Non-linear Time Series

A Dynamical System Approach

HOWELL TONG

Mathematical Institute University of Kent at Canterbury

OXFORD STATISTICAL SCIENCE SERIES

A. C. Atkinson: *Plots, Transformations, and Regression*
M. Stone: *Coordinate-Free Multivariable Statistics*
W. J. Krzanowski: *Principles of Multivariate Analysis: A User's Perspective*
M. Aitkin, D. Anderson, B. Francis, and J. Hinde: *Statistical Modelling in GLIM*
Peter J. Diggle: *Time Series: A Biostatistical Introduction*
Howell Tong: *Non-linear Time Series: A Dynamical System Approach*

Non-linear Time Series

A Dynamical System Approach

HOWELL TONG

Mathematical Institute, University of Kent at Canterbury

CLARENDON PRESS OXFORD
1990

Oxford University Press, Walton Street, Oxford OX2 6DP

Oxford New York Toronto
Delhi Bombay Calcutta Madras Karachi
Petaling Jaya Singapore Hong Kong Tokyo
Nairobi Dar es Salaam Cape Town
Melbourne Auckland

and associated companies in
Berlin Ibadan

Oxford is a trade mark of Oxford University Press

Published in the United States
by Oxford University Press, New York

British Library Cataloguing in Publication Data
Tong, Howell
Non-linear time series: a dynamical system approach.
1. Non-linear time series. Analysis
I. Title
519.5'5
ISBN 0-19-852224-X

Library of Congress Cataloging in Publication Data
Tong, Howell.
Non-linear time series: a dynamical system approach / Howell Tong.
p. cm. -- (Oxford statistical science series; 6)
1. Time-series analysis. 2. Nonlinear theories. I. Title.
II. Series.
QA280.T597 1990 519.5'5--dc20 89-29697
ISBN 0-19-852224-X

Set by
The Universities Press (Belfast) Ltd
Printed in Great Britain by
Biddles Ltd.
Guildford & King's Lynn

To Mary, Simon, and Anna, my Lyapunov functions.

Preface

Non-linear time series analysis is a rapidly developing subject. Of necessity, it draws on deeper aspects of probability theory and more sophisticated tools of statistical inference. It demands greater degrees of ingenuity and common sense in model building. In return, once freed from the shackles of linearity, the analyst has the opportunity of gaining a fuller appreciation of the beauty of the real world. Thus, I have included real data sets from animal populations, solar activities, economics, finance, medical sciences, hydrology, environmental sciences, and others. A user-friendly package is available which implements a fair proportion of the modelling and forecasting techniques described in this book.

Although a few specialized books have recently appeared, I believe that there is a need for a fairly comprehensive account, covering the fundamentals of probability theory, statistical inference, model building, and prediction of non-linear time series. In tune with modern developments in other scientific disciplines, I have adopted the dynamical system approach, in that whenever possible I emphasize links with dynamical systems. This is based on my unswerving belief that without well-tried physical underpinnings, no sustainable edifice can be erected. As a reflection of the experimental nature of this undertaking, readers will probably observe non-uniform motions in places in the book. These are also partly due to my desire to keep the level of mathematics as modest as possible, although I cannot claim that I have never yielded to the beauty of mathematics. Readers are usually warned of such instances, which are indicated by the symbols (¶ and ¶¶).

I am enormously indebted to all non-linear time series enthusiasts but would like to mention, in particular, Kung-sik Chan, with whom I have made many exciting excursions over the truly fascinating non-linear terrains. Without his selfless assistance, I am sure that I would have made many more mistakes in my navigation. All remaining errors are naturally mine alone. I am also grateful to Doyne Farmer, Russell Gerrard, Ian Jolliffe, Rahim Moeanaddin, Pham Dinh Tuan, Richard Smith, Akiva Yaglom, and Zhu Zhao-xuan. Mrs Mavis Swain deserves my sincere thanks for rendering my sometimes inscrutable writing intelligible under very difficult conditions. Last but not least, I would like to take this

opportunity of thanking my father, whose Chinese calligraphy has adorned this volume.

H.T.

Canterbury
April 1989

Acknowledgements

Grateful acknowledgements are made to the following:

The *Journal of Time Series Analysis* and Professor Peter M. Robinson for permission to reproduce Fig. 5.6.
The American Water Resources Association for permission to reproduce §7.4.
Springer Verlag for permission to reproduce material from the monograph entitled *Threshold models in non-linear time series* by H. Tong.
The Royal Society and Dr. D. A. Jones for permission to reproduce Figs. 4.1, 4.2, and 4.11.

Contents

Some suggested set meals for the readers

(A) Fast food (for applied statisticians in a hurry or with a limited mathematical background):

Chapter 1
Sections 2.1 and 2.2
Chapter 3
Sections 5.1, 5.2, 5.3.3, 5.3.5.3, 5.4, and 5.6
Chapter 7

(B) Vegetarian food (for those with minimal statistical background):

Chapter 1
Chapter 2
Chapter 3
A personal selection from the rest

(C) Gourmet food (for those looking for potential research problems):

Chapter 2
Chapter 4
Chapter 5
Chapter 7
Exercises and complements of all seven chapters

(D) Banquet (for those looking for a fairly comprehensive treatment):

Enjoy your seven-course meal!

The STAR PC package

A user-friendly floppy disk STAR may be purchased from Microstar Software (using the order form provided at the end of the book) which will provide a comprehensive statistical package for threshold modelling. It may be used in conjunction with Chapter 7 to gain hands-on experience. The package may be run on an IBM PC/XT or PC/AT or their compatibles with MS-DOS Version 3 or PC-DOS Version 3. It has extensive graphics.

1 Introduction

Non-linearity *begets completeness*;
Misjudgment creates linearity.

Ch. XXII Lao Tzu (circa 600 BC)

1.1 Time series model building

In our endeavours to understand the changing world around us, observations of one kind or another are frequently made sequentially over time. The record of sunspots is a classic example, which may be traced as far back as 28 BC (see e.g. Needham 1959, p. 435).

Let us tentatively call such records *time series*. Possibly the most important objective in our study of a time series is to help to uncover the dynamical law governing its generation. Obviously, a complete uncovering of the law demands a complete understanding of the underlying physics, chemistry, biology, etc. When the underlying theory is non-existent or far from being complete, and we are presented with not much more than the data themselves, we may adopt the following paradigm:

(1) recognize important features of the observed data;

(2) construct an empirical time series model, incorporating as much available background theory as possible;

(3) check that the constructed model is capable of capturing the features in (1) and look for further improvement if necessary.

Fundamentally, an empirical time series model represents a *hypothesis* concerning the probability transition over time, that is the dynamics. Some authors have used the word 'model' in a different sense from the one adopted here. For example, it has sometimes been used to mean a forecast algorithm, the form of which is completely specified except for some defining parameters to be determined from data. Stage (1) in the above model-building paradigm dictates the 'shape' of things to come and stage (3) judges the 'goodness of fit' of the delivered product. Stage (2) may be facilitated by specifying a fairly wide class of models, denoted by C, within which some optimal search technique, that is identification, may then be employed. An obvious requirement is that C should be wide

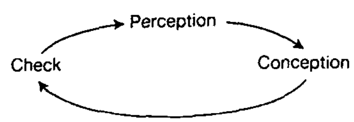

Fig. 1.1. A loop in model building

enough to include models capable of capturing the recognized features in (1). Equally obvious is the fact that the size of C is constrained by the amount of computation at our disposal for the search. A recognition of these two aspects reinforces the belief that model building is as much an art as it is a science.

Philosophically speaking, every specification of a times series model is coloured by some subjective judgement. What we have described in the above paradigm is best viewed as just one loop in a spiral of many and each loop should lead to an empirical model closer to the objective reality, in which the more important features are incorporated and the less important ones discarded (cf. Box 1980) (Fig. 1.1).

The other important function of an empirical model should not be overlooked, and that is it sharpens the perception in the next loop. Whilst an empirical model can never replace the underlying theory, the former can assist the development of the latter. At the same time, each advance in the latter can help bring about a more satisfactory empirical model. It may be argued that statistical modelling in general, and time series modelling in particular, should not be divorced from the underlying scientific discipline that the final product (a statistical model) is supposed to serve.

1.2 Stationarity

Let X_t denote a real-valued random variable representing the observation made at time t. For most of the book we confine our study to observations made at a regular time intervals, and, without loss of generality, we assume that the basic time interval is of duration one unit of time. We may now state the following definition:

Definition 1.1: A *time series*, $\{X_t\}$, is a family of real-valued random variables indexed by $t \in \mathbf{Z}$, where $\mathbf{Z}$ denotes the set of integers.

The more elaborate term 'discrete parameter time series' is not used because we shall study almost exclusively the case with $t \in \mathbf{Z}$. At any rate, the subscript t is reserved exclusively for this case.

In this volume, mainly those important features with time-invariant properties are considered.

Definition 1.2: The time series $\{X_t\}$ is said to be *stationary* if, for any $t_1, t_2, \ldots, t_n \in \mathbf{Z}$, any $k \in \mathbf{Z}$, and $n = 1, 2, \ldots,$

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, \ldots x_n) = F_{X_{t_1+k}, X_{t_2+k}, \ldots, X_{t_n+k}}(x_1, \ldots, x_n) \qquad (1.1)$$

where F denotes the distribution function of the set of random variables which appear as suffices.

The term 'strictly stationary' is more often used to describe the above situation, while the term 'weakly stationary', 'second-order stationary', 'covariance stationary', or 'wide-sense stationary' is used to describe the theoretically less restricted situation in which

$$E(X_{t_1}) = E(X_{t_1+k}) \qquad \operatorname{cov}(X_{t_1}, X_{t_2}) = \operatorname{cov}(X_{t_1+k}, X_{t_2+k}) \qquad (1.2)$$

for all $t_1, t_2, k \in \mathbf{Z}$, the covariances being assumed to exist. Strict stationarity implies weak stationarity provided $\operatorname{var} X_t$ exists. In the Gaussian case, they are equivalent. Unless otherwise stated, we use the terms 'stationary time series' and 'strictly stationary time series' interchangeably. In the main, we consider model building for stationary time series or for time series which may be made stationary after some simple transformation, such as taking differences of consecutive observations, subtracting a polynomial or a trigonometric trend, etc.

Consider a stationary time series $\{X_t\}$ with finite variance. It follows from (1.2) that $\operatorname{cov}(X_{t_1}, X_{t_2})$ is simply a function of $|t_1 - t_2|$. This function is called the *autocovariance function of* $\{x_t\}$ *at lag*$(t_2 - t_1)$. We denote it by $\gamma_{t_2 - t_1}$. It has the following properties (see e.g. Priestley 1981, pp. 108–10):

(1) $\gamma_0 = \operatorname{var} X_t$
(2) $|\gamma_\tau| \le \gamma_0, \ \forall \tau \in \mathbf{Z}$
(3) $\gamma_{-\tau} = \gamma_\tau, \ \forall \tau \in \mathbf{Z}$
(4) $\forall t_1, t_2, \ldots, t_n \in \mathbf{Z}$, $\forall$ positive $n \in \mathbf{Z}$, and $\forall$ real $z_1, z_2, \ldots, z_n$,

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

The ratio γ_τ / γ_0, $\tau \in \mathbf{Z}$, is called the *autocorrelation function of* $\{X_t\}$ *of lag* τ. It is denoted by ρ_τ. Properties (2), (3), and (4) still hold if the γ are replaced by the ρ with corresponding subscripts. It is well known that ρ_τ may be interpreted as a measure of *linear* association between X_t and $X_{t \pm \tau}$.

Property (4) is that of positive semi-definiteness and the following theorem describes the positive semi-definite function defined by $\{\rho_\tau : \tau =$

$0, \pm 1, \pm 2, \ldots\}$ as a Fourier transform.

Theorem 1.1: A function defined by $\{\rho_\tau : \tau = 0, \pm 1, \pm 2, \ldots\}$, $\rho_0 < \infty$, is positive semi-definite if and only if it can be expressed in the form

$$\rho_\tau = \int_{-\pi}^{\pi} e^{i\omega\tau} \, dF(\omega) \tag{1.3}$$

where F (defined for $|\omega| \le \pi$) is monotonic non-decreasing.

For a proof see for example Doob (1953, p. 474).

The Fourier transform F is called the (normalized) *integrated spectrum*. If $\{\rho_\tau\}$ is absolutely summable, then F has the continuous derivative f, given by

$$f(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho_\tau e^{-i\omega\tau} \quad \text{almost all } \omega \in [-\pi, \pi]. \tag{1.4}$$

The function f is called the (normalized) *spectral density function*. The analogous equations for γ_τ are

$$\gamma_\tau = \int_{-\pi}^{\pi} e^{i\omega\tau} \, dH(\omega) \tag{1.3'}$$

and

$$h(\omega) = \frac{dH(\omega)}{d\omega} = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_\tau e^{-i\omega\tau} \quad \text{almost all } \omega \in [-\pi, \pi] \tag{1.4'}$$

where H and h are called the (*non-normalized*) *integrated spectrum* and the (*non-normalized*) *spectral density function* respectively. Obviously H and F are related by

$$H(\omega) = \gamma_0 F(\omega), \quad \text{all } \omega. \tag{1.5}$$

(See, for example, Priestley (1981) for a detailed discussion of the branch of time series analysis called *spectral analysis* which is centred around the spectral functions.)

1.3 Linear Gaussian models

It is a remarkable fact that linear Gaussian models have dominated the development of time series model building for the past five decades. It may be said that the era of linear time series modelling began with such linear models as Yule's *autoregressive* (*AR*) *models* (1927), first introduced in the study of sunspot numbers. Specifically, the class of AR models consists of models of the form

$$X_t = a_0 + \sum_{j=1}^{k} a_j X_{t-j} + \varepsilon_t \tag{1.6}$$

where the a_j are real constants ($a_k \ne 0$), k is a finite positive integer referred to as the *order* of the AR model, and the ε_t are zero-mean uncorrelated random variables, called *white noise*, with a common variance, $\sigma_\varepsilon^2 (< \infty)$. Symbolically, we express (1.6) by $X_t \sim \text{AR}(k)$. A more general class of linear models is obtained by replacing ε_t by a weighted average of $\varepsilon_t, \varepsilon_{t-1}, \ldots, \varepsilon_{t-l}$, that is

$$X_t = a_0 + \sum_{j=1}^{k} a_j X_{t-j} + \sum_{j=0}^{l} b_j \varepsilon_{t-j} \tag{1.7}$$

where the b_j are real constants ($b_l \ne 0$) and b_0 may be set equal to unity without loss of generality. This is the so-called class of *autoregressive/moving average* (*ARMA*) *models*. Symbolically, we express (1.7) by $X_t \sim \text{ARMA}(k, l)$. Here, l is a finite non-negative integer referred to as the *order* of the moving-average part of the ARMA model. The special case of ARMA$(0, l)$ is referred to as the *moving-average* (*MA*) model of order l, denoted by MA(l).

We now introduce two conditions on the ARMA models. At the expense of some slight loss of theoretical generality, these two conditions lead to sharper results and some simplification of discussion. In any case, it seems that they are often made in practice, either explicitly or implicitly.

Condition A; The roots of the polynomials

$$A(z) = z^k - \sum_{j=1}^{k} a_j z^{k-j} \tag{1.8}$$

$$B(z) = \sum_{j=0}^{l} b_j z^{l-j} \quad (b_0 = 1) \tag{1.9}$$

all have modulus less than one. A and B will be called the *autoregressive generating function* and *moving-average generating function* respectively.

Condition B: $\{\varepsilon_t\}$ is a sequence of independent identically distributed random variables (an i.i.d. sequence), each with the distribution $\mathcal{N}(0, \sigma_\varepsilon^2)$. $\{\varepsilon_t\}$ is referred to as a *Gaussian white noise*.

Let

$$\mu_X = a_0 \Big/ \left(1 - \sum_{j=1}^{k} a_j\right) \quad \text{and} \quad \sigma_X^2 = \frac{\sigma_\varepsilon^2}{2\pi} \int_{-\pi}^{\pi} \left| \frac{B(e^{-i\omega})}{A(e^{-i\omega})} \right|^2 d\omega.$$

Under these two conditions, and subject to X_0 having a $\mathcal{N}(\mu_X, \sigma_X^2)$ distribution:

1. $\{X_t : t = 1, 2, \ldots\}$ is stationary.
2. $\forall t_1, t_2, \ldots, t_k \in \mathbf{Z}_+$, the set of non-negative integers, and $\forall k$ belonging to the set of positive integers, $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is jointly Gaussian. $\{X_t : t = 1, 2, \ldots\}$ is called a *Gaussian sequence*.

3. X_t admits the *linear (one-sided) model/linear representation*

$$X_t = \mu_X + \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j} \quad \text{with} \quad \sum_{j=0}^{\infty} \beta_j^2 < \infty,\ \beta_0 = 1, \text{ and } \mu_X = \mathrm{E}X_t. \tag{1.10}$$

3′. Likewise ε_t admits a linear model in terms of X_s, $s \le t$. This is sometimes called the *invertibility* of $\{X_t\}$ (see e.g. Rosenblatt 1979).

Henceforth, unless otherwise stated, all ARMA models are assumed to satisfy Conditions A and B. We may sometimes emphasize this fact by referring to them as stationary Gaussian ARMA models. On the other hand, if a model for $\{X_t\}$ is of the general form (1.10) which possesses properties (1) and (2) and in which $\{\varepsilon_t\}$ is an i.i.d. sequence (and necessarily Gaussian), it is called a *linear Gaussian model.* Henceforth, by an abuse of terminology, we do not distinguish between a time series model and the time series defined by it. Now, a well-defined linear Gaussian model for $\{X_t\}$ is completely specified by the mean, μ_X, and the autocovariances, γ_t, of $\{X_t\}$, or equivalently by μ_X and the (non-normalized) spectral density function, h. Note that

$$h(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \left| \sum_{j=0}^{\infty} \beta_j e^{-ij\omega} \right|^2 \tag{1.11}$$

which is a continuous integrable function of ω. On the other hand, an ARMA model has a (non-normalized) spectral density function of the form

$$\frac{\sigma_\varepsilon^2}{2\pi} \left| \frac{B(e^{-i\omega})}{A(e^{-i\omega})} \right|^2 \tag{1.12}$$

which is a rational function of $e^{-i\omega}$. It has been said that ARMA models enjoy the same degree of generality among the class of linear Gaussian models as rational functions among the class of continuous integrable functions (see e.g. Priestley 1980, p. 283). Of course, an ARMA model has the significant property of consisting of only a finite number of parameters. From the point of view of model building, we may conclude that *if and only if the autocovariances are considered an important feature, the class of ARMA models constitutes a useful choice of C.*

1.4 Some advantages and some limitations of ARMA models

The strengths and weaknesses of stationary Gaussian ARMA models are, in fact, already subsumed in the conclusion of the last section. We elaborate them as follows. Proofs of some of the results cited in this section will be given in full in Chapter 4.

1.4.1 SOME ADVANTAGES

In the following discussion, we merely highlight some of the significant achievements of the ARMA models.

1. Mathematically, linear difference equations are the simplest type of difference equations and a complete theory is available. Probabilistically, the theory of Gaussian sequences is readily understood. The theory of statistical inference is the most developed for linear Gaussian models.

The class of stationary Gaussian ARMA models has an elegant and fundamental geometric characterization in terms of the concepts of a *predictor space* and a *Markovian representation* introduced by Akaike (1974*a*). These concepts are rooted in control systems theory. He has shown that *a stationary Gaussian time series has a stationary Gaussian ARMA representation if and only if its predictor space is finite-dimensional.*

2. The computation time required for obtaining a parsimonious ARMA model for the data is well within the reach of most practitioners. Ready-made packages are available. Over the years, much experience has been accumulated in the application of ARMA models (see e.g. Box and Jenkins 1976).

3. These models have been reasonably successful as a practical tool for analysis, forecasting, and control (see e.g. Box and Jenkins 1976). They have not survived 60-odd years for nothing! We must conclude that they represent the objective world to a good first approximation.

1.4.2 SOME LIMITATIONS

Once again, in the discussion that follows we merely highlight some of the current interests in the subject of time series modelling. The discussion will be interspersed with the introduction of terminology, concepts, and theoretical results mostly relevant to later exposition.

1. On setting the innovation ε_t to a constant for all t (or equivalently on setting var ε_t to zero), eqn (1.7) becomes a *deterministic* linear difference equation in X. Under condition A, X_t will always tend to a unique finite constant, independent of the initial value, as t tends to infinity. The situation is described as a *stable limit point.* If $A(z)$ has one root greater than unity in modulus, $|X_t|$ will tend to infinity with t, and the situation is described as being *unstable.* If $A(z)$ has some roots equal to unity in modulus and the others less than unity in modulus, X_t will eventually osciallate among a set of points whose values depend on the initial value. The situation is described as being *neutrally stable.* We have

here merely restated the well-known result that a linear difference equation does not permit stable periodic solutions independent of initial value. This point will be developed further in Chapter 2.

2. Having symmetric joint distributions, stationary Gaussian ARMA models are not ideally suited for data exhibiting strong asymmetry. Figure 1.2 gives one typical set of hydrological data, a Gaussian model for which would be of limited value.

3. ARMA models are not ideally suited for data exhibiting sudden bursts of very large amplitude at irregular time epochs. This is clear in view of the normality of ARMA models. Essentially, ARMA models are more suitable for data with negligible probability of very high level crossings (and there are plenty of these about). We may recall the elementary result that if the kth order moment of a random variable X exists, then

$$P[|X|>c]=O(|c|^{-k}) \quad \text{as} \quad c\to\infty. \tag{1.13}$$

Thus, the probability of large excursions is connected with the existence of moments. If the set of all possible values of X is bounded from both sides, then the distribution of X has the *moment property,* that is moments of all orders exist (see e.g. Fisz 1963, p. 68). However, boundedness is not necessary since, for example, it is obvious that a Gaussian random variable (and hence a stationary Gaussian sequence) has the moment property.

It is interesting to note that models without the moment property may be constructed by some kind of stochastic 'perturbation' of the ARMA models. Consider a stochastic perturbation of an AR(1) model in the form of

$$X_t=(a+b\varepsilon_t)X_{t-1}+\varepsilon_t \tag{1.14}$$

where $\{\varepsilon_t\}$ is a Gaussian white noise with zero mean and, without loss of generality, unit variance. The coefficient of X_{t-1} is no longer a constant but is a random variable, which is a linear function of ε_t. This model is a special case of the general class of models called *bilinear models,* which will be discussed in more detail later.

Figure 1.3 illustrates the phenomenon of sudden bursts of the above class of models. We may prefer these models to the ARMA models when dealing with a situation of this type.

4. Since the autocovariances, γ_j $(j\in\mathbf{Z})$, are only one aspect of the joint distributions of (X_t, X_{t-j}), $(j\in\mathbf{Z})$, other aspects may contain vital information missed by the γ_j. One such aspect is, for example, *the regression function at lag* (j), that is $E(X_t \mid X_{t-j})$, $(j\in\mathbf{Z})$. For the ARMA models, these are all linear because of the joint normality. This

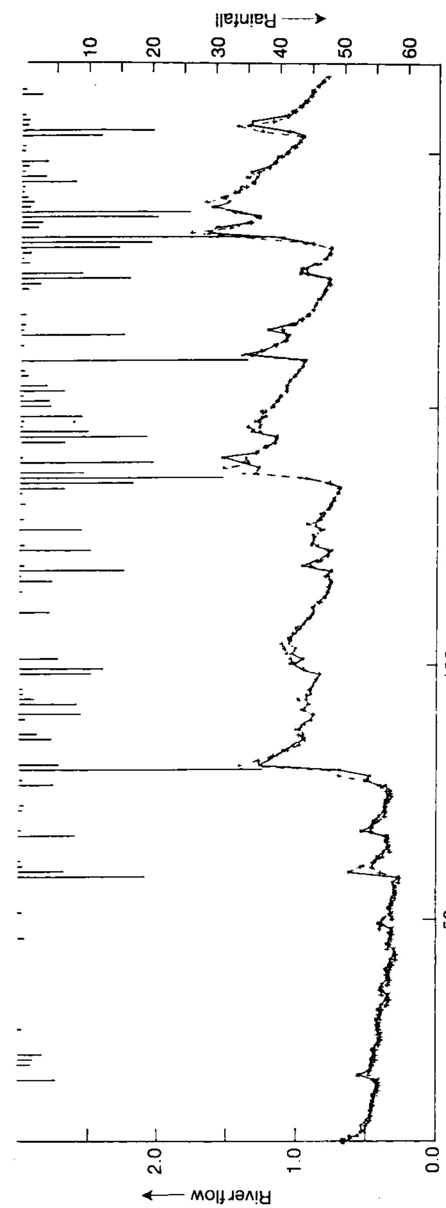

Fig. 1.2. Daily river flow data of Kanna River, Japan (in mm per/day), in the year 1956: +, real data; * from fitted model.

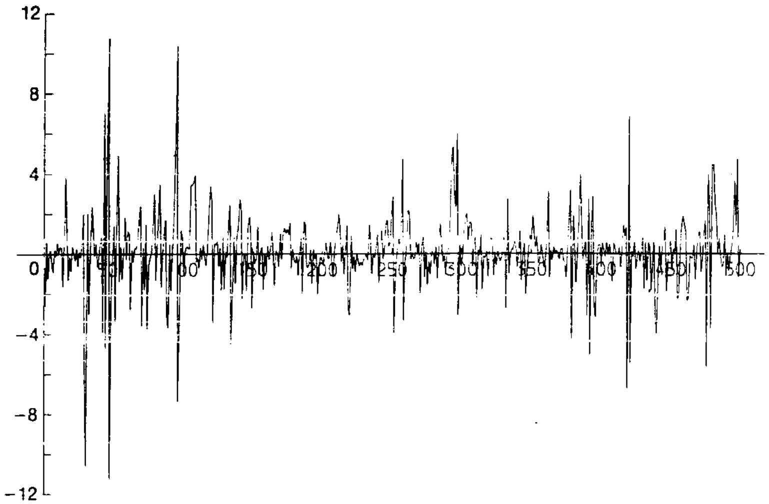

Fig. 1.3. A realization of the model $X_t = (-0.1 + 0.9\varepsilon_t)X_{t-1} + \varepsilon_t$, $\varepsilon_t \sim \mathcal{N}(0, 1)$. (The theoretical standard deviation of X_t is 2.357)

characteristic may sometimes weaken the usefulness of ARMA models for data exhibiting strong cyclicity. The following situation seems to have some practical relevance.

The autocorrelation function of strongly cyclical data is also strongly cyclical. At those lags for which the autocorrelation function is quite large in modulus, the corresponding regression functions may be sufficiently well approximated by linear functions. However, at those lags for which the autocorrelation function is quite small in modulus, a linear approximation for the corresponding regression functions is not always unquestionable. Indeed, it is conceivable that the strong cyclicity of the data may be linked with a strong association (not necessarily measurable by the autocorrelation function which measures only linear association) between X_t and X_{t-j}, $j = \pm 1, \pm 2, \ldots, \pm L$, for some finite integer L. In this case, a *non-linear* approximation for the regression functions may well be more appropriate for those lags with small autocorrelations. We may illustrate the above situation with the classic annual Canadian lynx data (1982–1934) (See Figs 1.4–1.6.) We shall return to a more comprehensive analysis of these data later.

In fact, the use of sample regression functions in the time series context goes right back to Yule (1927). The use of sample regression functions

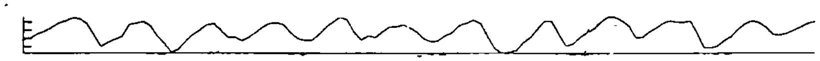

Fig. 1.4. Logarithmically transformed MacKenzie River series of annual Canadian lynx trappings from the years 1821–1934

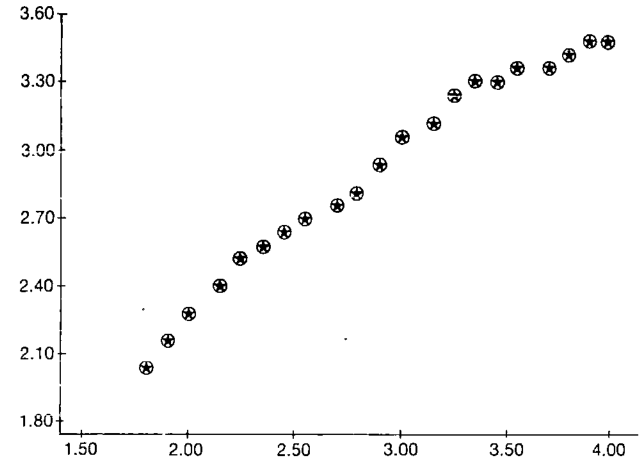

Fig. 1.5. Sample regression function of X_t on X_{t-1} for the data in Fig. 1.4. Sample autocorrelation function of lag 1 is 0.79. Sample estimate of $\gamma_0\rho_1^2/\mathrm{var}\{E(X_t \mid X_{t-1})\}$ is roughly 1. The linear fit for the regression of X_t on X_{t-1} implied by a linear Gaussian time series model is quite reasonable

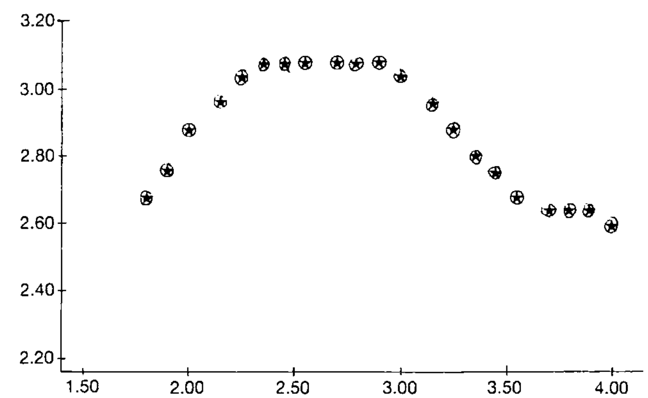

Fig. 1.6. Sample regression function of X_t on X_{t-3} for data in Fig. 1.4. Sample autocorrelation function of lag 3 is −0.13. Sample estimate of $\gamma_0\rho_3^2/\mathrm{var}\{E(X_t \mid X_{t-3})\}$ is roughly 0.04. A linear Gaussian time series model for the data would imply an *almost horizontal* linear fit for the regression of X_t on X_{t-3}, which is clearly poor

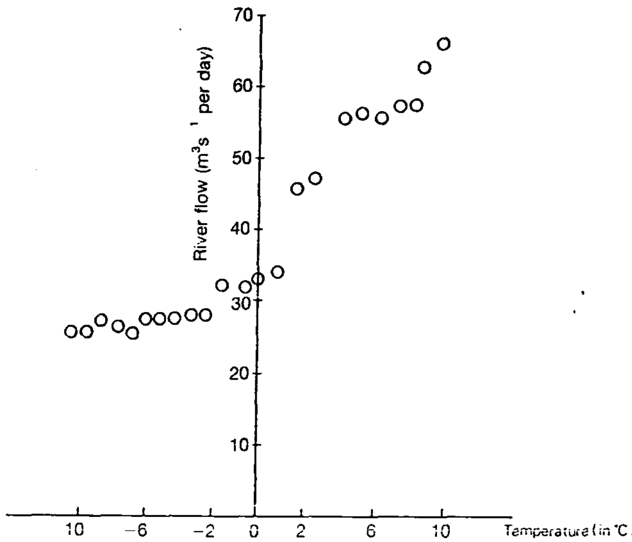

Fig. 1.7. Non-parametric regression, $E\{X \mid Z = z\}$, of river flow (X) on temperature (Z) for the River Jökulså Eystri, Iceland. The method does not assume an *a priori* functional form but is based on a (kernel) smoothing on the data histograms. Specifically, let $\{\delta_N(z)\}$ denote a sequence of symmetric and non-negative functions of z, of area 1, with the property that $\delta_N(z) \to$ Dirac delta function as $N \to \infty$. Then $E\{X \mid Z = z\} = \sum_{j=1}^{N} x_j \delta_N(z - z_j) / \sum_{j=1}^{N} \delta_N(z - z_j)$, where $(x_1, z_1), \ldots, (x_N, z_N)$ denote the N data points

may be easily extended beyond univariate time series. Figure 1.7 gives an illustration of regressing river flow on temperature. The non-linear effect due to the melting ice of a glacier within the catchment area of the river is clearly evident. More details about the analysis of these data will be given in the final chapter.

5. ARMA models are not ideally suited for data exhibiting *time irreversibility*. Figures 1.2 and 1.4 show examples of such data. A simple yet effective way of visualizing this is by tracing these data on a transparency and then turning it over.

One way of gaining further insights into the effect of time reversibility on the probabilistic structure of the time series $\{X_t\}$ is to introduce higher-order spectra. We will deal with this point in Chapter 4.

·1.5 What next?

After six decades of domination by linear Gaussian models, the time is certainly ripe for a serious study of ways of removing the many limitations of these models. Once we decide to incorporate features in addition to the autocovariances, the class of models would have to be greatly enlarged to include those besides the Gaussian ARMA models. We may either retain the general ARMA framework and allow the white noise to be non-Gaussian, or we may completely abandon the linearity assumption.

In the former case, limitations 2, 3, 4, 5 of Gaussian ARMA models can be removed, to some extent, by a *judicious* choice of the distributions of the ε_t. As a typical illustration, let us consider $E(X_t \mid X_{t-j})$ for the following non-Gaussian MA(1):

$$X_t = \varepsilon_t - a\varepsilon_{t-1} \tag{1.15}$$

where ε_t has a uniform distribution on $(-\sqrt{3}, \sqrt{3})$. After some non-trivial manipulation $E(X_t \mid X_{t-1})$ is shown to be non-linear as illustrated in Figs 1.8 and 1.9.

Shepp *et al.* (1980) have given a detailed study of the regression functions, $E(X_t \mid X_{t-1}, \ldots, X_{t-k})$, k being any positive integer, for a non-Gaussian MA(l). Another example, which is probably the simplest although a little extreme, is taken from Whittle (1963*a*, Section 2.6) and Rosenblatt (1979). Consider the stationary AR(1) model

$$x_t = \tfrac{1}{2}X_{t-1} + \varepsilon_t$$

where

$$\varepsilon_t = \begin{cases} \tfrac{1}{2} \text{ with probability } \tfrac{1}{2} \\ 0 \text{ with probability } \tfrac{1}{2} \end{cases}$$

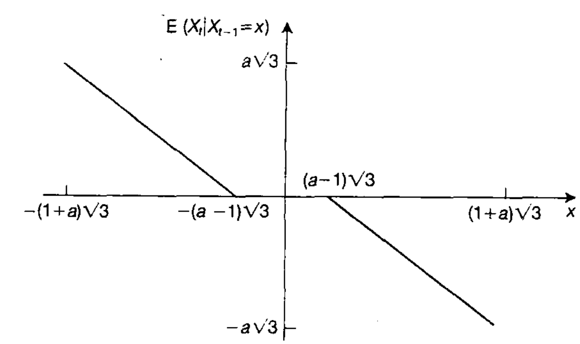

Fig. 1.8. Regression function of lag (1) of model $X_t = \varepsilon_t - a\varepsilon_{t-1}$, ε_t uniformly distributed on $(-\sqrt{3}, \sqrt{3})$; $(a \geq 1)$

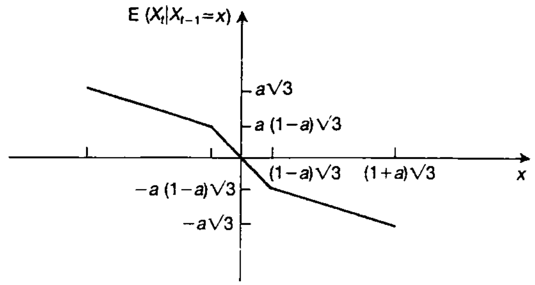

Fig. 1.9. Regression function of lag (1) of model of $X_t = \varepsilon_t - a\varepsilon_{t-1}$, ε_t uniformly distributed on $(-\sqrt{3}, \sqrt{3})$; $(a < 1)$

and ε_t is independent of X_s, $s < t$. X_t is then uniformly distributed on $[0, 1]$ and

$$2X_t = X_{t-1} + 2\varepsilon_t$$

$$= X_{t-1} \text{ (modulo 1).}$$

Obviously,

$$E(X_t \mid X_{t-1}) = \tfrac{1}{2}X_{t-1} + \tfrac{1}{4}$$

which is linear in X_{t-1}. However,

$$E(X_{t-1} \mid X_t) = 2X_t \quad \text{(modulo 1)}$$

which is non-linear!

There is no doubt that further exploration within the non-Gaussian ARMA framework may be quite fruitful. We may argue that the failure of non-Gaussian ARMA models to remove limitation 1 means that it would be appropriate to look elsewhere for models possessing much richer dynamical properties. The pages which follow will be devoted entirely to the removal of the linearity assumption. To end the chapter

the following classification summarizes the situation:

Linear Gaussian models e.g. ARMA models with Gaussian white noise	Linear non-Gaussian models e.g. ARMA models with non-Gaussian white noise
Non-linear Gaussian models e.g. Gaussian output with non-Gaussian white noise input through a non-linear filter	Non-linear non-Gaussian Models e.g. read on!

Bibliographical notes

The philosophical attitude adopted in this book is similar to that more eloquently discussed by Box (1980). It leans on dialectics. The basic theory of stationary time series is covered in Doob (1953) at a sophisticated level and in Priestley (1981) at a level more readily accessible to the non-specialists of probability theory. Akaike (1974*a*) is a remarkable paper, which completes the geometric delineation of ARMA models within the class of linear time models. The parallel picture in the class of non-linear time series models has no more than a few strokes on it (see Chapter 4). Serious limitations of linear Gaussian time series models in practical situations were mentioned earlier by Akaike, Cox, Galbraith, and Tunnicliffe-Wilson in the discussions of the papters by Campbell and Walker (1977) and Tong (1977*b*) on the analysis of the classic Canadian lynx data. More attention seems to be warranted in respect of time irreversibility.

Exercises and complements

(1) Let $\{X_t\}$ denote a stationary Gaussian time series. Prove that

$$E[X_s \mid X_t = x] \equiv E[X_t \mid X_s = x].$$

Generalize the result to higher-order conditional moments.

(2) Consider the strictly stationary AR(1) model

$$X_t = \tfrac{1}{2}X_{t-1} + \varepsilon \qquad t \geq 1$$

where X_0 is uniformly distributed on $[0, 1]$,

$$\varepsilon_t = \begin{cases} \tfrac{1}{2} \text{ with probability } \tfrac{1}{2} \\ 0 \text{ with probability } \tfrac{1}{2} \end{cases}$$

and ε_t is independent of X_s, $s<t$. Show that the joint distribution of (X_0, X_1) is different from that of (X_1, X_0). Let the times series $\{X_t\}$ be instantaneously transformed to the time series $\{Y_t\}$ via

$$Y_t = \Phi^{-1}(X_t), \text{ each } t$$

where Φ denotes the standard Gaussian distribution function. Show that, for each t, Y_t has a standard Gaussian distribution. Is $\{Y_t\}$ time reversible?

(3) Let $\{X_t\}$ be a strictly stationary time series all of whose joint distributions are symmetric about the origin. Prove that $\{X_t\}$ need not be time reversible. Does time reversibility imply symmetry of all joint distributions about the origin?

(4) Suppose that X_0 is uniformly distribution on (0, 1). Let

$$X_t = 2X_{t-1} \text{ (modulo 1)} \qquad t \geq 1$$

that is X_t is the fractional part of $2X_{t-1}$. Prove that the joint distributions of $(X_{t+\tau}, X_t)$ are degenerate and that

$$\operatorname{cov}(X_{t+\tau}, X_t) = \frac{2^{-|\tau|}}{12}.$$

Show that the linear least-squares predictor, $\hat{X}_t$, of X_t given all past values is given by

$$\hat{X}_t = \tfrac{1}{4} + \tfrac{1}{2}X_{t-1}$$

which has mean square error of prediction equal to $\frac{1}{16}$.
[Hint: Consider a binary representation of X_0].

(5) Let $\{X_t\}$ denote a Gaussian time series with zero mean and unit variance. Suppose that it has a (normalized) spectral density function f. Let

$$Y_t = X_t^2 - 1, \quad \text{for each } t.$$

Verify that $\{Y_t\}$ has normalized spectral density, h, given by

$$h(\omega) = \int_{-\pi}^{\pi} f(\omega - \theta) f(\theta)\, d\theta.$$

(6) Let

$$X_t = (a + b\varepsilon_t)X_{t-1}$$

where $\{\varepsilon_t\}$ is a sequence of independent identically distributed random variables with zero mean and unit variance. Suppose that $a^2 + b^2 < 1$. Prove that $X_t \to 0$ in probability as $t \to \infty$.

(7) Show that the sequence $\{X_1, X_2, \dots\}$ is strictly stationary if and only if there exists a sequence $\{Y_1, Y_2, \dots\}$ such that, for any n, the joint distribution of $(X_1, X_2, \dots, X_n)$ is the same as that of $(Y_n, Y_{n-1}, \dots, Y_1)$.

(Kingman and Taylor 1966, p. 394)