, Tanuary 1987.

Tong, H. and Moeanaddin, R. (1987b). A note on the conditional variance of the error of a multi-step non-linear least squares prediction. *Technical Report,* Mathematical Institute, Univ. of Kent at Canterbury, U.K.

Tong, H. and Moeanaddin, R. (1988). On multi-step non-linear least squares prediction. *The Statistician,* **37,** pp. 101-110.

Tong, H. and Wu, Z.M. (1982). Multi-step-ahead forecasting of cyclical data by threshold autoregression. In *Time Series Analysis: Theory and Practice I* (ed. O.D. Anderson), 733-753, Amesterdam: North Holland. Tuan, Pham-Dinh and Lanb, Tat Tran (1981). On the first order bilinear time series model. *J. Appl. Prob.,* **18,** 617-627.

Tsay, R.S. (1987). Testing and modelling threshold autoregressive processes. *Unpublished manuscript.*

Tsay, R.S. (1988). Non-linear Time Series Analysis of Blowfly population. *J. Time Series Anal.,* **9,** 247-263.

Wang, S.R. , An, H.Z. and Tong, H. (1983). On the distribution of a simple stationary bilinear Processes. *J. Time Series Anal.,* **4,** 209-216.

Wiener, N. (1958). *Nonlinear Problems in Random Theory.* John Wiley and Sons.

Weiss, G. (1975). Time-reversibility of linear stochastic processes. *J. Appl. Prob.,* **12,** 831-836.

Wilkinson, J.H. (1965). The Algebraic Eigenvalue Problem. Oxford University Press, *Amen House, London E.C.4*

Yule, G.U. (1927). On a method of investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Soc.,* London, Series A, **226,** 267-298.

```
C    IN ALL THESE PROGRAMMES SOME DIFFERENT NAG ROUTINE ARE
C    USED. C    THIS PROGRAM SIMULATES N DATA POINTS FROM EXPAR(1)
C       MODEL WHICH HAS BEEN USED IN CHAPTER 4
C    THE CONTOUR OF A PAIR OF COEFFICIENTS WHICH WILL BE
C       SPECIFIED IS PLOTTED
     REAL X(2000)
     REAL *8 XMIN,XMAX,YMIN,YMAX
     REAL *8 XM,SD,G05DDF,SURFCE(200,200),CHTS(200)
     INTEGER T
     CHARACTER *80 HEAD
     LOGICAL UNUSED(200,200)
     EXTERNAL J06GBY,J06GBV
C    N: SAMPLE SIZE
     READ(5,*)N
C    XM: MEAN OF ERROR TERM
C    SD: VARIANCE OF ERROR TERMS
     XM=0.0
     SD=1.0
C    K: PARAMETER WHICH CONTROLS THE RANDOM NUMBER GENERATOR
     READ(5,*)K
     PRINT*,K
C    TO INITIALIZE THE RANDOM NUMBER GENERATOR
     CALL G05CBF(K)
     XXX=999999999.0
     NOUT=6
     PRINT*,' N=',N
C    BC,CC,DC,GC: THE COEFFICIENTS OF EXPAR(1) MODEL
     READ(5,*)BC,CC,DC,GC
     PRINT*,BC,CC,DC,GC
     X(1)=0.0
     DO 1 I=2,1500+N+1
     F=EXP(-GC*X(I-1)*X(I-1))
     X(I)=(BC+(CC+DC*X(I-1))*F)*X(I-1)+G05DDF(XM,SD)
1    CONTINUE
     DO 13 I=1,N
13   X(I)=X(I+1499)
     SS=0.0
     SUM1=0.0
     DO 105 I=1,N
     SUM1=SUM1+X(I)**2
105  SUM=SUM+X(I)
     XB=SUM/FLOAT(N)
```

```
      XBB=SUM1/FLOAT(N)
      PRINT*,' XBAR=',XB
      PRINT*,' XBAR**2 =',XBB
      I=0
C     XMIN,XMAX,XINC: THE RANGE OF FIRST PARAMETER
      READ(5,*)XMIN,XMAX,XINC
      PRINT*,XMIN,XMAX,XINC
C     YMIN,YMAX,YINC: THE RANGE OF SECOND PARAMETER
      READ(5,*)YMIN,YMAX,YINC
      PRINT*,YMIN,YMAX,YINC
      B=BC
      C=CC
      G=GC
      D=DC
C     THE FIRST PARAMETER
      DO 2 D=XMIN,XMAX,XINC
      J=0
      I=I+1
C     THE SECOND PARAMETER
      DO 2 G=YMIN,YMAX,YINC
      J=J+1
      SUM=0.0
      DO 3 T=2,N
C     F=EXP(-G*X(I-1)*X)I-1)): NON-CENTRE MODEL
C     F=EXP(-G*(X(I-1)**2-XBB)): CENTRE MODEL
C     CHOOSE THE APROPRIATE ONE
C     F=EXP(-G*X(T-1)*X(T-1))
      F=EXP(-G*(X(T-1)**2-XBB))
      S=(B+(C+D*X(T-1))*F)*X(T-1)
      S=X(T)-S
      SUM=SUM+S*S
3     CONTINUE
C     PRINT*,' I=',I,' J=',J,' SUM=',SUM
C     THIS PRINT GIVE YOU A GOOD IDEA OF CHOOSING THE HEIGHTS
      SURFCE(I,J)=SUM
      IF(SUM.LT.XXX) THEN
      XXX=SUM
      P=G
      ENDIF
2     CONTINUE
      PRINT*,' MIN=',XXX,' G HAT=',P
C     MIN: THE MINIMUM HEIGHT THAT CAN BE PLOTTED
```

```
      PRINT*,' I=',I
      PRINT*,'J=',J
      MDIM=200
      MA=1
      MB=I
      NA=1
      NB=J
      ILAB=1
      IHIGH=0
      CALL X04AAF(1,NOUT)
      CALL XXXXXX
      CALL J06WAF
      CALL J06WBF(XMIN,XMAX,YMIN,YMAX,1)
      CALL J06WCF(0.0,1.0,0.0,1.0)
      CALL J06AAF
C     HEAD: THE TITLE OF CONTOUR
      READ(*,5)HEAD
5     FORMAT(A80)
      CALL J06AHF(HEAD,80)
C     d: THE LABEL OF FIRST AXIS
      CALL J06AJF(1,' d ')
C     g: THE LABEL OF SECOND AXIS
      CALL J06AJF(2,' g ')
C     NCHTS: NO. OF CURVES IN CONTOUR PLOT
      READ(5,*)NCHTS
      IGRID=0
C     ICH=0
C     ICH: A CODE FOR PLOT
C       =0 THE HEIGHTS OF CURVES WILL BE CALCULATED BY ROUTINE
C       =1 THE HEIGHTS OF CURVES MUST BE SPECIFIED BY USER
      READ(5,*)ICH
      PRINT*,'ICH=',ICH
      IF(ICH.NE.0) READ(5,*) (CHTS(I1),I1=1,NCHTS)
      IFAIL=3
      CALL J06GBF(SURFCE,MDIM,MA,MB,NA,NB,NCHTS,CHTS,ICH,
     1        J06GBY,ILAB,IHIGH,J06GBV,IGRID,UNUSED,IFAIL)
      PRINT*,' IFAIL=',IFAIL
      IFORM=2
      NWIDTH=8
      NDP=2
      CALL J06WDF
      CALL J06GZF(CHTS,NCHTS,IFORM,NWIDTH,NDP)
```

```
CALL J06WZF
STOP
END
```

```
C    THIS PROGRAMME ESTIMATES THE STATIONARY MARGINAL
C    PROBABILITY DENSITY FUNCTION OF SETAR(2;1,1) BY
C    USING THE CHAPMAN-KOLMOGOROV RELATION
C    THE MATRIX SQUARING METHOD IS EMPLOYED TO
C    ACCELERATE THE CONVERGENCE
C    THIS PROGRAMME ALSO CALCULATER THE EIGENVALUES
C    OF THE MATRIX WHICH IS EXPLAINED IN CHAPTER 5
     REAL *8 GH(200),AM,B,X02AAF,TOL
     REAL *8 RR(200),RI(200)
     INTEGER INTGER(200)
     REAL *8 GD(200,200),ZZ(200,200)
     REAL *8 W(200),X(200)
     REAL *8 Z(200),X01AAF,P
     REAL*8 WW(100),A(100)
     INTEGER H
     EXTERNAL D01BAX
     EXTERNAL D01BAW
     COMMON/AA/PP
     COMMON/CC/S
     COMMON/BB/A0,A1,B0,B1,R
     P=X01AAF(P)
     PP=1.0/SQRT(2.0*P)
     TOL=X02AAF(TOL)
     TOL=0.000001
C    TOL: THE CONVERGENCE CRITERION
     PRINT*,'  A0   A1   B0   B1'
C    A0 , A1 , B0 , B1 : THE SETAR COEFFICIENTS IN THE
C       FIRST AND SECOND REGIME RESPECTIVELY
     READ(5,*)A0,A1,B0,B1
     PRINT*,' A0=',A0,'A1=',A1,'B0=',B0,' B1=',B1
     PRINT*,'  R'
C    R: THRESHOLD
     READ(5,*)R
     PRINT*,' R=',R
     XXX=99999999999.0
     PRINT*,'  X0    S     M '
C    X0: X(0) THE GIVEN VALUE FOR CALCULATING THE CONDITIONAL
C              DENSITIES
C    S: THE STANDARD DEVIATION OF ERROR TERM
C    M: NO. OF POINTS/2
     READ(5,*)X0,S,M
     IH1=2000
```

```
      PRINT 93,X0,S,M
      PRINT*,'  B'
93    FORMAT(' INITIAL VALUE=',F6.3,' SD.=',F5.2,' NO. OF POINTS/2=',I5
C     B: LAG ROUTINE PARAMETER
      READ(5,*)B
      PRINT*,' B=',B
      PRINT*,'   AM'
C     AM: LOCATION PARAMETER TO SHIFT ALL THE POINTS
      READ(5,*)AM
      PRINT*,' AM=',AM
      ITYP=1
      IFAIL=0
C     CALL D01BBF(D01BAW,AM,B,ITYP,M,W,X,IFAIL)
      CALL D01BBF(D01BAX,AM,B,ITYP,M,WW,A,IFAIL)
      PRINT*,' IFAIL=',IFAIL
C     PRINT 1,(I,W(I),X(I),I=1,M)
1     FORMAT(1X,I4,5X,F8.3,5X,F8.3)
      DO 116 I=1,M
      IM=M-I+1
      X(IM)=A(I)
      W(IM)=WW(I)
116   CONTINUE
      B=-B
      CALL D01BBF(D01BAX,AM,B,ITYP,M,WW,A,IFAIL)
      PRINT*,' D01BBF IFAIL ',IFAIL
      DO 118 I=1,M
      IM=I+M
      X(IM)=A(I)
      W(IM)=WW(I)
118   CONTINUE
      N=M
C     CALL D01BBF(D01BAW,AM,B,ITYP,M,W,X,IFAIL)
      M=2*M
      PRINT 2,X(1),X(M)
2     FORMAT(1X,'X(1)=',F8.3,5X,'X(M)=',F8.3)
      CALL SIMU
C     THIS SUBROUTINE GIVES A ROUGH IDEA ABOUT THE CUT OFF
C     POINTS FOR CHOOSING B AND AM
      T1=SECNDS(0.0)
      ANS=G(X0)
C     G(.): THE SKELETON OF SETAR MODEL
      PRINT*,' ANS=',ANS
```

```
C    ANS: THE ONE STEP-AHEAD PREDICTION GIVEN X(0)=X0
     DO 10 I=1,M
     FB=(X(I)-ANS)/S
     GH(I)=PHI(FB)/S
     Z(I)=GH(I)
10   CONTINUE
C
     DO 901 I=1,M
     XI=X(I)
     ANS=G(XI)
     DO 902 J=1,M
     FB=(X(J)-ANS)/S
     GD(J,I)=PHI(FB)/S
902  CONTINUE
901  CONTINUE
C
C
     DO 903 I=1,M
     SUM=0.0
     DO 904 J=1,M
     SUM=SUM+W(J)*GD(J,I)
904  CONTINUE
     DO 905 J=1,M
     SUM=1.0
     GD(J,I)=GD(J,I)/SUM
     ZZ(J,I)=GD(J,I)
905  CONTINUE
903  CONTINUE
C
C    CALCULATE THE EIGENVALUES OF MATRIX GD
     IFAIL=0
     IA=200
     CALL F02AFF(ZZ,IA,M,RR,RI,INTGER,IFAIL)
     PRINT*,' IFAIL=',IFAIL
     PRINT 864,(RR(I),RI(I),I=1,M)
864  FORMAT(1X,F9.6,5X,F9.6)
C
C
     DO 40 H=2,IH1
     SS=0.0
     DO 50 I=1,M
     SS=SS+W(I)*GH(I)
```

```
50   CONTINUE
     XM=0.0
     PRINT*,' SS=',SS
     DO 60 I=1,M
     GH(I)=GH(I)/SS
     IF(GH(I).GT.XM) XM=GH(I)
60   CONTINUE
C
C
C
     ANS=0.0
     DO 80 I=1,M
     ANS=ANS+W(I)*G(X(I))*GH(I)
80   CONTINUE
     SUM=0.0
     DO 90 I=1,M
     SUM=SUM+W(I)*(G(X(I))-ANS)**2*GH(I)
90   CONTINUE
     S1=S**2+SUM
     S1=SQRT(S1)
C
C
C
     PRINT*,' MEAN =',ANS,' SD=',S1
     IF (H.LE.2) GO TO 441
     DO 442 I=1,M
       IF(ABS(GH(I)-Z(I))/XM .GT.TOL) GO TO 443
442    CONTINUE
     GO TO 132
443    CONTINUE
441    CONTINUE
     DO 444 I=1,M
444    GH(I)=Z(I)
C
     DO 100 I=1,M
     DO 100 K=1,M
     SUM=0.0
     DO 110 J=1,M
     SUM=SUM+W(J)*GD(I,J)*GD(J,K)
110  CONTINUE
     ZZ(I,K)=SUM
100  CONTINUE
```

```
      DO 111 I=1,M
      SUM=0.0
      DO 112 J=1,M
      SUM=SUM+W(J)*ZZ(J,I)
112   CONTINUE
      DO 113 J=1,M
      GD(J,I)=ZZ(J,I)/SUM
113   CONTINUE
111   CONTINUE
      DO 114 I=1,M
      Z(I)=GD(I,N)
114   CONTINUE
C
      GO TO 32
132   CONTINUE
      T2=SECNDS(T1)
      PRINT 33,(X(II),W(II),GH(II),II=1,M)
      WRITE(7,33)(X(II),W(II),GH(II),II=1,M)
      PRINT*,' SD =',S1
33    FORMAT(1X,F12.6,5X,5X,F10.6,5X,F10.6)
C
C
C
      SUM=0.0
      DO 51 I1=1,M
51    SUM=SUM+X(I1)*W(I1)*GH(I1)
      WRITE(7,*),' MEAN XT IS ',SUM
      PRINT*,' MEAN XT IS ',SUM
      XBAR=SUM
      SUM=0.0
      DO 52 I1=1,M
52    SUM=SUM+(X(I1)-XBAR)**2*W(I1)*GH(I1)
      WRITE(7,*) ' VARIANCE XT IS ',SUM
      PRINT*,' VARIANCE XT IS ',SUM
      SIGMA=SQRT(SUM)
      SUM=0.0
      DO 53 I1=1,M
53    SUM=SUM+(X(I1)-XBAR)**3*W(I1)*GH(I1)
      WRITE(7,*)' E(X-XBAR)**3=',SUM
      XKOR=SUM/(SIGMA**3)
      WRITE(7,*) ' SKEWNESS IS ',XKOR
      PRINT*,' SKEW IS ',XKOR
```

```
      SUM=0.0
      DO 54 I1=1,M
54    SUM=SUM+(X(I1)-XBAR)**4*W(I1)*GH(I1)
      SK=SUM/(SIGMA**4)-3.0
      WRITE(7,*)' E(X-XBAR)**4=',SUM
      WRITE(7,*) ' KURTESISS IS ',SK
      PRINT*,' KUR IS ',SK
      PRINT*,' TIME=',T2
      STOP
32    CONTINUE
C
C
40    CONTINUE
      STOP
      END
C
C
      REAL FUNCTION PHI(Z)
      COMMON/AA/PP
      PHI=EXP(-Z**2/2.0)*PP
      RETURN
      END
C
C
C
      REAL FUNCTION G(X)
      COMMON/BB/A0,A1,B0,B1,R
      IF(X.LE.R) THEN
          G=A0+A1*X
      ELSE
          G=B0+B1*X
      ENDIF
      RETURN
      END
C
C
C
C
C
      SUBROUTINE SIMU
      REAL*8 X(7000),WT(5000),SD,XM,XMEAN,XMIN,XMAX,S2,S3,S4,WTSUM
      COMMON/CC/S
```

```
      COMMON/BB/A0,A1,B0,B1,R
      XM=0.0
      SD=S
      X(1)=0.0
      N=5000
      CALL G05CCF
      DO 10 I=2,N+2000
      IF(X(I-1).LE.R) THEN
        X(I)=A0+A1*X(I-1)+G05DDF(XM,SD)
      ELSE
        X(I)=B0+B1*X(I-1)+G05DDF(XM,SD)
      ENDIF
10    CONTINUE
      DO 20 I=1,N
      X(I)=X(I+1999)
20    CONTINUE
      IFAIL=0
      J=0
      CALL G01AAF(N,X,J,WT,XMEAN,S2,S3,S4,XMIN,XMAX,WTSUM,IFAIL)
      PRINT*,' G01AAF IFAIL=',IFAIL
      PRINT 1,XMEAN,S2,XMIN,XMAX
1     FORMAT(1X,'MEAN=',F8.4,5X,' SD=',F8.4,5X,' MIN=',F8.4,
     1     5X,' MAX=',F8.4)
      RETURN
      END
```

```
C     THIS PROGRAMME GIVES THE DATA, THE PERIOD OF FORECASTING
C     AND THE SETAR(2,P1,P2) MODEL AND CALCULATE THE MEAN OF
C     M-STEP-AHEAD FORECASTING ERROR, M=1,2,...,
C     THIS PROGRAM CAN BE USED TO CALCULATE THE M-STEP AHEAD
C     FORECASTING BY ADDING SUFFICIENT ZERO TO THE END OF DATA SET
C     THEN IN EACH STEP, THE FIRST PREDICTED VALUE IS THE POINT
C     FORECAST FOR THAT GIVEN STEP
      IMPLICIT DOUBLE PRECISION (A-H,O-Y)
      INTEGER T,LO,HO,PER,CODE,D
      DIMENSION X(300),E(300),A(0:12),B(0:12),WT(1000)
      DIMENSION P(12,35)
C     THE MAXIMUM ORDER IS 11 AND THE MAXIMUM STEP IS 35
C     CLEARLY THESE RESTRICTION CAN BE REMOVED BY CHANGING
C     THE SIZE OF ARRAIES
      EXTERNAL XINV
C     XINV: THE INVERSE FUNCTION OF SQUARE ROOT TRANSFORMATION
C        FOR SUNSPOT NUMBERS
      COMMON/BB/CODE,XBAR
      COMMON/AA/A,B,R,D,LO,HO,IOR,S1,S2,PER
      READ(5,*)LO
C     LO: THE ORDER OF LOWER REGIME
      READ(5,*)(A(I),I=0,LO)
C     A(I): THE COEFFICIENTS OF THE FIRST REGIME
      READ(5,*) HO
C     HO: THE ORDER OF SECOND REGIME
      READ(5,*) (B(I),I=0,HO)
C     B(I): THE COEFFICIENTS OF SECOND REGIME
      PRINT 1,(A(I),I=0,LO)
1     FORMAT(20X,'LOWER REGION '/,5X,13(F7.4,2X))
      WRITE(1,1) (A(I),I=0,LO)
      PRINT 2,(B(I),I=0,HO)
2     FORMAT(20X,' UPPER REGION '/,5X,13(F7.4,2X))
      WRITE(1,2)(B(I),I=0,HO)
      READ(5,*)R,S1,S2,D
C     R: THRESHOLD
C     S1: THE STANDARD DEVIATION OF ERROR TERM IN THE FIRST REGIME
C     S2: SIMILAR TO S1 IN SECOND REGIME
C     D: DELAY
      PRINT 3,R,S1,S2,D
3     FORMAT(1X,'THRESHOLD=',F7.4,5X,'S1=',F7.3,4X,'S2=',F7.3,
     1 5X,'DELAY=',I2)
      WRITE(1,3)R,S1,S2,D
```

```
      READ(5,*)N,PER,CODE
C     N: NO. OF DATA
C     PER: NO. OF FORECASTING STEPS
C     CODE: TRANSFORMATION CODE
C       0: NO TRANSFORMATION
C       1: SQUARE ROOT TRANSFORMATION
C       2: LOG IN BASE 10 TRANSFORMATION
      READ(5,*)(X(I),I=1,N+PER)
      IF(CODE.EQ.0) THEN
      PRINT *,' NO TRANSFORMATION'
      WRITE(1,6)
      GO TO 10
      ENDIF
6     FORMAT(1X,' NO TRANSFORMATION ')
      IF(CODE.EQ.1) THEN
      PRINT *,' SQRT TRANSFORMATION'
      WRITE(1,*) ' SQRT TRANSFORMATION'
      DO 20 I=1,N+PER
20    X(I)=2.0*(DSQRT(X(I)+1.0)-1.0)
      GO TO 10
      ENDIF
      IF(CODE.EQ.2) THEN
      PRINT *,' LOG10 TRANSFORMATION'
      WRITE(1,*) 'LOG10 TRANSFORMATION'
      DO 30 I=1,N+PER
30    X(I)=DLOG10(X(I))
      GO TO 10
      ENDIF
      PRINT*,'CODE=',CODE,'ERROR'
      STOP
10    CONTINUE
      IOR=MAX(LO,HO)
      PRINT*,' IOR=',IOR
      WRITE(1,7)IOR
7     FORMAT(1X,' ORDER=',I3)
      SUM=0.0
      DO 40 I=1,N
40    SUM=SUM+X(I)
      XBAR=SUM/FLOAT(N)
      PRINT 8,XBAR
      WRITE(1,8)XBAR
8     FORMAT(1X,' MEAN OF DATA=',F10.5)
```

```
      DO 50 T=IOR+1,N
      IF(X(T-D).LE.R) THEN
      SUM=0.0
      DO 60 I=1,LO
60    SUM=SUM+A(I)*X(T-I)
      Y=SUM+A(0)
      ELSE
      SUM=0.0
      DO 70 I=1,HO
70    SUM=SUM+B(I)*X(T-I)
      Y=SUM+B(0)
      ENDIF
      E(T-IOR)=X(T)-Y
50    CONTINUE
      IFAIL=0
      IJ=0
      M=N-IOR
      CALL G01AAF(M,E,IJ,WT,XM,S,S3,S4,XMIN,XMAX,WTS,IFAIL)
      PRINT *,' IFAIL=',IFAIL
      PRINT 4,XM,S**2
C     XM MUST BE CLOSE TO ZERO AND S**2 CLOSE TO THE POOLED VARIANCE
C       IF NOT SOMETHING IS WRONG IN INPUT DATA
      WRITE(1,4)XM,S**2
      PRINT *,' M=',M
      WRITE(1,9)M
9     FORMAT(1X,' NO. OF ERRORS=',I4)
4     FORMAT(1X,'MEAN OF ERRORS=',F10.5,5X,' VAR. OF ERRORS=',F15.5)
C     TO CALCULATE THE EXACT 1 TO D STEP-AHEAD FORECASTING
      DO 99 J=1,D
99    CALL FOR(X,N,P,J)
C     TO USE THE MONTE CARLO METHOD TO CALCULATE THE
C     D+1 TO PER STEP-AHEAD FORECASTING
      CALL FORE(X,N)
      STOP
      END
C
C
      SUBROUTINE FORE(X,N)
      IMPLICIT DOUBLE PRECISION (A-H,O-Y)
      DIMENSION X(300),A(0:12),B(0:12)
      EXTERNAL XINV
      INTEGER CODE
```

```
      INTEGER LO,HO,PER,D
      COMMON/AA/A,B,R,D,LO,HO,IOR,S1,S2,PER
      COMMON/BB/CODE,XBAR
      DIMENSION FOR1(1000),WT(2000),FOR(1000,12),XM(100,12)
     1 ,X0(-11:12),P(12,35),SM(12),DM(12),SD(100,12)
      S1=DSQRT(S1)
      S2=DSQRT(S2)
      NN=35
      MM=1000
C     MM: NO. OF REPLICATION IN MONTE CARLO METHOD
      N1=12
      CALL G05CCF
      DO 10 J=1,PER
      M=N+J-1
      DO 11 I=1,NN
      DO 11 K=1,N1
      XM(I,K)=0.0
      SD(I,K)=0.0
11    CONTINUE
      DO 12 I=0,-IOR+1,-1
      X0(I)=X(M+I)
12    CONTINUE
      DO 555 KK=1,NN
      DO 40 I=1,MM
      CALL GEN(X0,N1)
      DO 50 KJ=1,N1
      IF(CODE.EQ.1) THEN
      FOR(I,KJ)=XINV(X0(KJ))
      ELSE
      FOR(I,KJ)=X0(KJ)
      ENDIF
50    CONTINUE
40    CONTINUE
      DO 200 KJ=1,N1
      DO 300 I=1,MM
      FOR1(I)=FOR(I,KJ)
300   CONTINUE
      IFAIL=0
      IJ=0
      CALL G01AAF(MM,FOR1,IJ,WT,XMEAN,S,S3,S4,XMIN,XMAX,WTSUM,IFAIL)
      IF(IFAIL.NE.0) THEN
      PRINT*,' G01AAF FAIL',IFAIL
```

```
      STOP
      ENDIF
C     PRINT 111,KJ,XMEAN,S
111   FORMAT(2X,I3,3X,F10.4,5X,F10.6)
      XM(KK,KJ)=XMEAN
      SD(KK,KJ)=S
200   CONTINUE
555   CONTINUE
      DO 333 K=1,N1
      SM(K)=0.0
      DM(K)=0.0
333   CONTINUE
      DO 666 IJ=1,N1
      DO 777 I=1,NN
      SM(IJ)=SM(IJ)+XM(I,IJ)
      DM(IJ)=DM(IJ)+SD(I,IJ)
777   CONTINUE
666   CONTINUE
      DO 100 M2=1,12
100   P(M2,J)=SM(M2)/FLOAT(NN)
10    CONTINUE
      DO 15 J2=1,PER
      WRITE(1,19)J2,(P(I,J2),I=1,12)
15    CONTINUE
19    FORMAT(1X,I2,2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2,
     1 2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2,2X,F8.2)
      DO 102 J=1,12
      IF(CODE.EQ.1) THEN
      SUM=0.0
      SUM1=0.0
      SUM2=0.0
      L1=0
      L2=0
      DO 1001 I=N+J,N+PER
C     IF(X(I).LE.XBAR) THEN
      IF(P(J,I-N-J+1).LE.XINV(XBAR)) THEN
      L1=L1+1
      SUM1=SUM1+(XINV(X(I))-P(J,I-N-J+1))**2
      ELSE
      L2=L2+1
      SUM2=SUM2+(XINV(X(I))-P(J,I-N-J+1))**2
      ENDIF
```

```
1001 SUM=SUM+(XINV(X(I))-P(J,I-N-J+1))**2
     ELSE
     SUM=0.0
     SUM1=0.0
     SUM2=0.0
     L1=0
     L2=0
     DO 101 I=N+J,N+PER
C    IF(X(I).LE.XBAR) THEN
     IF(P(J,I-N-J+1).LE.XBAR) THEN
     L1=L1+1
     SUM1=SUM1+(X(I)-P(J,I-N-J+1))**2
     ELSE
     L2=L2+1
     SUM2=SUM2+(X(I)-P(J,I-N-J+1))**2
     ENDIF
101  SUM=SUM+(X(I)-P(J,I-N-J+1))**2
     ENDIF
     WRITE(1,21)J,SUM,SUM/FLOAT(PER-J+1)
21   FORMAT(1X,'SUM OF FOR',I2,'=',F15.5,5X,'SIGMA=',F15.5)
     WRITE(1,22)J,SUM1,SUM1/FLOAT(L1)
     WRITE(1,23)J,SUM2,SUM2/FLOAT(L2)
     WRITE(1,24)L1,L2
22   FORMAT(1X,' LOWER THAN MEAN',2X,'SUM OF FOR',I2,'=',F15.5,
    1 5X,' SIGMA=',F15.5)
23   FORMAT(1X,' UPPER THAN MEAN',2X,'SUM OF FOR',I2,'=',F15.5,
    1 5X,' SIGMA=',F15.5)
24   FORMAT(3X,'L1=',I3,4X,'L2=',I3)
102   CONTINUE
     RETURN
     END
C
C-------------------------------------------------------
C
     SUBROUTINE GEN(X,N)
     IMPLICIT DOUBLE PRECISION (A-H,O-Y)
     DIMENSION X(-11:12),A(0:12),B(0:12)
     INTEGER T,LO,HO,D,PER
     COMMON/AA/A,B,R,D,LO,HO,IOR,S1,S2,PER
     XM=0.0
     DO 10 T=1,N
     IF(X(T-D).LE.R) THEN
```

```
      SUM=0.0
      DO 20 I=1,LO
20    SUM=SUM+A(I)*X(T-I)
      X(T)=A(0)+SUM+G05DDF(XM,S1)
      ELSE
      SUM=0.0
      DO 30 I=1,HO
30    SUM=SUM+B(I)*X(T-I)
      X(T)=B(0)+SUM+G05DDF(XM,S2)
      ENDIF
10    CONTINUE
      RETURN
      END
C
C
C-----------------------------------------------------------------
C
      SUBROUTINE FOR(X,N,P,J)
      IMPLICIT DOUBLE PRECISION (A-H,O-Y)
      DIMENSION X(300),A(0:12),B(0:12)
      DIMENSION P(12,35)
      INTEGER LO,HO,PER,D,CODE
      COMMON/BB/CODE,XBAR
      EXTERNAL XINV
      COMMON/AA/A,B,R,D,LO,HO,IOR,S1,S2,PER
      S=0.0
      SUM1=0.0
      SUM2=0.0
      L1=0.0
      L2=0.0
      DO 10 I=N+J,N+PER
      T=I-J
      IF(X(T-D+J).LE.R) THEN
      SUM=0.0
      DO 20 II=J,LO
20    SUM=SUM+A(II)*X(T-II+J)
      DO 30 II=1,J-1
30    SUM=SUM+A(II)*P(J-II,I-N-J+1)
      Y=A(0)+SUM
      ELSE
      SUM=0.0
      DO 40 II=J,HO
```

```
40   SUM=SUM+B(II)*X(T-II+J)
     DO 50 II=1,J-1
50   SUM=SUM+B(II)*P(J-II,I-N-J+1)
     Y=B(0)+SUM
     ENDIF
     P(J,I-N-J+1)=Y
     IF(CODE.EQ.1) THEN
     S=S+(XINV(X(T+J))-XINV(Y))**2
C    IF(X(T+J).LE.XBAR) THEN
     IF(Y.LE.XBAR) THEN
     L1=L1+1
     SUM1=SUM1+(XINV(X(T+J))-XINV(Y))**2
     ELSE
     L2=L2+1
     SUM2=SUM2+(XINV(X(T+J))-XINV(Y))**2
     ENDIF
     PRINT 1,I-N-J+1,XINV(Y),XINV(X(T+J))
     WRITE(1,1)I-N-J+1,XINV(Y),XINV(X(T+J))
     ELSE
     S=S+(X(T+J)-Y)**2
C    IF(X(T+J).LE.XBAR) THEN
     IF(Y.LE.XBAR) THEN
     L1=L1+1
     SUM1=SUM1+(X(T+J)-Y)**2
     ELSE
     L2=L2+1
     SUM2=SUM2+(X(T+J)-Y)**2
     ENDIF
     PRINT 1,I-N-J+1,Y,X(T+J)
     WRITE(1,1)I-N-J+1,Y,X(T+J)
     ENDIF
10   CONTINUE
     PRINT 2,J,S,S/FLOAT(PER-J+1)
2    FORMAT(1X,'SUM OF FOR',I1,'=',F15.5,5X,'SIGMA=',F15.5)
     WRITE(1,2)J,S,S/FLOAT(PER-J+1)
     WRITE(1,3)J,SUM1,SUM1/FLOAT(L1)
     WRITE(1,4)J,SUM2,SUM2/FLOAT(L2)
     WRITE(1,5)L1,L2
3    FORMAT(1X,' LOWER THAN MEAN',' SUM OF FOR',I2,'=',F15.5,
    1 5X,'SIGMA=',F15.5)
4    FORMAT(1X,' UPPER THAN MEAN',' SUM OF FOR',I2,'=',F15.5,
    1 5X,'SIGMA=',F15.5)
```

```
5    FORMAT(1X,' L1=',I3,5X,'L2=',I3)
1    FORMAT(1X,I4,4X,F15.5,4X,F15.5)
     RETURN
     END
C
C
C-------------------------------------------------
C
     DOUBLE PRECISION FUNCTION XINV(X)
     DOUBLE PRECISION X
     XINV=(((X+2.0)/2.0)**2)-1.0
     RETURN
     END
```

KENT LIBRARY UNIVERSITY