Title	MODELING AND PARAMETER IDENTIFICATION FOR A CLASS OF NONSTATIONARY NONLINEAR SYSTEMS
Author(s)	Fukuda, Tokuo
Citation	大阪大学, 1987, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/1600
rights	
Note	

MODELING AND PARAMETER IDENTIFICATION FOR A CLASS OF NONSTATIONARY NONLINEAR SYSTEMS

(非定常非線形システムのモデリングとパラメータ同定に関する研究)

A DISSERTATION
PRESENTED TO
OSAKA UNIVERSITY

BY

TOKUO FUKUDA

FOR THE DEGREE OF
DOCTOR OF ENGINEERING

1987

ACKNOWLEDGMENTS

The author would like to express his sincere thanks to Professor Yoshifumi Sunahara of Kyoto Institute of Technology for his supervisions, suggestions and continuous encouragement during the long years that the preparation of this thesis was in progress.

Thanks are also extended to Prof. Shinzo Kodama, Prof. Yutaka Suzuki, Prof. Hiroaki Terada and Prof. Minoru Yamamoto of Osaka University for their comments on this thesis.

The author would like to thank many members of Control System Science Laboratory of Kyoto Institute of Technology for their helpful discussions and contributions.

This thesis is dedicated to my parents, Mr. Mamoru Fukuda and Mrs. Yukino Fukuda.

T. Fukuda

June 1987

ABSTRACT

In this thesis, some classes of stochastic system modeling are studied, where system models are described by stochastic differential equations or stochastic difference equations.

A systematic procedure is firstly derived to model a class of linear but nonstationary systems. The system dynamics and the observation mechanism are assumed to be described by linear stochastic differential equations with time-varying coefficients. The principal line of attack is to utilize the nonlinear filtering theory for parameter identification. The unknown system order is also determined by using multiple hypothesis testing method.

Secondly, the modeling problem is considered for a class of stationary but nonlinear systems whose outstanding feature lies in sporadically large peak values. The system model is assumed to be given by a nonlinear moving average (NMA) model, where nonlinear terms are described by a set of Hermite polynomials. Furthermore, in order to handle the data whose autocorrelations between time intervals longer than a single time step may not be zero, NMA model is extended to a nonlinear ARMA model consisting of linear AR terms and nonlinear MA ones.

Finally, the modeling problem for a class of nonstationary nonlinear system is considered. The system model is assumed to be described by a class of nonlinear time-varying difference equation. The estimators of unknown parameters are given from the maximum likelihood concept and their asymptotic properties are examined mathematically. Restricting the system model to the single-input single-output one, the criterion function for the structure determination is derived by evaluating the upper bound of the entropy associated with the estimation errors of both input noise and unknown parameters. Throughout the thesis, asymptotic properties of estimators are investigated theoretically and numerically.

CONTENTS

CHAPTER 1 INTRODUCTION

In the field of control theory, for example, it has been well known that the required knowledge to design a control system is seldom available a priori. Even if the equation representing a system is known in principle, the knowledge of particular parameters is often missing, and furthermore it is often encountered that the analytical model given by the theoretical consideration for a system is too complex for the intended use of the model. The situation mentioned above naturally occurs in other fields such as statistics, system theory, econometrics, information theory, etc. These are the reasons why the system modeling and identification problem have been so widely and intensively investigated in various fields of science and engineering.

Many physical systems show the following three aspects in common; (i) they are excited by the inevitable random noise or their environments are randomly disturbed; (ii) in general they exhibit various kinds of nonlinear behaviors; (iii) their behaviors depend deeply on the time evolution. Hence, the properties mentioned above should be reflected in the mathematical models of almost all physical systems.

In this thesis, the author will establish the modeling method for nonstationary and nonlinear systems and, asymptotic properties of the estimated system parameters and the determined system structure will be also investigated from both numerical and theoretical viewpoints.

1.1 Historical Background

For better understanding of the present status of studies on stochastic system modeling, the historical background is outlined in the following two versions.

1.1.A Parameter Identification for Nonlinear and Nonstationary Systems

Parameter identification was firstly investigated by K.F. Gauss in 1795 associated with the parameter identification of planetary orbits. In the last few decades, various identification problems have been studied based on the output signals (and input signals if they are available). One of the most important results before 1960 was given by Wald [W1], who worked on parametric statistics, i.e., parameters in the probability density function related to a family of independent and identically distributed random data. He obtained a set of estimators which converges to their true values, and further pointed out the maximum likelihood estimator (M.L.E.) belonging to this set of estimators.

Since physical data are seldom statistically independent, the researchers' interest after 1960 is turned into problems of parameter identification with respect to dependent but stationary observation data with the assumption that systems are described by linear models. Up to the present time, there is certainly overwhelming literature concerned with dependent observation data, and for a general survey, the reader may be referred to Åström and Eykhoff [A1], Box and Jenkins [B1], Eykhoff [E1], Goodwin and Payne [G1] and Mehra and Lainiotis [L1]. However, as observed in the geophysical data, many physical data encountered in various fields of engineering are nonstationary and reflect the nonlinearity of the system. Recently from the viewpoint mentioned above, many authors pay great attention to the identification problem for nonstationary or nonlinear systems.

In relation to nonlinear models, Ivakhnenko [I1] has developed a general approach known as the Group Method of Data Handling (GMDH) for the identification of complex nonlinear systems with unknown structure, based on the pioneering work of Wiener [W2] and Gabor [G5] for nonlinear system representations and the work of Rosenblatt [R5] for the perceptron concept. Using stochastic approximation technique [K7],[R6], Netravali and De Figueiredo [N2] proposed the recursive parameter identification method for a class of nonlinear time-invariant systems where statistical properties of the input disturbance were completely known a priori. They also showed that estimators of unknown parameters had the consistent property with probability one. With more natural condition that the input disturbance is not observable, the following works are representative. Ozaki and Oda [O1] presented a maximum likelihood estimator for a class of nonlinear AR type models and gave several numerical results applied to real time series of actual ship rolling which was known to exhibit some nonlinearity. Poznyak [P1] proved the consistency of the least squares estimators of unknown parameters in a class of nonlinear regression models. For a class of nonlinear moving average models, Robinson [R7] proposed estimators of unknown system parameters by using the moment method and mathematically investigated their asymptotic properties. Sunahara, Ohsumi and the author

[S9] proposed another type of nonlinear moving average models and presented consistent estimators by using the moment method and the asymptotic normality was also shown theoretically. Furthermore they extended the nonlinear moving average model to a nonlinear ARMA models which had linear AR terms and nonlinear MA terms [S10] and presented consistent estimators by using the modified extended Kalman filtering approach [L2].

On the other hand, in relation to nonstationary models, Gran [G3] proposed a method of parameter identification for a class of time-varying linear systems represented by a stochastic differential equation. He assumed that the time-varying unknown parameters in the system model can be approximated by the finite series of polynomials with unknown constant parameters, and augmenting unknown parameters to the state vector, the estimates were obtained by using an approximated nonlinear filtering technique. Using the same notion as in Gran's paper for the parameterization of time-varying coefficients, Lee [L4] and Nakajima and Kozin [N3] presented consistent estimators by using the maximum likelihood method and the least squares method respectively for a class of time-varying linear systems. For nonstationary nonlinear systems, Sunahara and the author [S11],[S12], established a method of identification, and asymptotic properties of the estimators were also investigated mathematically and numerically.

1.1.B Structure Determination

Due to the lack of the complete understanding of the underlying system or the complexity of available model, we are often enforced to select an appropriate system structure or order among an admissible class of models only from the observable output data (and the input data if it is available).

The problem mentioned above is said to be "structure determination" or "order determination" problem. In this section, we shall survey studies on the structure determination problem.

The hypothesis testing theory has been widely applied after 1960 to the structure determination of system models by many investigators [B1],[B2],[C1],[G1],[M1],[M2],[T1],[V1],[W2],[S13]. For example, Mehra [M1] proposed a method which tests the whiteness of the innovation process of an enhancement of a measured product-moment matrix, Woodside [W2] proposed three procedures for noisy stationary linear systems. With the help of the minimal realization theory given by Kalman [K1], Tse and Weinert [T1] considered the problem of estimating the system structure of multivariate stationary linear systems by using the innovation representation for the output process. Sunahara, Ohsumi and the author [S13] derived a systematic procedure to identify the system order for a class of nonstationary continuous systems described by linear time-varying stochastic differential equations.

Recently, several efforts have been made on the structure determination that are somewhat different from the references mentioned above. Akaike [A2],[A3] proposed the criterion function called AIC, which was based on the concept of minimization of the distance between estimated distributions and the true one measured by Kullback-Leibler Information, and the asymptotic properties of AIC were intensively investigated by Shibata [S1]$\sim$[S3]. Although AIC is originally derived for linear stationary models such as AR, MA, ARMA models, Kozin and Nakajima [K2], and Ozaki and Oda [01] respectively applied it to a class of nonstationary models and nonlinear models. In order to removing the defect of over-fitting possibility of AIC, Schwarz [S8] and Akaike [A5] proposed independently a modified criterion function called BIC for the independently and identical distributed observation data, and further improved by Hannan and Quinn [H1] so as to over- and under-estimate the system order of AR models with a less probability. Rissanen [R1],[R2] proposed a new criterion function based on the shortest data description concept and he showed that the criterion function could be derived without assuming the linearity of system models and that the identified system order is consistent when the system is described by a AR or ARMA model. Evaluating the upper bound of the entropy associated with estimation errors of both input noise and unknown parameters in the system model, Sunahara and the author [S14],[S15],[S16] proposed a criterion function for the order determination of nonstationary nonlinear models. They also showed that the identified system order is consistent for a class of nonstationary nonlinear systems.

1.2 Problem Statement

Since many physical data exhibit nonstationarity and reflect nonlinearities of real systems, the author considers the problem of modeling nonstationary, nonlinear, and nonstationary nonlinear systems.

In building mathematical models from one sample path of observation data, there are three main problems to be solved, that is,

Problem 1: structure determination of a model

Problem 2: identification of unknown system parameters

and

Problem 3: investigation of the quality of estimators.

In this thesis, structure determination is tried to solve by using Bayesian decision theory and information theoretic approach where estimators of unknown parameters are obtained by using nonlinear filtering technique, the moment method and the maximum likelihood method, etc.

Asymptotic properties of estimators such as consistency and asymptotic normality are investigated since it is almost impossible to research into their transient properties.

1.3 Summary of Contents

In this thesis, some classes of stochastic system modeling subjected to random inputs are studied, i.e., (i) modeling of nonstationary systems, (ii) modeling of nonlinear systems, and (iii) modeling of nonstationary nonlinear systems.

Chapter 2 is devoted to mathematical preliminaries related to the theory of stochastic processes which will be used in the succeeding developments.

Based on the assumption that the system is linear but nonstationary, a method is presented in Chapter 3 for the nonstationary modeling from noisy data. The model with respect to the unknown system is specified by an n-th order linear stochastic differential equation with time-varying coefficients. The goal of this chapter is to estimate the system order n and identify the time-varying coefficients. Decision rule for the system order is established based on the notion of the multi-hypothesis testing and a procedure to identify the unknown system order and unknown coefficients is given within the framework of estimation theory. Asymptotic properties of estimators are also investigated numerically.

On the other hand, Chapter 4 describes the modeling for stationary but nonlinear systems. As a mathematical model, a class of nonlinear MA (moving average) models is proposed, where nonlinear terms in the system model are described by a set of Hermite orthogonal functions. The proposed nonlinear MA model is expected to be a good basic model in case of fitting the data whose outstanding feature lies in its sporadically large values. A method for identification of unknown system parameters is given by using the moment method. Furthermore, the nonlinear MA model is extended to the nonlinear ARMA (autoregressive moving average) model and a parameter identification method is given by the modified extended Kalman filtering approach.

In Chapter 5, from practical viewpoints that most of all physical data exhibit both nonstationarity and nonlinearity, a method is presented for modeling nonstationary nonlinear systems. The system model proposed here is a type of nonlinear time-varying difference equations. Assuming that the nonlinear time-varying function in the system model can be expanded into finite sets of known functions with unknown constant coefficients, unknown coefficients are identified by using the maximum likelihood concept. The structure determination problem for nonlinear nonstationary system is considered in Chapter 6. The objective system is a single-input single-output system described by a scalar nonlinear time-varying difference equation which is, in some sense, a particular one in the class of models proposed in Chapter 5. The key notion for the structure determination is to minimize the upper bound of the estimation error entropy associated with both input noise and unknown parameters.

The remainder is devoted to discussing a summary of results and some suggestions for areas of further researches.

Throughout all chapters except Chapters 2, 3 and 7, asymptotic properties of estimators are investigated from both theoretical and numerical viewpoints.

CHAPTER 2 MATHEMATICAL PRELIMINARY

2.1 Basic Definitions and Symbolic Conventions

Before presenting the key aspect of this thesis, several basic definitions and symbolic conventions are presented.

Let R^n denotes an n-dimensional Euclidean space. If x is an element of R^n ($x \in R^n$), then x' denotes the transpose of vector x. Similarly, if M is a matrix, then M' and det M denote its transpose and determinant respectively. As a rule, vector and matrix notations follow the usual manner, that is, lower case letters a, b, c ⋯ denote column vectors with i-th components a_i, b_i, c_i ⋯ . Capital letters A, B, C, ⋯ denote matrices with (i, j)-th components a_{ij}, b_{ij}, c_{ij}, ⋯ respectively. The mathematical expectation of a random variable x is denoted by $E\{x\}$ and its conditional one conditioned with respect to C is denoted by $E\{x \mid C\}$.

The following background knowledges are important [L1], [L3].

(1) Stochastic process: A stochastic process $\{x(t,\omega); t \in T, \omega \in \Omega\}$, $x(t,\omega) \in R^n$ is a family of random variables and defined on a common probability space ($\Omega, \mathscr{F}, P$), where Ω is the sample space, $\mathscr{F}$ some minimum σ-algebra, P a probability measure defined on $\mathscr{F}$ and ω a generating point of $\mathscr{F}$. For

each t, $x(t,\omega)$ is an $\mathcal{F}$-measurable function, and for each ω, $x(t,\omega)$, $t \in T$ is called a sample function or realization of the process.

The stochastic process is said to be continuous if T is a connected subset of R^1 and said to be discrete if T is a finite set from R^1, i.e., $\{t_1, t_2, \cdots, t_n\}$. In many applications, we shall think of $\{t_1, t_2, \cdots, t_n\}$ as being the points in time at which observations of the process are available. For economy of description, we omit to write the symbol ω in the following chapters because no confusion will result, and the discrete stochastic process $x(t_k)$ $(k=0, 1, 2, \cdots)$ is usually described by x_k, i.e., $x_k = x(t_k)$.

(2) Stationary process: A stochastic process $\{x(t), t \in T\}$ is said to be a stationary if, for any $\{t_1, t_2, \cdots, t_n\}$, i.e., any finite subset of T, the joint probability distribution of $\{x(t_1+\tau), x(t_2+\tau), \cdots, x(t_n+\tau)\}$ does not depend upon τ.

(3) Wide-sense stationary process: A process $\{x(t), t \in T\}$ is said to be wide-sense stationary or weak stationary if

(a) $E\{x(t)\} = \mu$ (μ: a finite constant)

(b) the covariance matrix

$$P(t, s) \triangleq E\{(x(t) - \mu)(x(s) - \mu)'\}$$

exists, and

(c) $P(t, s)$ depends on only $t-s$.

(4) Independent sequence: A discrete-time stochastic process is said to be an independent sequence if for any set $\{t_1, t_2, \cdots, t_N\} \subset T$, the corresponding random variables x_1, x_2, $\cdots$, x_N $(x_k = x(t_k))$ are independent, i.e., the joint distribution F can be factored as follows:

$$F(x_1, x_2, \cdots, x_N) = F_1(x_1)F_2(x_2)\cdots F_N(x_N),$$

where $F_k(x_k)$ is the marginal distribution function of x_k. Further, if F_1,

$F_2, \cdots, F_N$ are identical functions, then the sequence is said to be an independent and identically distributed (i. i. d.) sequence.

(5) Martingale: A stochastic process $\{x_t, \ t \in T\}$ is said to be a martingale with respect to a sequence of σ-algebras $\{\mathcal{F}_t; \ t \in T\}$ if

(a) $\{\mathcal{F}_t\}$ is increasing,

(b) $E\{x_t \mid \mathcal{F}_t\} = x_t$,

(c) $E\{\mid x_t \mid\} < \infty$,

(d) $x_s = E\{x_t \mid \mathcal{F}_s, \ s<t\}$.

If the property (d) is replaced by

(e) $x_s \leq E\{x_t \mid \mathcal{F}_s, \ s<t\}$,

then $\{x_t\}$ is said to be a submartingale with respect to $\{\mathcal{F}_t\}$ and if the property (d) is replaced by

(f) $x_s \geq E\{x_t \mid \mathcal{F}_s, \ s<t\}$,

then $\{x_t\}$ is said to be a supermartingale with respect to $\{\mathcal{F}_t\}$.

For convenience of the present description, the principal symbols used here are listed below:

t: time variable, particularly present time

k: the variable representing the present time step for a discrete process

$x(t)$, $y(t)$: stochastic processes representing system states and observations in continuous-time systems, respectively

x_k, y_k: stochastic processes representing system states and observations in discrete-time systems, respectively

$w(t)$, $v(t)$: Brownian motion processes representing the system noise and the observation noise, respectively

v_k: discrete white noise sequence representing the observation noise

$\mathring{\theta}$: unknown system parameters to be identified

$A(t)$, $B(t)$, ⋯ : matrices whose components depend on time t

A_k, B_k, ⋯ : matrices whose components depend on step k

$\| x \|$: the Euclidean norm of an n-dimensional vector defined by

$$\| x \| = (\sum_{i=1}^{n} x_i^2)^{1/2} = x'x \qquad (n\text{: dimension of } x)$$

$\| A \|$: the matrix norm of a matrix A defined by

$$\| A \| = [\lambda_{max}(A'A)]^{1/2}$$

where $\lambda_{max}(*)$ is the maximum eigenvalue of *

$x_k \to x$ w.p. 1: x_k converges with probability one to x

$x_k \to x$ in prob.: x_k converges in probability to x

$x_k \overset{law}{\to} x$: x_k converges in law to x

$x \sim N(*, **)$: the probability distribution of the random variable x is Gaussian with mean * and the variance **

$I\{\cdot\}$: characteristic function

Θ: parameter space such that $\theta \in \Theta \subset R^n$ where n is the dimension of Θ

2.2 Mathematical Models for Dynamical Systems

The objective of this thesis is to model various kinds of systems reflecting nonlinear or nonstationary features. With the help of Table 2.1, mathematical models for dynamical systems are classified as follows:

Table 2.1 Features of Dynamical Models

a	Physical	Non-physical
b	Deterministic	Stochastic
c	External	Internal
d	Continuous-time	Discrete-time
e	Time-invariant	Time-variant
f	Linear	Nonlinear

(a) Physical and non-physical models: If the structure of the mathematical model is determined by considerations based on well-known physical properties such as Newtonian mechanics, Kirchhoff's law, etc., the model is said to be a physical model or a mechanical model. Alternatively, in the case where regarding the underlying system as a black box, the system model is obtained by fitting an adequate model to the real data, the obtained model is said to be a non-physical model.

(b) Deterministic and stochastic models: The system model is said to be stochastic if the input process is random or there exists some randomness in the system elements. Otherwise the system is said to be deterministic. It should be noted that deterministic models are often inadequate to describe real systems and we are naturally led to consider the system output as being a realization of a stochastic process.

(c) External and internal descriptions of system model: When the system is represented by the input/output relation as

$$f(y(t), \dot{y}(t), \cdots ; u(t), \dot{u}(t), \cdots) = 0$$

$$(y(t): \text{output}, \ u(t): \text{input})$$

or

$$f(y_k, y_{k-1}, \cdots ; u_k, u_{k-1}, \cdots) = 0$$

$$(y_k\text{: output, } u_k\text{: input}),$$

then it is said that the system is represented by the external description. Alternatively when the system is represented by the state space form as

$$\begin{cases} \frac{dx}{dt} = f(x, u, t, x_0) & (x_0\text{: initial state}) \\ y(t) = g(x, u, t) \end{cases}$$

or

$$\begin{cases} x_{k+1} = f(x_k, u_k, k, x_0) & (x_0\text{: initial state}) \\ y_k = g(x_k, u_k, k), \end{cases}$$

then, the system is represented by the internal description. It should be noted that the internal description is converted uniquely to the external description, whereas the inverse is not true, and this fact causes the difficulty of finding unique estimators in the internal model.

(d) Continuous-time and discrete-time models: If the time variable is a set of real numbers, then the system model is said to be a continuous-time model. Otherwise, if the variable is a set of discontinuous numbers, the model is said to be discrete. In case of discrete models, the time-variable is considered as a set of integers without loss of generality.

(e) Time-invariant and time-variant models: The system model whose all system parameters are time-independent is said to be time-invariant, whereas it is said to be time-variant when the several system parameters vary with time evolution. Time-invariant model is often said to be a stationary model.

(f) Linear and nonlinear models: When system parameters are independent of the input, output and their derivations, the system model is said to be linear. Alternatively when there exist parameters which depend on the input, output or their derivative, the model is said to be nonlinear.

The classification of various kinds of models is depicted in Fig. 2.1. In this thesis, the author deals with models which are classified as non-physical stochastic models, especially time-variant or nonlinear and further time-variant nonlinear ones.

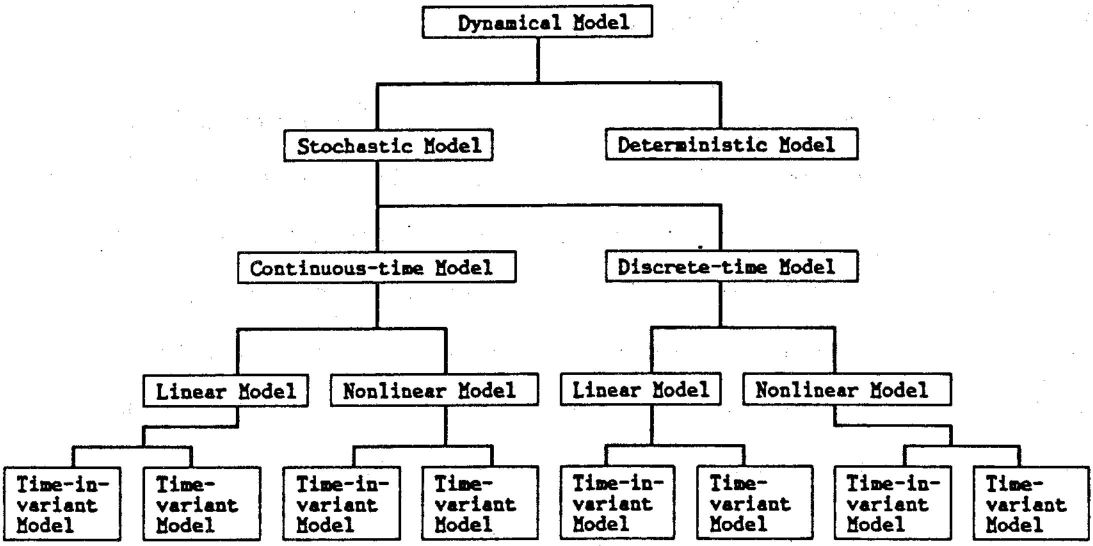


Fig. 2.1 Classification of various kinds of dynamical models where the classification for deterministic models is omitted because of economy of space

2.3 Methods for Parameter Identification

Let Y_N be a collection of the output process y_k $(k=0,1,2,\cdots,N)$, i.e., $Y_N \triangleq \{y_0, y_1, \cdots, y_N\}$. The realization of Y_N is called the observation data or simply the data. With the preassigned model structure, unknown system parameters must be identified by using a single record of the observation data Y_N. Let $\overset{\circ}{\theta}$ be the vector representing unknown parameters, then the estimator $\hat{\theta}(Y_N)$ is defined as a function of Y_N. For a given data (a realization of of Y_N), $\hat{\theta}(Y_N)$ is called an estimate. In the followings, $\hat{\theta}(Y_N)$ is usually abbreviated as $\hat{\theta}_N$. Commonly used estimators are as follows [G1].

[Definition 2.1] Given a function $f: R^N \times R^n \to R^N$, then a least squares estimator $\hat{\theta}_N$ is defined by

$$(2.1) \quad \| f(Y_N, \hat{\theta}_N) \|_Q \leq \| f(Y_N, \theta_N^*) \|_Q$$

for any other function $\theta_N^* = \theta^*(Y_N)$ where $\| x \|_Q = x'Qx$ with a matrix $Q > 0$.

Since the least squares estimator minimizes the weighted sum of squares of vector $f(Y_N, \theta)$, the least squares approach is independent of any underlying probabilistic description of the data-generating mechanism.

[Definition 2.2] Given a parametric family of $\{p(\cdot \mid \theta);\ \theta \in \Theta \subset R^n\}$ of probability density functions on the same sample space Ω, then the maximum likelihood estimator (M.L.E.) $\hat{\theta}: \Omega \to \Theta$ is defined by

$$(2.2) \quad p(Y_N \mid \hat{\theta}_N) \geq p(Y_N \mid \theta_N^*)$$

where θ_N^* is any other function.

The function $p(Y_N \mid \cdot)$ is said to be the likelihood function. The heuristic justification of maximum likelihood estimators is that the likelihood function can be interpreted as giving a measure of the plausibility of the data under different parameters. Thus, the maximum likelihood estimator $\hat{\theta}_N$ is the value of θ which makes the data most plausible as measured by the likelihood function.

[Definition 2.3] A maximum a posteriori estimator is the mode of the a posteriori probability density function $p(\hat{\theta}_N \mid Y_N)$.

Since the a posteriori probability density function $p(\theta \mid Y_N)$ can be represented by

$$p(\theta \mid Y_N) = \frac{p(\theta)}{p(Y_N)} \, p(Y_N \mid \theta),$$

the maximum a posteriori estimator coincides with the maximum likelihood estimator, provided that a priori distribution of θ is uniform.

Let statistical moments of the stationary output process $\{y_k\}$ be defined by

$$(2.3) \quad \psi(\ell_0, \ell_1, \cdots, \ell_p) \triangleq E\{y_k^{\ell_0} y_{k-1}^{\ell_1} \cdots y_{k-p}^{\ell_p}\}$$

and assume that the unknown system parameter $\dot{\theta}$ is represented by a function of statistical moments of the output process.

[Definition 2.4] The estimator by the moment method is given as a function of estimators of $\psi(\ell_0, \ell_1, \cdots, \ell_p)$.

The estimator of $\psi(\ell_0, \ell_1, \cdots, \ell_p)$ is usually given by

$$(2.4) \quad \hat{\psi}_N(\ell_0, \ell_1, \cdots, \ell_p) = \frac{1}{N-p} \sum_{k=p}^{N} y_k^{\ell_0} y_{k-1}^{\ell_1} \cdots y_{k-p}^{\ell_p}.$$

Estimators defined above are usually used to identify the unknown parameters in the model of external description. The representative method for identifying unknown parameters in the state space model (internal description) is to augment state vectors by unknown parameters and one of approximated nonlinear filtering techniques is utilized.

Clearly we would like to get "good" estimate of the unknown parameter $\mathring{\theta}$. Hence the meaning of the "good" estimate has to be exactly defined. Properties that are commonly used to describe estimators are defined below. However, it should be noted that holding one or more of these properties does not necessarily imply that the estimator is "best" for the given purpose.

[Definition 2.5] An estimator $\hat{\theta}_N$ for $\mathring{\theta}$ is said to be unbiased if $\hat{\theta}_N$ has the expected value $\mathring{\theta}$, i.e.,

$$(2.5)\quad E\{\hat{\theta}_N\} = \mathring{\theta}.$$

[Definition 2.6] An estimator $\hat{\theta}_N$ for $\mathring{\theta}$ is said to be a uniformly minimum mean square if

$$(2.6)\quad E\{(\hat{\theta}_N - \mathring{\theta})(\hat{\theta}_N - \mathring{\theta})'\} \leq E\{(\theta^*_N - \mathring{\theta})(\theta^*_N - \mathring{\theta})'\}$$

for all $\mathring{\theta}$ in the parameter space Θ and every other estimator θ^*_N.

[Definition 2.7] An estimator $\hat{\theta}_N$ is said to be a minimum variance unbiased estimator if it has the minimum mean square error uniformly in $\mathring{\theta}$ among the class of unbiased estimator.

[Definition 2.8] An unbiased estimator is said to be efficient if its covariance is equal to the Cramér-Rao lower bound, where the Cramér-Rao lower bound is given by the inverse of Fisher's information matrix defined by

$$(2.7)\quad M(\theta) \triangleq E\{(\frac{\partial \log p(Y_N \mid \theta)}{\partial \theta})'(\frac{\partial \log p(Y_N \mid \theta)}{\partial \theta})\}.$$

Asymptotic versions of definitions 2.5 to 2.8 are also given, i.e., asymptotically unbiased estimator, asymptotically minimum variance estimator, etc.

2.4 Stochastic Convergence

In order to investigate theoretically asymptotic properties of estimators described in the previous section, the stochastic convergence concept should be investigate in some detail as the mathematical base [L1], [L3]. A stochastic process $\{x_k\}$ is a sequence of functions $x_k=x_k(\omega)$ k=1, 2, $\cdots$. Hence in order to define the convergence of stochastic processes, the existence of probability measure must be taken into account. Commonly used concepts of stochastic convergence are given as follows.

[Definition 2.9] The stochastic process $\{x_k\}$ is said to converge to x with probability one (w.p. 1) if

$$(2.8) \quad P\{\lim_{k\to\infty} x_k = x\} = 1.$$

[Definition 2.10] The stochastic process $\{x_k\}$ is said to converge to x in probability if for any $\varepsilon > 0$

$$(2.9) \quad \lim_{k\to\infty} P\{ | x_k - x | > \varepsilon\} = 0.$$

[Definition 2.11] A sequence of random variables $\{x_k\}$ is said to converge in the r-th mean to x if all x_k and x have finite moments of order $r > 0$ and if

$$(2.10) \quad \lim_{k\to\infty} E\{ | x_k - x |^r\} = 0.$$

In case of r=2, the convergence is usually said to be the convergence in the quadratic mean.

It should be noted that if the estimator $\hat{\theta}_N$ of $\dot{\theta}$ converges to its true value as $N\to\infty$ with probability one or in probability, $\hat{\theta}_N$ is said to be a strong consistent estimator or a weak consistent estimator respectively.

The remaining commonly used concept of stochastic convergence is the weak convergence of distribution functions, where "weak" means that a sequence of distribution functions converges to a limiting function at all continuity points.

[Definition 2.12] If the sequence of distribution functions $\{F_k\}$ of $\{x_k\}$ converges to F at continuity points, $\{x_k\}$ is said to converge in law to a random variable x whose distribution function is given by F.

The mutual relation between concepts of stochastic convergence defined in this section is shown in Fig. 2.2.

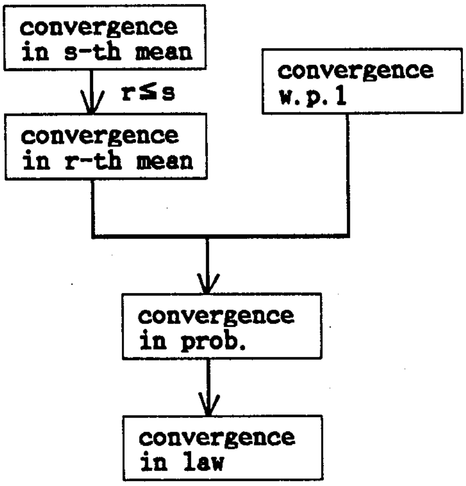


Fig. 2.2 The relation between concepts of stochastic convergence

2.5 Convergence of Martingales and Dependent Random Variables

In the remainder of this chapter, theorems of law of large numbers for martingales and central limit theorems for both martingales and dependent random variables are presented, which play a key role in the analysis of properties of estimators proposed in this thesis.

[Theorem 2.1] Let $\{x_k\}$ be a scalar martingale with respect to $\{\mathscr{F}_k\}$ such that $E\{x_k^2\} < \infty$ for all k. Then, there exists a random variable x having the bounded variance such that

$$(2.11) \quad x_k \to x \quad \text{w.p. } 1 \quad \text{as } N \to \infty.$$

(Proof) See e.g. [N1], [R3].

[Theorem 2.2] (Khazminskii) [K6] Assume that $\{x_k\}$ is a scalar martingale with respect to $\{\mathscr{F}_k\}$ and suppose that $E\{x_k^2\} \leq ck$. Then,

$$(2.12) \quad \frac{x_k}{k} \to 0 \quad \text{w.p. } 1 \quad \text{as } k \to \infty.$$

[Theorem 2.3] (Brown) [B2] For the martingale $\{x_k, \mathscr{F}_k\}$ (k=1,2,⋯) with $x_0=0$, assume that

$$(2.13) \quad X_k \triangleq \left(\sum_{j=1}^{k} E\{(x_j - x_{j-1})^2 \mid \mathscr{F}_{j-1}\}\right)^{1/2}$$

and

$$(2.14) \quad u_k \triangleq (E\{x_k^2\})^{1/2}.$$

Let further following conditions hold:

$$(2.15) \quad \frac{X_k^2}{u_k^2} \to 1 \quad \text{in prob.} \quad \text{as } k \to \infty$$

and for any $\varepsilon > 0$,

$$(2.16) \quad \frac{1}{u_k^2} \sum_{j=1}^{k} E\{(x_j - x_{j-1})^2 I\{|x_j - x_{j-1}| > \varepsilon u_k\}\} \to 0$$

in prob. as $k \to \infty$.

Then,

$$\frac{x_k}{u_k} \overset{law}{\to} x \qquad \text{as} \quad k \to \infty.$$

where

$$x \sim N(0,1).$$

[Theorem 2.4] (Rosen) [R4] Consider the sum of doubly indexed random variables $\{x_{k,N}\}$ such that

$$(2.17) \quad x_N = \sum_{k=1}^{N} x_{k,N}$$

$$(2.18) \quad E\{x_{k,N}\} = 0$$

$$(2.19) \quad E\{x_N^2\} = 1,$$

where $\{x_{1,N}, x_{2,N}, \cdots, x_{s,N}\}$ and $\{x_{k,N}, x_{k+1,N}, \cdots, x_{n,N}\}$ are independent for $k-s > \nu$. Assume that the following two conditions hold:

$$(2.20) \quad \lim_{N\to\infty} \sup \sum_{k=0}^{N} E\{x_{k,N}^2\} < \infty$$

$$(2.21) \quad \lim_{N\to\infty} \sum_{k=0}^{N} \int_{|x_{k,N}| > \varepsilon} |x_{k,N}|^2 dP(x_{k,N}) = 0 \qquad \text{for every } \varepsilon > 0.$$

Then the distribution function of x_N converges to $N(0,1)$.

CHAPTER 3 MODELING FOR NONSTATIONARY LINEAR SYSTEMS

3.1 Introductory Remarks

It is often observed that the random data obtained from real systems exhibit more or less the nonstationarity, i.e., the mean value, the variance or the covariance of the observed data shows time dependence. Hence in order to treat the data mentioned above, one has to use an appropriate class of nonstationary models, in which time-varying system parameters are included. In this chapter, the author proposes a method of the system modeling for a class of nonstationary linear systems from noisy observation data, where the determination problem of system order is also investigated.

The underlying system is assumed to be adequately modeled by a linear stochastic differential equation of Itô-type with unknown time-varying coefficients as described in Section 3.2. Furthermore, since the output data available to the system modeling is usually corrupted by random noise, the model of the observation mechanism is also given by a stochastic differential equation as well as the system dynamics. To formulate the order determination problem within the framework of Bayesian decision theory, necessary hypotheses are set in Section 3.3, and the likelihood-ratio function associated with the multiple-decision theoretic approach for the order determination is given in Section 3.4. Augmenting state variables of the system model by unknown system parameters, the identification problem is formulated as the nonlinear filtering problem and this is described in Section 3.5. Simulation results are presented in Section 3.6.

3.2 Nonstationary Linear System Models

It is of interest to consider whether a sample path of the stochastic system has a well-defined representation as the solution to the n-dimensional stochastic differential equation of the form,

$$(3.1) \quad dx(t) = A(t)x(t)dt + G(t)dw(t), \qquad x(0) = x_0$$

where $A(t)$, $G(t)$ are respectively $n \times n$ and $n \times p$ matrices and $w(t)$ is the p-dimensional standard Brownian motion process. We have a continuous record of a realization of the m-dimensional observation process $y(t)$ through the dynamics represented by

$$(3.2) \quad dy(t) = H(t)x(t)dt + R(t)dv(t), \qquad y(0) = 0$$

where $H(t)$, $R(t)$ are respectively $m \times n$ and $m \times q$ matrices; and $v(t)$ is the q-dimensional standard Brownian motion process independent of $w(t)$. We refer to such a set of systems defined by (3.1) and (3.2) as the system Σ_0.

In the system Σ_0, if we set

$$(3.3) \quad A(t) = \begin{bmatrix} 0 & 1 & & & 0 \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & 0 & & 0 & 1 \\ -a_n(t) & -a_{n-1}(t) & \cdots & -a_2(t) & -a_1(t) \end{bmatrix},$$

then the mathematical model (3.1) expresses a time-variant linear system whose output is determined by a solution process of the n-th order linear differential equation excited by an appropriate number of white Gaussian random processes.

The system Σ_0 given by (3.1) and (3.2) may not be completely known. For instance, the system model given above is specified by the set of unknown parameters $a_1(t), \cdots, a_n(t)$ which are assumed to be time-variant. In addition, if we have only the observation data $\{y(t),\ 0 \leq t \leq T\}$, it is natural to consider that the system order n is unknown.

We set the matrix H(t) as

$$H(t) = \left[\begin{array}{c|c} \underbrace{\begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_m(t) \end{matrix}}_{m} & \underbrace{\;0\;}_{n-m} \end{array}\right] \Big\} m \qquad \text{if } n \geq m \tag{3.4a}$$

$$H(t) = \underbrace{\left[\begin{array}{c} \begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_m(t) \end{matrix} \\ \hline 0 \end{array}\right]}_{n} \begin{matrix} \Big\} n \\ \} m-n \end{matrix} \qquad \text{if } n < m. \tag{3.4b}$$

The $m \times q$ matrix R(t) is known and $n \times p$ matrix G(t) has the structure

$$G(t) = \underbrace{\left[\begin{array}{c} 0 \\ \hline \begin{matrix} g_p(t) & & 0 \\ & \ddots & \\ 0 & & g_1(t) \end{matrix} \end{array}\right]}_{p} \begin{matrix} \} n-p \\ \Big\} p \end{matrix} \qquad \text{if } n \geq p \tag{3.5a}$$

$$G(t) = \left[\begin{array}{c|c} \underbrace{\;0\;}_{p-n} & \underbrace{\begin{matrix} g_n(t) & & 0 \\ & \ddots & \\ 0 & & g_1(t) \end{matrix}}_{n} \end{array}\right] \Big\} n \qquad \text{if } n < p. \tag{3.5b}$$

In (3.4) and (3.5) the components $\{h_i(t);\ i=1,\cdots,m\}$ and $\{g_i(t);\ i=1,\cdots,p\}$ are assumed to be known a priori for fixed m and p, while $\{a_i(t);\ i=1,\cdots,n\}$ of the matrix A(t) are unknown but they are of the form

$$(3.6) \qquad a_i(t) = \sum_{k=1}^{N} \frac{1}{(k-1)!} a_{ik} t^{k-1}, \qquad i=1,2,\cdots,n$$

where N is the fixed integer and coefficients $\{a_{ik}\}$ are unknown constants. When we model the time-variant system, the assumption expressed by (3.6) is often used in the literature, see e.g. [G3]. Then, our objectives for the modeling of dynamical systems are twofold: (i) determination of the system order n, and (ii) identification of unknown parameters, based on the observation data $Y_0^T \triangleq \{y(t);\ 0 \leq t \leq T\}$.

3.3 <u>Basic Hypotheses</u>

Guided by the decision making concept, the following hypothesis H_j ($j=1,2,\cdots,k$) is set:

H_j: hypothesis that the order of the system (3.1) is j, i.e., $n=j$.

Then, with the hypothesis H_j, the system model $\Sigma^{(j)}$ is given by

$$\Sigma^{(j)} \begin{cases} dx^{(j)}(t) = A(t)^{(j)}x^{(j)}(t)dt + G^{(j)}(t)dw(t), & x^{(j)}(0) = x_0^{(j)} \\ \\ dy(t) = H^{(j)}(t)x^{(j)}(t)dt + R(t)dv(t), & y(0) = 0, \end{cases} \tag{3.7}$$

where $x^{(j)}(t)$ is the j-dimensional state vector; and matrices $A^{(j)}(t)$, $H^{(j)}(t)$ and $G^{(j)}(t)$ are respectively $j\times j$, $m\times j$ and $j\times p$ matrices given by

$$A^{(j)}(t) = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -a_j^{(j)}(t) & -a_{j-1}^{(j)}(t) & \cdots\cdots & -a_1^{(j)}(t) \end{bmatrix}, \tag{3.8}$$

$$H^{(j)}(t) = \left.\begin{bmatrix} \underbrace{\begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_m(t) \end{matrix}}_{m} & \vdots & \underbrace{\begin{matrix} \\ 0 \\ \\ \end{matrix}}_{j-m} \end{bmatrix}\right\} m \qquad \text{if } j \geq m \tag{3.9a}$$

$$H^{(j)}(t) = \underbrace{\begin{bmatrix} \begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_j(t) \end{matrix} \\ \cdots\cdots\cdots\cdots\cdots \\ 0 \end{bmatrix}}_{j} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ b \\ c \end{matrix}}\right\} j \\ \left.\vphantom{0}\right\} m-j \end{matrix} \qquad \text{if } j<m \tag{3.9b}$$

$$(3.10\text{a})\qquad G^{(j)}(t) = \underbrace{\left[\begin{array}{ccc} & 0 & \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ g_p(t) & 0 & \\ & \ddots & \\ 0 & & g_1(t) \end{array}\right]}_{p} \begin{array}{l} \left.\begin{array}{c} \\ \\ \end{array}\right\} j-p \\ \left.\begin{array}{c} \\ \\ \\ \end{array}\right\} p \end{array} \qquad \text{if } j \geq p$$

$$(3.10\text{b})\qquad G^{(j)}(t) = \left[\begin{array}{c:ccc} & g_j(t) & & \\ & & \ddots & 0 \\ \underbrace{\quad 0 \quad}_{p-j} & & & \\ & 0 & \ddots & \\ & \underbrace{\qquad\qquad\qquad}_{j} & & g_1(t) \end{array}\right] \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} j \qquad \text{if } j < p.$$

Time-varying parameters $\{a_i^{(j)}(t);\ i=1,2,\cdots,j\}$ in (3.8) are given by

$$(3.11)\qquad a_i^{(j)}(t) = \sum_{k=1}^{N} \frac{1}{(k-1)!}\, a_{ik}^{(j)} t^{k-1}, \qquad i=1,2,\cdots,j.$$

Let $z^{(j)}(t)$ be a new $(j+jN)$-dimensional vector defined by

$$(3.12)\qquad \left\{\begin{array}{ll} z_i^{(j)}(t) \triangleq x_i^{(j)}(t) & \text{for } i=1,2,\cdots,j \\ z_i^{(j)}(t) \triangleq \displaystyle\sum_{k=i-j}^{N} \frac{1}{(k+j-i)!}\, a_{1k}^{(j)} t^{k+j-i} & \text{for } i=j+1,\cdots,j+N \\ z_i^{(j)}(t) \triangleq \displaystyle\sum_{k=i-(j+N)}^{N} \frac{1}{(k+j+N-i)!}\, a_{2k}^{(j)} t^{k+j+N-i} & \\ \quad\vdots & \text{for } i=j+N+1,\cdots,j+2N \\ \quad\vdots & \\ z_i^{(j)}(t) \triangleq \displaystyle\sum_{k=i-\{j+(j-1)N\}}^{N} \frac{1}{\{k+j+(j-1)N-i\}!}\, a_{jk}^{(j)} t^{k+j+(j-1)N-i} & \\ & \text{for } i=j+(j-1)N,\cdots,j+jN \end{array}\right.$$

With $z_i^{(j)}(t)$ defined by (3.12), the system model $\Sigma^{(j)}$ is represented by

$$
(3.13a)\quad \begin{cases}
dz_i^{(j)}(t) = z_{i+1}^{(j)}(t)dt + g_{j-i+1}(t)dw_i(t) & \text{for } i=1,2,\cdots,j-1 \\
dz_j^{(j)}(t) = -\sum_{i=1}^{j} z_{j+(i-1)N+1}^{(j)}(t)z_i^{(j)}(t)dt + g_1(t)dw_j(t) & \\
dz_i^{(j)}(t) = z_{i+1}^{(j)}(t)dt & \text{for } i=j+1,j+2,\cdots,j+jN \\
 & \text{except for } i=j+N,j+2N,\cdots,j+jN \\
dz_i^{(j)}(t) = 0 & \text{for } i=j+N,j+2N,\cdots,j+jN
\end{cases}
$$

$$
(3.13b)\quad dy_i(t) = h_i(t)z_i^{(j)}(t)dt + \sum_{k=1}^{q} r_{ik}(t)dv_k(t) \qquad \text{for } i=1,2,\cdots,m
$$

where if $j>p$, then $g_i(t)=0$ $(i=p+1,\cdots,j)$ and if $j<m$, $h_j(t)=0$ $(i=j+1,\cdots,m)$.

Furthermore, (3.13) can be rewritten in a vector form as

$$
(3.14)\quad \tilde{\Sigma}^{(j)} \begin{cases}
dz^{(j)}(t) = F^{(j)}[z^{(j)}(t)]z^{(j)}(t) + C^{(j)}(t)dw(t), \\
\qquad\qquad\qquad\qquad z^{(j)}(0) = z_0^{(j)} \\
dy(t) = M^{(j)}(t)z^{(j)}(t) + R(t)dv(t), \qquad y(0) = 0
\end{cases}
$$

where $F^{(j)}[z^{(j)}(t)]$, $C^{(j)}(t)$ and $M^{(j)}(t)$ are respectively $(j+jN)\times(j+jN)$-, $(j+jN)\times p$- and $m\times(j+jN)$-dimendional matrices whose structures are given in Appendix 3.A.

3.4 Order Determination of Nonstationary Linear Models

3.4.1 Radon-Nikodym Derivative and Likelihood-Ratio Function

In this section, a method for order determination is presented by adopting the multiple alternative hypothesis testing [S4], [V2]. Based on the system model $\tilde{\Sigma}^{(j)}$, the hypothesis H_j is

$$(3.15) \quad H_j: dy(t) = M^{(j)}(t)z^{(j)}(t)dt + R(t)dv(t), \quad y(0) = 0$$

where $j=1,2,\cdots,K$; and $z^{(j)}(t)$ is the solution process of (3.13a). Our method is to accept one of hypotheses $\{H_j;\ j=1,2,\cdots,K\}$. To fix the idea, introduce the process $\{\tilde{y}(t)\}$ defined by

$$(3.16) \quad H_0: d\tilde{y}(t) \triangleq R(t)dv(t), \quad \tilde{y}(0) = 0$$

and consider the likelihood-ratio function

$$(3.17) \quad \Lambda^{(j)}(t) = \frac{p\{Y_0^t \mid H_j\}}{p\{\tilde{Y}_0^t \mid H_0\}}, \quad (j=1,2,\cdots,K)$$

where

$$\begin{cases} Y_0^t \triangleq \{y(s);\ 0 \le s \le t\} \\ \tilde{Y}_0^t \triangleq \{\tilde{y}(s);\ 0 \le s \le t\}. \end{cases}$$

Let P_j and P_0 be the measures induced respectively in the space of continuous functions by the observation Y_0^t under H_j and $\tilde{Y}_0^t$.

[Lemma 3.1] The measure P_j is absolutely continuous with respect to the measure P_0, i.e., $p_j \ll P_0$, and the Radon-Nikodym derivative of P_j with respect to P_0 is given by

$$(3.18)\quad \frac{dP_j}{dP_0} = \exp\{\int_0^t [\hat{z}^{(j)}(t \mid t)]'[M^{(j)}(t)]'\{R(t)R'(t)\}^{-1}dy(t)$$

$$- \frac{1}{2} \int_0^t \| M^{(j)}(t)\hat{z}^{(j)}(t \mid t) \|^2_{\{R(t)R'(t)\}^{-1}} dt\},$$

where

$$(3.19)\quad \hat{z}^{(j)}(t \mid t) \triangleq E\{z^{(j)}(t) \mid Y_0^t, H_j\}.$$

The proof of Lemma 3.1 is straightforward by using results of [S5], [K3] and [K6].

Defining the vector $\hat{\xi}^{(j)}(t \mid t)$ and the matrix $H_m(t)$ by

$$(3.20)\quad \hat{\xi}^{(j)}(t \mid t) \triangleq [\hat{x}_1^{(j)}(t \mid t) \cdots \hat{x}_m^{(j)}(t \mid t)]'$$

$$= [\hat{z}_1^{(j)}(t \mid t) \cdots \hat{z}_m^{(j)}(t \mid t)]'$$

and

$$(3.21)\quad H_m(t) \triangleq \text{diag.}[h_1(t)\quad h_2(t) \cdots h_m(t)]$$

respectively, where if $j<m$, $h_j(t)=0$ for $i=j+1, \cdots, m$ and noting the structure of (A.3a), we can rewrite (3.18) as

$$(3.22)\quad \frac{dP_j}{dP_0} = \exp\ \{\int_0^t [\hat{\xi}^{(j)}(t \mid t)]'H_m(t)\{R(t)R'(t)\}^{-1}dy(t)$$

$$- \frac{1}{2} \int_0^t \| H_m(t)\hat{\xi}^{(j)}(t \mid t) \|^2_{\{R(t)R'(t)\}^{-1}} dt\}.$$

Recalling the relation between Radon-Nikodym derivative and the likelihood-ratio function [K3],

$$(3.23)\quad \frac{dP_j}{dP_0} = \frac{p\{Y_0^T \mid H_j\}}{p\{\tilde{Y}_0^T \mid H_0\}} = \Lambda^{(j)}(T), \qquad (j=1,2,\cdots,K)$$

we have

$$(3.24)\quad \Lambda^{(j)}(t) = \exp\{\int_0^t [\hat{\xi}^{(j)}(s \mid s)]' H_m(s)\{R(s)R'(s)\}^{-1} dy(s) - \frac{1}{2}\int_0^t \| H_m(s)\hat{\xi}^{(j)}(s \mid s) \|^2_{\{R(s)R'(s)\}^{-1}} ds\}.$$

An application of the Itô-stochastic calculus to (3.24) gives that the likelihood-ratio function $\Lambda^{(j)}(t)$ satisfies the following stochastic differential equation:

$$(3.25)\quad \begin{cases} d\Lambda^{(j)}(t) = \Lambda^{(j)}(t)[\hat{\xi}^{(j)}(t \mid t)]' H_m(t)\{R(t)R'(t)\} dy(t) \\ \Lambda^{(j)}(0) = 1, \quad j=1,2,\cdots,K. \end{cases}$$

3.4.2 Decision Making

Consider the following average risk $\bar{C}$:

$$(3.26)\quad \bar{C} = \sum_{i=1}^{K}\sum_{j=1}^{K} c_{ij} P[H_j] \int_{S_i} p\{Y_0^T \mid H_j\} dY_0^T,$$

where c_{ij} is the cost associated with choosing hypothesis H_i when the hypothesis H_j is true; $P[H_j]$ is the a priori probability of H_j; and S_i is the family of Y_0^T where H_i is acceptable. Without loss of generality, let the costs $\{c_{ij};\ i,j=1,2,\cdots,K\}$ be $c_{ij}=1$ for $i \neq j$ and $c_{ij}=0$ for $i=j$. Then (3.26) can be represented by

$$(3.27)\quad \bar{C} = \sum_{j=1}^{K} P[H_j]\{\int_{S_1} p\{Y_0^T \mid H_j\}dY_0^T + \cdots + \int_{S_{i-1}} p\{Y_0^T \mid H_j\}dY_0^T$$

$$+ \int_{S_{i+1}} p\{Y_0^T \mid H_j\}dY_0^T + \cdots + \int_{S_K} p\{Y_0^T \mid H_j\}dY_0^T\}$$

$$= \sum_{j=1}^{K} P[H_j]\int_{S-S_j} p\{Y_0^T \mid H_j\}dY_0^T.$$

where S is the direct sum of S_i $(i=1,\cdots,K)$, i.e., $S \triangleq S_1 \oplus \cdots \oplus S_K$ and $\int_S p\{Y_0^T \mid H_j\}dY_0^T=1$ has been used. Rewriting as

$$(3.28)\quad \bar{C} = P[H_j]\int_{S-S_j} p\{Y_0^T \mid H_j\}dY_0^T + P[H_i]\int_{S-S_i} p\{Y_0^T \mid H_i\}dY_0^T$$

$$+ \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^{K} P[H_\ell]\int_{S-S_\ell} p\{Y_0^T \mid H_\ell\}dY_0^T,$$

we have

$$(3.29)\quad \bar{C} = \int_{S_j} [P[H_i]p\{Y_0^T \mid H_i\} - P[H_j]p\{Y_0^T \mid H_j\}]dY_0^T$$

$$+ P[H_j] + P[H_i]\int_{S-S_i \oplus S_j} p\{Y_0^T \mid H_i\}dY_0^T$$

$$+ \sum_{\substack{\ell=1 \\ \ell \neq i,j}}^{K} P[H_\ell]\int_{S-S_\ell} p\{Y_0^T \mid H_\ell\}dY_0^T.$$

In (3.29), if $S-S_i \oplus S_j$ is determined to be fixed, then terms except the first one are considered to be constants. Then it can be seen that the average risk $\bar{C}$ given by (3.27) is minimized by choosing S_j which satisfies

(3.30) $P[H_j]p\{Y_0^T \mid H_j\} > P[H_i]p\{Y_0^T \mid H_i\}$

for all $i=1,2,\cdots,K$ except for $j=i$. Then we have a statement of the decision making stated as follows: Accept H_j if the inequality (3.30) holds for $i=1,2,\cdots,K$ $(i \neq j)$; and reject H_j if otherwise.

We simply assume that the a priori probabilities are uniform, i.e., $P[H_1]=P[H_2]=\cdots=P[H_K]$. Hence, invoking the likelihood-ratio function $\Lambda^{(j)}(t)$ defined by (3.17), we have the following decision rule for system order determination, based on the observed data $Y_0^T=\{y(t),\ 0 \leq t \leq T\}$.

[Decision Rule 3.1] Decide that order of the system (3.1) is j which gives the maximum value of the likelihood-ratio function $\Lambda^{(j)}(T)$ for $j=1,2,\cdots,K$.

From (3.24) or (3.25) we see that, in order to compute likelihood-ratio function $\Lambda^{(j)}(t)$, it is necessary to compute $\hat{\xi}^{(j)}(t \mid t)=[\hat{x}_1^{(j)} \cdots \hat{x}_m^{(j)}]'$. In other words, the computation of $\Lambda^{(j)}(t)$ requires the state estimation of the system $\sum^{(j)}$. In the following section, the method of parameter identification together with the state estimation is developed.

3.5 Parameter Identification

Under the hypothesis H_j $(j=1, 2, \cdots, K)$, unknown parameters $\{a_i^{(j)}(t),\ i=1, 2, \cdots, j\}$ contained in the system model (3.7) are identified from the observation data Y_0^T. The system model $\tilde{\Sigma}^{(j)}$ defined by (3.14) is used as the basic equation for parameter identification.

Although we are free to choose the type of appropriate filter, we use here the well known truncated second-order filter [J1]. For $\tilde{\Sigma}^{(j)}$, the resulting filter equation is

$$
\tag{3.31}
\left\{
\begin{aligned}
d\hat{z}^{(j)}(t \mid t) &= [F^{(j)}[\hat{z}^{(j)}(t \mid t)]\hat{z}^{(j)}(t \mid t) + \frac{1}{2}\sum_{i_1=1}^{j+jN}\sum_{i_2=1}^{j+jN} e_{i_1 i_2}^{(j)} \\
&\quad \times p_{i_1 i_2}^{(j)}(t \mid t)]dt + P^{(j)}(t \mid t)[M^{(j)}(t \mid t)]' \\
&\quad \times \{R(t)R'(t)\}^{-1}\{dy(t) - M^{(j)}(t)\hat{z}^{(j)}(t \mid t)dt\}, \\
&\qquad\qquad\qquad\qquad \hat{z}^{(j)}(0 \mid 0) = \hat{z}_0^{(j)} \\
\frac{dP^{(j)}(t \mid t)}{dt} &= B^{(j)}[\hat{z}^{(j)}(t \mid t)]P^{(j)}(t \mid t) + P^{(j)}(t \mid t) \\
&\quad \times (B^{(j)}[\hat{z}^{(j)}(t \mid t)])' + C^{(j)}(t)[C^{(j)}(t)]' \\
&\quad - P^{(j)}(t \mid t)[M^{(j)}(t)]'\{R(t)R'(t)\}^{-1} \\
&\quad \times M^{(j)}(t)P^{(j)}(t \mid t), \qquad P^{(j)}(0 \mid 0) = P_0^{(j)},
\end{aligned}
\right.
$$

where $P^{(j)}(t \mid t) \triangleq \text{cov.}\{z^{(j)}(t) \mid Y_0^t, H_j\}$; the vector $e_{i_1 i_2} \in R^{j+jN}$ is defined by

$$(3.32)\quad \begin{cases} e^{(j)}_{i_1 i_2} \triangleq (\underbrace{0 \ \cdots \ 0 \ \ -1}_{j} \ \vdots \ \underbrace{0}_{jN})' & \\ e^{(j)}_{i_2 i_1} \triangleq e^{(j)}_{i_1 i_2} & \text{for } i_2 = j+(i_1-1)N+1 \\ e_{i_1 i_2} \triangleq 0 & \text{for } i_2 \neq j+(i_1-1)N+1 \end{cases}$$

and $B^{(j)}[\hat{z}^{(j)}(t \mid t)]$ is the $(j+jN) \times (j+jN)$ matrix whose structure is also given in Appendix 3.A.

3.6 Digital Simulation Studies

Example 3.1 First the author considers the simplest case, that is, the observation data $Y_0^T=\{y(t),\ 0\leq t\leq T\}$ is determined by

$$(3.33)\qquad dy(t) = hx_1(t)dt + rdv(t), \qquad y(0) = 0$$

where h and r are constants. The true system output $x_1(t)$ is assumed to be a component of the solution process of the following two-dimensional linear system modeled by

$$(3.34)\qquad \begin{cases} d\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_2 & -a_1 \end{bmatrix} dt + \begin{bmatrix} 0 \\ g \end{bmatrix} dw(t) \\ \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_{10} \\ x_{20} \end{bmatrix}. \end{cases}$$

With the help of Fig. 3.1, the procedure to modeling (3.34) is performed by the following steps:

(1) Collection of observation data. Obtain the observation data up to the preassigned time T, $Y_0^T=\{y(t),\ 0\leq t\leq T\}$, from the actual physical system at hand. In digital simulation studies, the observation data is obtained by simulating (3.33) and (3.34) on a digital computer. A typical sample run of the observation process y(t) is depicted in Fig. 3.2.

(2) Assignment of hypotheses and parameter identification. Assign K hypotheses H_j $(j=1,2,\cdots,K)$. With the hypothesis H_j, write the system model $\Sigma^{(j)}$ by (3.7). The system $\Sigma^{(j)}$ is written by

$$(3.35)\quad \begin{cases} d\begin{bmatrix} x_1^{(j)}(t)\\ x_2^{(j)}(t)\\ \vdots\\ x_j^{(j)}(t)\end{bmatrix} = \begin{bmatrix} 0 & 1 & & 0\\ & \ddots & \ddots & \\ 0 & & 0 & 1\\ -a_j^{(j)}(t) & -a_{j-1}^{(j)}(t) & \cdots & -a_1^{(j)}(t)\end{bmatrix} + \begin{bmatrix} 0\\ \\ 0\\ g\end{bmatrix} dw(t)\\ \qquad\qquad (x_1^{(j)}(0) \cdots x_j^{(j)}(0))' = (x_{10}^{(j)} \cdots x_{j0}^{(j)})'\\ dy(t) = hx_1^{(j)}(t)dt + rdv(t), \qquad y(0) = 0. \end{cases}$$

In (3.35), we assume that unknown parameters $\{a_i^{(j)},\ i=1,2,\cdots,j\}$ are constants. Since we are free to choose the number of hypotheses K, K is fixed as K=5 in simulation experiments.

(3) Computation of likelihood-ratio function. Using the observed data Y_0^T, solve the filtering equation (3.31), and compute the likelihood-ratio function $\Lambda^{(j)}(t)$ by (3.24) or (3.25) for j=1, 2, ⋯, K.

For the system model $\sum^{(j)}$ described by (3.35), define the vector $z^{(j)}(t)$ by

$$(3.36)\quad \begin{cases} z_i^{(j)}(t) \triangleq x_i^{(j)}(t) & \text{for } i=1,2,\cdots,j\\ z_i^{(j)}(t) \triangleq a_{i-j}^{(j)} & \text{for } i=j+1, j+2, \cdots, 2j. \end{cases}$$

Then the system $\tilde{\sum}^{(j)}$ can be written by

$$(3.37)\quad \begin{cases} dz^{(j)}(t) = F^{(j)}[z^{(j)}(t)]z^{(j)}(t)dt + C^{(j)}dw(t)\\ \qquad\qquad\qquad\qquad z^{(j)}(0) = z_0^{(j)}\\ dy(t) = m^{(j)}z^{(j)}(t) + rdv(t), \qquad y(0) = 0 \end{cases}$$

where

(3.38a)
$$F^{(j)}[z^{(j)}(t)] = \left[\begin{array}{cccc:c} 0 & 1 & & 0 & \\ & \ddots & \ddots & & 0 \\ 0 & & 0 & 1 & \\ -z^{(j)}_{2j}(t) & \cdots & & -z^{(j)}_{j+1}(t) & \\ \hdashline & & 0 & & 0 \end{array}\right] \begin{array}{l} \left.\begin{array}{c} \\ \\ \\ \\ \end{array}\right\} j \\ \left.\begin{array}{c} \\ \end{array}\right\} j \end{array}$$
$$\qquad\qquad\qquad\qquad\qquad \underbrace{\qquad\qquad\qquad\qquad}_{j} \; \underbrace{\qquad\quad}_{j}$$

(3.38b)
$$C^{(j)} = (\underbrace{0 \ \cdots \ 0 \ g}_{j} \vdots \underbrace{0 \ \cdots \ 0}_{j})' \qquad \text{(2j-dim. vector)}$$

(3.38c)
$$m^{(j)} = (h \ 0 \ \cdots \ 0)' \qquad \text{(2j-dim. vector).}$$

For the nonlinear system $\tilde{\tilde{\Sigma}}^{(j)}$, the truncated second-order filter is given by

(3.39)
$$\left\{\begin{array}{l} d\hat{z}^{(j)}(t \mid t) = [F^{(j)}[\hat{z}^{(j)}(t \mid t)]\hat{z}^{(j)}(t \mid t) + \frac{1}{2} \sum_{i_1=1}^{2j} \sum_{i_2=1}^{2j} e^{(j)}_{i_1 i_2} \\ \qquad \times p^{(j)}_{i_1 i_2}]dt + \frac{1}{r^2} P^{(j)}(t \mid t)[m^{(j)}]'\{dy(t) - m^{(j)}\hat{z}^{(j)}(t \mid t)dt\} \\ \qquad\qquad\qquad\qquad\qquad\qquad \hat{z}^{(j)}(0 \mid 0) = \hat{z}^{(j)}_0 \\ \frac{dP^{(j)}(t \mid t)}{dt} = B^{(j)}[\hat{z}^{(j)}(t \mid t)]P^{(j)}(t \mid t) + P^{(j)}(t \mid t) \\ \qquad \times (B^{(j)}[\hat{z}^{(j)}(t \mid t)])' + C^{(j)}[C^{(j)}]' - \frac{1}{r^2} P^{(j)}(t \mid t)m^{(j)}[m^{(j)}]' \\ \qquad\qquad\qquad\qquad\qquad\qquad P^{(j)}(0 \mid 0) = P^{(j)}_0 \end{array}\right.$$

where

$$(3.40a)\qquad B^{(j)}[\hat{z}^{(j)}(t \mid t)]=\left[\begin{array}{cccc:ccc} 0 & 1 & & & & & \\ & \ddots & \ddots & 0 & & & \\ & & & & & 0 & \\ 0 & & \ddots & \ddots & & & \\ & & 0 & 1 & & & \\ b_1^{(j)} & b_2^{(j)} & \cdots & b_j^{(j)} & b_{j+1}^{(j)} & \cdots & b_{2j}^{(j)} \\ \hdashline \multicolumn{4}{c:}{\underbrace{\qquad\qquad 0 \qquad\qquad}_{j}} & \multicolumn{3}{c}{\underbrace{\qquad 0 \qquad}_{j}} \end{array}\right]\begin{array}{l}\left.\vphantom{\begin{array}{c}0\\0\\0\\0\\0\\0\end{array}}\right\}j\\ \left.\vphantom{\begin{array}{c}0\\0\end{array}}\right\}j\end{array}$$

$$(3.40b)\qquad \begin{cases} b_i^{(j)} = -z_{j+(i-1)N+1}^{(j)} & \text{for } i=1,2,\cdots,j \\ b_{j+i}^{(j)} = -z_i^{(j)} & \text{for } i=1,2,\cdots,j. \end{cases}$$

$$(3.40c)\qquad e_{i_1 i_2}^{(j)} = \begin{cases} e_{i_2 i_1}^{(j)} = (\underbrace{0 \cdots 0\ -1}_{j}\ \underbrace{0 \cdots 0}_{j})' & \text{for } i_2=j+i_1 \\ 0 & \text{for } i_2 \neq j+i_1 \end{cases}$$

Using the running estimate $\hat{z}^{(j)}(t \mid t)$ obtained by solving (3.39), the likelihood-ratio function $\Lambda^{(j)}(t)$ is determined by

$$(3.41)\qquad d\Lambda^{(j)}(t) = \frac{h}{r^2}\Lambda^{(j)}(t)\hat{z}_1^{(j)}(t \mid t)dy(t), \qquad \Lambda^{(j)}(0) = 1.$$

In simulation experiments, the simple Eular method was used. The increment of time was set as $dt=1\times 10^{-4}$; the true values of a_1 and a_2 were set as $a_1=0.5$, $a_2=0.06$; and parameters h, r, and g were h=1, r=0.5 and g=0.1 or 0.01. A set of initial values were as follows: $x_{10}=3$, $x_{20}=-0.6$; $\hat{z}_{i0}^{(j)}=-0.2$ $(i=1,2,\cdots,2j)$; $p_{11,0}^{(j)}=3$, $p_{ii,0}^{(j)}=1$ $(i=1,2,\cdots,2j)$; and $p_{ij,0}^{(j)}=0$ $(i,j=1,2,\cdots,2j;i\neq j)$.

Figure 3.3 shows values $\Lambda^{(j)}(T)$ for two different observation intervals T= 1.5 and 2.0.

(4) Determination of system order. As the order of the system, accept j for which the likelihood-ratio function $\Lambda^{(j)}(T)$ takes its maximum value, and hence determine the unknown order n to be j. Adopt, then $\hat{z}^{(j)}(t \mid t)$ as estimates of states and unknown parameters of the system model.

From Figs. 3.3(a) and 3.3(b), we can conclude that the likelihood-ratio function $\Lambda^{(2)}(T)$ takes the maximum value for both T=1.5 and 2.0, and that the hypothesis H_2 may be accepted. Sample runs of $\hat{a}_1^{(2)}(t \mid t)$ and $\hat{a}_2^{(2)}(t \mid t)$ are given in Fig. 3.4, where dashed lines show true values of a_1 and a_2. Running estimates of $\hat{x}_1^{(2)}(t \mid t)$ and $\hat{x}_2^{(2)}(t \mid t)$ are also plotted in Fig.3.5.

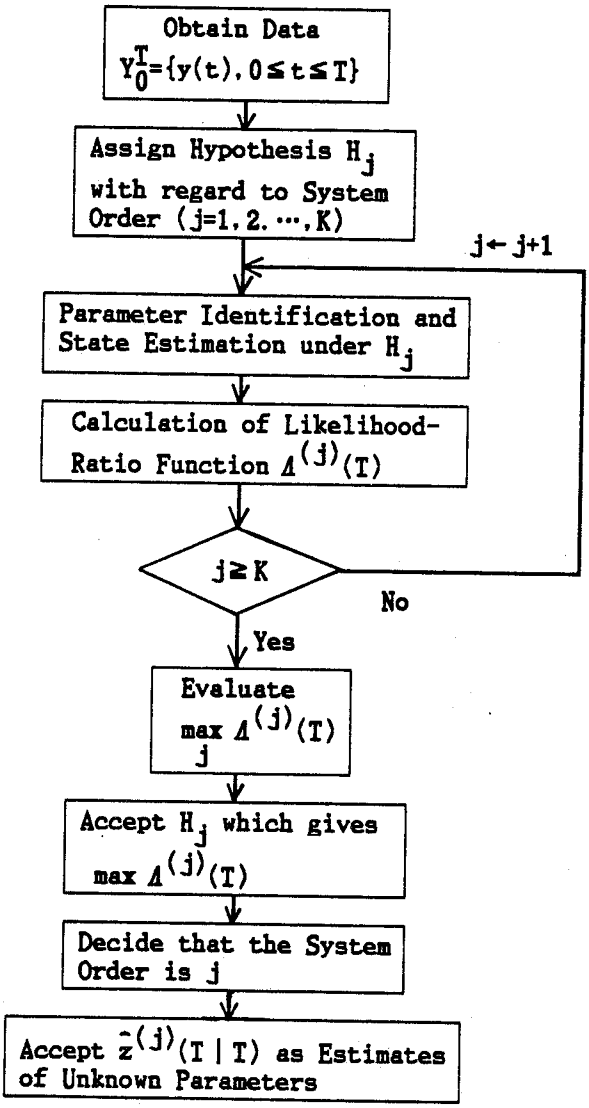


Fig. 3.1 Schematic illustration of system order determination and parameter identification

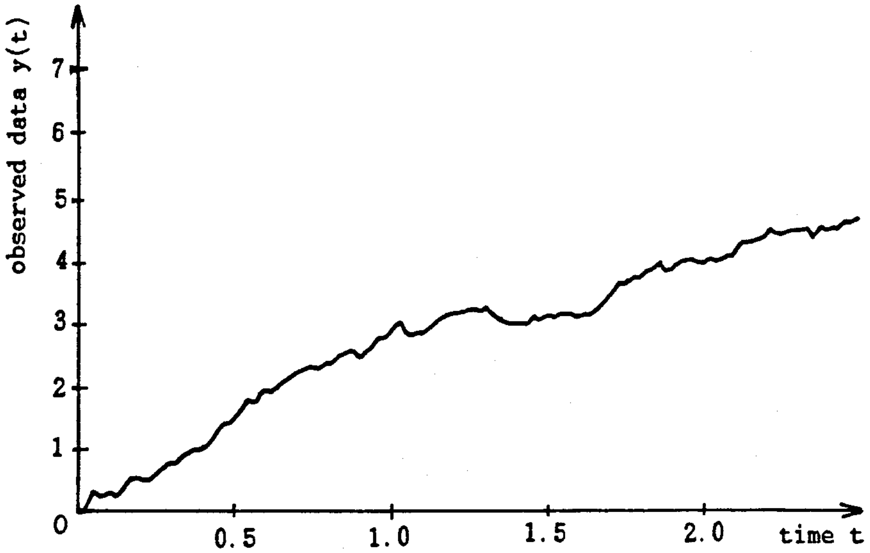
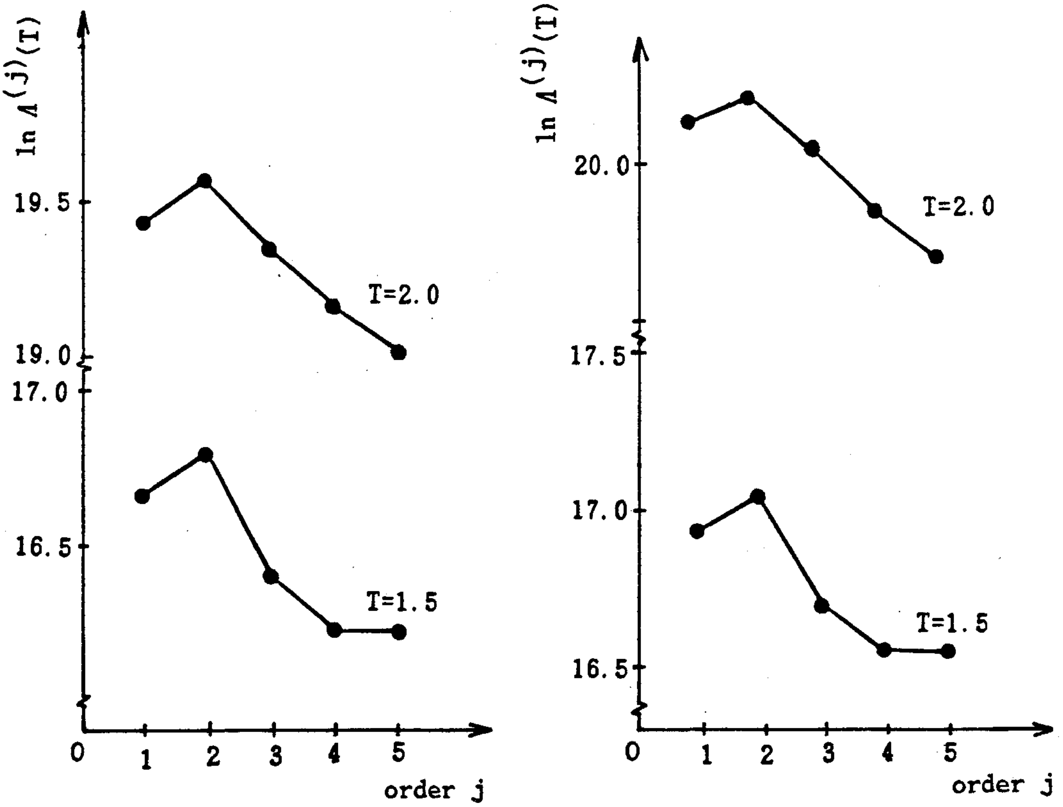


Fig. 3.2 A typical sample run of observed process y(t) (Example 3.1)



(a) System order test for g=0.1 (b) System order test for g=0.01

Fig. 3.3 System order test for two different values of system noise parameter (Example 3.1)

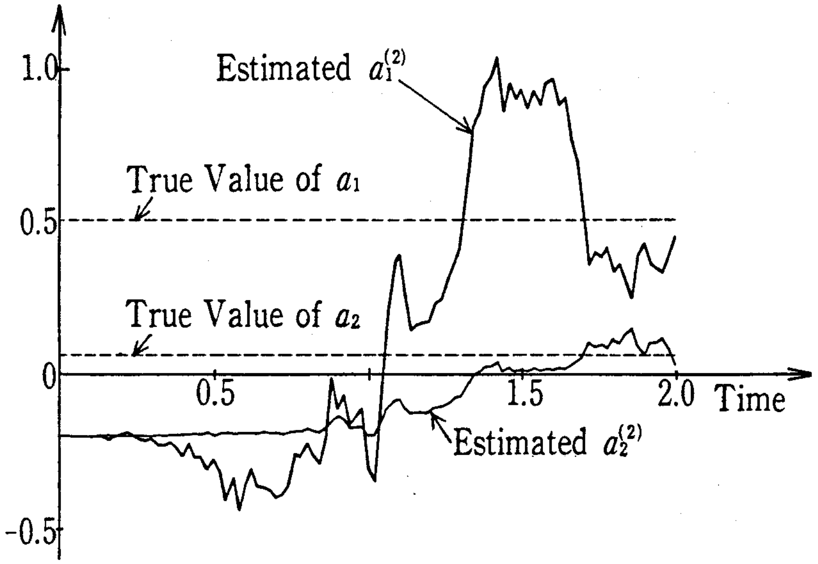


Fig. 3.4 Sample runs of estimated $a_1^{(j)}$ and $a_2^{(j)}$ (g=0.1, Example 3.1)

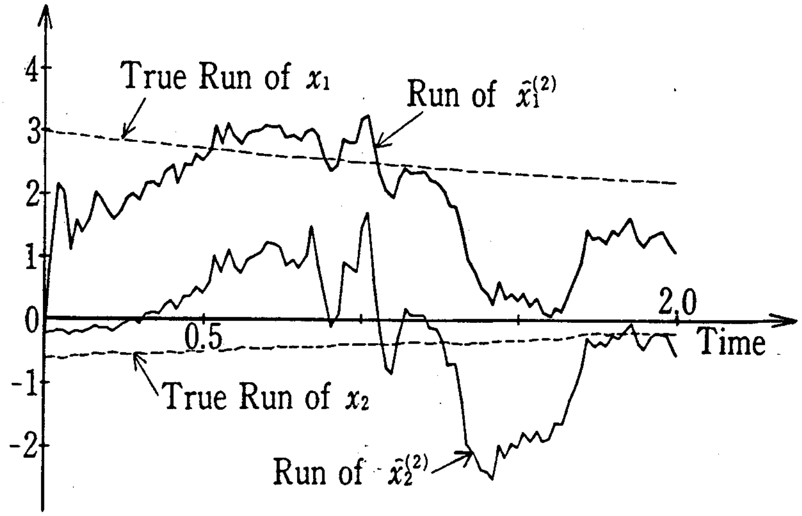


Fig. 3.5 Running estimates of $x_1^{(2)}$ and $x_2^{(2)}$ (g=0.1, Example 3.1)

Example 3.2 A slightly complex system is examined where unknown system Σ is given by

$$(3.42)\quad \Sigma \left\{ \begin{array}{l} d\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -a_2(t) & -a_1 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ g \end{bmatrix} dw(t) \\ (x_1(0) \quad x_2(0))' = (x_{10} \quad x_{20})' \\ \\ dy(t) = hx_1(t)dt + rdv(t), \qquad y(0) = 0 \end{array} \right.$$

where $a_2(t)$ is the time-varying unknown parameter given by

$$(3.43)\quad a_2(t) = a_{21} + a_{22}\, t \qquad (a_{21},\ a_{22}\text{: unknown constants})$$

whereas another unknown parameter a_1 is a constant. Then the system model $\Sigma^{(j)}$ with the hypothesis H_j is given by

$$(3.44)\quad \Sigma^{(j)} \left\{ \begin{array}{l} d\begin{bmatrix} x_1^{(j)}(t) \\ \vdots \\ x_j^{(j)}(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -a_j^{(j)}(t) & -a_{j-1}^{(j)} & \cdots & -a_1^{(j)} \end{bmatrix} dt + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dw(t) \\ (x_1^{(j)}(0) \cdots x_j^{(j)}(0))' = (x_{10}^{(j)} \cdots x_{j0}^{(j)})' \\ \\ dy(t) = hx_1^{(j)}(t)dt + rdv(t), \qquad y(0) = 0 \end{array} \right.$$

where $a_j^{(j)}(t)$ is given by

$$(3.45)\quad a_j^{(j)}(t) = a_{j1}^{(j)} + a_{j2}^{(j)}\, t$$

while other unknown parameters $\{a_1^{(j)} \cdots a_{j-1}^{(j)}\}$ are assumed to be constants. Hence by defining the state vector $z^{(j)}(t)$ as

$$(3.46) \qquad z_i^{(j)}(t) \triangleq \begin{cases} x_i^{(j)}(t) & \text{for } i=1,2,\cdots,j \\ a_{i-j}^{(j)} & \text{for } i=j+1,\cdots,2j-1 \\ a_{j1}^{(j)} + a_{j2}^{(j)}\, t & \text{for } i=2j \\ a_{j2}^{(j)} & \text{for } i=2j+1, \end{cases}$$

the system $\tilde{\Sigma}^{(j)}$ can be written by the same form as given by (3.37) except for the structure of $F^{(j)}[z^{(j)}(t)]$ given by

$$(3.47) \qquad F^{(j)}[z^{(j)}(t)] = \begin{array}{c} \left[\begin{array}{cccc|cccc} 0 & 1 & & 0 & & & & \\ & \ddots & \ddots & & & 0 & & \\ 0 & & 0 & 1 & & & & \\ -z_{2j}^{(j)}(t) & \cdots & \cdots & -z_{j+1}^{(j)}(t) & & & & \\ \hline & & & & 0 & 0 & & 0 \\ & 0 & & & & \ddots & \ddots & 0 \\ & & & & & 0 & & 1 \\ & & & & & & & 0 \end{array}\right] \\ \underbrace{\hphantom{xxxxxxxxxxxxxxxxxxxxxxxxxx}}_{j}\;\underbrace{\hphantom{xxxxxxxxxxxxxx}}_{j+1} \end{array} \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} j \\ \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} j+1 \\ \vphantom{\underbrace{x}_{j}} \end{array} .$$

The truncated second-order filter and the likelihood-ratio function are also given by same forms as (3.39) and (3.41) respectively except for

$$(3.48)\qquad B^{(j)}[\hat{z}^{(j)}(t \mid t)] = \left[\begin{array}{cccc:cccc} 0 & 1 & & 0 & & & & \\ & \ddots & \ddots & & & 0 & & \\ & 0 & 0 & 1 & & & & \\ b_1^{(j)} & b_2^{(j)} & \cdots & b_j^{(j)} & b_{j+1}^{(j)} & \cdots & b_{2j}^{(j)} & 0 \\ \hdashline & & & & 0 & 0 & & \\ & 0 & & & & \ddots & \ddots & 0 \\ & & & & & 0 & \ddots & 0 \\ \underbrace{\phantom{b_1^{(j)}\;b_2^{(j)}\;\cdots\;b_j^{(j)}}}_{j} & & & & \underbrace{\phantom{b_{j+1}^{(j)}\;\cdots\;b_{2j}^{(j)}\;0}}_{j+1} & & 0 & 1 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} j \\ \left.\vphantom{\begin{array}{c}0\\0\\0\\0\end{array}}\right\} j+1 \end{array}.$$

The same procedure as stated in Example 3.1 was used to identify the system order n and unknown parameters in the model (3.42). In simulation studies, Runge-Kutta method was used for solving the filter equation (3.39) with the time increment dt=0.001; True values of a_1, a_{21} and a_{22} were set as $a_1=1$, $a_{21}=0.4$ and $a_{22}=0.2$; the system parameter g was set as 2; initial values of x(t) and $\hat{z}^{(j)}(t \mid t)$ were set as $x(0)=(0\ 0)'$ and $\hat{z}_{i0}^{(j)}=0$ for $i=1,\cdots,2j-1$ and $\hat{z}_{i0}^{(j)}=-0.2$ for $i=2j$ and $2j+1$; the number of hypotheses was set as K=4; and other situations were set as the same in Example 3.1. A sample run of the observation process y(t) is depicted in Fig. 3.6 and Fig. 3.7 shows values of $\Lambda^{(j)}(T)$ at T=20. From Fig. 3.7, we can conclude that the likelihood-ratio function $\Lambda^{(2)}(T)$ takes the maximum value at T=20 and hence the Hypothesis H_2 may be accepted. Sample runs of $\hat{a}_1^{(2)}(t \mid t)$, $\hat{a}_{21}^{(2)}(t \mid t)$ and $\hat{a}_{22}^{(2)}(t \mid t)$ are given in Fig. 3.8.

From results of Examples 3.1 and 3.2, we may fairly say that the parameter identification is well achieved as well as the determination of the system order.

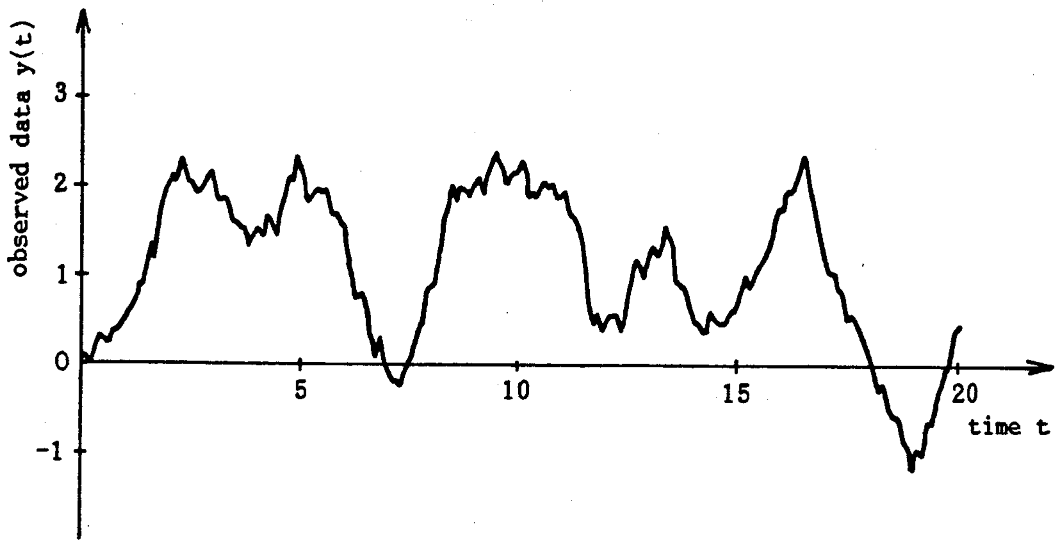


Fig. 3.6 A typical sample run of observation process (Example 3.2)

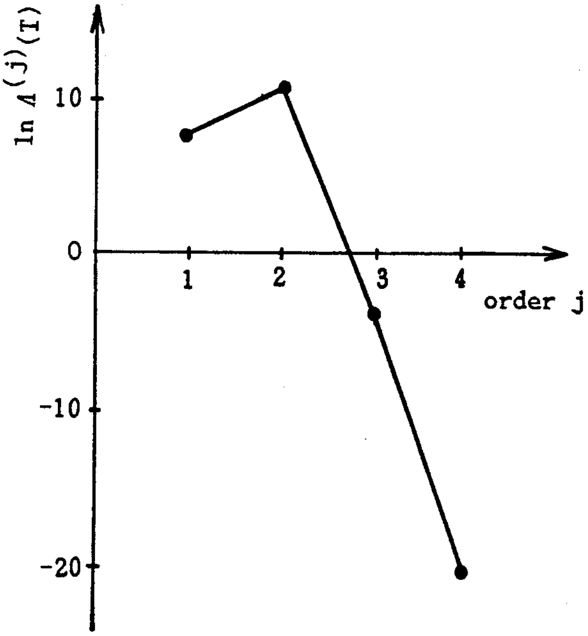


Fig. 3.7 System order test for T=20 (Example 3.2)

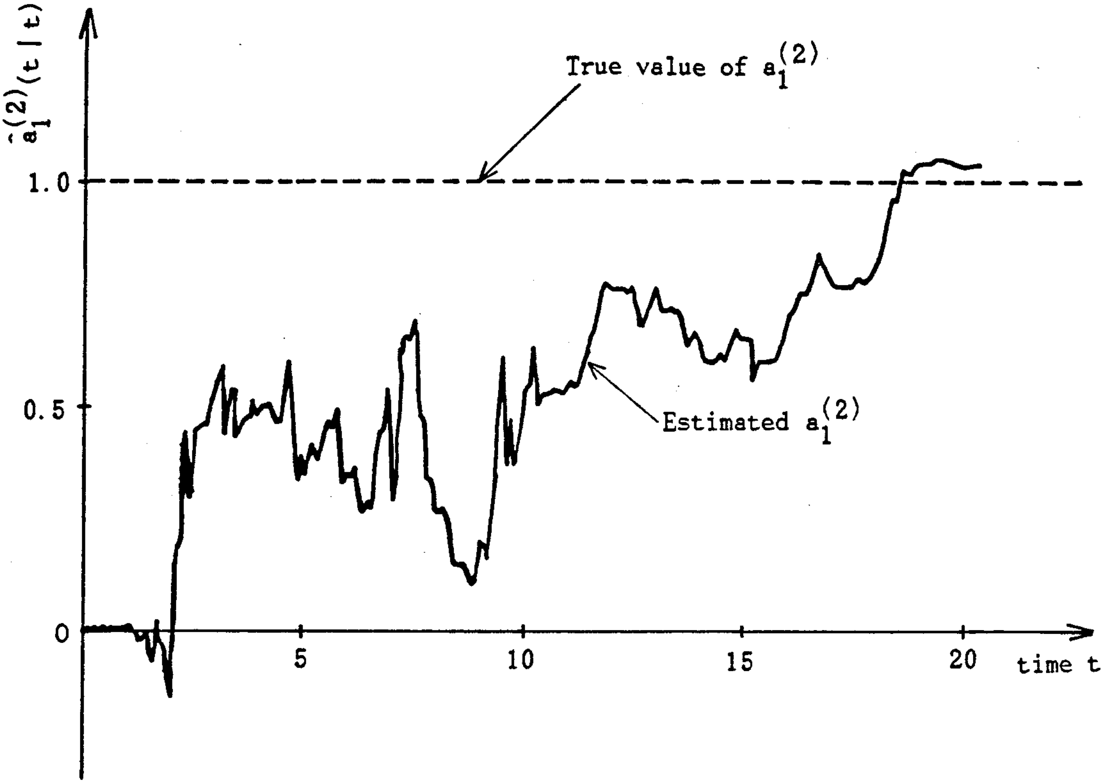


Fig. 3.8 (a) A sample run of $a_1^{(2)}$

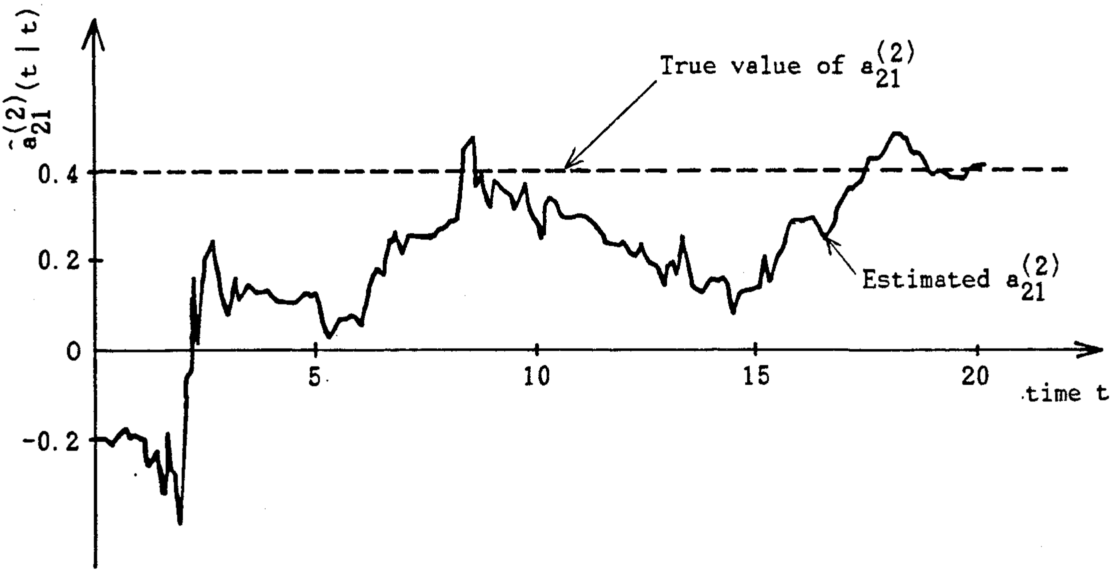
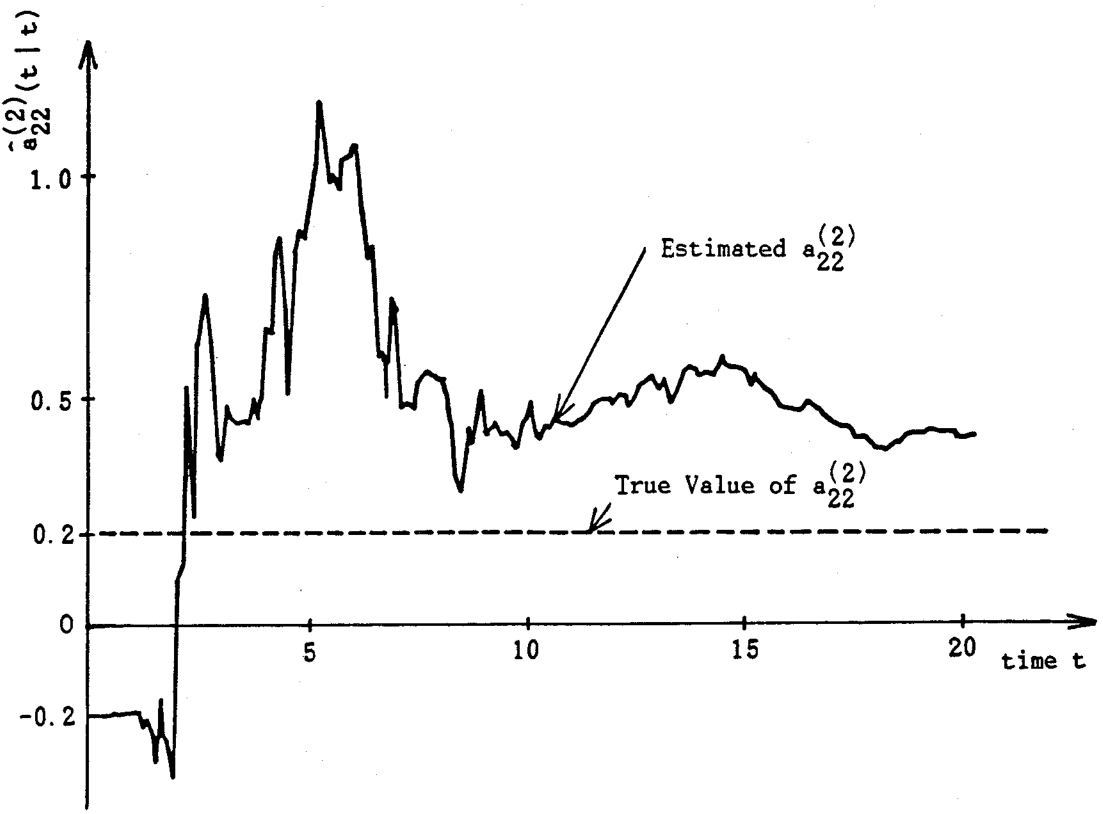


Fig. 3.8 (b) A sample run of $a_{21}^{(2)}$



(c) A sample run of $a_{22}^{(2)}$

Fig. 3.8 Sample runs of $a_{1}^{(2)}$, $a_{21}^{(2)}$ and $a_{22}^{(2)}$ (Example 3.2)

3.7 Discussions

In this chapter, a method of system modeling has been presented for a class of nonstationary linear systems. The key assumption is that unknown time-varying parameters in the system are well approximated by polynomials with respect to time, where coefficients of polynomials are unknown constants.

Regarding unknown system parameters as augmented state variables, the nonlinear filtering technique has been applied to identify unknown parameters. Based on the likelihood-ratio function derived from the multi-hypothesis testing theory, the unknown system order has been also determined.

Appendix 3.A Structural Scope of Matrices

The matrices $F^{(j)}[z^{(j)}(t)]$, $C^{(j)}(t)$ and $M^{(j)}(t)$ have the following structures:

(A.1)
$$F^{(j)}[z^{(j)}(t)] = \begin{array}{cc} \left[\begin{array}{c|c} \begin{matrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -z^{(j)}_{j+(j-1)N+1} & \cdots & & -z^{(j)}_{j+1} \end{matrix} & 0 \\ \hline 0 & \begin{matrix} \underbrace{\left.\begin{matrix} 0 & 1 & 0 \\ & \ddots & 1 \\ 0 & & 0 \end{matrix}\right\}N}_{N} & & \\ & \ddots & \\ & & N\left\{\overbrace{\begin{matrix} 0 & 1 & 0 \\ & \ddots & 1 \\ 0 & & 0 \end{matrix}}^{N}\right. \end{matrix} \end{array}\right] & \begin{matrix} \}\ j \\ \\ \}\ jN \end{matrix} \\ \begin{matrix} \underbrace{\hphantom{xxxxxxxx}}_{j} & \underbrace{\hphantom{xxxxxxxx}}_{jN} \end{matrix} & \end{array} ;$$

(A.2a)
$$C^{(j)}(t) = \begin{array}{cc} \left[\begin{array}{c} 0 \\ \hline \begin{matrix} g_p(t) & & 0 \\ & \ddots & \\ 0 & & g_1(t) \end{matrix} \\ \hline 0 \end{array}\right] & \begin{matrix} \}\ j-p \\ \}\ p \\ \}\ jN \end{matrix} \\ \underbrace{\hphantom{xxxxxxxx}}_{p} & \end{array} \qquad \text{if } j \geq p$$

$$\text{(A.2b)}\quad C^{(j)}(t) = \begin{array}{c@{}l} \left[\begin{array}{c:c} 0 & \begin{matrix} g_j(t) & & 0 \\ & \ddots & \\ 0 & & g_1(t) \end{matrix} \\ \hdashline 0 & 0 \end{array}\right] & \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} j \\ \left.\vphantom{a}\right\} jN \end{array} \\ \begin{array}{cc} \underbrace{\qquad\quad}_{p-j} & \underbrace{\qquad\qquad\qquad}_{j} \end{array} & \end{array} \quad \text{if } j<p:$$

$$\text{(A.3a)}\quad M^{(j)}(t) = \begin{array}{c@{}l} \left[\begin{array}{c:c} \begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_m(t) \end{matrix} & 0 \end{array}\right] & \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} m \\ \begin{array}{cc} \underbrace{\qquad\qquad\qquad}_{m} & \underbrace{\qquad\quad}_{j+jN-m} \end{array} & \end{array} \quad \text{if } j \geq m$$

$$\text{(A.3b)}\quad M^{(j)}(t) = \begin{array}{c@{}l} \left[\begin{array}{c:c} \begin{matrix} h_1(t) & & 0 \\ & \ddots & \\ 0 & & h_j(t) \end{matrix} & 0 \\ \hdashline 0 & 0 \end{array}\right] & \begin{array}{l} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} j \\ \left.\vphantom{a}\right\} m-j \end{array} \\ \begin{array}{cc} \underbrace{\qquad\qquad\qquad}_{j} & \underbrace{\qquad\quad}_{jN} \end{array} & \end{array} \quad \text{if } j<m.$$

Furthermore, the matrix $B^{(j)}[z^{(j)}(t)]$ in (3.31) is given by

(A.4) $B^{(j)}[z^{(j)}(t)]$

$$
= \begin{array}{cl}
\left[\begin{array}{cccc:cccc:c:cccc}
0 & 1 & & 0 & & & & & & & & & \\
 & \ddots & \ddots & & & & & 0 & & & & & \\
0 & & 0 & 1 & & & & & & & & & \\
b_1^{(j)} & b_2^{(j)} & \cdots & b_j^{(j)} & b_{j+1}^{(j)} & 0 & \cdots & 0 & \cdots & b_{2j}^{(j)} & 0 & \cdots & 0 \\ \hdashline
 & & & & 0 & 1 & & 0 & & & & & \\
 & & & & & \ddots & \ddots & & & & 0 & & \\
 & & & & 0 & & & 1 & & & & & \\
 & & & & & & & 0 & & & & & \\ \hdashline
 & & 0 & & & & & & \ddots & & & & \\ \hdashline
 & & & & & & & & & 0 & 1 & & 0 \\
 & & & & & & & & & & \ddots & \ddots & \\
 & & & & & & & & & 0 & & & 1 \\
 & & & & & & & & & & & & 0
\end{array}\right]
&
\begin{array}{l}
\left.\begin{matrix}\\ \\ \\ \\ \end{matrix}\right\} j \\
\left.\begin{matrix}\\ \\ \\ \\ \\ \\ \\ \\ \\ \end{matrix}\right\} jN
\end{array}
\\
\underbrace{\hphantom{b_1^{(j)}\ b_2^{(j)}\ \cdots\ b_j^{(j)}}}_{j}\ \underbrace{\hphantom{b_{j+1}^{(j)}\ 0\ \cdots\ 0\ \cdots\ b_{2j}^{(j)}\ 0\ \cdots\ 0}}_{jN} &
\end{array}
$$

Each diagonal block in the lower right: $N\left\{\begin{matrix}0 & 1 & & 0\\ & \ddots & \ddots & \\ 0 & & & 1\\ & & & 0\end{matrix}\right.$, $\underbrace{\hphantom{0\ 1\ \ \ 0}}_{N}$

where $b_i^{(j)}=-\hat{z}_{j+(i-1)N+1}^{(j)}(t\mid t)$ and $b_{j+1}^{(j)}=-\hat{z}_i^{(j)}(t\mid t)$ for $i=1,2,\cdots,j$.

CHAPTER 4 PARAMETER IDENTIFICATION OF STATIONARY NONLINEAR MA SYSTEMS

4.1 Introductory Remarks

In this chapter, a class of nonlinear moving average (MA) models is proposed which is expected to be a basic model when the observed data exhibits sporadic large values. Unknown parameters in the nonlinear MA model is identified by using the moment method. Furthermore, the basic nonlinear MA model is extended to the nonlinear ARMA model by adding linear AR terms to the original nonlinear MA model.

In Section 4.2, a class of nonlinear MA models is proposed whose nonlinear terms is described by a set of Hermite polynomials. The estimation procedure of unknown model parameters is discussed in Section 4.3, and asymptotic properties of estimators are investigated in Section 4.4. In Section 4.5, the nonlinear MA model given in Section 4.2 is extended to the nonlinear ARMA model, where linear AR terms are added to the original nonlinear MA model. Finally, in Section 4.6, salient features of proposed nonlinear MA, nonlinear ARMA models and estimators of unknown parameters are discussed from numerical viewpoints.

4.2 Nonlinear Moving Average (MA) Models

Let $\{y_k; k=1, 2, \cdots, N\}$ be the discretely observable scalar stochastic process, and consider the following nonlinear model described by

$$(4.1) \quad y_k = v_k + \alpha_1 H_1(v_{k-1}; \sigma^2) + \alpha_2 H_2(v_{k-1}; \sigma^2),$$

where the function $H_1(\cdot, \cdot)$ and $H_2(\cdot, \cdot)$ are Hermite polynomials in the forms of

$$(4.2a) \quad H_1(v_{k-1}; \sigma^2) = \frac{v_{k-1}}{\sigma}$$

$$(4.2b) \quad H_2(v_{k-1}; \sigma^2) = \frac{v_{k-1}^2}{\sigma^2} - 1.$$

The input sequence $\{v_k\}$ is the stochastic process satisfying the following basic condition:

(C-1) Let $\mathcal{F}_k$ (k=1, 2, ⋯, N) be the increasing σ-algebra generated from $\{v_0, v_1, \cdots, v_k\}$. Then, for each k, $\{v_k\}$ satisfies

$$(4.3) \quad E\{v_k^p \mid \mathcal{F}_{k-1}\} = \begin{cases} 0 & \text{for } p=1, 3 \text{ and } 5 \\ \sigma^2 & \text{for } p=2 \\ 3\sigma^4 & \text{for } p=4 \end{cases} \qquad \text{w.p. 1.}$$

The nonlinear model (4.1) has the following properties. First, the interest property lies in its potential to generate large values of y_k. Since as being observed from (4.2), the model (4.1) includes the nonlinear term v_{k-1}^2, its contribution to the output y_k is striking when v_{k-1} has a large value, whereas the contribution may be ignored when v_{k-1} has relatively a small value. Secondly, in view of (4.1), the proposed model can be understood as an approximated one from the more general nonlinear model of the form

$$(4.4) \quad y_k = v_k + f(v_{k-1})$$

with the expansion

$$(4.5) \quad f(v_{k-1}) = \sum_{i=1}^{m} \alpha_i H_i(v_{k-1};\sigma^2).$$

For the nonlinear model (4.1), the problem is to find consistent estimators of unknown constant parameters α_1, α_2 and σ^2 from the given observation data $Y_N \triangleq \{y_1, y_2, \cdots, y_N\}$.

4.3 Parameter Identification by Using the Moment Method

Unknown parameters α_1, α_2 and σ^2 are estimated by using the moment method. For this purpose, define the following three statistical moments:

$$(4.6)\quad \begin{cases} a \triangleq E\{y_k y_{k-1}\} \\ b \triangleq E\{y_k^2 y_{k-1}\} \\ c \triangleq E\{y_k y_{k-1}^2\}. \end{cases}$$

Then, invoking that

$$(4.7a)\quad E\{H_i(v_k;\sigma^2) \mid \mathcal{F}_{k-1}\} = 0 \qquad \text{w.p. } 1$$

$$(4.7b)\quad E\{H_i(v_k;\sigma^2)H_j(v_k;\sigma^2) \mid \mathcal{F}_{k-1}\} = i!\delta_{ij} \qquad \text{w.p. } 1$$

for i=1,2 and j=1,2, it is easily verified that statistical moments a, b and c are evaluated under the basic condition (C-1) as follows:

$$(4.8)\quad \begin{cases} a = \alpha_1\sigma \\ b = 4\alpha_1\alpha_2\sigma \\ c = 2\alpha_2\sigma^2. \end{cases}$$

Then unknown parameters in the nonlinear model (4.1), α_1, α_2 and σ^2 can be represented by functions of three statistical moments a, b and c defined in (4.6), i.e.,

$$(4.9)\quad \begin{cases} \alpha_1 = a\{(2ac)^{-1}b\}^{1/2} \\ \alpha_2 = (4a)^{-1}b \\ \sigma^2 = 2acb^{-1}. \end{cases}$$

Since the observation data $Y_N=\{y_1, \cdots, y_N\}$ is given, it may be reasonable to estimate statistical moments a, b and c by

$$(4.10) \quad \begin{cases} \hat{a}_N = \frac{1}{N} \sum_{k=2}^{N} y_k y_{k-1} \\ \hat{b}_N = \frac{1}{N} \sum_{k=2}^{N} y_k^2 y_{k-1} \\ \hat{c}_N = \frac{1}{N} \sum_{k=2}^{N} y_k y_{k-1}^2 . \end{cases}$$

The consistency of estimators of a, b and c given by (4.10) will be shown below in Theorem 4.1. Then replacing a, b and c in (4.9) by $\hat{a}_N$, $\hat{b}_N$ and $\hat{c}_N$ respectively, we obtain estimators of unknown model parameters as follows:

$$(4.11) \quad \begin{cases} \hat{\alpha}_1(N) = \hat{a}_N\{(2\hat{a}_N\hat{c}_N)^{-1}\hat{b}_N\}^{1/2} \\ \hat{\alpha}_2(N) = (4\hat{a}_N)^{-1}\hat{b}_N \\ \hat{\sigma}^2(N) = 2\hat{a}_N\hat{c}_N\hat{b}_N^{-1} . \end{cases}$$

It should be noted that we are free to choose statistical moments a, b and c. However the choice of the triplet (a b c) may be the best in the sense that α_1, α_2 and σ^2 can not be expressed uniquely by any other combinations of statistical moments. In fact, any other combinations of statistical moments such that $E\{y_k^i y_{k-1}^j\}$ limiting the argument to the case $i+j \leq 3$ $(i \geq 0, j \geq 0)$, do not give us the unique triplet of $(\alpha_1, \alpha_2, \sigma^2)$ as the function of selected statistical moments. For instance, choose three statistical moments as $a=E\{y_k^2\}$, $b=E\{y_k y_{k-1}\}$ and $c=E\{y_k^2 y_{k-1}\}$. Then, we have

$$\alpha_1 = b\sigma^{-1}$$

$$\alpha_2 = (4b)^{-1}c$$

$$\sigma^2 = \frac{(8b^2)^{-1}c^2 - a \pm \sqrt{\{(8b^2)^{-1}c^2 - a\}^2 - 4b^2}}{2} .$$

Since σ^2 has two solutions, the triplet $(\alpha_1, \alpha_2, \sigma^2)$ is not represented uniquely by the function of a, b and c, and consequently uniqueness properties of estimators $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ can not be guaranteed.

In the next section, it will also be shown that $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ are consistent estimators of α_1, α_2 and σ^2 respectively and their asymptotic probability distributions are Gaussian.

4.4 Asymptotic Properties of Estimators

4.4.1 Consistent Properties

Since the triplet of estimators $(\hat{\alpha}_1(N), \hat{\alpha}_2(N), \hat{\sigma}^2(N))$ consists of estimators $\hat{a}_N$, $\hat{b}_N$ and $\hat{c}_N$, let us examine first asymptotic properties of $\hat{a}_N$, $\hat{b}_N$ and $\hat{c}_N$. Define the two vectors,

$$(4.12) \quad \begin{cases} \mathring{\xi} \triangleq (a \quad b \quad c)' \\ \hat{\xi}_N \triangleq (\hat{a}_N \quad \hat{b}_N \quad \hat{c}_N)'. \end{cases}$$

Then, for the convergence of $\hat{\xi}_N$ to $\mathring{\xi}$, we have the following theorem.

[Theorem 4.1] Assume that the basic condition (C-1) holds, and the following conditions hold:

(C-2) The sequence $\{ | v_k |^{2p}; k=1,2,\cdots\}$ is uniformly integrable for some p $(5 \leq p < 10)$;

(C-3) $E\{ | v_k |^{2p} | \mathscr{F}_{k-1}\} < c$ (const.) w.p. 1.

Then, the estimator $\hat{\xi}_N$ is consistent in probability with the convergent rate $o(N^{5/p-1})$, i.e.,

$$(4.13) \quad \hat{\xi}_N - \mathring{\xi} = o(N^{5/p-1}) \qquad \text{in prob.} \qquad \text{as} \quad N \to \infty.$$

To prove this theorem, the following lemma is required.

[Lemma 4.1] Let $\{e_k; k=1,2,\cdots\}$ be a sequence of $\mathscr{F}_k$-measurable random variables, and let $\{e_k\}$ satisfy following conditions:

$$(C\text{-}4) \quad E\{e_k^{i(j)} | \mathscr{F}_{k-1}\} = h_j < +\infty \qquad \text{w.p. } 1$$

for $j=0,1,\cdots,m$, where $h_j=0$ for at least one j, and both m and $\{i(j); j=0,1,\cdots,m\}$ are integers;

(C-5) $E\{ | e_k |^{2p} | \mathcal{F}_{k-1}\} < c$ (const.) w.p. 1

for some p such that $i_{max} \leq p < 2i_{max}$, where $i_{max} \triangleq \max_j i(j)$ $(j=0,1,\cdots,m)$; and

(C-6) The sequence $\{ | e_k |^{2p}; k=1,2,\cdots\}$ is uniformly integrable.

Then, for the random variable defined by

$$S_N \triangleq \frac{1}{N} \sum_{k=m+1}^{N} e_k^{i(0)} e_{k-1}^{i(1)} \cdots e_{k-m}^{i(m)}, \tag{4.14}$$

it follows that

$$E\{ | S_N |^{p/i_{max}}\} = o(N^{1-p/i_{max}}) \quad \text{as} \quad N \to \infty. \tag{4.15}$$

The proof of this lemma is given in Appendix 4.A.

(Proof of Theorem 4.1) Recalling definitions of $\hat{a}_N$, $\hat{b}_N$ and $\hat{c}_N$ given by (4.10), and using (4.1) and (4.8) it can readily be seen that each of $(\hat{a}_N-a)$, $(\hat{b}_N-b)$ and $(\hat{c}_N-c)$ is expressed by linear combinations of following three quantities:

$$\begin{cases} \gamma_1(N) \triangleq \frac{1}{N} \sum_{k=2}^{N} v_k^{i(0)} v_{k-1}^{i(1)} v_{k-2}^{i(2)} \\ \gamma_2(N) \triangleq \frac{1}{N} \sum_{k=2}^{N} (v_k^2 - \sigma^2)^{k(0)} (v_{k-1}^2 - \sigma^2)^{k(1)} \\ \gamma_3(N) \triangleq \frac{1}{N} \sum_{k=2}^{N} (v_k^4 - 3\sigma^2), \end{cases} \tag{4.16}$$

where $i(0)=0,1,2,3$, $i(1)=0,1,2,\cdots,5$, $i(2)=0,1,2$ but, for some j $(j=0,1,2)$, there exists at least one $i(j)$ such that $i(j)=1,3$ or 5; and $k(0)$ and $k(1)$ take these values either $k(0)=1$, $k(1)=0$ or $k(0)=0,1,2$, $k(1)=1$.

First, consider $\gamma_1(N)$ in (4.16). From (C-1), we see that

$$(4.17)\quad E\{v_k^{i(j)} \mid \mathcal{F}_{k-1}\} = \begin{cases} 0 & \text{for } i(j) = 0,1,3,5 \\ \sigma^2 & \text{for } i(j) = 2 \\ 3\sigma^4 & \text{for } i(j) = 4, \end{cases}$$

where j=0,1,2. This implies that $\{v_k\}$ satisfies the condition (C-4) of Lemma 4.1. Since we have $i_{max}=\max_j i(j)=5$, the conditions for p, $i_{max} \leq p < 2i_{max}$ of Lemma 4.1 are also satisfied. Thus, $\{v_k\}$ satisfies automatically the conditions (C-4), (C-5) and (C-6) of Lemma 4.1. Then, we can conclude from Lemma 4.1 that

$$(4.18)\quad E\{ | \gamma_1(N) |^{p_0}\} = o(N^{1-p_0}) \qquad \text{as} \quad N \to \infty,$$

where $p_0=p/i_{max}$. It is also easily shown that $(v_k^2-\sigma^2)^{k(i)}$ (i=0,1) and $(v_k^4-3\sigma^4)$ in (4.16) satisfy all conditions of Lemma 4.1. Consequently, we can prove that the p_0-th moments of second and third quantities of (4.16) are in the order of $o(N^{1-p_0})$ as N goes to infinity.

Recalling the Markov and Hölder inequalities, the following estimate is obtained.

$$\begin{aligned}(4.19)\quad P\{ | \hat{a}_N - a | > \varepsilon\} &\leq \frac{1}{\varepsilon} E\{ | \hat{a} - a | \} \\ &= \frac{1}{\varepsilon} E\{ | \text{linear combinations of quantities in(4.16)} | \} \\ &\leq \frac{1}{\varepsilon} \sum E\{ | \text{one of quantities in (4.16)} | \} \\ &\leq \frac{1}{\varepsilon} \sum [E\{ | \text{one of quantities in (4.16)} |^{p_0}\}]^{1/p_0}.\end{aligned}$$

Since the term in the square bracket in the last ineqaulity in (4.19) are in the order $o(N^{1-p_0})$, it follows that

$$(4.20)\qquad \hat{a}_N - a = o(N^{(1-p_0)/p_0}) = o(N^{1/p_0-1}) \qquad \text{in prob.} \qquad \text{as} \quad N \rightarrow \infty.$$

Similar arguments yield that $(\hat{b}_N-b)$ and $(\hat{c}_N-c)$ are in the order $o(N^{1/p_0-1})$ in probability as $N\rightarrow\infty$. The replacement of p_0 by $p/5$ gives (4.13), which completes the proof.

Here define

$$(4.21)\qquad \begin{cases} \dot{\zeta} \triangleq (\alpha_1 \quad \alpha_2 \quad \sigma^2) \\ \hat{\zeta}_N \triangleq \ (\hat{\alpha}_1(N) \quad \hat{\alpha}_2(N) \quad \hat{\sigma}^2(N))'. \end{cases}$$

Then, the following theorem states the consistency of estimator $\hat{\zeta}_N$.

[Theorem 4.2] Assume that same conditions as those in Theorem 4.1 are satisfied, and furthermore assume that $\alpha_1\alpha_2\neq 0$ in (4.1). Then the estimator $\hat{\zeta}_N$ of the unknown parameter $\dot{\zeta}$ converges to its true value in probability with convergence rate $o(N^{5/p-1})$, i.e.,

$$(4.22)\qquad \hat{\zeta}_N - \dot{\zeta} = o(N^{5/p-1}) \qquad \text{in prob.} \qquad \text{as} \quad N \rightarrow \infty.$$

(Proof of Theorem 4.2) Note that from (4.8), $\dot{\xi}=(a\ b\ c)'$ is given by a function of the unknown parameter vector $\dot{\zeta}=(\alpha_1\ \alpha_2\ \sigma^2)'$, i.e., $\dot{\xi}=g(\dot{\zeta})$. Hence, by the mean value theorem, we have

$$(4.23)\qquad \hat{\zeta}_N - \dot{\zeta} = A_N(\ \hat{\xi} - \dot{\xi}\ ),$$

where the Jacobian matrix $A_N \triangleq [\partial g(\bar{\zeta}_N)/\partial\zeta]^{-1}$ is evaluated from (4.8) as follows:

$$(4.24)\qquad A_N = \begin{bmatrix} \bar\sigma(N) & 4\bar\alpha_2(N)\bar\sigma(N) & 0 \\ 0 & 4\bar\alpha_1(N)\bar\sigma(N) & 2\bar\sigma(N)^2 \\ \dfrac{\bar\alpha_1(N)}{2\bar\sigma(N)} & \dfrac{2\bar\alpha_1(N)\bar\alpha_2(N)}{\bar\sigma(N)} & 2\bar\alpha_2(N) \end{bmatrix}^{-1}$$

where $\bar\zeta_N \triangleq (\bar\alpha_1(N)\ \ \bar\alpha_2(N)\ \ \bar\sigma^2(N))'$ is evaluated through the relation (4.8) at $\bar\xi_N \triangleq (\bar a_N\ \ \bar b_N\ \ \bar c_N)'$ which satisfies $\|\bar\xi_N-\mathring\xi\| \le \|\hat\xi_N-\mathring\xi\|$. Since $\hat\xi_N-\mathring\xi\to 0$ in prob. as $N\to\infty$ from Theorem 4.1, we have $(\bar\xi_N-\mathring\xi)\to 0$ in prob. as $N\to\infty$. Now, note that Jacobian matrix A_N can also be evaluated in terms of $\bar\xi_N$ because of $A_N=[\partial g(\bar\zeta_N)/\partial\zeta]^{-1}=\partial g^{-1}(\bar\xi_N)/\partial\xi$ where $g^{-1}(\cdot)$ is the inverse mapping of $g(\cdot)$. Hence A_N converges in probability with the same rate of convergence as that of $(\hat\xi_N-\mathring\xi)$ to the limit

$$(4.25)\qquad A = \begin{bmatrix} \sigma & 4\alpha_2\sigma & 0 \\ 0 & 4\alpha_1\sigma & 2\sigma^2 \\ \dfrac{\alpha_1}{2\sigma} & \dfrac{2\alpha_1\ \alpha_2}{\sigma} & 2\alpha_2 \end{bmatrix}^{-1},$$

where the existence of A is guaranteed by the the assumption $\alpha_1\alpha_2\neq 0$. Therefore, in view of (4.13), we can conclude that $\hat\zeta_N-\mathring\zeta$ converges in probability to zero with the same rate as that of $\hat\xi_N-\mathring\xi$.

Let $\{e_k;k=1,2,\cdots\}$ be a sequence of independent and identically distributed (i.i.d.) random variables. Then, the following lemma is obtained which is the "almost surely" version of Lemma 4.1.

[Lemma 4.2] Let $\{e_k\}$ be a sequence of i.i.d. random variables, and let $\{e_k\}$ satisfy the following conditions:

(C-4)' $E\{e_k^{i(j)}\} = h_j < +\infty.$

for some $j=0, 1, \cdots, m$, where $h_j=0$ for at least one j, and m and $i(j)$ are integers;

(C-5)' $E\{ | e_k | ^p\} < c$ (const.)

for some p such that $i_{max} \leq p < 2i_{max}$.

Then, for S_N defined by (4.14), we have

(4.26) $S_N = o(N^{i_{max}/p-1})$ w.p. 1 as $N \to \infty$.

The proof of Lemma 4.2 is given in Appendix 4.B.

Using Lemma 4.2 instead of Lemma 4.1, we have the following two theorems which are "almost surely" versions of Theorems 4.1 and 4.2.

[Theorem 4.3] Let $\{v_k\}$ satisfy that

(C-7) $\{v_k; k=0, 1, 2, \cdots\}$ is the sequence of i.i.d. random variables where

$$E\{v_k^p\} = \begin{cases} 0 & \text{for } p=1, 3 \text{ and } 5 \\ \sigma^2 & \text{for } p=2 \\ 3\sigma^4 & \text{for } p=4 \end{cases}$$

and

$$E\{ | v_k | ^p\} < c \text{ (const.) for some } p \ (5 \leq p < 10).$$

Then,

(4.27) $\hat{\xi}_N - \mathring{\xi} = o(N^{5/p-1})$ w.p. 1 as $N \to \infty$.

[Theorem 4.4] Assume that the same conditions as those in Theorem 4.3 hold, and further assume $\alpha_1\alpha_2 \neq 0$. Then

$$(4.28) \quad \hat{\zeta}_N - \mathring{\zeta} = o(N^{5/p-1}) \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty.$$

Proofs of Theorems 4.3 and 4.4 are almost same as those of Theorems 4.1 and 4.2 and hence omitted to write here.

4.4.2 Asymptotic Normality of Estimators

In this section, we shall show the asymptotic normality of the estimator $\hat{\theta}_N$ of the unknown parameter $\mathring{\theta}$. For this purpose the asymptotic normality $\hat{\xi}_N$ will be firstly shown. Define

$$(4.29) \qquad B \triangleq \lim_{N \to \infty} N\, E\{(\hat{\xi}_N - \mathring{\xi})(\hat{\xi}_N - \mathring{\xi})'\}.$$

[Theorem 4.5] Assume that the condition (C-1) holds and let the sequence $\{v_k\}$ satisfy the following conditions:

$$(\text{C-8}) \qquad E\{v_k^{2i} \mid \mathcal{F}_{k-1}\} = E\{v_k^{2i}\} < +\infty \qquad \text{w.p. } 1.$$

for all $k=0,1,\cdots$, where i is an integer such that $3 \leq i \leq 10$;

(C-9) There exists a random variable $\tilde{v}$ such that $E\{\tilde{v}^{20}\} < +\infty$, which satisfies, for all $\lambda > 0$,

$$(4.30) \qquad P\{\,|v_k| > \lambda\} \leq P\{\,|\tilde{v}| > \lambda\}.$$

Then

$$(4.31) \qquad \sqrt{N}(\hat{\xi}_N - \mathring{\xi}) \xrightarrow{\text{law}} x \qquad \text{as} \quad N \to \infty,$$

where

$$x \sim N(0, B).$$

In order to prove Theorem 4.5, we need Lemma 4.3.

[Lemma 4.3] Let $\{e_k; k=1,2,\cdots\}$ be a sequence of $\mathcal{F}_k$-measurable random variables which satisfies (C-4) and the following conditions:

$$(\text{C-10}) \qquad E\{e_k^{\ell i(j)} \mid \mathcal{F}_{k-1}\} = E\{e_k^{\ell i(j)}\} < +\infty \qquad \text{w.p. } 1.$$

for $\ell=2,4$ and $j=0,1,\cdots,m$;

(C-11) $s^2_{jN} \to \infty$ as $N \to \infty$,

where

(4.32) $s^2_{jN} \triangleq \sum_{k=1}^{N} [\prod_{s=0}^{m} E\{(e_{k(m+1)+j-s})^{2i(s)}\}];$

and

(C-12) there exists a random variable $\tilde{e}$ such that $E\{\tilde{e}^{4i_{max}}\} < +\infty$, which satisfies, for all $\lambda_0 > 0$,

(4.33) $P\{ | e_k | > \lambda_0\} \leq P\{ | \tilde{e} | > \lambda_0\}.$

Then, for the random variable S defined by (4.14),

(4.34) $\sqrt{N}\, S_N \overset{law}{\to} S$ as $N \to \infty$,

where

$$S \sim N(0, NE\{S_N^2\}).$$

The proof of Lemma 4.3 is given in Appendix 4.C.

(Proof of Theorem 4.5) Express again each component of vector $(\hat{\xi}_N - \mathring{\xi})$ by linear combinations of quantities in (4.16). Consider first, the first quantity $\gamma_1(N)$ in (4.16). Regarding v_k as e_k in Lemma 4.3, it can be shown all conditions in Lemma 4.3 are satisfied. Indeed, conditions (C-4) and (C-10) are readily followed from conditions (C-1) and (C-8). Furthermore, recalling the elementary inequality $E\{ | y |^s\} \leq E\{ | y |^r\}^{s/r}$ for $0 < s \leq r$ where y is any random variable with $E\{ | y |^r\} < +\infty$ (see, e.g. [L1, p. 34]), we see by identifying $s=2$, $r=2i(j)$ (≥ 2) and $y=v_k$ that

(4.35) $E\{ | v_k |^{2i(j)}\} \geq E\{ | v_k |^2\}^{i(j)} = \sigma^{2i(j)} > 0.$

It is obvious that $E\{ | v_k | ^{2i(j)}\}=1$ for $i(j)=0$. Hence, in any case, it follows that

$$(4.36) \qquad E\{ | v_{3k+j-k} | ^{2i(k)}\} \geq c \text{ (const.)},$$

and it can be seen by setting $m=2$ in (4.32) that

$$(4.37) \qquad s_{jN}^2 = \sum_{k=1}^{N} [\prod_{\ell=0}^{2} E\{(v_{3k+j-\ell})^{2i(\ell)}\}] \geq \sum_{k=1}^{N} c^3 = N\, c^3 \rightarrow \infty \quad \text{as} \quad N \rightarrow \infty$$

which verifies the condition (C-11). Finally, from (C-9) in Theorem 4.5, we see that $E\{ | \tilde{v} | ^{4i_{\max}}\}=E\{ | \tilde{v} | ^{20}\}<+\infty$. Then, regarding $\tilde{v}$ as $\tilde{e}$ in Lemma 4.3, the condition (C-12) is satisfied. Hence, the first quantity in (4.16) satisfies all conditions of Lemma 4.3. For other quantities in (4.16), using a similar procedure to that for the first quantity, it can be verified that they also satisfy all conditions in Lemma 4.3. Therefore, it can be conclude that $\sqrt{N}(\hat{\xi}_N-\mathring{\xi})$ converges in law to the Gaussian random variable with zero mean and the variance B.

Invoking that $A_N \rightarrow A$ in prob. as $N \rightarrow \infty$, it is almost obvious that $\sqrt{N}(\hat{\zeta}_N-\mathring{\zeta})$ converges in law to the Gaussian variable with zero mean and the variance ABA'.

[Theorem 4.6] Assume that the same conditions as those in Theorem 4.5 hold, and furthermore assume that $\alpha_1\alpha_2 \neq 0$. Then, it follows that

$$(4.38) \qquad \sqrt{N}(\hat{\zeta}_N - \mathring{\zeta}) \xrightarrow{\text{law}} x \qquad \text{as} \quad N \rightarrow \infty,$$

where

$$x \sim N(0, ABA').$$

4.5 Extension to Nonlinear ARMA Models

It is one of the remarkable properties of the proposed nonlinear MA model that the autocorrelation between the output y_k and $y_{k+\ell}$ of the nonlinear model equals completely zero when $\ell \geq 2$. Hence, in handling the data whose autocorrelations between the time interval longer than a single time unit may not be zero, it is insufficient to use the model (4.1) directly. Two possible methods for improving the original nonlinear MA model are (i) to add some kinds of MA terms with time lags longer than that of (4.1), and (ii) to adopt the nonlinear ARMA model such as

$$(4.39) \qquad y_k + \sum_{i=1}^{m} \mathring{\theta}_i y_{k-i} = v_k + \alpha_1 H_1(v_{k-1};\sigma^2) + \alpha_2 H_2(v_{k-1};\sigma^2).$$

If the method (i) is used, we shall encounter the difficulty of finding estimators of unknown parameters uniquely and explicitly. Hence, it is appropriate to adopt (4.39) as an extended one of (4.1).

In the followings, for the model (4.39), an identification method of unknown parameters $\{\mathring{\theta}_i; i=1,2,\cdots,m\}$, α_1, α_2 and σ^2 is proposed and the consistency of estimators is investigated. As a basic assumption, we assume for simplicity of discussions that the following condition holds:

(C-1)′ The sequence $\{v_k\}$ is of an independent Gaussian random variables with zero mean and the variance σ^2.

4.5.1. Parameter Identification Using Extended Kalman Filter Approach

The moment method used for identifying unknown parameters in the nonlinear MA model is also applicable for those in the nonlinear ARMA model (4.39). However, it is afraid that a direct application of the moment method may be the cause of slow convergence rates of estimators to true values of unknown parameters. Hence, in this section, another method for the identification is proposed, where the extended Kalman Filtering approach is used.

First, in order to estimate unknown parameters in the linear AR terms in (4.39), it is assumed that the output data $\{y_k; k=1, 2, \cdots\}$ are given through

$$(4.40)\quad y_k + \sum_{i=1}^{m} \mathring{\theta}_i y_{k-i} = \varepsilon_k + \mathring{\theta}_{m+1}\varepsilon_{k-1},$$

where the sequence $\{\varepsilon_k\}$ is mutually independent of $\{v_k\}$ with the properties $E\{\varepsilon_n\}=0$ and $E\{\varepsilon_n\varepsilon_m\}=\sigma_\varepsilon^2\delta_{nm}$; and $\mathring{\theta}_{m+1}$ is the dummy parameter which is determined so that all autocorrelation functions of the output process of (4.40) coincide completely with those of (4.39) (see e.g. [D1]). In other words, (4.40) is adopted as the "model" of (4.39).

For convenience of discussions, we shall represent (4.39) and (4.40) by the state space forms. Define the state vector x_k by

$$(4.41)\quad x_k \triangleq \Big(-\sum_{i=1}^{m} \mathring{\theta}_i y_{k-i},\ -\sum_{i=2}^{m} \mathring{\theta}_i y_{k-i+1},\ \cdots,\ -\sum_{i=m-1}^{m} \mathring{\theta}_i y_{k-i+m},\ -\mathring{\theta}_m y_k\Big)'.$$

Then, the state space representation of (4.39) and (4.40) is

$$(4.42)\quad \begin{cases} x_{k+1} = A(\mathring{\theta})x_k + G(\mathring{\theta})w_k \\ y_k = Cx_k + v_k \end{cases}$$

and

$$(4.43)\quad \begin{cases} x_{k+1} = A(\mathring{\theta})x_k + K(\mathring{\theta})\varepsilon_k \\ y_k = Cx_k + \varepsilon_k \end{cases}$$

respectively, where $\mathring{\theta}$ is defined by

$$(4.44) \quad \mathring{\theta} \triangleq (\, \mathring{\theta}_1 \;\; \mathring{\theta}_2 \;\; \cdots \;\; \mathring{\theta}_m \;\; \mathring{\theta}_{m+1} \,)' ;$$

and other matrices appeared in (4.42) and (4.43) are given by

$$(4.45) \quad \left\{ \begin{array}{lll}
A(\mathring{\theta}) \triangleq \begin{bmatrix} -\mathring{\theta}_1 & 1 & & & 0 \\ -\mathring{\theta}_2 & & \ddots & & \\ \vdots & & & \ddots & \\ \vdots & & 0 & & 1 \\ -\mathring{\theta}_m & 0 & \cdots & \cdots & 0 \end{bmatrix} & & (m \times m \text{ matrix}) \\
 & & \\
G(\mathring{\theta}) \triangleq \left[\begin{array}{c:c} \alpha_1 - \mathring{\theta}_1 \sigma & \alpha_2 \\ \hdashline -\mathring{\theta}_2 \sigma & \\ \vdots & 0 \\ -\mathring{\theta}_m \sigma & \end{array} \right] & & (m \times 2 \text{ matrix}) \\
 & & \\
K(\mathring{\theta}) \triangleq (\mathring{\theta}_{m+1} - \mathring{\theta}_1 \;\; \mathring{\theta}_2 \;\; \cdots \;\; \mathring{\theta}_m)' & & (m \text{ vector}) \\
 & & \\
w_k \triangleq \begin{bmatrix} H_1(v_k ; \sigma^2) \\ H_2(v_k ; \sigma^2) \end{bmatrix} & & \\
 & & \\
C \triangleq [1 \;\; 0 \;\; \cdots \;\; 0]' & & (m \text{ vector}).
\end{array} \right.$$

Hence, by using the modified extended Kalman filtering approach proposed by Ljung [L2], the following algorithm for the identification of $\mathring{\theta}$ is obtained:

$$(4.46)\quad \begin{cases} \hat{\theta}_{k+1} = \hat{\theta}_k + L_k(y_k - C\hat{x}_k), & \hat{\theta}_0 = \bar{\theta} \\ \hat{x}_{k+1} = A(\hat{\theta}_k)\hat{x}_k + K(\hat{\theta}_k)(y_k - C\hat{x}_k), & \hat{x}_0 = x_0 \\ L_k = \hat{\sigma}_\varepsilon^{-2}(k)P_1(k)'C' & \\ P_1(k+1) = A(\hat{\theta}_k)P_1(k) + M(\hat{\theta}_k)P_2(k) - K(\hat{\theta}_k)L_k'\hat{\sigma}_\varepsilon^2(k), & P_1(0) = 0 \\ P_2(k+1) = P_2(k) - L_kL_k'\hat{\sigma}_\varepsilon^2(k) - \delta P_2^2(k), & P_2(0) = \varepsilon I \quad (\varepsilon > 0) \\ \hat{\sigma}_\varepsilon^2(k+1) = \hat{\sigma}_\varepsilon^2(k) + (k+1)^{-1}\{(y_k - C\hat{x}_k)^2 - \hat{\sigma}_\varepsilon^2(k)\}, & \hat{\sigma}_\varepsilon^2(0) = \bar{\sigma}_\varepsilon^2, \end{cases}$$

where $\hat{\theta}_k$ and $\hat{\sigma}_\varepsilon^2(k)$ are respectively estimators of $\dot{\theta}$ and σ_ε^2; $P_1(k)$ is an $m\times(m+1)$ matrix and $P_2(k)$ is an $(m+1)\times(m+1)$ positive definite matrix; δ is an arbitrary small positive number; and $M(\hat{\theta}_k)$ is the $m\times(m+1)$ matrix given by

$$(4.47)\quad M(\hat{\theta}_k) \triangleq \left[\begin{array}{ccc:c} y_k & & 0 & y_k - C\hat{x}_k \\ & \ddots & & 0 \\ 0 & & y_k & \end{array}\right].$$

In (4.46), the value of $\hat{\theta}_{m+1,k+1}$ of the estimator $\dot{\theta}_{m+1}$ is set at an arbitrary point in $|\hat{\theta}_{m+1,k+1}| < 1$ if $|\hat{\theta}_{m+1,k+1}| \geq 1$.

In order to estimate α_1, α_2 and σ^2 in the nonlinear MA terms, define that

$$(4.48)\quad \tilde{y}_k \triangleq y_k + \dot{\theta}_1 y_{k-1} + \cdots + \dot{\theta}_m y_{k-m}.$$

Then, $\tilde{y}_k$ can be represented from (4.39) by

$$(4.49)\quad \tilde{y}_k = v_k + \alpha_1 H_1(v_{k-1};\sigma^2) + \alpha_2 H_2(v_{k-1};\sigma^2)$$

which may be regarded as the nonlinear MA model in Section 4.2. Hence using the same procedure as that in Section 4.3, it follows that

$$(4.50)\quad \begin{cases} \alpha_1 = \eta_1\{(2\eta_1\eta_3)^{-1}\eta_2\}^{1/2} \\ \alpha_2 = (4\eta_1)^{-1}\eta_2 \\ \sigma^2 = 2\eta_1\eta_3\eta_2^{-1} \end{cases}$$

where

$$(4.51)\quad \begin{cases} \eta_1 \triangleq E\{\tilde{y}_k\tilde{y}_{k-1}\} = \alpha_1\sigma \\ \eta_2 \triangleq E\{\tilde{y}_k^2\ \tilde{y}_{k-1}\} = 4\alpha_1\alpha_2\sigma \\ \eta_3 \triangleq E\{\tilde{y}_k\tilde{y}_{k-1}^2\} = 2\alpha_2\sigma^2. \end{cases}$$

Since only the observation data $Y_N=\{y_1, \cdots, y_N\}$ and the estimator $\hat{\theta}_N$ and $\hat{\sigma}_\varepsilon^2(N)$ at the final step N are given, the three statistical moments η_1, η_2 and η_3 are respectively estimated by

$$(4.52)\quad \begin{cases} \hat{\eta}_1(N) = \hat{\theta}_{m+1,N}\hat{\sigma}_\varepsilon^2(N) \\ \hat{\eta}_2(N) = \frac{1}{N}\sum\limits_{k=2}^{N}\tilde{y}_k^2(\hat{\theta}_N)\tilde{y}_{k-1}(\hat{\theta}_N) \\ \hat{\eta}_3(N) = \frac{1}{N}\sum\limits_{k=2}^{N}\tilde{y}_k(\hat{\theta}_N)\tilde{y}_{k-1}^2(\hat{\theta}_N) \end{cases}$$

where

$$(4.53)\quad \tilde{y}_k(\hat{\theta}_N) = y_k + \hat{\theta}_{1,N}y_{k-1} + \cdots + \hat{\theta}_{m,N}y_{k-m}$$

and in deriving the estimator $\hat{\eta}_1(N)$, an easily verified relation $\dot{\theta}_{m+1}\sigma_{\varepsilon}^2=\alpha_1\sigma$ has been used. Then, replacing η_1, η_2 and η_3 by $\hat{\eta}_1(N)$, $\hat{\eta}_2(N)$ and $\hat{\eta}_3(N)$ respectively, we can obtain estimators of α_1, α_2 and σ^2 as follows:

$$(4.54) \quad \begin{cases} \hat{\alpha}_1(N) = \hat{\eta}_1(N)\{(2\hat{\eta}_1(N)\hat{\eta}_3(N))^{-1}\hat{\eta}_2(N)\}^{1/2} \\ \hat{\alpha}_2(N) = (4\eta_1(N))^{-1}\eta_2(N) \\ \hat{\sigma}^2(N) = 2\hat{\eta}_1(N)\hat{\eta}_3(N)\hat{\eta}_2^{-1}(N). \end{cases}$$

As observed in (4.46) and (4.54), unknown parameters in linear AR terms are estimated recursively while those in nonlinear MA terms are not estimated recursively. However, replacing $\hat{\theta}_N$ in $\hat{\eta}_i(N)$ (i=1, 2, 3) by the current estimate $\hat{\theta}_k$, α_1, α_2 and σ^2 may be estimated recursively.

4.5.2 Asymptotic Properties of Estimators

[Theorem 4.7] Assume that the condition (C-1)' and the following conditions hold:

(C-13) The polynomial

$$(4.55) \quad G(q^{-1};\mathring{\theta}) \triangleq 1+ \mathring{\theta}_1 q^{-1} + \cdots + \mathring{\theta}_m q^{-m}$$

has all its zeros inside the unit circle;

$$(4.56) \quad D(q^{-1};\mathring{\theta}) \triangleq 1 + \mathring{\theta}_{m+1} q^{-1}$$

has its zeros inside the unit circle, i.e., $|\mathring{\theta}_{m+1}| < 1$ where q^{-1} is the backward shift operator; and

(C-14) $G(q^{-1};\mathring{\theta})$ and $D(q^{-1};\mathring{\theta})$ do not have common factors and $\mathring{\theta}_m$ and $\mathring{\theta}_{m+1}$ are not zero.

Then

$$(4.57) \quad \hat{\theta}_N \rightarrow \mathring{\theta} \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty$$

$$(4.58) \quad (\hat{\alpha}_1(N) \ \ \hat{\alpha}_2(N) \ \ \hat{\sigma}^2(N))' \rightarrow (\alpha_1 \ \ \alpha_2 \ \ \sigma^2)' \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty.$$

In order to prove Theorem 4.7, the following two lemmas are required.

[Lemma 4.4] (Ljung [L2]) Consider that the observation data is generated by (4.42) (or (4.39)) and let the model be given by (4.43) (or (4.40)). Suppose that

(C-15) the algorithm given by (4.46) is complemented with a projection facility to keep $\hat{\theta}_k$ in a compact subset of

$$(4.59) \quad D_s \triangleq \{ \theta \mid \| (A(\theta) - K(\theta)C)^k \| \leq c\lambda^k \ \text{ for all } k=1,2,\cdots \}$$

where c is a positive constant and λ is also a positive constant but $\lambda < 1$.

Then, the estimator $(\hat{\theta}_N \ \hat{\sigma}_\varepsilon^2(N))$ converges with probability one to a stationary point of the function

$$(4.60) \quad V(\theta, \sigma_\varepsilon^2) = E\{\varepsilon_k^2(\theta)\}\sigma_\varepsilon^{-2} + \log \sigma_\varepsilon^2$$

where

$$(4.61) \quad \begin{cases} \varepsilon_k(\theta) = y_k - Cx_k(\theta) \\ x_{k+1}(\theta) = A(\theta)x_k + K(\theta)\varepsilon_k(\theta) \end{cases}$$

for a fixed point $\theta \in D_s$.

[Lemma 4.5] (Åström and Söderström [A6]) Let δ_k be a sequence defined by the operator form

$$(4.62) \quad D(q^{-1};\theta)G(q^{-1};\mathring{\theta})\delta_k = G(q^{-1};\theta)D(q^{-1};\mathring{\theta})\varepsilon_k.$$

Assume that conditions (C-13) and (C-14) hold. Then, the evaluation

$$(4.63) \quad E\{\delta_k^2\} \geq \sigma_\varepsilon^2$$

is obtained, where the inequality holds if and only if $\theta_i = \mathring{\theta}_i$ for all $i=1, 2, \cdots, m+1$.

(Proof of Theorem 4.7) From definitions of $A(\theta)$, $K(\theta)$ and C in (4.45), we have

$$(4.64) \quad A(\hat{\theta}_N) - K(\hat{\theta}_N)C = \left[\begin{array}{c:cccc} -\hat{\theta}_{m+1}(N) & 1 & & & 0 \\ 0 & 0 & \ddots & & \\ \vdots & & \ddots & \ddots & 1 \\ \vdots & & 0 & & \\ 0 & & & & 0 \end{array}\right]$$

and hence eigenvalues of $A(\hat{\theta}_N)-K(\hat{\theta}_N)C$ are zeros and $-\hat{\theta}_{m+1}(N)$. Since $|\hat{\theta}_{m+1}(N)|<1$, the absolute values of all eigenvalues are less than 1 (see e.g. [K4]). Then, it is concluded from Lemma 4.4 that $(\hat{\theta}_N\ \ \hat{\sigma}_\varepsilon^2(N))$ converges with probability one to a stationary point of $V(\theta,\sigma_\varepsilon^2)$ defined by (4.60). Furthermore, it is easily verified from (4.40), (4.43) and (4.61) that

$$(4.65)\qquad \varepsilon_k(\theta) = \frac{G(q^{-1};\theta)D(q^{-1};\mathring{\theta})}{G(q^{-1};\mathring{\theta})D(q^{-1};\theta)}\,\varepsilon_k$$

and from (4.39) and (4.40) that

$$(4.66)\qquad \begin{cases} \sigma^2 + \alpha_1^2 + 2\alpha_2^2 = \sigma_\varepsilon^2(1 + \mathring{\theta}_{m+1}^2) \\ \alpha_1\sigma = \mathring{\theta}_{m+1}\sigma_\varepsilon^2. \end{cases}$$

Since it is easily verified that $(\sigma^2+\alpha_1^2+2\alpha_2^2)^2>4\alpha_1^2\sigma^2$, we can always choose $\mathring{\theta}_{m+1}$ such that $|\mathring{\theta}_{m+1}|<1$, and, hence, all conditions of Lemma 4.5 are satisfied by identifying $\varepsilon_k(\theta)$ as δ_k. Then, it is concluded from (4.65) that $E\{\varepsilon_k^2(\theta)\}$ has only one stationary point at $\theta=\mathring{\theta}$ and $V(\theta,\sigma_\varepsilon^2)$ has also only one statinoary point at $\theta=\mathring{\theta}$ and $\sigma_\varepsilon^2=E\{\sigma_\varepsilon^2(\theta)\}$. Then, we have

$$(4.67)\qquad \begin{cases} \hat{\theta}_N \to \mathring{\theta} & \text{w.p. } 1 \quad \text{as } N \to \infty \\ \hat{\sigma}_\varepsilon^2(N) \to \sigma_\varepsilon^2 & \text{w.p. } 1 \quad \text{as } N \to \infty. \end{cases}$$

For the consistency of estimators $\hat{\eta}_i(N)$ $(i=1,2,3)$, it is easily shown by using the ergodic property of y_k and (4.67) that

$$(4.68)\qquad \begin{cases} \hat{\eta}_1(N) \to \eta_1 = \alpha_1\sigma & \text{w.p. } 1 \quad \text{as } N \to \infty \\ \hat{\eta}_2(N) \to \eta_2 = 4\alpha_1\alpha_2\sigma & \text{w.p. } 1 \quad \text{as } N \to \infty \end{cases}$$

$$\left. \hat{\eta}_3(N) \rightarrow \eta_3 = 2\alpha_2\sigma^2 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty. \right.$$

Therefore, using the same procedure as that in Theorem 4.2, we have

$$(4.69) \qquad \begin{cases} \hat{\alpha}_1(N) \rightarrow \alpha_1 & \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty \\ \hat{\alpha}_2(N) \rightarrow \alpha_2 & \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty \\ \hat{\sigma}^2(N) \rightarrow \sigma^2 & \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty. \end{cases}$$

4.6 Digital Simulation Studies

4.6.1 Comparison of Nonlinear MA Models with Linear MA Models

In this section, the features of the proposed nonlinear MA model is examined from the numerical point of view. Figure 4.1(a) depicts a typical sample run of the input sequence $\{v_k; k=1,2,\cdots\}$ where $\{v_k\}$ is independent standard Gaussian random variables generated by a digital computer. The simulated record of $\{y_k\}$ is depicted in Fig. 4.1(b), which comes through the input/output relation (4.1) with the use of $\{v_k\}$, where the parameter values are set as $\alpha_1=0.5$ and $\alpha_2=1.0$, i.e.,

$$(4.70) \qquad y_k = v_k + 0.5v_{k-1} + (v_{k-1}^2 - 1).$$

For convenience of discussion, a sample path of the linear MA model corresponding to (4.70), i.e.,

$$(4.71) \qquad y_k^* = v_k + 0.5v_{k-1}$$

is shown in Fig. 4.1(c). Comparing y_k with y_k^*, we know that the nonlinear term in (4.70) is important in producing sporadically large values of y_k.

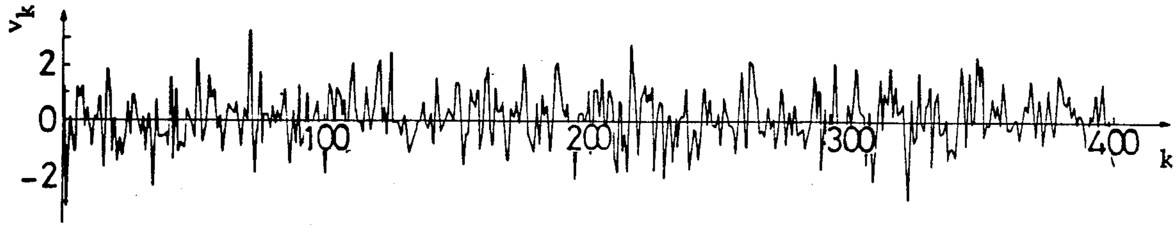
Next, we shall try to fit the linear MA model of order ℓ, i.e.,

$$(4.72) \qquad y_k = v_k + \sum_{i=1}^{\ell} \beta_i v_{k-i}$$

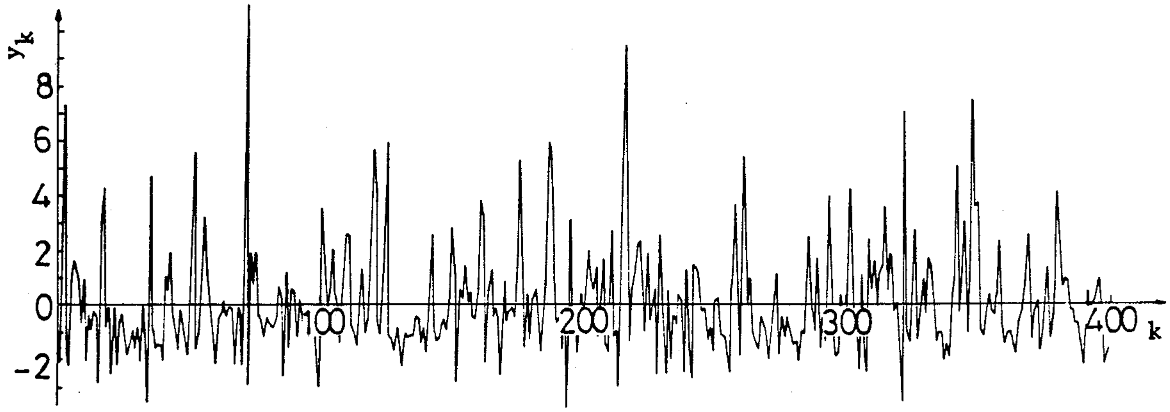
to the data obtained by (4.70). Since it is hard to determine the reasonable order ℓ a priori, we used AIC proposed by Akaike [A2] and its result was shown in Fig. 4.2. The determined order $\hat{\ell}$ based on 2000 observation data is 4 and estimates of unknown parameters of (4.72) with $\ell=4$ were given by $\hat{\beta}_1=0.2$, $\hat{\beta}_2=-0.19$, $\hat{\beta}_3=-0.08$, $\hat{\beta}_4=0.2$ and $\hat{\sigma}^2=3.54$, where the Dabidon's variance algorithm was used [A4]. Figure 4.3 shows a simulated run of (4.72) by using estimated order and parameters, i.e.,

$$(4.73)\qquad y_k = v_k + \sum_{i=1}^{4} \hat{\beta}_i v_{k-i}$$

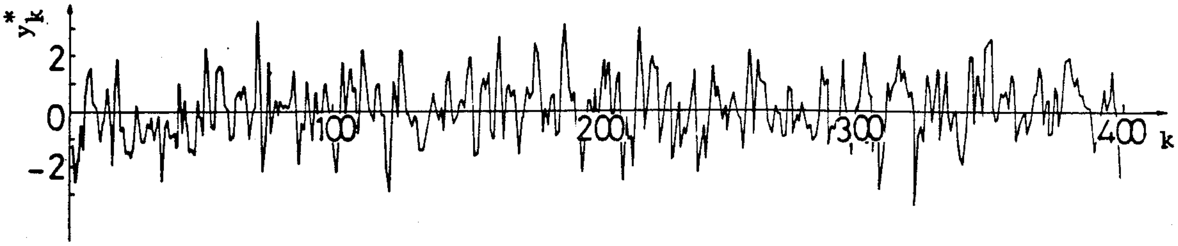
where the input $\{v_k\}$ is the same sequence as that depicted in Fig. 4.1(a). It may be numerically concluded from Fig. 4.3 that linear models are not adequate for the data whose remarkable feature is its sporadically peak values, and, hence, it may be fair to say that the proposed nonlinear MA model will be useful to model such random data as mentioned above.



(a) A sample run of v_k process



(b) A sample run of y_k process



(c) A sample run of y_k^* process

Fig 4.1 Sample runs of v_k, y_k and y_k^* processes

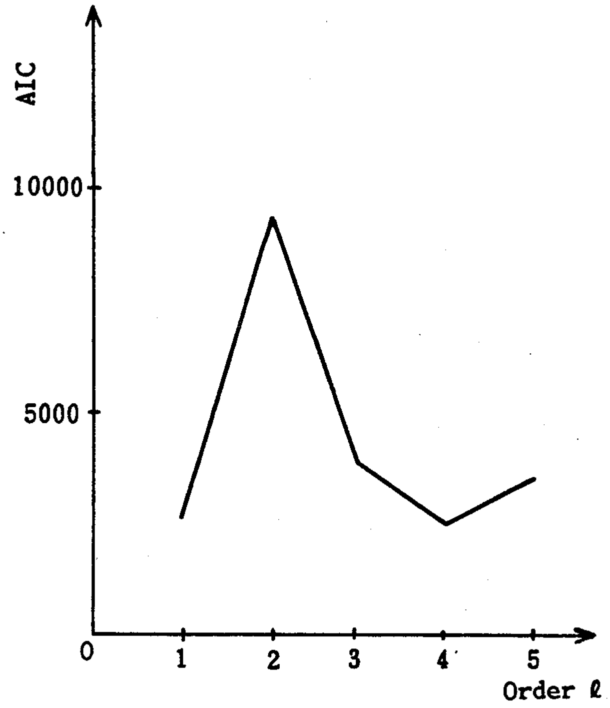


Fig. 4.2 Values of AIC at N=2000

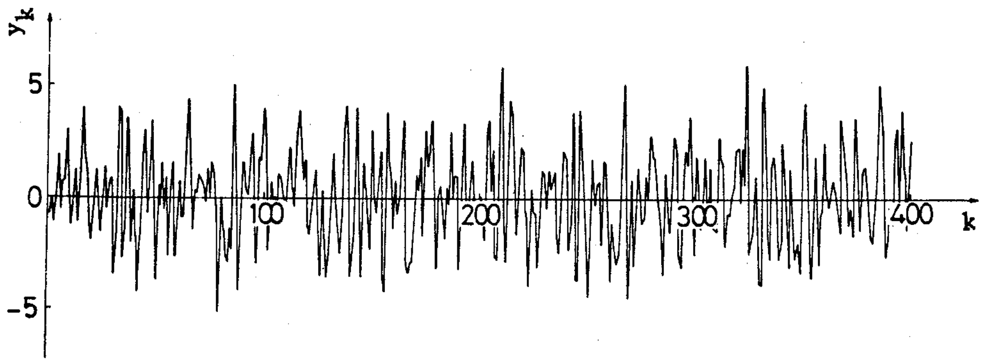


Fig. 4.3 A typical sample run of the misfitted linear MA model whose order is determined as 4 via the AIC approach

4.6.2 Parameter Identification of Nonlinear MA model

Several simulation studies were given in order to investigate estimators of unknown parameters in the nonlinear MA model. The convergence feature of estimators $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ given by (4.11) is observed by sample runs shown in Fig. 4.4, where the input sequence $\{v_k\}$ is the same as that depicted in Fig. 4.1(a). To examine detailed aspects of the convergence rate of estimators, consider the following measure:

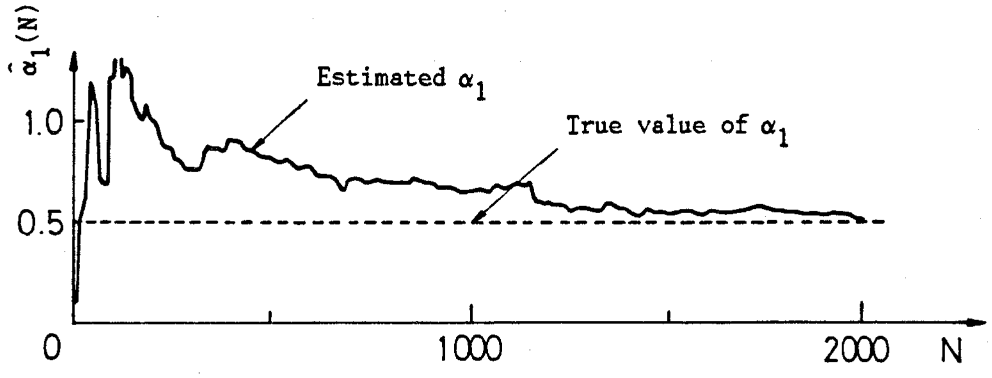
$$(4.74) \qquad C_L(N,\hat{z}) \triangleq \frac{1}{L} \sum_{\ell=1}^{L} (\hat{z}^{(\ell)}(N) - z)^2$$

where $z=\alpha_1$, α_2 or σ^2 and $\hat{z}(N)=\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ or $\hat{\sigma}^2(N)$; the superposition (ℓ) denotes the ℓ-th sample run of $\hat{z}(N)$; and L is the number of sample runs to be averaged. Figure 4.5 illustrates a result of Monte Carlo trials for sample runs of $\{y_k\}$'s that are simulated by

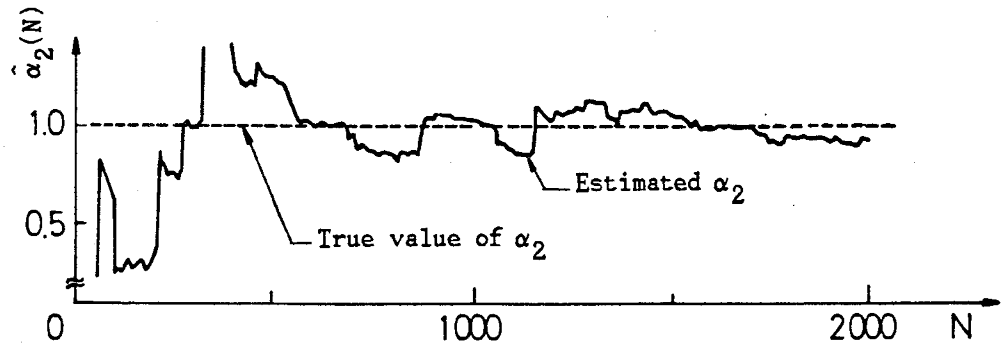
$$(4.75) \qquad y_k^{(\ell)} = v_k^{(\ell)} + 0.5v_{k-1}^{(\ell)} + \{(v_{k-1}^{(\ell)})^2 - 1\} \qquad \ell = 1, 2, \cdots, 100,$$

where $v_k^{(\ell)}$ is the ℓ-th sample process of input disturbances. Convergences of estimators of α_1, α_2 and σ^2 are well achieved as the observation data N increases as shown in Fig. 4.5.

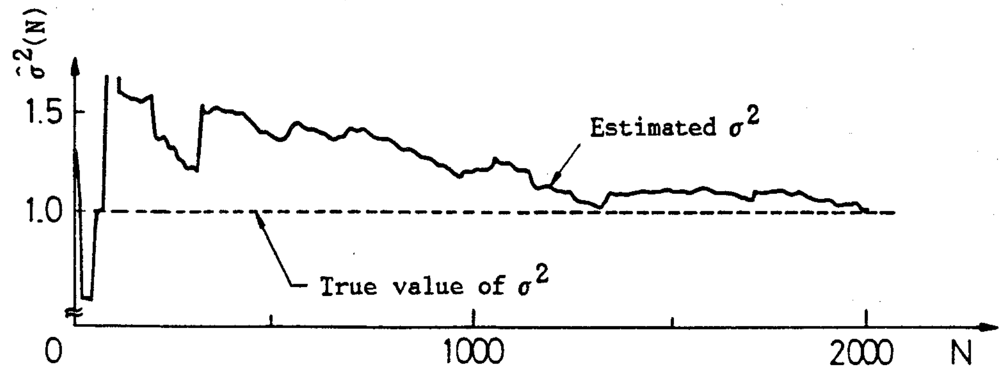
Histograms of the 100 sample runs of estimators $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ at N=2000 are shown in Fig. 4.6, where fitted normal curves are also plotted. Using histograms shown in Fig. 4.6, cumulative frequency curves are plotted on normal-probability papers, and they are shown in Fig. 4.7. Furthermore, we make use of Chi-square test with 5 per cent level of significance in order to check more precisely the normality of sampling distributions of estimators. Results of Chi-square test is shown in Table 4.1. It may be fair to say from Figs. 4.6, 4.7 and Table 4.1 that distributions of estimators $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ at N=2000 are approximately normal.



(a) A sample run of $\hat{\alpha}_1(N)$

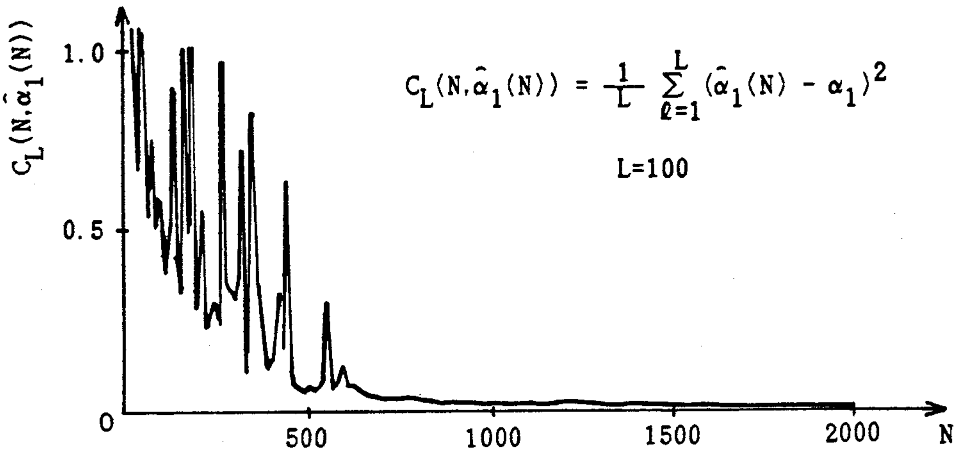


(b) A sample run of $\hat{\alpha}_2(N)$

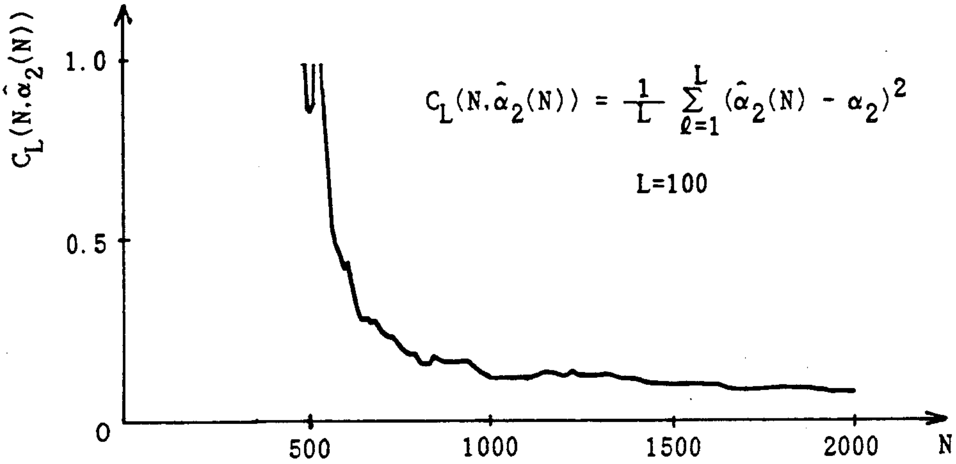


(c) A sample run of $\hat{\sigma}^2(N)$

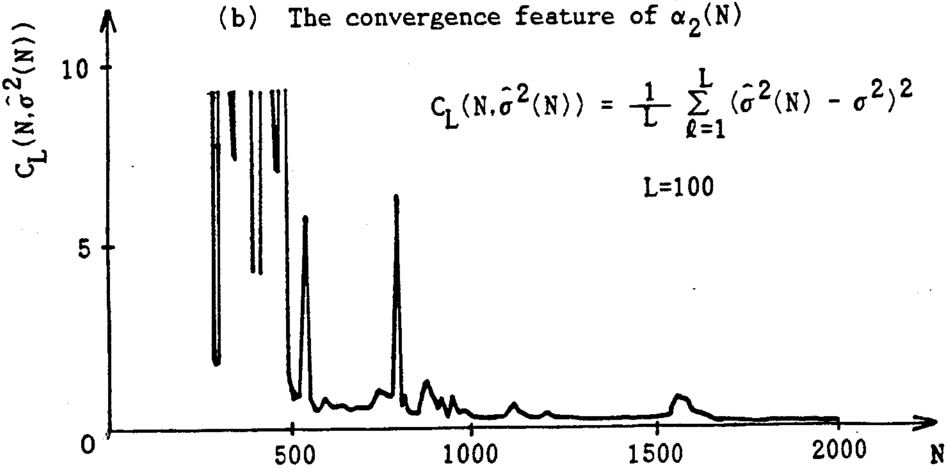
Fig. 4.4 Convergence features of estimated α_1, α_2 and σ^2



(a) The convergence feature of of $\hat{\alpha}_1(N)$

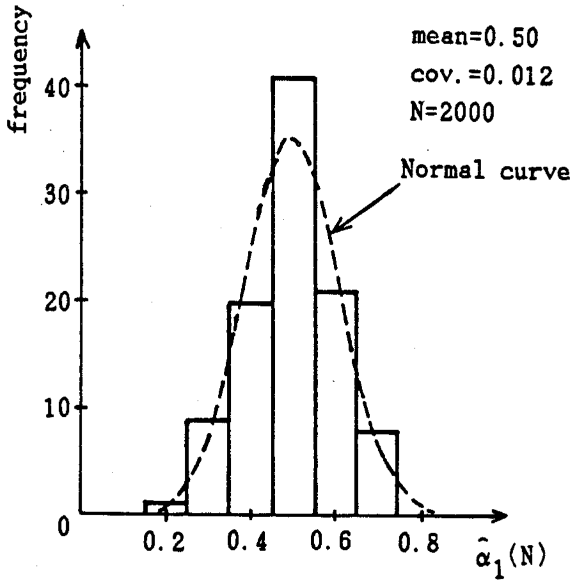


(b) The convergence feature of $\hat{\alpha}_2(N)$

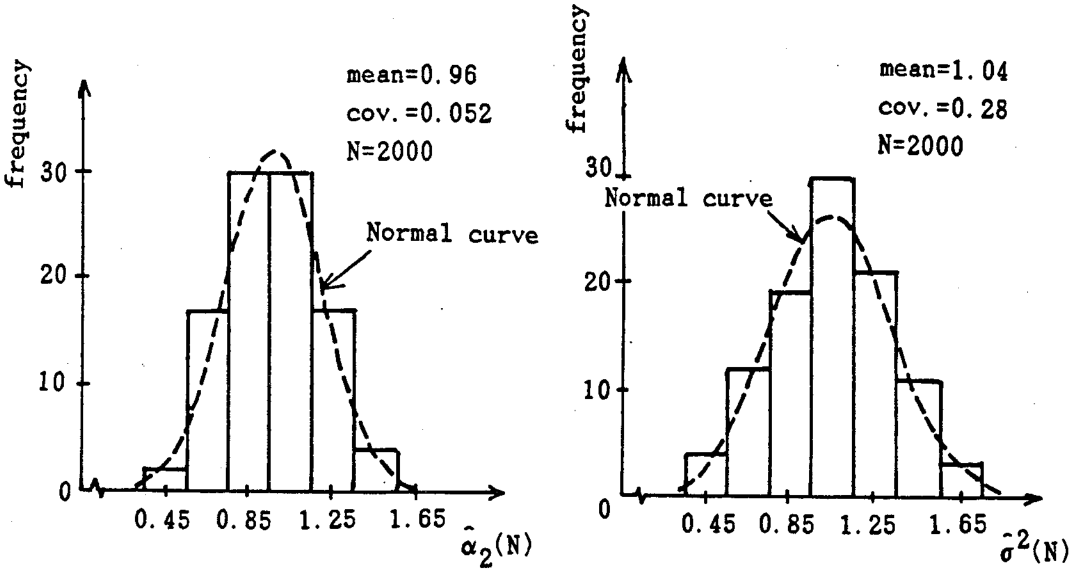


(c) The convergence feature of $\hat{\sigma}^2(N)$

Fig. 4.5 Convergence features of $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$



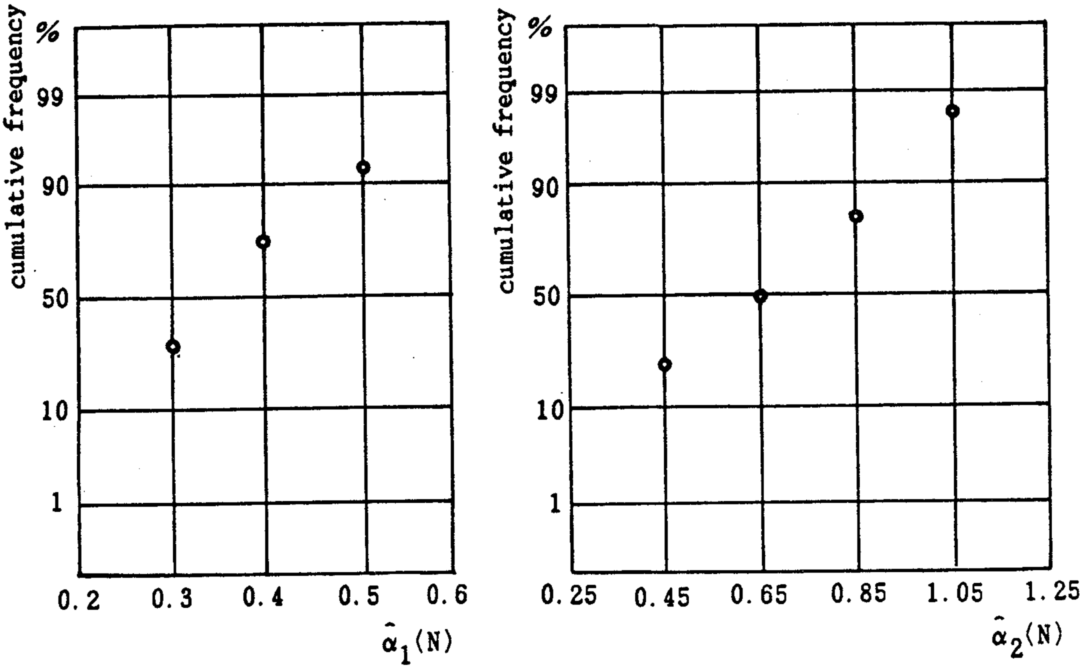
(a) Histogram of $\hat{\alpha}_1(N)$



(b) Histogram of $\hat{\alpha}_2(N)$

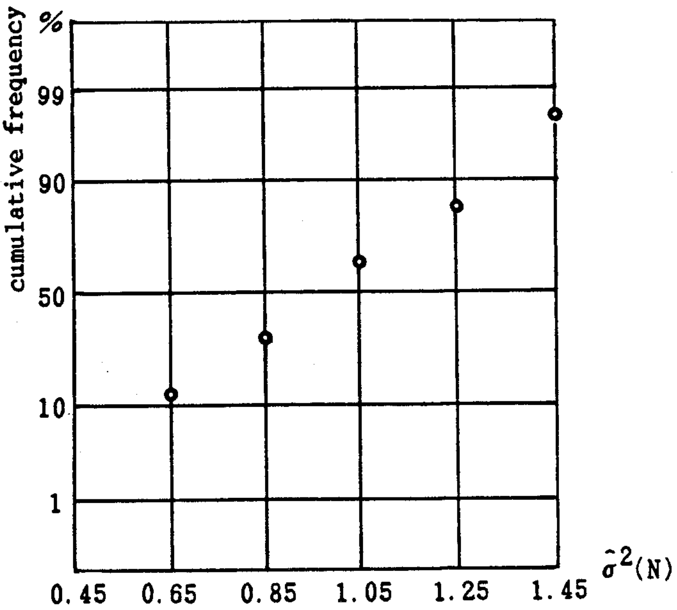
(c) Histogram of $\hat{\sigma}^2(N)$

Fig. 4.6 Histograms of 100 sample runs of estimates $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ at N=2000, where fitted normal curves are also depicted



(a) Cumulative frequency curve of $\hat{\alpha}_1(N)$

(b) Cumulative frequency curve of $\hat{\alpha}_2(N)$



(c) Cumulative frequency curve of $\hat{\sigma}^2(N)$

Fig. 4.7 Cumulative frequency curves of $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ at N=2000

Table 4.1 Chi-square tests for the check of normality (N=2000)

(a) Chi-square test for $\hat{\alpha}_1(N)$

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above 0.65	8	8.08	0.0008
0.55~0.65	21	22.77	0.1376
0.45~0.55	41	34.69	1.1478
0.35~0.45	20	25.61	1.2289
Below 0.35	10	8.85	0.1494
Total	100	100.00	2.6645
$\chi^2_{0.95}(2) = 5.99 > 2.6645$			

(b) Chi-square test for $\hat{\alpha}_2(N)$

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above 1.35	4	4.46	0.0474
1.15~1.35	17	16.73	0.0044
0.95~1.15	30	26.82	0.3770
0.75~0.95	30	33.58	0.3816
0.55~0.75	17	14.82	0.3206
Below 0.55	2	3.59	0.7042
Total	100	100.00	1.8352
$\chi^2_{0.95}(3) = 7.81 > 1.8352$			

(c) Chi-square test for $\hat{\sigma}^2(N)$

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above 1.55	3	3.59	0.0970
1.35~1.55	11	9.98	0.1042
1.15~1.35	21	20.89	0.0379
0.95~1.15	30	27.33	0.2608
0.75~0.95	19	23.52	0.8686
0.55~0.75	12	11.10	0.0730
Below 0.55	4	3.59	0.0468
Total	100	100.00	1.4883
$\chi^2_{0.95}(4) = 9.49 > 1.4883$			

4.6.3 Simulation Studies for Nonlinear ARMA Models

In this section, our purpose is to study numerically the nonlinear ARMA model given by (4.39) and estimators given by (4.46) and (4.54). Figure 4.8 (a) shows a sample run of input sequence which was used to generate the record $\{y_k\}$ of the simple first order nonlinear ARMA model. The simulated y_k is depicted in Fig. 4.8(b) through the input/output relation,

$$(4.76) \quad y_k + \overset{\circ}{\theta}_1 y_{k-1} = v_k + \alpha_1 H_1(v_{k-1};\sigma^2) + \alpha_2 H_2(v_{k-1};\sigma^2)$$

where parameter values were set as $\overset{\circ}{\theta}=-0.8$, $\alpha_1=0.1$, $\alpha_2=0.2$ and $\sigma^2=0.1$. Figure 4.8(c) shows a sample path of the nonlinear MA model associated with (4.76), i.e.,

$$(4.77) \quad y_k^{**} = v_k + \alpha_1 H_1(v_{k-1};\sigma^2) + \alpha_2 H_2(v_{k-1};\sigma^2).$$

Comparing y_k with y_k^{**}, we know the fact that the existence of the AR term in (4.76) plays an important role in exhibiting smoothing shapes of the output process y_k, although the ability of generating sporadic peak values is sufficient which is the fundamental property of the nonlinear MA model (4.1). Hence, it may be concluded that the proposed nonlinear ARMA model will be useful to model such random data which exhibits sporadic peak values and seems to have the serial dependency much longer than the single time unit.

Regarding the model (4.76) as the real input/output relation, the model of (4.76) is given by

$$(4.78) \quad y_k + \overset{\circ}{\theta}_1 y_{k-1} = \varepsilon_k + \overset{\circ}{\theta}_2 \varepsilon_{k-1},$$

and identifying $\overset{\circ}{\theta}$, $A(\overset{\circ}{\theta})$, $G(\overset{\circ}{\theta})$, $K(\overset{\circ}{\theta})$ and C in (4.45) with

$$
\left\{
\begin{array}{l}
\mathring{\theta} \triangleq (\mathring{\theta}_1 \quad \mathring{\theta}_2)' \\
A(\mathring{\theta}) \triangleq -\mathring{\theta}_1 \\
G(\mathring{\theta}) \triangleq (\alpha_1 - \mathring{\theta}_1 \sigma \quad \alpha_2) \\
K(\mathring{\theta}) \triangleq \mathring{\theta}_2 - \mathring{\theta}_1 \\
C \triangleq 1
\end{array}
\right.
\tag{4.79}
$$

respectively, the state space representations of (4.76) and (4.78) are given by (4.42) and (4.43). Hence, the algorithm for unknown parameter $\mathring{\theta}$ is given by

$$
\left\{
\begin{array}{l}
\hat{\theta}_{k+1} = \hat{\theta}_k + L_k(y_k - \hat{x}_k), \qquad \hat{\theta}_0 = \bar{\theta} \\
\hat{x}_{k+1} = -\hat{\theta}_{1k} x_k + (\hat{\theta}_{2k} - \hat{\theta}_{1k})(y_k - \hat{x}_k) \\
L_k = \hat{\sigma}_\varepsilon^{-2}(k) P_1(k)' \\
P_1(k+1) = -\hat{\theta}_{1k} P_1(k) + (y_k \quad y_k - \hat{x}_k) P_2(k) - (\hat{\theta}_{2k} - \hat{\theta}_{1k}) L_k' \hat{\sigma}_\varepsilon^2(k) \\
\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad P_1(0) = 0 \\
P_2(k+1) = P_2(k) - L_k L_k' \hat{\sigma}_\varepsilon^2(k) - \delta P_2^2(k), \qquad P_2(0) = \varepsilon I \quad (\varepsilon > 0) \\
\hat{\sigma}_\varepsilon^2(k+1) = \hat{\sigma}_\varepsilon^2(k) + (k+1)^{-1}\{(y_k - \hat{x}_k)^2 - \hat{\sigma}_\varepsilon^2(k)\}, \quad \hat{\sigma}_\varepsilon^2(0) = \bar{\sigma}_\varepsilon^2,
\end{array}
\right.
\tag{4.80}
$$

where $P_1(k)$ is the 2-dimensional low vector and $P_2(k)$ is the 2-dimensional positive definite matrix.

In case of (4.76), the $\tilde{y}_k$ process defined by (4.48) becomes

(4.81) $$\tilde{y}_k = y_k + \mathring{\theta}_1 y_{k-1}$$

and hence $\tilde{y}_k(\hat{\theta}_N)$ defined by (4.53) is given by

(4.82) $$\tilde{y}_k(\hat{\theta}_N) = y_k + \hat{\theta}_{1,N} y_{k-1}.$$

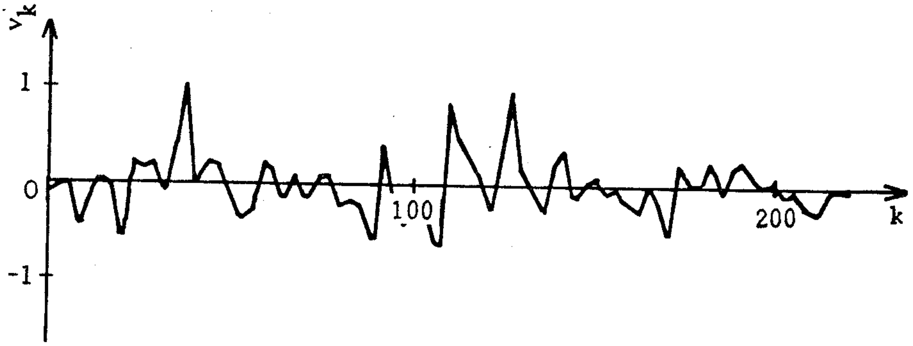
Therefore $\hat{\eta}_1(N)$, $\hat{\eta}_2(N)$ and $\hat{\eta}_3(N)$ are given by (4.52), i.e.,

(4.83) $$\begin{cases} \hat{\eta}_1(N) = \hat{\theta}_{2,N}\hat{\sigma}_\varepsilon^2(N) \\ \hat{\eta}_2(N) = \frac{1}{N} \sum_{k=2}^{N} \tilde{y}_k^2(\hat{\theta}_N)\tilde{y}_{k-1}(\hat{\theta}_N) \\ \hat{\eta}_3(N) = \frac{1}{N} \sum_{k=2}^{N} \tilde{y}_k(\hat{\theta}_N)\tilde{y}_{k-1}^2(\hat{\theta}_N) \end{cases}$$

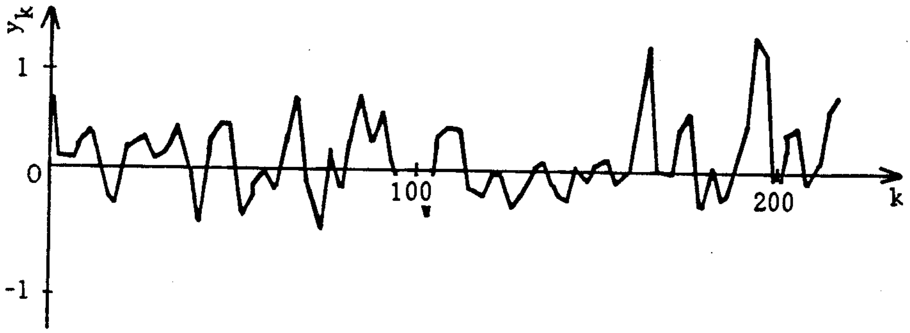
and then unknown parameters in the nonlinear MA terms of (4.76) are obtained by

(4.84) $$\begin{cases} \hat{\alpha}_1(N) = \hat{\eta}_1(N)\{(2\hat{\eta}_1(N)\hat{\eta}_3(N))^{-1}\hat{\eta}_2(N)\}^{1/2} \\ \hat{\alpha}_2(N) = (4\hat{\eta}_1(N))^{-1}\hat{\eta}_2(N) \\ \hat{\sigma}^2(N) = 2\hat{\eta}_1(N)\hat{\eta}_3(N)\hat{\eta}_2(N)^{-1}. \end{cases}$$

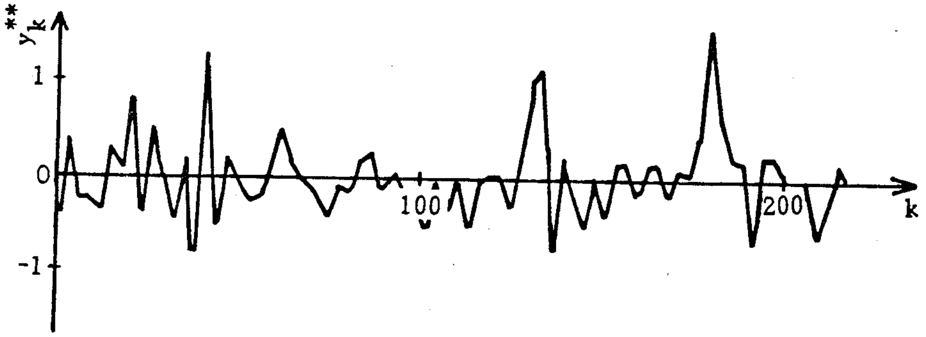
In digital simulation studies, the initial values of $\hat{\theta}_k$, $\hat{x}_k$ and $P_2(k)$ were set as $\hat{\theta}_0=(-0.5 \quad 0)'$, $\hat{x}_0=0$ and $P_2(0)=10I$. Convergences of estimators $\hat{\theta}_1(N)$, $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ are observed by the typical sample runs shown in Fig. 4.9. Convergences of estimators $\hat{\theta}_1(N)$, $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$ are well achieved as the observation data N increases as shown in Fig. 4.9.



(a) A sample run of the input noise v_k



(b) A sample run of the output process of the nonlinear ARMA model



(c) A sample run of the output process of the nonlinear MA model

Fig. 4.8 Sample runs of the input noise, the outputs of the nonlinear ARMA model and the nonlinear MA model

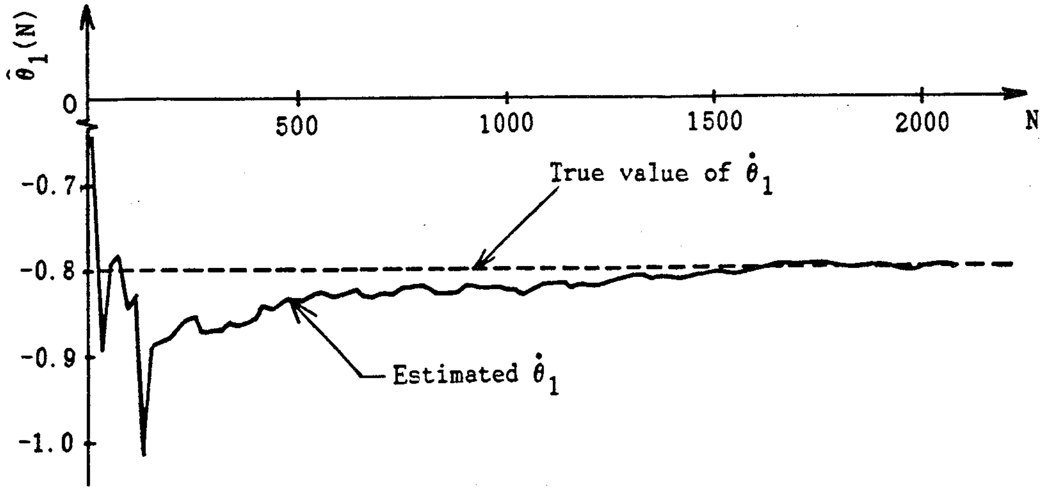


Fig. 4.9 (a) A sample run of $\hat{\theta}_1(N)$

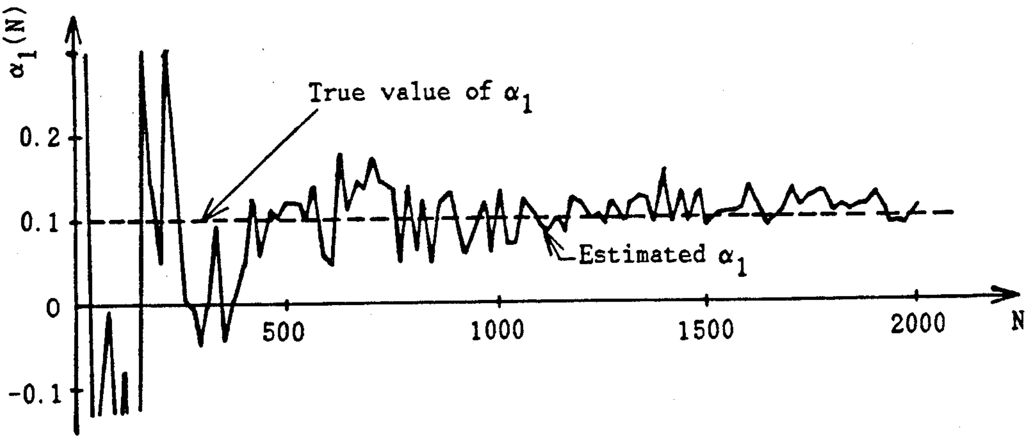


Fig. 4.9 (b) A sample run of $\hat{\alpha}_1(N)$ (every 20 steps)

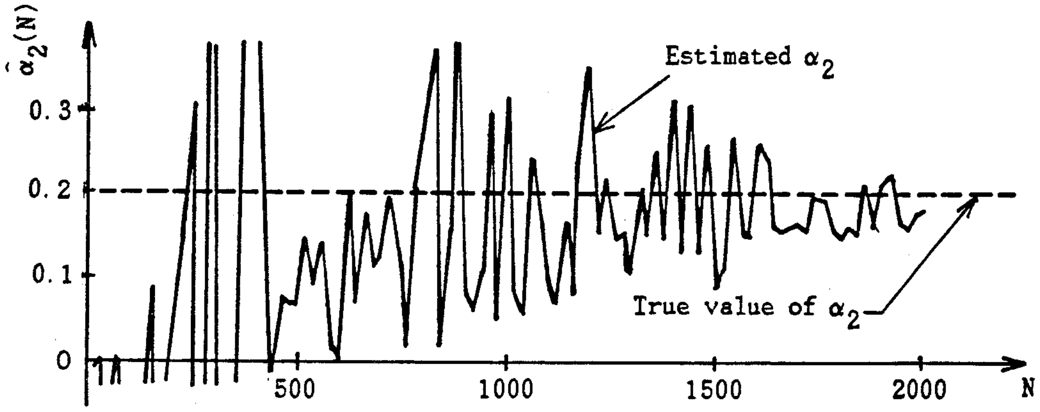


Fig. 4.9 (c) A sample run of $\hat{\alpha}_2(N)$ (every 20 steps)

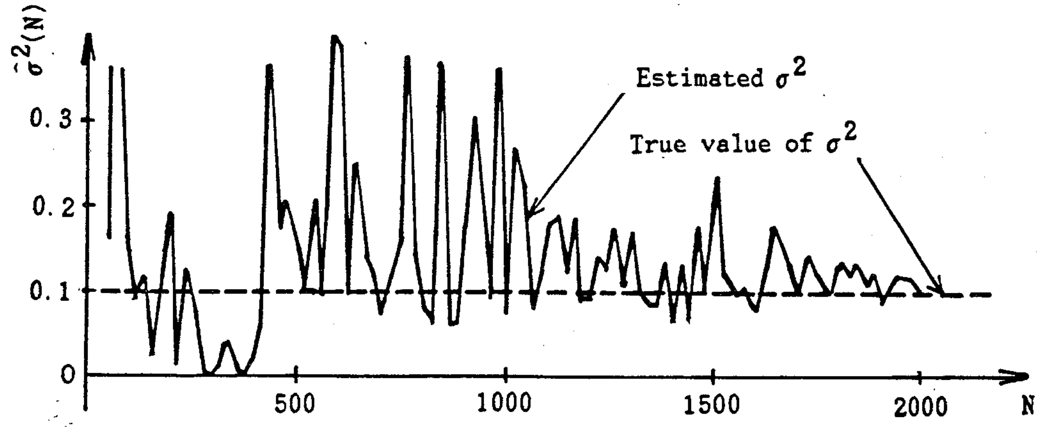


Fig. 4.9 (d) A sample run of $\hat{\sigma}^2(N)$ (every 20 steps)

Fig. 4.9 Convergence features of $\hat{\theta}_1(N)$, $\hat{\alpha}_1(N)$, $\hat{\alpha}_2(N)$ and $\hat{\sigma}^2(N)$

4.7 Discussions

In this chapter, a class of nonlinear MA model has been introduced and using the moment method, estimators of unknown system parameters have been derived. Proposed estimators are consistent and asymptotically normal.

Using the fact that unknown parameters included in the proposed model are uniquely described as functions of the second and the third moments of the output process, estimators of unknown parameters have been obtained. The consistency and the asymptotic normality of estimators have been shown by using the martingale properties of quantities given by (4.16).

The proposed nonlinear MA model has following properties compared with the Robinson's model which is another representative one; (i) since the proposed model utilizes the orthogonal polynomials, it has a possibility to extent to more general one; (ii) in the proposed model we can find a set of statistics (a b c) that can uniquely determine the triplet of estimators $(\hat{\alpha}_1(N)\ \hat{\alpha}_2(N)\ \hat{\sigma}^2(N))$ as shown in Section 4.3, while, in the Robinson's model, estimators of unknown parameters are not uniquely determined; (iii) the proposed model has been the possibility to produce larger values than those by Robinson's model because of the fact that $v_n v_{n-1} \leq [\max(v_n, v_{n-1})]^2$. The distinction between estimators of the two models mentioned above are listed in Table 4.2.

The autocorrelation of the output process generated by the proposed nonlinear MA model completely equals to zero when its time lag is longer than the single time unit. Hence, in the case of autocorrelation between the observed data y_k and $y_{k+\ell}(\ell \geq 2)$ may not be ignored, it is appropriate use of the nonlinear ARMA model given in Section 4.5.

Table 4.2 Comparison of the two nonlinear MA models

	Proposed model	Robinson's model
models	$y_k = v_k + \alpha_1 H_1(v_{k-1};\sigma^2) + \alpha_2 H_2(v_{k-1};\sigma^2)$	$y_k = v_k + \alpha_1 v_{k-1} + \alpha_2 v_k v_{k-1}$
Three statistical moments	$\hat{a}_N = \frac{1}{N} \sum_{k=2}^{N} y_k y_{k-1}$ $\hat{b}_N = \frac{1}{N} \sum_{k=2}^{N} y_k^2 y_{k-1}$ $\hat{c}_N = \frac{1}{N} \sum_{k=2}^{N} y_k y_{k-1}^2$	$\hat{a}_N = \frac{1}{N} \sum_{k=2}^{N} y_k^2$ $\hat{b}_N = \frac{1}{N} \sum_{k=2}^{N} y_k y_{k-1}$ $\hat{c}_N = \frac{1}{N} \sum_{k=3}^{N} y_k y_{k-1} y_{k-2}$
Estimators	$\hat{\alpha}_1(N) = \hat{a}_N \left(\frac{\hat{b}_N}{2\hat{a}_N \hat{c}_N} \right)^{1/2}$ $\hat{\alpha}_2(N) = \frac{\hat{b}_N}{4\hat{a}_N}$ $\hat{\sigma}^2(N) = \frac{2\hat{a}_N \hat{c}_N}{\hat{b}_N}$	$\hat{\alpha}_1(N) = \frac{1}{2} \{ \frac{\hat{a}_N}{\hat{b}_N} - \frac{\hat{c}_N^2}{\hat{b}_N^3} \pm ((\frac{\hat{a}_N}{\hat{b}_N} - \frac{\hat{c}_N^2}{\hat{b}_N^3})^2 - 4)\}^{1/2}$ $\hat{\alpha}_2(N) = \frac{\hat{c}_N}{\hat{\alpha}_1(N)\hat{\sigma}^2(N)}$ $\hat{\sigma}^2(N) = \frac{\hat{b}_N}{\hat{\alpha}_1(N)}$

Appendix 4.A Proof of Lemma 4.1

Define

$$\text{(A.1)} \quad r_k \triangleq \prod_{j=0}^{m} e_{k-j}^{i(j)},$$

then we can rewrite S_N defined by (4.14) as follows:

$$\text{(A.2)} \quad S_N = \frac{1}{N} \sum_{k=m+1}^{N} r_k$$

$$= \frac{1}{N} \{\sum_{n=m+1}^{M(m+1)+m} r_n - \sum_{n=N+1}^{M(m+1)+m} r_n\}$$

$$= \frac{1}{N} \{\sum_{j=0}^{m} \sum_{k=1}^{M} r_{k(m+1)+j} - \sum_{j=L+1}^{m} r_{M(m+1)+j}\}$$

$$= \frac{1}{N} \{\sum_{j=0}^{L} \sum_{k=1}^{M} r_{k(m+1)+j} + \sum_{j=L+1}^{m} (\sum_{k=1}^{M} r_{k(m+1)+j} - r_{M(m+1)+j})\}$$

$$= \sum_{j=0}^{L} \{\frac{1}{N} \sum_{k=1}^{M} r_{k(m+1)+j}\} + \sum_{j=L+1}^{m} \{\frac{1}{N} \sum_{k=1}^{M-1} r_{k(m+1)+j}\}$$

where $M=[N/(m+1)]$, $L=N-M(m+1)$ and the square bracket [ ] denotes the largest integer not greater than $N/(m+1)$. Letting

$$\text{(A.3)} \quad S_{j,M} = \begin{cases} \frac{1}{N} \sum_{k=1}^{M} r_{k(m+1)+j} & \text{for } M \neq 0 \\ 0 & \text{for } M=0, \end{cases}$$

it follows from (A.2) that

$$\text{(A.4)} \quad S_N = \sum_{j=0}^{L} S_{j,M} + \sum_{j=L+1}^{m} S_{j,M-1}.$$

Now, letting $\alpha(k,j) \triangleq k(m+1)+j$ $(j=0,1,\cdots,m)$, we have from the condition (C-4) that

$$(\text{A.5}) \quad E\{r_{k(m+1)+j} \mid \mathscr{F}_{(k-1)(m+1)+j}\}$$

$$= E\{r_{\alpha(k,j)} \mid \mathscr{F}_{\alpha(k,j)-(m+1)}\}$$

$$= E\{(e_{\alpha(k,j)-m})^{i(m)} E\{(e_{\alpha(k,j)-(m-1)})^{i(m-1)} \cdots$$

$$\times E\{(e_{\alpha(k,j)})^{i(0)} \mid \mathscr{F}_{\alpha(k,j)-1}\} \mid \cdots\} \mid \mathscr{F}_{\alpha(k,j)-(m+1)}\}$$

$$= h_m h_{m-1} \cdots h_0$$

$$= 0.$$

Furthermore, the uniform integrability of $\{|r_k|^{2p_0}\}$ $(p_0=p/i_{max},\ 1 \leq p_0 < 2)$ will be shown in order to evaluate the rate of convergence of $S_{j,M}$. Letting $\lambda > 0$, we have

$$(\text{A.6}) \quad \{\omega: |r_k| \geq \lambda\} = \{\omega: |\prod_{j=0}^{m} e_{k-j}^{i(j)}| \geq \lambda\}$$

$$\subset \{\omega: \frac{1}{m+1} \sum_{j=0}^{m} |e_{k-j}|^{i(j)} \geq \lambda^{1/(m+1)}\}$$

$$\subset \bigcup_{j=0}^{m} \{\omega: |e_{k-j}|^{i(j)} \geq \lambda^{1/(m+1)}\}.$$

Hence, it follows that

$$(\text{A.7}) \quad \int_{\{\omega: |r_k| \geq \lambda\}} |r_k|^{p_0} dP$$

$$\leq \int_{\Omega} \prod_{j=0}^{m} |e_{k-j}|^{i(j)p_0} (\sum_{\ell=0}^{m} I\{|e_{k-\ell}|^{i(\ell)} \geq \lambda^{1/(m+1)}\}) dP$$

$$\le \sum_{\ell=0}^{m}[E\{\prod_{\substack{j=0\\j\neq\ell}}^{m}|e_{k-j}|^{2i(j)p_0}\}]^{1/2}[\int_\Omega |e_{k-\ell}|^{2i(\ell)p_0}$$

$$\times I\{|e_{k-\ell}|^{i(\ell)} \ge \lambda^{1/(m+1)}\}dP]^{1/2}$$

$$\le \sum_{\ell=0}^{m}[E\{E\{|e_{k-m}|^{2i(m)p_0}E\{|e_{k-m+1}|^{2i(m-1)p_0}\cdots$$

$$\times E\{|e_k|^{2i(0)p_0}|\mathcal{F}_{k-1}\}|\cdots|\mathcal{F}_{k-m}\}|\mathcal{F}_{k-m-1}\}\}]^{1/2}$$

$$\times[\int_\Omega |e_{k-\ell}|^{2i(\ell)p_0}I\{|e_{k-\ell}|^{i(\ell)} > \lambda^{1/(m+1)}\}dP]^{1/2}$$

$$\le c^{m/2}\sum_{\ell=0}^{m}[\int_\Omega |e_{k-\ell}|^{2i(\ell)p_0}I\{|e_{k-\ell}|^{i(\ell)} > \lambda^{1/(m+1)}\}dP]^{1/2}.$$

where the condition (C-5) has been used. From (A.7), the uniform integrability of the variable $\{|r_k|^{p_0}\}$ follows from that of $|e_k|^{2i(\ell)p_0}$, which is guaranteed by the condition (C-6). Hence, we know that $\{|r_{\alpha(k,j)}|^{p_0}\}$ is uniformly integrable.

Recall the following lemma due to Chow [C2].

[Lemma 4.A](Chow [C2]) Let $\{u_k;\ k=1,2,\cdots\}$ be a sequence of random variables and let

$$\text{(A.8)}\quad \tilde{S}_N \triangleq \sum_{k=1}^{N} u_k.$$

Suppose that the sequence $\{|u_k|^{\nu};\ k=1,2,\cdots\}$ is uniformly integrable for some $1\le\nu<2$. Then

$$\text{(A.9)}\quad E\{|\tilde{S}_N - \tilde{u}_N|^{\nu}\} = o(N) \qquad \text{as} \quad N \to \infty,$$

where

(A.10) $\tilde{u}_N \triangleq \sum_{k=1}^{N} E\{u_k \mid u_1, u_2, \cdots, u_{k-1}\}$.

Then, regarding $NS_{j,M}$ and $r_{\alpha(k,j)}$ as $\tilde{S}_N$ and u_k in Lemma 4.A respectively, the following evaluation is obtained:

$$E\{\mid NS_{j,M} \mid^{p_0}\} = o(M) \qquad \text{as} \quad M \to \infty,$$

or equivalently

$$\text{(A.11)} \quad E\{\mid S_{j,M} \mid^{p_0}\} = o(MN^{-p_0}) = o(N^{1-p_0}) \qquad \text{as} \quad N \to \infty.$$

The same evaluation as (A.11) holds for $S_{j,M-1}$. Therefore, using elementary inequalities

$$\mid z_1 + z_2 \mid^{\nu} \le 2^{\nu}(\mid z_1 \mid^{\nu} + \mid z_2 \mid^{\nu}) \qquad \text{for } \nu > 0$$

and

$$\mid \sum_{i=0}^{N} z_i \mid^{\nu} \le (N+1)^{\nu-1} \sum_{i=0}^{N} \mid z_i \mid^{\nu} \qquad \text{for } \nu > 0,$$

and (A.4), we have

$$\text{(A.12)} \quad E\{\mid S_N \mid^{p_0}\} \le 2^{p_0} E\{\mid \sum_{j=0}^{L} S_{j,M} \mid^{p_0} + \mid \sum_{j=L+1}^{m} S_{j,M-1} \mid^{p_0}\}$$

$$\le 2^{p_0}[(L+1)^{p_0-1} E\{\sum_{j=0}^{L} \mid S_{j,M} \mid^{p_0}\}$$

$$+ (m-L)^{p_0-1} E\{\sum_{j=L+1}^{m} \mid S_{j,M-1} \mid^{p_0}\}].$$

Hence, from (A.11) and (A.12), we have that for $1 \leq p_0 < 2$

$$(\mathrm{A}.13) \quad E\{ | S_N |^{p_0} \} = o(N^{1-p_0}) \qquad \text{as} \quad N \to \infty$$

and it follows (4.15) by replacing p_0 by p/i_{max} (where $i_{max} \leq p < 2i_{max}$).

Appendix 4.B Proof of Lemma 4.2

Since $\{e_k\}$ is the i.i.d. sequence, we have

$$\text{(B.1)} \quad E\{r_k\} = E\{\prod_{j=0}^{m} e_{k-j}^{i(j)}\}$$

$$= \prod_{j=0}^{m} E\{e_{k-j}^{i(j)}\}$$

$$= h_m h_{m-1} \cdots h_0 = 0$$

and, for $p_0 = p/i_{max}$,

$$\text{(B.2)} \quad E\{|r_k|^{p_0}\} \leq \prod_{j=0}^{m} E\{|e_{k-j}|^{i(j)p_0}\}$$

$$= \prod_{j=0}^{m} E\{|e_{k-j}|^{i(j)p/i_{max}}\}$$

$$\leq \prod_{j=0}^{m} [E\{|e_{k-j}|^{p}\}]^{i_{max}/i(j)} < c \text{ (const.)}$$

because of the condition (C-5)'.

Recall the following lemma due to Marcinkiewicz (see e.g. [L3, p.255]).

[Lemma 4.B] (Marcinkiewicz) Let $\{u_k: k=1,2,\cdots\}$ be a sequence of i.i.d. random variables and let $1 \leq \nu < 2$. If $E\{|u_k|^{\nu}\} < \infty$, then

$$\text{(B.3)} \quad \sum_{k=0}^{N} (u_k - \tilde{u}) \rightarrow o(N^{1/\nu}) \qquad \text{w.p. } 1 \qquad \text{as } N \rightarrow \infty.$$

where

$$\tilde{u} = E\{u_k\}$$

Then, regarding r_k and p_0 as u_k and λ respectively, we have from $S_{j,M}$

defined by (A.3) that

$$(B.4)\quad S_{j,M} = o(N^{-1}M^{1/p_0}) = o(N^{1/p_0-1}) \qquad \text{w.p. } 1 \qquad \text{as } N \to \infty.$$

The same evaluation as in (B.4) holds for $S_{j,M-1}$. Hence, it follows that

$$(B.5)\quad S_N = o(N^{1/p_0-1}) \qquad \text{w.p. } 1 \qquad \text{as } N \to \infty$$

and replacing p_0 by p/i_{max}, we have (4.26).

Appendix 4.C **Proof of Lemma 4.3**

Noting that S_N defined by (4.14) is expressed as (A.4) in Appendix 4.A, we have from the definition of $S_{j,M}$ that

$$(C.1) \quad E\{S_{j,M} \mid \mathcal{F}_{(M-1)(m+1)+j}\}$$

$$= \frac{1}{N} \sum_{k=1}^{M} E\{r_{k(m+1)+j} \mid \mathcal{F}_{(M-1)(m+1)+j}\}$$

$$= \frac{1}{N} [\sum_{k=1}^{M-1} E\{r_{\alpha(k,j)} \mid \mathcal{F}_{\alpha(M-1,j)}\} + E\{r_{\alpha(M-1,j)+m+1} \mid \mathcal{F}_{\alpha(M-1,j)}\}],$$

where $\alpha(k,j)=k(m+1)+j$. From the definition of r_k and (C-4), we see that

$$(C.2) \quad E\{r_{\alpha(M-1,j)+m+1} \mid \mathcal{F}_{\alpha(M-1,j)}\}$$

$$= E\{(e_{\alpha(M-1,j)+m+1})^{i(0)}(e_{\alpha(M-1,j)+m})^{i(1)}$$

$$\cdots (e_{\alpha(M-1,j)+1})^{i(m)} \mid \mathcal{F}_{\alpha(M-1,j)}\}$$

$$= 0.$$

Hence, we can see from (C.1) and (C.2) that

$$(C.3) \quad E\{S_{j,M} \mid \mathcal{F}_{(M-1)(m+1)+j}\} = S_{j,M-1},$$

and this means that $\{S_{j,M}, \mathcal{F}_{M(m+1)+j}\}$ $(M=1,2,\cdots)$ is a martingale.

Since, from the condition (C-10), it is easily verified that the variance of $S_{j,M}$ is finite, the proof of Lemma 4.3 will be completed if we can establish that for each j, $M^{-1/2}S_{j,M}$ converges to a Gaussian random variable with zero mean and the variance $M^{-1}E\{S_{j,M}^2\}$ as $M \to \infty$. For this purpose we need the help of Theorem 2.3 concerning a central limit theorem for martingales due to Brown [B2]. Identifying $S_{j,M}$ $(M=1,2,\cdots)$ with the variable x_k $(k=1,2,\cdots)$ in Theorem 2.3, and using s_{jM}^2 defined by (4.32), we have

$$(\text{C.4})\quad E\{S^2_{j,M}\} = \frac{1}{N^2} \sum_{p=1}^{M} \sum_{q=0}^{M} E\{r_{p(m+1)+j} r_{q(m+1)+j}\}$$

$$= \frac{1}{N^2} \sum_{p=1}^{M} \prod_{k=0}^{m} E\{(e_{p(m+1)+j-k})^{2i(k)}\}$$

$$= \frac{1}{N^2} s^2_{jM},$$

which implies that $N^{-2}s^2_{jM}$ can be identified as u^2_k in Theorem 2.3. Hence, in order to apply the result of Theorem 2.3, we have to show that

$$(\text{C.5})\quad \left(\frac{1}{N^2} s^2_{jM} \right)^{-1} \sum_{k=1}^{M} E\{(S_{j,k} - S_{j,k-1})^2 \mid \mathcal{F}_{(k-1)(m+1)+j}\}$$

$$\rightarrow 1 \quad \text{in prob.} \quad \text{as } M \rightarrow \infty,$$

and for $\varepsilon > 0$,

$$(\text{C.6})\quad \left(\frac{1}{N^2} s^2_{jM} \right)^{-1} \sum_{k=1}^{M} E\{(S_{j,k} - S_{j,k-1})^2 I\{ \mid S_{j,k} - S_{j,k-1} \mid \geq \varepsilon(\frac{s_{jM}}{N}))\}$$

$$\rightarrow 0 \quad \text{in prob.} \quad \text{as } M \rightarrow \infty.$$

Using (A.3) and (A.4), these conditions can be rewritten as follows:

$$(\text{C.7})\quad s^{-2}_{jM} \sum_{k=1}^{M} E\{(r_{\alpha(k,j)})^2 \mid \mathcal{F}_{\alpha(k-1,j)}\} \rightarrow 1 \quad \text{in prob.} \quad \text{as } M \rightarrow \infty,$$

$$(\text{C.8})\quad s^{-2}_{jM} \sum_{k=1}^{M} E\{(r_{\alpha(k,j)})^2 I\{ \mid r_{\alpha(k,j)} \mid \ > \varepsilon s_{jM}\}\} \rightarrow 0$$

$$\text{in prob.} \quad \text{as } M \rightarrow \infty.$$

It is easily derived from the condition (C-10) that

$$(\text{C.9})\quad E\{(r_{\alpha(k,j)})^2 \mid \mathcal{F}_{\alpha(k-1,j)}\} = E\{\prod_{s=0}^{m} (e_{\alpha(k,j)-s})^{2i(s)} \mid \mathcal{F}_{\alpha(k-1,j)}\}$$

$$= \prod_{s=0}^{m} E\{(e_{\alpha(k,j)-s})^{2i(s)}\},$$

and hence (C.7) is guaranteed.

Furthermore by a similar procedure to the derivation of (A.7), it follows that

(C.10) $$E\{(r_{\alpha(k,j)})^2 I\{\,|r_{\alpha(k,j)}| > \varepsilon s_{jM}\}\}$$

$$= \int_{\Omega} (r_{\alpha(k,j)})^2 I\{\,|r_{\alpha(k,j)}| > \varepsilon s_{jM}\} dP$$

$$\leq \int_{\Omega} \prod_{s=0}^{m} (e_{\alpha(k,j)-s})^{2i(s)} [\sum_{p=0}^{m} I\{\,|e_{\alpha(k,j)-p}|^{i(p)} \geq (\varepsilon s_{jM})^{1/(m+1)}\}] dP$$

$$\leq \sum_{p=0}^{m} [\int_{\Omega} \prod_{\substack{s=0 \\ s \neq p}}^{m} (e_{\alpha(k,j)-s})^{4i(s)} dP]^{1/2}$$

$$\times [\int_{\Omega} (e_{\alpha(k,j)-p})^{4i(p)} I\{\,|e_{\alpha(k,j)-p}|^{i(p)} \geq (\varepsilon s_{jM})^{1/(m+1)}\} dP]^{1/2}.$$

Using the notation $e_0 \triangleq (e_{\alpha(k,j)-p})^{i(p)}$ and $\varepsilon_0 \triangleq (\varepsilon s_{jM})^{1/(m+1)}$ for simplicity and letting $F(\lambda) \triangleq P(\,|e_0| > \lambda)$, it can be obtained that

(C.11) $$\int_{\Omega} (e_{\alpha(k,j)-p})^{4i(p)} I\{\,|e_{\alpha(k,j)-p}|^{i(p)} > (\varepsilon s_{jM})^{1/(m+1)}\} dP$$

$$= \int_{\Omega} e_0^4 I\{\,|e_0| > \varepsilon_0\} dP$$

$$= -\int_0^{\infty} \lambda^4 I\{\lambda > \varepsilon_0\} dF(\lambda)$$

$$= -\lim_{\lambda \to \infty} \lambda^4 F(\lambda) I\{\lambda > \varepsilon_0\} + 4\int_0^{\infty} \lambda^3 F(\lambda) I\{\lambda > \varepsilon_0\} d\lambda$$

$$= -\lim_{\lambda \to \infty} \lambda^4 F(\lambda) + 4\int_{\varepsilon_0}^{\infty} \lambda^3 P(\,|e_0| > \lambda) d\lambda.$$

Since $E\{|\tilde{e}|^{4i_{\max}}\} < \infty$, it follows from the condition (C-12) that

$$\text{(C.12)} \quad \lim_{\lambda\to\infty} \lambda^4 F(\lambda) = \lim_{\lambda\to\infty} \lambda^4 P(|e_0| > \lambda)$$

$$\le \lim_{\lambda\to\infty} \lambda^4 P(|\tilde{e}| > \lambda)$$

$$= 0.$$

Hence (C.11) can be represented by

$$\text{(C.13)} \quad \int_\Omega (e_{\alpha(k,j)-p})^{4i(p)} I\{|e_{\alpha(k,j)-p}|^{i(p)} > (\varepsilon s_{jM})^{1/(m+1)}\} dP$$

$$= 4\int_{(\varepsilon s_{jM})^{1/(m+1)}}^{\infty} \lambda^3 P(|e_{\alpha(k,j)-p}|^{i(p)} > \lambda) d\lambda$$

and it is obvious from (C.10) and (C.13) that

$$\text{(C.14)} \quad E\{(r_{\alpha(k,j)})^2 I\{|r_{\alpha(k,j)}| > \varepsilon s_{jM}\}\}$$

$$\le 2\sum_{p=0}^{m} [\int_\Omega \prod_{\substack{s=0 \\ s\neq p}}^{m} (e_{\alpha(k,j)-s})^{4i(s)} dP]^{1/2}$$

$$\times [\int_{(\varepsilon s_{jM})^{1/(m+1)}}^{\infty} \lambda^3 P(|e_{\alpha(k,j)-s}|^{i(p)} > \lambda) d\lambda]^{1/2}$$

$$\to 0 \quad \text{as} \quad M \to \infty,$$

because $s_{jM} \to \infty$ as $M \to \infty$ by the condition (C.11).

CHAPTER 5 PARAMETER IDENTIFICATION OF NONSTATIONARY NONLINEAR SYSTEMS

5.1 Introductory Remarks

Based on the assumption that the system is linear whose model was given by a linear stochastic differential equation with time-varying coefficients, the nonstationary statistics of the data was taken into account in Chapter 3. In Chapter 4, although the data was assumed to be stationary, a class of nonlinear models was adopted. In this chapter, from more practical viewpoints, the author proposes a class of nonstationary nonlinear model where a type of difference equations derived directly from stochastic differential equations is adopted. The principal line of attack is to assume that the nonlinear time varying function in the system model can be expanded into M known functions with unknown constant coefficients.

The mathematical model is given in Section 5.2 including the problem statement. Based on the maximum likelihood principle, the identification procedure is derived in Section 5.3. Sections 5.4 and 5.5 are devoted to investigating asymptotic properties of estimators. In the final section, experiments are presented to demonstrate asymptotic properties of proposed estimators.

5.2 Nonstationary Nonlinear System Models

Let $\{y_k;k=0,1,2,\cdots\}$ be the observed n-dimensional discrete output data and assume that $\{y_k\}$ comes actually from the following nonstationary nonlinear input/output relation:

$$(5.1)\quad \begin{cases} y_{k+1} - y_k = \phi_k(y_k,\theta) + B_{k+1}v_{k+1} \\ \\ y_0 = \xi \end{cases}$$

where B_{k+1} is the $n \times m$ known matrix whose structure is assumed to be

$$(5.2)\quad B_{k+1} = \underbrace{\begin{bmatrix} 0 \\ \bar{B}_{k+1} \end{bmatrix}}_{m} \begin{matrix} \}n-m \\ \}m \end{matrix} ,$$

and $\bar{B}_{k+1}$ is the bounded square matrix such that rank $\bar{B}_{k+1}=m$ for $k=0,1,\cdots$. The input sequence $\{v_{k+1}\}$ is an m-dimensional unobservable random variables which satisfy the following basic condition:

(C-1) Let $\mathcal{F}_k$ $(k=1,2,\cdots,N)$ be the increasing σ-algebra generated from $\{v_1, v_2, v_3, \cdots, v_k\}$. Then, for each k, $\{v_k\}$ satisfies

$$(5.3)\quad \begin{cases} E\{ v_k \mid \mathcal{F}_{k-1}\} = 0 & \text{w.p. } 1 \\ E\{ v_k v_k' \mid \mathcal{F}_{k-1}\} = \sigma^2 I & \text{w.p. } 1 \\ E\{(v_k v_k')^2 \mid \mathcal{F}_{k-1}\} \leq cI & \text{w.p. } 1, \end{cases}$$

where $\sigma^2 I$ is the unknown variance of $\{v_k\}$.

The nonlinear time-varying function $\phi_k(y_k,\theta)$ in (5.1) may be expanded into

$$(5.4)\quad \phi_k(y_k,\mathring{\theta}) = \phi_{0k}(y_k) + \sum_{i=1}^{M} \mathring{\theta}_i \phi_{ik}(y_k),$$

where $\{ \phi_{ik}(\cdot); i=0,1,\cdots,M \}$ are known time-varying functions and $\mathring{\theta}_i$ is the unknown constant parameter such that

$$(5.5)\quad \mathring{\theta} \triangleq [\mathring{\theta}_1,\mathring{\theta}_2,\cdots,\mathring{\theta}_M]'.$$

The problem in this chapter is to find consistent estimators of unknown parameters $\mathring{\theta}$ and σ^2 from the given observed data, and investigate asymptotic properties of estimators from both theoretical and numerical viewpoints.

5.3 Parameter Identification by Using the M.L.E. Approach

In order to estimate the unknown parameter θ and σ^2 from the observed data $Y_N \triangleq \{y_1, \cdots, y_N\}$, the following criterion function is adopted:

$$(5.6) \quad \ell_N(\theta, \sigma^2) \triangleq Nm \log \sigma^2 + \frac{1}{\sigma^2} \sum_{k=0}^{N} \{y_{k+1} - y_k - \phi_k(y_k, \theta)\}' \times \tilde{B}_{k+1}^2 \{y_{k+1} - y_k - \phi_k(y_k, \theta)\}.$$

where

$$(5.7) \quad \tilde{B}_{k+1} \triangleq \begin{array}{cc} \left[\begin{array}{cc} 0 & 0 \\ 0 & (\bar{B}_{k+1}\bar{B}'_{k+1})^{-1/2} \end{array} \right] & \begin{array}{l} \}n-m \\ \}m \end{array} \\ \begin{array}{cc} \underbrace{\quad}_{n-m} & \underbrace{\qquad\qquad\qquad}_{m} \end{array} & \end{array} \quad (n \times n \text{ matrix})$$

The above criterion function is derived from the maximum likelihood identification method when the input sequence $\{v_k\}$ are independent and identically distributed Gaussian random variables.

By setting $\partial \ell_N(\theta, \sigma^2)/\partial \sigma^2 = 0$, it is easily verified that the minimization of $\ell_N(\theta, \sigma^2)$ is achieved by minimizing

$$(5.8) \quad \bar{\ell}_N(\theta) \triangleq \sum_{k=0}^{N} \{y_{k+1} - y_k - \phi_k(y_k, \theta)\}' \tilde{B}_{k+1}^2 \{y_{k+1} - y_k - \phi_k(y_k, \theta)\}$$

and the estimate of σ^2 is given by

$$(5.9) \quad \hat{\sigma}^2(N) = \frac{1}{mN} \sum_{k=0}^{N} \{y_{k+1} - y_k - \phi_k(y_k, \hat{\theta}(N))\}' \tilde{B}_{k+1}^2 \{y_{k+1} - y_k - \phi_k(y_k, \hat{\theta}(N))\}$$

where $\hat{\theta}(N)$ is the estimate of θ, which is obtained by minimizing (5.8), and this is also easily obtained by setting $\partial \bar{\ell}_N(\theta)/\partial \theta = 0$. From (5.4) and (5.8), we have

(5.10) $$\frac{\partial\bar{\ell}(N)}{\partial\theta_j} = 2\sum_{k=0}^{N}\{y_{k+1} - y_k - \phi_k(y_k,\theta)\}'\tilde{B}_{k+1}^2 \frac{\partial}{\partial\theta_j}\{y_{k+1} - y_k - \phi_k(y_k,\theta)\}$$

and

(5.11) $$\frac{\partial}{\partial\theta_j}\{y_{k+1} - y_k - \phi(y_k,\theta)\} = -\phi_{jk}(y_k)$$

respectively. Hence, it follows that

(5.12) $$\begin{aligned}\frac{\partial\bar{\ell}_N(\theta)}{\partial\theta_j} &= -2\sum_{k=0}^{N}\{y_{k+1} - y_k - \phi_k(y_k,\theta)\}'\tilde{B}_{k+1}^2\phi_{jk}(y_k)\\ &= -2\sum_{k=0}^{N}\{y_{k+1} - y_k - \phi_{0k}(y_k,\theta)\}'\tilde{B}_{k+1}^2\phi_{jk}(y_k)\\ &\qquad + 2\sum_{k=0}^{N}\{\sum_{i=1}^{M}\phi'_{ik}(y_k)\tilde{B}_{k+1}^2\phi_{jk}(y_k)\}\theta_i.\\ &= -2\sum_{k=0}^{N}\phi'_{jk}(y_k)\tilde{B}_{k+1}^2\{y_{k+1} - y_k - \phi_{0k}(y_k)\}\\ &\qquad + 2\sum_{k=0}^{N}\{\sum_{i=1}^{M}\phi'_{ik}(y_k)\tilde{B}_{k+1}^2\phi_{jk}(y_k)\}\theta_i.\end{aligned}$$

Then by setting $\partial\bar{\ell}_N(\theta)/\partial\theta_j=0$ $(j=1,2,\cdots,M)$, we have

(5.13) $$\begin{aligned}&\sum_{k=0}^{N}\phi'_{jk}(y_k)\tilde{B}_{k+1}^2\{y_{k+1} - y_k - \phi_{0k}(y_k)\}\\ &\qquad = \sum_{k=0}^{N}\{\sum_{i=1}^{M}\phi'_{ik}(y_k)\tilde{B}_{k+1}^2\phi_{jk}(y_k)\}\theta_i.\end{aligned}$$

Hence defining $F_k(y_k)$, $s(N)$ and $Q(N)$ by

(5.14) $$F_k(y_k) \triangleq [\phi_{1k}(y_k) \quad \phi_{2k}(y_k) \quad \cdots \quad \phi_{Mk}(y_k)]'\tilde{B}_{k+1} \qquad (M\times n \text{ matrix})$$

(5.15) $$s(N) \triangleq \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}(y_{k+1} - y_k - \phi_{0k}(y_k)) \qquad \text{(M-dim. vector)}$$

(5.16) $$Q(N) \triangleq \sum_{k=0}^{N} F_k(y_k)F_k'(y_k) \qquad (M \times M \text{ matrix})$$

respectively, (5.13) can be rewritten by

(5.17) $$Q(N)\hat{\theta}(N) = s(N).$$

In order to avoid numerical difficulties due to the singularity of Q(N), we introduce the matrix $\Gamma(N)$ defined by

(5.18) $$\Gamma(N) \triangleq (Q(N) + \rho I)^{-1},$$

where ρ is an arbitrary small positive constant given a priori. Using $\Gamma(N)$ for Q(N), the unknown parameter $\mathring{\theta}$ is uniquely estimated by

(5.19) $$\hat{\theta}(N) = \Gamma(N)s(N).$$

Invoking the matrix inversion lemma [S7], we have the recursive version of (5.19) as follows:

(5.20a) $$\hat{\theta}(N) = \hat{\theta}(N-1) + \Gamma(N-1)F_N(y_N)\{ I + F_N'(y_N)\,\Gamma(N-1)F_N(y_N)\}^{-1}$$
$$\times \{\tilde{B}_{N+1}(y_{N+1} - y_N - \phi_{0N}(y_N)) - F_N'(y_N)\hat{\theta}(N-1)\}, \qquad \hat{\theta}(0) = \Gamma(0)s(0)$$

(5.20b) $$\Gamma(N) = \Gamma(N-1) - \Gamma(N-1)F_N(y_N)\{ I + F_N'(y_N)\,\Gamma(N-1)F_N(y_N)\}^{-1}$$
$$\times F_N'(y_N)\,\Gamma(N-1), \qquad \Gamma(0) = (Q(0) + \rho I)^{-1}.$$

Furthermore, replacing $\phi_k(y_k, \hat{\theta}(N))$ in (5.9) by $\phi_k(y_k, \hat{\theta}(k))$, the recursive estimate of σ^2 is given by

(5.20c) $$\hat{\sigma}_*^2(N) = \hat{\sigma}_*^2(N-1) + \frac{1}{N}\,\{\frac{1}{m}\,(y_{N+1} - y_N - \phi_N(y_N, \hat{\theta}(N)))'\tilde{B}_{N+1}^2$$
$$\times (y_{N+1} - y_N - \phi_N(y_N, \hat{\theta}(N))) - \hat{\sigma}_*^2(N-1)\}$$
$$\hat{\sigma}_*^2(0) = \frac{1}{m}\,(y_1 - y_0 - \phi_0(y_0, \hat{\theta}(0)))'\tilde{B}_1^2(y_1 - y_0 - \phi_0(y_0, \hat{\theta}(0))).$$

5.4 Asymptotic Properties of Estimators

5.4.1 Consistency of Estimators

First, define the estimation error $\tilde{\theta}(N)$ by

(5.21) $\tilde{\theta}(N) \triangleq \mathring{\theta} - \hat{\theta}(N).$

Then, from (5.19), it follows that

(5.22) $\tilde{\theta}(N) = \mathring{\theta} - \Gamma(N)s(N).$

Since, from (5.1), (5.4), (5.14), (5.16) and (5.18), s(N) defined by (5.15) is represented by

(5.23)
$$\begin{aligned} s(N) &= \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}\{ \sum_{i=1}^{M} \mathring{\theta}_i\phi_{ik}(y_k) + B_{k+1}v_{k+1}\} \\ &= \sum_{k=0}^{N} F_k(y_k)F_k'(y_k)\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} \\ &= Q(N)\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{K+1} \\ &= (\Gamma^{-1}(N) - \rho I)\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}, \end{aligned}$$

the estimation error $\tilde{\theta}(N)$ can be rewritten by

(5.24) $\tilde{\theta}(N) = \rho\,\Gamma(N)\mathring{\theta} - \Gamma(N) \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}.$

In the followings, we prove that $\tilde{\theta}(N)$ converges to zero with probability one in two cases.

(A) Single-Parameter Case. In the case when $\mathring{\theta}$ is scalar, $\phi_k(y_k, \mathring{\theta})$, $F_k(y_k)$ and $\Gamma(N)$ defined respectively by (5.4), (5.14) and (5.18) can be given by

(5.25) $\phi_k(y_k, \mathring{\theta}) = \phi_{0k}(y_k) + \mathring{\theta}\phi_{1k}(y_k)$ (n-dim. column vector)

(5.26) $F_k(y_k) = \phi_{1k}'(y_k)\tilde{B}_{k+1}$ (n-dim. row vector)

(5.27) $\Gamma(N) = \{ \sum_{k=0}^{N} \phi'_{1k}(y_k)\tilde{B}_{k+1}^{2}\phi_{1k}(y_k) + \rho\}^{-1}$ (scalar).

Then the following theorem is obtained.

[Theorem 5.1] Assume that the condition (C-1) and the following condition hold:

(C-2) the quantity defined by (5.27) converges to zero w.p. 1, i.e.,

(5.28) $\Gamma(N) \to 0$ w.p. 1 as $N \to \infty$

Then,

(5.29) $\hat{\theta}(N) \to \mathring{\theta}$ w.p. 1 as $N \to \infty$.

In order to prove the theorem, we need the following lemmas.

[Lemma 5.1] (Toeplitz Lemma) Let $\{a_{nk}\}$ $(k=1, 2, \cdots, n)$ be a sequence of numbers such that, for every fixed k,

(5.30) $a_{nk} \to 0$ as $n \to \infty$

and for all n,

(5.31) $\sum_{k=1}^{n} | a_{nk} | \leq c < +\infty$;

let $\{\tilde{x}_n\}$ be the sequence as

(5.32) $\tilde{x}_n = \sum_{k=1}^{n} a_{nk}x_k$.

Then, if $x_k \to 0$ as $k \to \infty$,

(5.33) $\tilde{x}_n \to 0$ as $n \to \infty$.

The proof of Lemma 5.1 is given in ref. [L3] and hence omitted here.

[Lemma 5.2] Let the conditions (C-1) and (C-2) hold. Then, the following scalar quantity h(N) defined by

$$(5.34) \qquad h(N) = \begin{cases} \sum\limits_{k=0}^{N} \Gamma(k)F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} & \text{for } N = 0, 1, \cdots \\ 0 & \text{for } N = -1, -2, \cdots \end{cases}$$

converges with probability one to a certain random variable h with the property $E\{h^2\} < +\infty$.

The proof of Lemma 5.2 is given in Appendix 5.A.

(Proof of Theorem 5.1) Using the random variable h whose existence is guaranteed by Lemma 5.2, the second term in the R.H.S. of (5.24) may be rewritten as

$$\begin{aligned}(5.35) \qquad \Gamma(N) \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} &= \sum_{k=0}^{N} \Gamma(N)\,\Gamma^{-1}(k)(h(k) - h(k-1)) \\ &= \sum_{k=0}^{N} \Gamma(N)\,\Gamma^{-1}(k)\{(h(k) - h) - (h(k-1) - h)\} \\ &= \sum_{k=0}^{N} \Gamma(N)\,\Gamma^{-1}(k)(h(k) - h) - \sum_{k=1}^{N} \Gamma(N)\,\Gamma^{-1}(k)(h(k-1) - h) \\ &\qquad + \Gamma(N)\,\Gamma^{-1}(0)h \\ &= \sum_{k=0}^{N} \Gamma(N)\,\Gamma^{-1}(k)(h(k) - h) - \sum_{k=1}^{N} \Gamma(N)(\,\Gamma^{-1}(k-1) \\ &\qquad + F_k(y_k)F_k'(y_k))(h(k-1) - h) + \Gamma(N)\,\Gamma^{-1}(0)h \\ &= \sum_{k=1}^{N} \Gamma(N)F_k(y_k)F_k'(y_k)(h - h(k-1)) + (h(N) - h) \\ &\qquad + \Gamma(N)\,\Gamma^{-1}(0)h.\end{aligned}$$

Hence, substituting (5.35) into (5.24), it follows that

$$(5.36)\quad \tilde{\theta}_N = \rho\,\Gamma(N)\mathring{\theta} - \sum_{k=1}^{N} \Gamma(N)F_k(y_k)F_k'(y_k)(h - h(k-1))$$

$$- (h(N) - h) - \Gamma(N)\Gamma^{-1}(0)h.$$

In the sequel, let us examine each term of (5.36). First we shall show the second term of (5.36) converges to zero. For this purpose, define p(N) by

$$(5.37)\quad p(N) = \sum_{k=1}^{N} \Gamma(N)F_k(y_k)F_k'(y_k)(h - h(k-1)).$$

We have from (5.26) and (5.27) that

$$(5.38)\quad \sum_{k=1}^{N} |\,\Gamma(N)F_k(y_k)F_k'(y_k)\,| = \sum_{\ell=1}^{N} \frac{F_\ell(y_\ell)F_\ell'(y_\ell)}{\sum_{k=0}^{N} F_k(y_k)F_k'(y_k) + \rho} < 1 \qquad \text{w.p. } 1$$

and also

$$(5.39)\quad \Gamma(N)F_k(y_k)F_k'(y_k) \to 0 \qquad \text{w.p. } 1 \quad \text{as} \quad N \to \infty.$$

if $\Gamma(N) \to 0$ w.p. 1 as $N \to \infty$, which is guaranteed by (C-2). Hence identifying $\Gamma(N)F_k(y_k)F_k'(y_k)$, $h-h(k-1)$, $p(N)$ and N with a_{nk}, x_k, $\tilde{x}_n$ and n respectively in Lemma 5.1, we may conclude from Lemmas 5.1 and 5.2 that

$$(5.40)\quad p(N) \to 0 \qquad \text{w.p. } 1 \quad \text{as} \quad N \to \infty.$$

It is also clear that the first and fourth terms in (5.36) converge to zero and from Lemma 5.2 the third term in (5.36) tends to zero. Hence, $\tilde{\theta}(N)$ converges to zero with probability one as $N \to \infty$.

(B) Multi-parameter Case. Define first

(5.41) $\tilde{\phi}_k(z,\theta) \triangleq z + \phi_k(z,\theta)$.

Then, following conditions are sufficient to prove the consistency of the estimator $\hat{\theta}(N)$ in the multi-parameter case:

(C-3) $\tilde{\phi}_k(z,\mathring{\theta})$ satisfies

(5.42) $\| \tilde{\phi}_k(z,\mathring{\theta}) \| \leq c_1 \| z \| + c \quad (c_1 < 1)$ for all $z \in R^n$

(C-4) The functions $\{\phi_{ik}(\cdot);i=0,1,2,\cdots,M\}$ satisfy

(5.43) $\| \phi_{ik}(z,\mathring{\theta}) \| \leq c(\| z \| + 1)$ for all $z \in R^n$

and

(C-5) $\liminf_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\} > cI > 0$ w.p. 1.

[Theorem 5.2] Let the conditions (C-1), (C-3), (C-4) and (C-5) hold. Then

(5.44) $\hat{\theta}(N) \to \mathring{\theta}$ w.p. 1 as $N \to \infty$.

In order to prove Theorem 5.2, we need the following lemma.

[Lemma 5.3] Assume that the same conditions as in Theorem 5.2 hold. Then

(5.45) $E\{ \| y_k \|^4\} \leq c$ for $k=1,2,\cdots$.

The proof of Lemma 5.3 is given in Appendix 5.B.

(Proof of Theorem 5.2) Define a random variable q(N) by

$$(5.46)\qquad q(N) = \sum_{k=0}^{N} x'F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1},$$

where x is an arbitrary vector such as $x \neq 0$. Using the martingale convergence of Theorem 2.2, we shall show that q(N) converges to zero with probability one as $N \to \infty$. First, since

$$(5.47)\qquad E\{q(N) \mid \mathcal{F}_N\} = E\{ \sum_{k=0}^{N-1} x'F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} + x'F_N(y_N)\tilde{B}_{N+1}B_{N+1}v_{N+1} \mid \mathcal{F}_N\}$$

$$= \sum_{k=0}^{N-1} x'F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}$$

$$= q(N-1),$$

$\{q(N), \mathcal{F}_N\}$ is a martingale. For the second moment of q(N), the following evaluation is obtained:

$$(5.48)\qquad E\{q^2(N)\} = x'E\{ \sum_{i=0}^{N} \sum_{j=0}^{N} F_i(y_i)\tilde{B}_{i+1}B_{i+1}v_{i+1}v'_{j+1}B'_{j+1}\tilde{B}_{j+1}F'_j(y_j)\}x$$

$$= x'(\sum_{i=0}^{N} E\{F_i(y_i)\tilde{B}_{i+1}B_{i+1}v_{i+1}v'_{i+1}B'_{i+1}\tilde{B}_{i+1}F'_i(y_i)\})x$$

$$+ 2x'(E\{ \sum_{i>j}^{N} F_i(y_i)\tilde{B}_{i+1}B_{i+1}v_{i+1}v'_{j+1}B'_{j+1}\tilde{B}_{j+1}F'_j(y_j)\})x$$

$$= \sigma^2 x'(\sum_{i=0}^{N} E\{F_i(y_i)\tilde{B}_{i+1}B_{i+1}B'_{i+1}\tilde{B}_{i+1}F'_i(y_i)\})x$$

$$\leq \sigma^2 \| x \|^2 \sum_{i=0}^{N} E\{ \| F_i(y_i) \|^2\}.$$

For the (ℓ, m)-th component $\phi_{\ell i}(y_i)\tilde{B}_{i+1}^2\ \phi'_{mi}(y_i)$ of the matrix $F(y_i)F'(y_i)$, we have

$$(5.49)\quad E\{\,|\,\phi_{\ell i}(y_i)\tilde{B}_{i+1}^2\ \phi'_{mi}(y_i)\,|\,\}$$

$$\leq \|\tilde{B}_{i+1}\|^2 E\{\,\|\phi_{\ell i}(y_i)\|\ \|\phi_{mi}(y_i)\|\,\}$$

$$\leq \|\tilde{B}_{i+1}\|^2 [E\{\,\|\phi_{\ell i}(y_i)\|^2\}E\{\,\|\phi_{mi}(y_i)\|^2\}]^{1/2},$$

where Schwarz inequality has been used.

Since from the condition (C-4) and Lemma 5.3, it follows that

$$(5.50)\quad E\{\,\|\phi_{\ell i}(y_i)\|^2\} \leq c^2 E\{(\,\|y_i\| + 1)^2\}$$

$$\leq 2c^2(E\{\,\|y_i\|^2\} + 1)$$

$$\leq 2c^2(E\{\,\|y_i\|^4\}^{1/2} + 1)$$

$$\leq 2c^2(c^{1/2} + 1) < +\infty,$$

and this implies that

$$(5.51)\quad E\{\,\|F_i(y_i)\|^2\} < c.$$

Substituting (5.51) into (5.48), we have

$$(5.52)\quad E\{q^2(N)\} \leq cN.$$

Hence identifying $q(N)$ with x_k in Theorem 2.2, we can see that $\{q(N), \mathscr{F}_N\}$ is the martingale which satisfies the conditions in Theorem 2.2, and then it follows that

$$(5.53)\quad \frac{1}{N}\, q(N) \to 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty,$$

or equivalently

$$(5.54)\quad \frac{1}{N} \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} \to 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty.$$

Next, we shall show that r(N) defined by

$$(5.55)\qquad r(N) \triangleq \sum_{k=0}^{N} x'[F_k(y_k)F_k'(y_k) - E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}]x$$

converges to zero with probability one. The procedure to perform this is almost same as that of (5.54). First, since

$$(5.56)\ E\{r(N) \mid \mathcal{F}_{N-1}\} = E\{\sum_{k=0}^{N} x'[F_k(y_k)F_k'(y_k) - E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}]x \mid \mathcal{F}_{N-1}\}$$

$$= \sum_{k=0}^{N-1} x'[F_k(y_k)F_k'(y_k) - E\{F_k(y_k) \mid \mathcal{F}_{k-1}\}]x$$

$$= r(N-1),$$

it can be concluded that $\{r(N), \mathcal{F}_{N-1}\}$ is a martingale. For the second moment $r^2(N)$ is evaluated as follows:

$$(5.57)\qquad E\{r^2(N)\} = E\{(\sum_{i=0}^{N} x'[F_i(y_i)F_i'(y_i) - E\{F_i(y_i)F_i'(y_i) \mid \mathcal{F}_{i-1}\}]x)^2\}$$

$$= \sum_{i=0}^{N} E\{(x'[F_i(y_i)F_i'(y_i) - E\{F_i(y_i)F_i'(y_i) \mid \mathcal{F}_{i-1}\}]x)^2\}$$

$$+ 2\sum_{i>j}^{N} E\{x'[F_i(y_i)F_i'(y_i) - E\{F_i(y_i)F_i'(y_i) \mid \mathcal{F}_{i-1}\}]x$$

$$\times x'[F_j(y_j)F_j'(y_j) - E\{F_j(y_j)F_j'(y_j) \mid \mathcal{F}_{j-1}\}]x\}$$

$$= \sum_{i=0}^{N} E\{(x'[F_i(y_i)F_i'(y_i) - E\{F_i(y_i)F_i'(y_i) \mid \mathcal{F}_{i-1}\}]x)^2\}$$

$$\leq 2\sum_{i=0}^{N} (E\{(x'F_i(y_i)F_i'(y_i)x)^2\}$$

$$+ E\{(x'E\{F_i(y_i)F_i'(y_i) \mid \mathcal{F}_{i-1}\}x)^2\})$$

$$\leq 2\|x\|^4\sum_{i=0}^{N} (E\{\|F_i(y_i)\|^4\} + E\{(E\{\|F_i(y_i)\|^2 \mid \mathcal{F}_{i-1}\})^2\})$$

$$\leq 4\|x\|^4 \sum_{i=0}^{N} E\{\|F_i(y_i)\|^4\}$$

where Jensen's inequality has been used in the derivation of the last inequality. By using the same procedure as that in (5.51), we have from the condition (C-4) and Lemma 5.3

$$(5.58)\quad E\{\| F_i(y_i) \|^4\} < c.$$

Consequently, it follows from (5.57) and (5.58) that

$$(5.59)\quad E\{r^2(N)\} \leq cN,$$

and by using Theorem 2.2, we can conclude that

$$(5.60)\quad \frac{r(N)}{N} \to 0 \qquad \text{w.p. } 1 \qquad \text{as } N \to \infty$$

or equivalently

$$(5.61)\quad \frac{1}{N} x'\{ \sum_{k=0}^{N} F_k(y_k)F_k'(y_k) - E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}\}x \to 0 \qquad \text{w.p. } 1 \quad \text{as } N \to \infty.$$

Recalling the elementary inequality

$$(5.62)\quad \inf_{n>N} a_n - \inf_{n>N} b_n \geq \inf_{n>N} (a_n - b_n),$$

we have from (5.61) that

$$(5.63)\quad \lim_{N\to\infty} (\inf_{n>N} \frac{1}{n} \sum_{k=0}^{n} x'F_k(y_k)F_k'(y_k)x - \inf_{n>N} \frac{1}{n} \sum_{k=0}^{n} x'E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}x)$$

$$\geq \lim_{N\to\infty} \inf_{n>N} (\frac{1}{n} \sum_{k=0}^{n} x'F_k(y_k)F_k'(y_k)x - \frac{1}{n} \sum_{k=0}^{n} x'E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}x)$$

$$= \lim_{N\to\infty} \inf(\frac{1}{N}\sum_{k=0}^{N} x'F_k(y_k)F_k'(y_k)x - \frac{1}{N}\sum_{k=0}^{N} x'E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}x)$$

$$= 0,$$

where the last equality came from (5.61). Hence from the condition (C-5), it follows that

$$(5.64)\quad \liminf_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} x'F_k(y_k)F_k'(y_k)x$$

$$\geq \liminf_{N\to\infty} \frac{1}{N}\sum_{k=0}^{N} x'E\{F_k(y_k)f_k'(y_k) \mid \mathscr{F}_{k-1}\}x \geq cI \quad \text{w.p. } 1.$$

Hence $\Gamma(N)$ defined by (5.18) can be evaluated as

$$(5.65)\quad \liminf_{N\to\infty} \frac{1}{N}(\Gamma^{-1}(N) - \rho I) > cI \qquad \text{w.p. } 1,$$

which implies

$$(5.66)\quad \limsup_{N\to\infty} N\Gamma(N) < cI < +\infty \qquad \text{w.p. } 1.$$

For the first and second terms of (5.24), they are evaluated from (5.54) and (5.66) as

$$(5.67)\quad \rho\Gamma(N)\mathring{\theta} = \frac{1}{N}\rho(N\Gamma(N))\mathring{\theta} \to 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty,$$

and

$$(5.68)\quad \Gamma(N)\sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1} = N\Gamma(N)\frac{1}{N}\sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}$$

$$\to 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty,$$

which complete the proof.

We note that if the nonlinear function $\tilde{\phi}_k(z,\mathring{\theta})$ can be represented by

$$(5.69)\quad \tilde{\phi}_k(z,\mathring{\theta}) = A_k(z,\mathring{\theta})z,$$

the condition (C-3) is replaced by the following condition:

(C-3)' For any $z \in R^n$,

$$(5.70)\quad \max_i |\lambda_i\{A(z,\mathring{\theta})\}| < 1,$$

where $\lambda_i\{*\}$ denotes the eigenvalue of $*$.

For the consistency of the estimator $\hat{\sigma}^2(N)$ of σ^2, the following theorem is obtained.

[Theorem 5.3] Assume that same conditions as those of Theorem 5.2 hold. Then,

$$(5.71) \quad \hat{\sigma}^2(N) \rightarrow \sigma^2 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty .$$

(Proof of Theorem 5.3) Define

$$(5.72) \quad \begin{cases} \xi_1(k) \triangleq \frac{1}{m} v_k v_k' \\ \xi_2(k) \triangleq \frac{1}{m} \tilde{\theta}'(N) F_k(y_k) \tilde{B}_k B_k v_k \\ \xi_3(k) \triangleq \frac{1}{m} \tilde{\theta}'(N) F_k(y_k) F_k'(y_k) \tilde{\theta}(N), \end{cases}$$

then $\hat{\sigma}^2(N)$ defined by (5.9) can be rewritten by

$$(5.73) \quad \hat{\sigma}^2(N) = \frac{1}{N} \sum_{k=0}^{N} \xi_1(k) + \frac{1}{N} \sum_{k=0}^{N} \xi_2(k) + \sum_{k=0}^{N} \xi_3(k).$$

It is easily shown by using the same procedure as that for the convergence of q(N) defined by (5.46) that

$$(5.74) \quad \frac{1}{N} \sum_{k=0}^{N} \xi_1(k) \rightarrow \sigma^2 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty$$

$$(5.75) \quad \frac{1}{N} \sum_{k=0}^{N} \xi_2(k) \rightarrow 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty .$$

For the last term of (5.73), from (5.18), (5.19) and (5.24), we have

$$(5.76) \quad \frac{1}{N} \sum_{k=0}^{N} \xi_3(k) = \frac{1}{m} [\eta'(N) N \Gamma(N) \eta(N) - \frac{2}{N} \rho \mathring{\theta} N \Gamma(N) \eta(N)$$
$$- \frac{1}{N} \rho \eta'(N) (N \Gamma(N))^2 \eta(N) + \frac{2}{N^2} \rho^2 \mathring{\theta}' (N \Gamma(N))^2 \eta(N)$$
$$- \frac{1}{N^3} \rho^3 \mathring{\theta}' (N \Gamma(N))^2 \mathring{\theta} + \frac{1}{N^2} \rho^2 \mathring{\theta}' N \Gamma(N) \mathring{\theta}]$$

where

$$\eta(N) = \frac{1}{N} \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}.$$

Hence from (5.54) and (5.66), it follows that

$$\frac{1}{N} \sum_{k=0}^{N} \xi_3(k) \rightarrow 0 \qquad \text{w.p. } 1 \qquad \text{as } N \rightarrow \infty,$$

and with (5.74) and (5.75), we can conclude (5.71).

5.4.2 Asymptotic Normality of Estimators

In this section, in order to evaluate the asymptotic accuracy of estimators, we shall show the asymptotic normality of the estimator $\hat{\theta}(N)$. For this purpose, we define first that

$$(5.77) \quad P(N) \triangleq \frac{1}{N} \Gamma^{-1}(N)$$

$$(5.78) \quad U(N) \triangleq \frac{\sigma^2}{N} \sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k)\}.$$

Then, the following theorem is obtained.

[Theorem 5.4] Assume that the conditions (C-1) and (C-5) hold. Furthermore assume that the following conditions hold:

(C-3)'' $\tilde{\phi}_k(z_1, \mathring{\theta})$ satisfies the uniform Lipschtz condition and it is uniformly bounded, i.e., for any z_1 and $z_2 \in R^n$,

$$(5.79) \quad \begin{cases} \| \tilde{\phi}_k(z_1, \mathring{\theta}) - \tilde{\phi}_k(z_2, \mathring{\theta}) \| \leq c_1 \| z_1 - z_2 \| \quad (c_1 < 1) \\ \| \tilde{\phi}_k(0\ , \mathring{\theta}) \| \leq c; \end{cases}$$

(C-4)' The functions $\{\tilde{\phi}_{ik}(\cdot) ; i=0,1,2,\cdots,M\}$ satisfy the uniform Lipschitz condition and they are unformly bounded, i.e., for any z_1 and $z_2 \in R^n$,

$$(5.80) \quad \begin{cases} \| \phi_{ik}(z_1) - \phi_{ik}(z_2) \| \leq c_1 \| z_1 - z_2 \| \quad (c_1 < 1) \\ \| \phi_{ik}(0) \| \leq c \text{ for } i=0,1,\cdots,M. \end{cases}$$

Then,

$$(5.81) \quad \sqrt{N} U^{-1/2}(N) P(N)(\mathring{\theta} - \hat{\theta}(N)) \overset{law}{\rightarrow} z \qquad \text{as } N \rightarrow \infty,$$

where

$$(5.82) \quad z \sim N(0, I).$$

In order to prove the above theorem, we need the following lemmas.

[Lemma 5.4](Anderson [A7]) Let

(5.83) $x_N = a_{\ell N} + b_{\ell N} \qquad \ell, N = 1, 2, \cdots$

such that

(5.84) $E\{b^2_{\ell N}\} \leq d_\ell , \quad \lim_{\ell \to \infty} d_\ell = 0$

(5.85) $P\{a_{\ell N} < a\} = D_{\ell N}(a)$

(5.86) $\lim_{N \to \infty} D_{\ell N}(a) = D_\ell(a)$

(5.87) $\lim_{\ell \to \infty} D_\ell(a) = D(a).$

Then

(5.88) $\lim_{N \to \infty} P\{x_N \leq a\} = D(a).$

[Lemma 5.5] (Knopp [K5]) If a sequence $\{a_k\}$ has a limit a, then

(5.89) $\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} a_k = a = \lim_{k \to \infty} a_k.$

[Lemma 5.6] Assume that the same conditions as those in Theorem 5.4 hold. Then

(5.90) $E\{ \| y_k - y_k(k-\ell) \|^2\} < c\, c_1^\ell \quad (c_1 < 1) \quad$ for $\ell = 0, 1, 2, \cdots,$

where $y_k(k-\ell)$ is the sequence which is generated by

(5.91)
$$\begin{cases} y_{i+1}(k-\ell) - y_i(k-\ell) = \phi_i(y_i(k-\ell), \hat{\theta}) + B_{i+1} v_{i+1} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } i = k-\ell, k-\ell+1, \cdots, k-1 \\ y_{k-\ell}(k-\ell) = 0. \end{cases}$$

(Proof of Lemma 5.6) From (5.1), (5.41) and (5.91), we have

(5.92) $\| y_k - y_k(k-\ell) \| = \| \tilde{\phi}_{k-1}(y_{k-1}, \hat{\theta}) - \tilde{\phi}_{k-1}(y_{k-1}(k-\ell), \hat{\theta}) \| .$

Hence, by using the condition (C-3)'', we can evaluate (5.91) as

$$(5.93)\quad \begin{aligned} \| y_k - y_k(k-\ell) \| &\leq c_1 \| y_{k-1} - y_{k-1}(k-\ell) \| \\ &\leq c_1^2 \| y_{k-2} - y_{k-2}(k-\ell) \| \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ &\leq c_1^{\ell} \| y_{k-\ell} - y_{k-\ell}(k-\ell) \| \\ &= c_1^{\ell} \| y_{k-\ell} \| . \end{aligned}$$

Therefore, by using Lemma 5.3, the following evaluation is given:

$$(5.94)\quad \begin{aligned} E\{ \| y_k - y_k(k-\ell) \|^2 \} &\leq c_1^{2\ell} E\{ \| y_{k-\ell} \|^2 \} \\ &\leq c_1^{2\ell} [E\{ \| y_{k-\ell} \|^4 \}]^{1/2} \leq c\, c_1^{\ell}, \end{aligned}$$

where c_1^2 is re-defined as c_1.

(Proof of Theorem 5.4) Define the following quantities:

$$(5.95)\quad Q(\ell,N) \triangleq \sum_{k=0}^{N} F_k(y_k(k-\ell)) F_k'(y_k(k-\ell))$$

$$(5.96)\quad \Gamma(\ell,N) \triangleq [Q(\ell,N) + \rho I]^{-1}$$

$$(5.97)\quad P(\ell,N) \triangleq \frac{1}{N} \Gamma^{-1}(\ell,N)$$

$$(5.98)\quad z(N) \triangleq \sqrt{N} U^{-1/2}(N) P(N) \tilde{\theta}(N)$$

$$(5.99)\quad \alpha(\ell,N) \triangleq \sqrt{N} U^{-1/2}(N) P(\ell,N) \bar{\theta}(\ell,N)$$

$$(5.100)\quad \beta(\ell,N) \triangleq \sqrt{N} U^{-1/2}(N) [P(N) \tilde{\theta}(N) - P(\ell,N) \bar{\theta}(\ell,N)]$$

where

$$(5.101)\quad \tilde{\theta}(\ell,N) \triangleq \mathring{\theta} - \Gamma(\ell,N) s(\ell,N)$$

$$(5.102)\quad \bar{\theta}(\ell,N) \triangleq \tilde{\theta}(\ell,N) - \rho\, \Gamma(\ell,N) \mathring{\theta} \qquad (N \geq \ell)$$

$$(5.103)\quad s(\ell,N) \triangleq \sum_{k=0}^{N} F_k(y_k(k-\ell)) \tilde{B}_{k+1} [y_{k+1}(k-\ell) - y_k(k-\ell) - \phi_{0k}(y_k(k-\ell))] \qquad (N \geq \ell).$$

By using (5.99) and (5.100), z(N) can be represented by

(5.104) $z(N) = \alpha(\ell,N) + \beta(\ell,N).$

Define further

(5.105) $R(\ell,N) \triangleq E\{\alpha(\ell,N)\alpha'(\ell,N)\},$

then it follows that

(5.106) $z(N) = R^{-1/2}(\ell,N)\alpha(\ell,N) + (I - R^{-1/2}(\ell,N))\alpha(\ell,N) + \beta(\ell,N).$

We apply Lemma 5.4 to prove the asymptotic normality of z(N). This will be done by showing that the first term of (5.106) is asymptotically normal according to Theorem 2.4 and two other terms of (5.106) have the vanishing second moments as ℓ tends to infinity. First, we shall show that the asymptotic normality of $R^{-1/2}(\ell,N)\alpha(\ell,N)$. From (5.91), (5.4) and (5.103), we have

(5.107)
$$\begin{aligned} s(\ell,N) &= \sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}\{\sum_{i=1}^{M} \mathring{\theta}_i\phi_{ik}(y_k(k-\ell)) + B_{k+1}v_{k+1}\} \\ &= \sum_{k=0}^{N} F_k(y_k(k-\ell))F_k'(y_k(k-\ell))\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1} \\ &= Q(\ell,N)\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1} \\ &= (\Gamma^{-1}(\ell,N) - \rho I)\mathring{\theta} + \sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1}. \end{aligned}$$

Substituting (5.107) and (5.101) into (5.102), it follows that

(5.108)
$$\begin{aligned} \bar{\theta}(\ell,N) &= \mathring{\theta} - \Gamma(\ell,N)s(\ell,N) - \rho\,\Gamma(\ell,N)\mathring{\theta} \\ &= \mathring{\theta} - \Gamma(\ell,N)\{(\Gamma^{-1}(\ell,N) - \rho I)\mathring{\theta} \\ &\quad + \sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1}\} - \rho\,\Gamma(\ell,N)\mathring{\theta} \\ &= -\Gamma(\ell,N)\sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1}. \end{aligned}$$

Hence, we have from (5.97) and (5.108) that

(5.109) $$\alpha(\ell,N) = -\frac{1}{\sqrt{N}} U^{-1/2}(N)\sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1}.$$

Define

(5.110) $$x(k,N) \triangleq -\frac{1}{\sqrt{N}} R^{-1/2}(\ell,N)U^{-1/2}(N)F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1},$$

then it follows that

(5.111) $$R^{-1/2}(\ell,N)\alpha(\ell,N) = \sum_{k=0}^{N} x(k,N).$$

For the quantity x(k,N) defined above, it is easily shown that

(5.112) $$\begin{aligned} E\{x(k,N)\} &= -\frac{1}{\sqrt{N}}R^{-1/2}(\ell,N)U^{-1/2}(N)E\{F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}E\{v_{k+1}|\mathcal{F}_k\}\} \\ &= 0 \end{aligned}$$

(5.113) $$\begin{aligned} &E\{(\sum_{k=0}^{N} x(k,N))(\sum_{k=0}^{N} x(k,N))'\} \\ &= R^{-1/2}(\ell,N)E\{\alpha(\ell,N)\alpha'(\ell,N)\}R^{-1/2}(\ell,N) \\ &= I, \end{aligned}$$

which means that

(5.114) $$\begin{cases} E\{ \dfrac{x'x(k,N)}{\|x\|} \} = 0 \\ E\{(\dfrac{x'(\sum_{k=0}^{N} x(k,N))}{\|x\|})^2\} = 1 \end{cases}$$

where x is an arbitrary vector such that $x \neq 0$. It is easily verified that

$$(5.115)\quad E\{(\frac{x'x(k,N)}{\| x \|})^2\}$$

$$= \frac{1}{\| x \|^2}\cdot\frac{1}{N}\ x'R^{-1/2}(\ell,N)U^{-1/2}(N)E\{F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}$$

$$\times E\{v_{k+1}v_{k+1}' \mid \mathscr{F}_k\}B_{k+1}'\tilde{B}_{k+1}F_k'(y_k(k-\ell))\}U^{-1/2}(N)R^{-1/2}(\ell,N)x$$

$$= \frac{1}{\| x \|^2}\cdot\frac{\sigma^2}{N}\ x'R^{-1/2}(\ell,N)U^{-1/2}(N)E\{F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}$$

$$\times B_{k+1}'\tilde{B}_{k+1}F_k'(y_k(k-\ell))\}U^{-1/2}(N)R^{-1/2}(\ell,N)x$$

$$= \frac{1}{\| x \|^2}\cdot\frac{\sigma^2}{N}\ x'R^{-1/2}(\ell,N)U^{-1/2}(N)E\{F_k(y_k(k-\ell))F_k'(y_k(k-\ell))\}$$

$$\times U^{-1/2}(N)R^{-1/2}(\ell,N)x.$$

Therefore, we have

$$(5.116)\quad \limsup_{N\to\infty}\sum_{k=0}^{N}E\{(\frac{x'x(k,N)}{\| x \|})^2\}$$

$$= \limsup_{N\to\infty}\frac{1}{\| x \|^2}\cdot\frac{\sigma^2}{N}\ x'R^{-1/2}(\ell,N)U^{-1/2}(N)$$

$$\times[\sum_{k=0}^{N}E\{F_k(y_k(k-\ell))F_k'(y_k(k-\ell))\}]U^{-1/2}(N)R^{-1/2}(\ell,N)x$$

$$\le \limsup_{N\to\infty}\frac{1}{\| x \|^2}\ \sigma^2E\{\| F_k(y_k(k-\ell))\|^2\}x'R^{-1/2}(\ell,N)$$

$$\times U^{-1}(N)R^{-1/2}(\ell,N)x$$

$$\le \frac{c\,\sigma^2}{\| x \|^2}\limsup_{N\to\infty}x'R^{-1/2}(\ell,N)U^{-1}(N)R^{-1/2}(\ell,N)x$$

where (5.51) has been used. From Fatou-Lebesgue Theorem (see e.g. [L3]) and the condition (C-5), the quantity U(N) defined by (5.78) is evaluated as

(5.117) $$\liminf_{N\to\infty} U(N) = \liminf_{N\to\infty} \frac{\sigma^2}{N}\sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k)\}$$

$$= \liminf_{N\to\infty} E\{\frac{\sigma^2}{N}\sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}\}$$

$$\geq E\{\liminf_{N\to\infty} \frac{\sigma^2}{N}\sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k) \mid \mathcal{F}_{k-1}\}\}$$

$$\geq c\,\sigma^2 I$$

or, equivalently, for the sufficiently large N

(5.118) $$U^{-1}(N) \leq cI.$$

Simular procedure yields that for the sufficiently large N,

(5.119) $$R^{-1/2}(\ell,N) \leq cI.$$

Therefore, from (5.116), (5.118) and (5.119), the following evaluation is obtained:

(5.120) $$\limsup_{N\to\infty} \sum_{k=0}^{N} E\{(\frac{x'x(k,N)}{\| x \|})^2\} < +\infty.$$

We have for $0 \leq k \leq N$ that

(5.121) $$\{\omega; \frac{x'x(k,N)}{\| x \|} > \varepsilon\}$$

$$= \{\omega; \frac{| x'R^{-1/2}(\ell,N)U^{-1/2}(N)F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1} |}{\sqrt{N}\,\| x \|} > \varepsilon\}$$

$$\subseteq \{\omega; \frac{| x'R^{-1/2}(\ell,N)U^{-1/2}(N)F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1} |}{\| x \|} > \sqrt{k}\,\varepsilon\}.$$

Then by defining

(5.122) $$\gamma(k,N) \triangleq \frac{1}{\| x \|}\, x'R^{-1/2}(\ell,N)U^{-1/2}(N)F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1},$$

it follows that

$$(5.123)\quad \sum_{k=0}^{N} E\{(\frac{x'x(k,N)}{\| x \|})^2 I(\left| \frac{x'x(k,N)}{\| x \|} \right| > \varepsilon)\}$$

$$\leq \frac{1}{N} \sum_{k=0}^{N} E\{ \gamma^2(k,N) I(| \gamma(k,N) | > \varepsilon \sqrt{k})\}.$$

Furthermore, it is easily shown from (5.51), (5.118) and (5.119) that

$$(5.124)\quad E\{\gamma^2(k,N)\} = \frac{\sigma^2}{\| x \|} x' R^{-1/2}(\ell,N) U^{-1/2}(N) E\{F_k(y_k(k-\ell))F_k'(y_k(k-\ell))\}$$

$$\times U^{-1/2}(N) R^{-1/2}(\ell,N) x \leq c.$$

Hence, we have

$$(5.125)\quad E\{\gamma^2(k,N) I(| \gamma(k,N) | > \varepsilon \sqrt{k})\} \to 0 \quad \text{as} \quad k \to \infty$$

and, by identifying $\gamma^2(k,N)$ with a_k in Lemma 5.5, it follows that

$$(5.126)\quad \sum_{k=0}^{N} E\{(\frac{x'x(k,N)}{\| x \|})^2 I(\left| \frac{x'x(k,N)}{\| x \|} \right| > \varepsilon)\} \to 0 \quad \text{as} \quad N \to \infty.$$

Since it is easily verified that $\{x(1,N),\cdots,x(s,N)\}$ and $\{x(k,N),\cdots,x(p,N)\}$ are independent for $| k-s | \geq \ell$, we see from (5.114), (5.120) and (5.126) that all conditions of Theorem 2.4 are satisfied by regarding $x'R^{-1/2}(\ell,N)\alpha(\ell,N)/\| x \|$ and $x'x(k,N)/\| x \|$ as x_N and x_{kN} respectively. Therefore, we can conclude that

$$(5.127)\quad R^{-1/2}(\ell,N)\alpha(\ell,N) \xrightarrow{law} z \quad (z \sim N(0,1)) \quad \text{as} \quad N \to \infty,$$

Next, we shall show that second moments of the last two terms of (5.106) converge to zero. For the second term of (5.106), we find that

$$(5.128)\quad E\{x'(I - R^{-1/2}(\ell,N))\alpha(\ell,N)\alpha'(\ell,N)(I - R^{-1/2}(\ell,N))x\}$$

$$= x'(I - R^{-1/2}(\ell,N))E\{\alpha(\ell,N)\alpha'(\ell,N)\}(I - R^{-1/2}(\ell,N))x$$

$$= x'(I - R^{-1/2}(\ell,N))R(\ell,N)(I - R^{-1/2}(\ell,N))x$$

$$= x'(I - R^{1/2}(\ell,N))^2 x$$

Let λ_i be the i-th eigenvalue of $R^{1/2}(\ell,N)$, then λ_i is non-negative since $R^{1/2}(\ell,N)$ is a semi-positive definite matrix. Hence, the i-th eigenvalue $\tilde{\lambda}_i$ of

(5.129) $$(I - R(\ell,N))^2 - (I - R^{1/2}(\ell,N))^2 = R^2(\ell,N) - 3R(\ell,N) + 2R^{1/2}(\ell,N)$$

can be represented by

(5.130) $$\tilde{\lambda}_i = \lambda_i^4 - 3\lambda_i^2 + 2\lambda_i$$

$$= \lambda_i(\lambda_i - 1)^2(\lambda_i + 2) \geq 0 \qquad \text{for } i=1,2,\cdots,M,$$

and this implies that

(5.131) $$(I - R(\ell,N))^2 \geq (I - R^{1/2}(\ell,N))^2.$$

Then, the following evaluation holds:

(5.132) $$E\{x'(I - R^{-1/2}(\ell,N))\alpha(\ell,N)\alpha'(\ell,N)(I - R^{-1/2}(\ell,N))x\}$$

$$\leq x'(I - R(\ell,N))^2 x$$

$$\leq \| I - R(\ell,N) \|^2 \| x \|^2.$$

Using (5.78), (5.105) and (5.109), we have

(5.133) $$\| U^{1/2}(N)(I - R(\ell,N))U^{1/2}(N) \|$$

$$= \| U(N) - \frac{1}{N} E\{(\sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1})$$

$$\times (\sum_{k=0}^{N} F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}v_{k+1})'\} \|$$

$$= \| U(N) - \frac{\sigma^2}{N} \sum_{k=0}^{N} E\{F_k(y_k(k-\ell))\tilde{B}_{k+1}B_{k+1}B'_{k+1}\tilde{B}_{k+1}F'_k(y_k(k-\ell))\} \|$$

$$= \| U(N) - \frac{\sigma^2}{N} \sum_{k=0}^{N} E\{F_k(y_k(k-\ell))F'_k(y_k(k-\ell))\} \|$$

$$\leq \frac{\sigma^2}{N} \sum_{k=0}^{N} E\{ \| F_k(y_k)F'_k(y_k) - F_k(y_k(k-\ell))F'_k(y_k(k-\ell)) \| \}.$$

Furthermore, we obtain

$$(5.134)\quad \| F_k(y_k)F_k'(y_k) - F_k(y_k(k-\ell))F_k'(y_k(k-\ell)) \|$$

$$\leq E\{ \| F_k(y_k) \| \; \| F_k(y_k) - F_k(y_k(k-\ell)) \| \}$$

$$+ E\{ \| F_k(y_k(k-\ell)) \| \; \| F_k(y_k) - F_k(y_k(k-\ell)) \| \}$$

$$\leq [E\{ \| F_k(y_k) \|^2\}]^{1/2}[E\{ \| F_k(y_k) - F_k(y_k(k-\ell)) \|^2\}]^{1/2}$$

$$+ [E\{ \| F_k(y_k(k-\ell)) \|^2\}]^{1/2}[E\{ \| F_k(y_k) - F_k(y_k(k-\ell)) \|^2\}]^{1/2},$$

and, from (C-1), (C-4) and Lemma 5.3, it is easily verified that

$$(5.135)\quad \begin{cases} E\{ \| F_k(y_k) \|^2\} \leq c \\ E\{ \| F_k(y_k(k-\ell)) \|^2\} \leq c, \end{cases}$$

and, from the condition (C-4) and Lemma 5.6, we have

$$(5.136)\quad E\{ \| F_k(y_k) - F_k(y_k(k-\ell)) \|^2\} \leq c\, c_1^\ell \qquad (c_1 < 1).$$

Then, it follows that

$$(5.137)\quad E\{ \| F_k(y_k)F_k'(y_k) - F_k(y_k(k-\ell))F_k'(y_k(k-\ell)) \| \} < c\, c_1^\ell$$

and the following evaluation for (5.133) is obtained:

$$(5.138)\quad \| U^{1/2}(N)(I - R(\ell,N))U^{1/2}(N) \| \leq \sigma^2 c\, c_1^\ell \to 0 \qquad \text{as} \quad \ell \to \infty.$$

Since, from (5.118), U(N) is evaluated for a sufficiently large N as

$$(5.139)\quad U^{-1/2}(N) \leq cI,$$

it can be concluded that

$$(5.140)\quad \| I - R(\ell,N) \| = \| U^{-1/2}(N)U^{1/2}(N)(I - R(\ell,N))U^{1/2}(N)U^{-1/2}(N) \|$$

$$\leq \| U^{-1/2}(N) \|^2 \| U^{1/2}(N)(I - R(\ell,N))U^{1/2}(N) \|$$

$$\to 0 \qquad \text{as} \quad \ell \to \infty,$$

which implies that

$$(5.141)\quad E\{(I - R^{-1/2}(\ell,N))\alpha(\ell,N)\alpha'(\ell,N)(I - R^{-1/2}(\ell,N))\} \rightarrow 0 \quad \text{as } \ell \rightarrow \infty.$$

Finally, we shall show that the third term of (5.106) converges to zero as ℓ tends to infinity. From (5.24) and (5.77), it follows that

$$(5.142)\quad \sqrt{N}U^{-1/2}(N)P(N)\tilde{\theta}(N)$$

$$= \frac{1}{\sqrt{N}} U^{-1/2}(N)(\rho\mathring{\theta} - \sum_{k=0}^{N} F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}).$$

Then $\beta(\ell,N)$ defined by (5.100) can be represented by

$$(5.143)\quad \beta(\ell,N) = \frac{1}{\sqrt{N}} U^{-1/2}(N)(\rho\mathring{\theta} - \sum_{k=0}^{N}\{F_k(y_k) - F_k(y_k(k-\ell))\}\tilde{B}_{k+1}B_{k+1}v_{k+1}),$$

where (5.99) and (5.109) have been used. Then, we have

$$(5.144)\quad E\{x'\beta(\ell,N)\beta'(\ell,N)x\}$$

$$= \frac{1}{N} x'U^{-1/2}(N)E\{[\rho\mathring{\theta} - \sum_{k=0}^{N}\{F_k(y_k) - F_k(y_k(k-\ell))\}\tilde{B}_{k+1}B_{k+1}v_{k+1}]$$

$$\times [\rho\mathring{\theta} - \sum_{k=0}^{N}\{F_k(y_k) - F_k(y_k(k-\ell))\}\tilde{B}_{k+1}B_{k+1}v_{k+1}]'\}U^{-1/2}(N)x$$

$$= \frac{\rho^2}{N} x'U^{-1/2}(N)\mathring{\theta}\mathring{\theta}'U^{-1/2}(N)x$$

$$+ \frac{\sigma^2}{N} x'U^{-1/2}(N)\sum_{k=0}^{N} E\{[F_k(y_k) - F_k(y_k(k-\ell))][F_k(y_k) - F_k(y_k(k-\ell))]'\}$$

$$\times U^{-1/2}(N)x$$

$$\leq \frac{\rho^2}{\ell} \| U^{-1/2}(N) \|^2 \| \mathring{\theta} \|^2 \| x \|^2$$

$$+ \frac{\sigma^2}{N} \| U^{-1/2}(N) \|^2 \| x \|^2 \sum_{k=0}^{N} E\{ \| F_k(y_k) - F_k(y_k(k-\ell)) \|^2\},$$

where $\ell \leq N$ has been used to derive the last inequality. Hence, it is concluded from (5.136) and (5.139) that

$$(5.145)\quad E\{x'\beta(\ell,N)\beta'(\ell,N)x\}$$

$$\leq \frac{\rho^2}{\ell} c^2 \| \mathring{\theta} \|^2 \| x \|^2 + \sigma^2 c^3 \| x \|^2 c_1^{\ell} \to 0 \quad \text{as} \quad \ell \to \infty,$$

or equivalently

$$(5.146)\quad E\{\beta(\ell,N)\beta'(\ell,N)\} \to 0 \quad \text{as} \quad \ell \to \infty.$$

Consequently from (5.127), (5.141), (5.146) and Lemma 5.4, we can see that $z(N)$ defined by (5.98) converges to a Gaussian random variable with zero mean and the unit variance, which means

$$(5.147)\quad \sqrt{N} U^{-1/2}(N)P(N)(\mathring{\theta} - \hat{\theta}(N)) \overset{\text{law}}{\to} z \quad \text{as} \quad N \to \infty,$$

where

$$z \sim N(0,I).$$

5.5 Digital Simulation Studies

Consider

$$(5.148)\quad \begin{cases} \ddot{x}(t) + \alpha\,\dot{x}(t) + \beta(1 + \gamma x^2(t))x(t) = \xi(t) \\ x(0) = x_0\ ,\ \dot{x}(0) = \dot{x}_0 \end{cases}$$

where $\xi(t)$ is a white Gaussian noise process. The model (5.148) is well known as a model of rolling motion of ships[01]. By setting $x_1(t)=x(t)$, $x_2(t)=\dot{x}_1(t)$, we have

$$(5.149)\quad \begin{cases} d\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\beta(1 + \gamma x_1^2(t)) & -\alpha \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} dt + \begin{bmatrix} 0 \\ 1 \end{bmatrix} dw(t) \\ \begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} x_0 \\ \dot{x}_0 \end{bmatrix} \end{cases}$$

where the white Gaussian noise process $\xi(t)$ was identified with the Brownian motion process $w(t)$ as $\xi(t)dt=dw(t)$. Furthermore, the following discrete form of (5.149) is obtained by setting the sampling time t_k $(k=0,1,\cdots)$ and sampling time interval $\delta=t_{k+1}-t_k$:

$$(5.150)\quad \begin{cases} y_{k+1} - y_k = \delta \begin{bmatrix} 0 & 1 \\ -\beta(1 + \gamma y_{1k}^2) & -\alpha \end{bmatrix} y_k + bv_{k+1} \\ y_0 = \begin{bmatrix} x_0 \\ \dot{x}_0 \end{bmatrix} \end{cases}$$

where

$$(5.151)\quad \begin{cases} y_k = [y_{1k}\ \ y_{2k}]' \triangleq [x_1(t_k)\ \ x_2(t_k)]' \\ b \triangleq [0\ \ 1]' \\ v_{k+1} \triangleq w(t_{k+1}) - w(t_k) \end{cases}$$

In simulation studies, the parameter γ is assumed as $\gamma > 0$.

(A) Case 1 (Single-unknown-parameter case). First we consider the case where the coefficient β is unknown, while the coefficient α is known. Set as

$$(5.152) \quad \begin{cases} \mathring{\theta} = -\beta \\ \phi(y_k, \mathring{\theta}) = \phi_0(y_k) + \mathring{\theta}\phi_1(y_k) \\ \phi_0(y_k) = \begin{bmatrix} \delta y_{2k} \\ -\delta\alpha y_{2k} \end{bmatrix} \\ \phi_1(y_k) = \begin{bmatrix} 0 \\ \delta(1 + \gamma y_{1k}^2) y_{1k} \end{bmatrix}, \end{cases}$$

then, the system model (5.150) can be rewritten by

$$(5.153) \quad \begin{cases} y_{k+1} - y_k = \phi_0(y_k) + \mathring{\theta}\phi_1(y_k) + bv_{k+1} \\ y_0 = [x_0 \quad \dot{x}_0]'. \end{cases}$$

Then, the estimator of $\mathring{\theta}$ is given by (5.20), i.e.,

$$(5.154a) \quad \hat{\theta}(N) = \hat{\theta}(N-1) + \frac{\delta(1 + \gamma y_{1N}^2) y_{1N} F(N-1)}{1 + \delta^2 \Gamma(N-1)(1 + \gamma y_{1N}^2)^2 y_{1N}^2} \times (y_{2,N+1} - y_{2N} + \delta\alpha y_{2N} - \delta(1 + \gamma y_{1N}^2) y_{1N} \hat{\theta}(N-1))$$

$$(5.154b) \quad \Gamma(N) = \Gamma(N-1) - \frac{\delta^2(1 + \gamma y_{1N}^2)^2 y_{1N}^2 \Gamma^2(N-1)}{1 + \delta^2 \Gamma(N-1)(1 + \gamma y_{1N}^2)^2 y_{1N}^2}.$$

Since $\{v_k\}$ is an independent Gaussian noise sequence, the basic condition (C-1) is satisfied. Hence, if the condition (C-2) is satisfied, the estimator $\hat{\theta}(N)$ given by (5.154) is a consistent one. For the system model (5.153), the matrix $\tilde{B}_{k+1}$ defined by (5.7) is given by

(5.155) $$\tilde{B}_{k+1} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then the condition (C-2) can be rewritten as

(5.156) $$\begin{aligned} \Gamma(N) &= \{ \sum_{k=0}^{N} \phi_{1k}'(y_k)\tilde{B}_{k+1}^2\phi_{1k}(y_k) + \rho\}^{-1} \qquad \text{(scalar)}. \\ &= \{ \sum_{k=0}^{N} \phi_{1k}'(y_k)\phi_{1k}(y_k) + \rho\}^{-1} \\ &= \{ \sum_{k=0}^{N} (1 + \gamma y_{1k}^2)y_{1k}^2 + \rho\}^{-1} \\ &\to 0 \qquad \text{w.p. } 1 \qquad \text{as } N \to \infty, \end{aligned}$$

where (5.27) and (5.152) have been used. For the output process y_{1k}, it can be easily verified that

(5.157) $$y_{1k} = f_1(v_1, v_2, \cdots, v_{k-2}) + \delta v_{k-1}$$

where

(5.158) $$\begin{cases} f_1(v_1, v_2, \cdots, v_{k-2}) \triangleq y_{1,k-1} + \delta y_{2,k-2} + \delta f_2(v_1, v_2, \cdots, v_{k-2}) \\ f_2(v_1, v_2, \cdots, v_{k-1}) \triangleq (1 + \delta\hat{\theta}_2)y_{2,k-1} + \hat{\theta}_1\delta(1 + \gamma y_{1,k-2}^2)y_{1,k-2} \end{cases}.$$

Hence by using the simular procedure as that of Theorem 5.3, it can be easily verified that

(5.159) $$\begin{aligned} \liminf_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} (1 + \gamma y_{1k}^2)^2 y_{1k}^2 &\geq \liminf_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N} y_{1k}^2 \\ &\geq \delta\sigma^2 > 0 \qquad \text{w.p. } 1 \end{aligned}$$

and this implies that

$$\sum_{k=1}^{N} (1 + \gamma y_{1k}^2)^2 y_{1k}^2 \to \infty \qquad \text{w.p. } 1 \quad \text{as } N \to \infty.$$

and hence the condition (C-2) is satisfied.

In digital simulation studies, the true value of the unknown parameter was set as $\mathring{\theta}=-8.0$; other known parameters were $\alpha=4.0$, $\sigma^2=0.433$, $\gamma=5.0$ and $\rho=0.001$; and the sampling interval and the initial state y_0 were respectively set as $\delta=0.01$ and $y_0=[0\ 0]'$. With these true values, the output sequence of y_{1k} and y_{2k} were obtained by simulating (5.154) on a digital computer. Figure 5.1(a) shows a typical sample run of the input sequence $\{v_k;k=1,2,\cdots\}$ of independent Gaussian random variables generated by a digital computer. The y_k process is depicted in Fig. 5.1(b) and 5.1(c). A typical sample run of the estimator $\hat{\theta}(N)$ is shown in Fig. 5.2. By using the same measure as that defined by (4.74) in the previous chapter, i.e.,

$$(5.160) \qquad C_L(N,\hat{\theta}(N)) = \frac{1}{L}\sum_{\ell=1}^{L}(\hat{\theta}^{(\ell)}(N) - \mathring{\theta})^2,$$

the convergence of $\hat{\theta}(N)$ was examined numerically. Figure 5.3 illustrates a result of Monte Carlo experiments for 100 sample runs of $\{y_k\}$.

(B) Case 2 (Multi-unknown-parameter case). Next we consider the case where the coefficients α and β are both unknown. In this case, the system model can be represented by

$$(5.161) \qquad \begin{cases} y_{k+1} - y_k = \phi(y_k,\mathring{\theta}) + bv_{k+1} \\ y_0 = [x_0\ \ \dot{x}_0]' \end{cases}$$

where

$$\mathring{\theta} \triangleq [\mathring{\theta}_1\ \ \mathring{\theta}_2]' = [-\beta\ \ -\alpha]'$$

$$(5.162)\quad \begin{cases} \phi(y_k,\mathring{\theta}) = \phi_0(y_k) + \mathring{\theta}_1\phi_1(y_k) + \mathring{\theta}_2\phi_2(y_k) \\ \phi_0(y_k) = \delta\begin{bmatrix} y_{2k} \\ 0 \end{bmatrix} \qquad \phi_1(y_k) = \delta\begin{bmatrix} 0 \\ g(y_{1k})y_{1k} \end{bmatrix} \\ \phi_2(y_k) = \delta\begin{bmatrix} 0 \\ y_{2k} \end{bmatrix} . \end{cases}$$

In (5.162), in order to satisfy the condition (C-3)' and (C-4), $(1+\gamma y_{1k}^2)$ was replaced by the function $g(y_{1k})$ defined by

$$(5.163)\quad g(y_{1k}) = \begin{cases} 1 + \gamma\mu & \text{for } y_{1k}^2 \geq \mu \\ 1 + \gamma y_{1k}^2 & \text{for } y_{1k}^2 < \mu, \end{cases}$$

where μ is a preassigned constant with a sufficient large value, and γ is assumed to be a positive constant. From (5.162), it is easily shown that $\tilde{\phi}_K(y_K)$ given by (5.69) can be represented by

$$(5.164)\quad \tilde{\phi}_k(z,\mathring{\theta}) = A(z,\mathring{\theta})z$$

where

$$(5.165)\quad A(z,\mathring{\theta}) = \begin{bmatrix} 1 & \delta \\ \mathring{\theta}_1\delta g(z_1) & 1 + \delta\mathring{\theta}_2 \end{bmatrix} .$$

Hence, if two eigenvalues $\lambda_i (i=1,2)$ of $A(z,\mathring{\theta})$ are $|\lambda_i| < 1$, the condition (C-3)' is satisfied. Since the characteristic equation of (5.165) is given by

$$(5.166)\quad \lambda^2 - (2 + \delta\mathring{\theta}_2)\lambda + \{1 - \delta^2 g(z_1)\mathring{\theta}_1 + \delta\mathring{\theta}_2\} = 0,$$

it is easily verified that $|\lambda_i| < 1$ $(i=1,2)$, if

$$(5.167)\quad \begin{cases} \mathring{\theta}_1 < 0 & \\ 2\delta\mathring{\theta}_2 + 4 > \delta^2\mathring{\theta}_1 g(z_1) & \text{for all } z_1 \\ \delta g(z_1)\mathring{\theta}_1 > \mathring{\theta}_2 & \text{for all } z_1. \end{cases}$$

It is almost obvious that the system model (5.163) satisfies the conditions (C-1) and (C-4). Hence, all conditions of Theorems 5.2, 5.3 satisfied except for the condition (C-5). We shall show below that the system model (5.161) satisfies the condition (C-5). From (5.155) and (5.162), $F_k(y_k)$ defined by (5.14) can be rewritten by

$$(5.168) \qquad F_k(y_k) = \delta \begin{bmatrix} 0 & g(y_{1k})y_{1k} \\ 0 & y_{2k} \end{bmatrix}.$$

Then, it follows that

$$(5.169) \qquad F_k(y_k)F_k'(y_k) = \delta^2 \begin{bmatrix} g^2(y_{1k})y_{1k}^2 & g(y_{1k})y_{1k}y_{2k} \\ g(y_{1k})y_{1k}y_{2k} & y_{2k}^2 \end{bmatrix}.$$

By the same procedure as that for (5.157), we have

$$(5.170) \qquad y_{1k} = f_3(v_1, v_2, \cdots, v_{k-2}) + \delta v_{k-1}$$

$$(5.171) \qquad y_{2k} = f_4(v_1, v_2, \cdots, v_{k-1}) + v_k$$

where

$$(5.172) \qquad f_3(v_1, v_2, \cdots, v_{k-2}) \triangleq y_{1,k-1} + \delta y_{2,k-2} + \delta f_4(v_1, v_2, \cdots, v_{k-2})$$

$$(5.173) \qquad f_4(v_1, v_2, \cdots, v_{k-1}) \triangleq (1 + \delta\mathring{\theta}_2)y_{2,k-1} + \mathring{\theta}_1\delta g(y_{1,k-1})y_{1,k-1}.$$

Hence, it follows that

$$(5.174) \qquad \frac{1}{N}\sum_{k=0}^{N} E\{F_k(y_k)F_k'(y_k) \mid \mathscr{F}_{k-1}\}$$

$$= \delta^2 \begin{bmatrix} \frac{1}{N}\sum_{k=0}^{N} g^2(y_{1k})y_{1k}^2 & \frac{1}{N}\sum_{k=0}^{N} g(y_{1k})y_{1k}f_4(v_1, \cdots, v_{k-1}) \\ \frac{1}{N}\sum_{k=0}^{N} g(y_{1k})y_{1k}f_4(v_1, \cdots, v_{k-1}) & \frac{1}{N}\sum_{k=0}^{N} f_4^2(v_1, \cdots, v_{k-1}) + \sigma^2 \end{bmatrix}$$

and

$$(5.175)\quad x'\{\frac{1}{N}\sum_{k=0}^{N}E\{F_k(y_k)F_k'(y_k)\mid \mathcal{F}_{k-1}\}\}x$$

$$= \delta^2\frac{1}{N}\sum_{k=0}^{N}\{g(y_{1k})y_{1k}x_1 + f_4(v_1,\cdots,v_{k-1})x_2\}^2 + \delta^2\sigma^2x_2^2$$

where $x \triangleq (x_1, x_2)'$ is an arbitrary vector such that $x \neq 0$. By using the same procedure as that of (5.159), we have

$$(5.176)\quad \liminf_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N}y_{1k}^2 \geq \delta^2\sigma^2 > 0 \qquad \text{w.p. } 1$$

Hence, it follows that

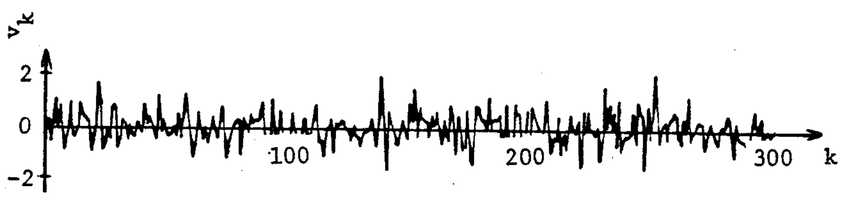
$$(5.177)\quad \liminf_{N\to\infty} x'\{\frac{1}{N}\sum_{k=0}^{N}E\{F_k(y_k)F_k'(y_k)\mid \mathcal{F}_{k-1}\}\}x$$

$$\geq \begin{cases} \delta^2\sigma^2x_2^2 & \text{if } x_2 \neq 0 \\ \liminf_{N\to\infty}\delta^2\frac{1}{N}\sum_{k=0}^{N}g^2(y_{1k})y_{1k}^2x_1^2 \geq \delta^2\sigma^2x_1^2 & \text{if } x_2 = 0, \end{cases}$$

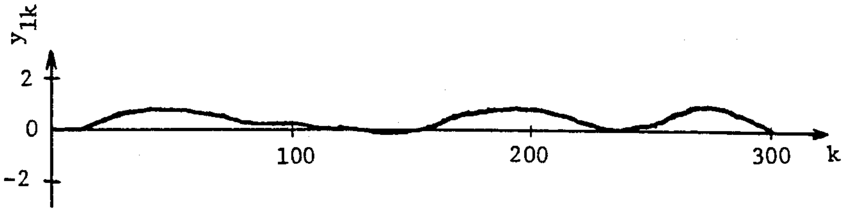
which implies that the condition (C-5) is satisfied.

In digital simulation studies, the parameter μ was set as $\mu=10$, and other parameters were set as the same values as in Case 1. Typical sample runs of the estimator $\hat{\theta}(N)=[\ \hat{\theta}_1(N)\ \ \hat{\theta}_2(N)]'$ and $\hat{\sigma}^2(N)$ are shown in Fig. 5.4. Monte-Carlo experiments for 100 sample runs of $\{y_k\}$'s is illustrated in Fig. 5.5. Histograms of 100 sample runs of estimators $\hat{\theta}_1(N)$, $\hat{\theta}_2(N)$ and $\hat{\sigma}^2(N)$ at N=3000 are shown in Fig. 5.6, where fitted normal curves are also depicted. By using Fig. 5.6, cumulative frequency curves for each 100 sample runs of estimators are plotted on the normal-probability papers, and they are depicted in Fig. 5.7. Furthermore we make use of Chi-square test with the 5 percent level of significance in order to check more precisely the normality of sample distributions of estimators. Results of Chi-square test are shown in Table 5.1.

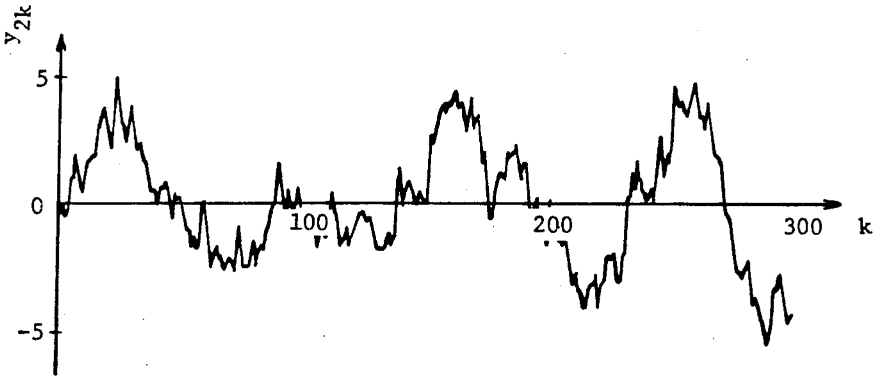
Throughout both cases of Case 1 and Case 2, we may conclude that estimators of unknown parameters well converge to their true values, and that, in Case 2, sample distributions of the estimators at N=3000 are approximately Gaussian.



(a) A sample run of the input noise v_k



(b) A sample run of y_{1k}



(c) A sample run of y_{2k}

Fig 5.1 Typical sample runs of the input noise v_k and the output sequence y_k

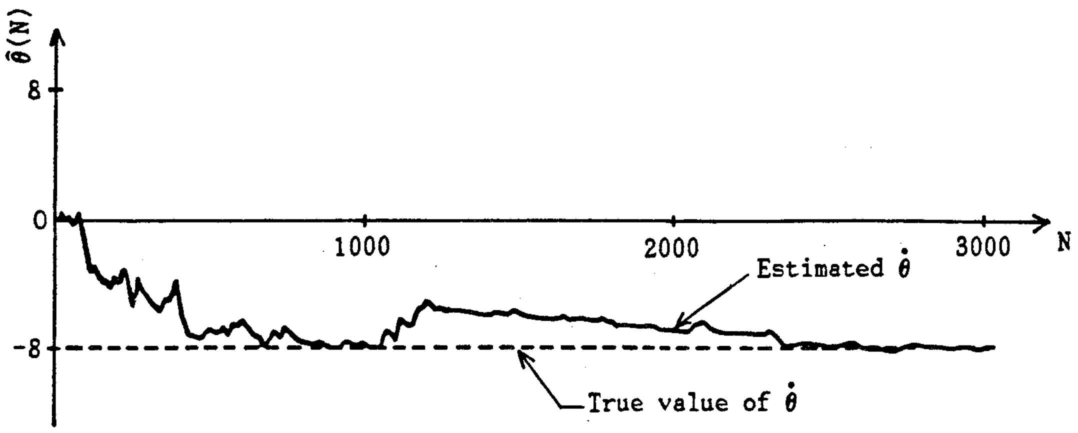


Fig 5.2 A typical sample run of the estimate $\hat{\theta}(N)$

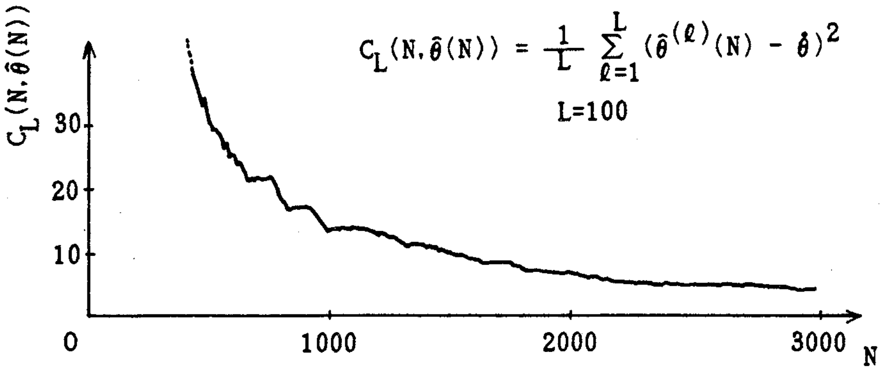
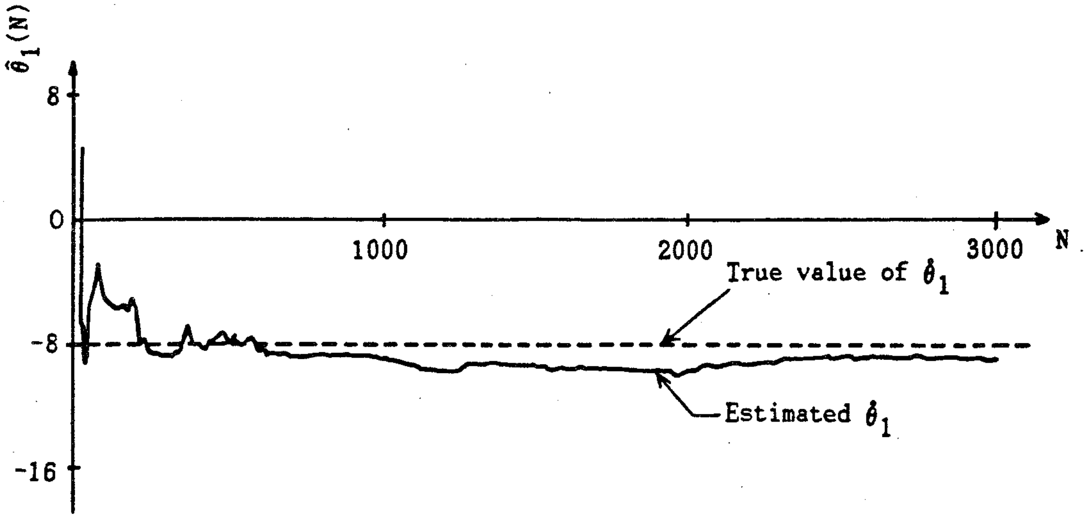
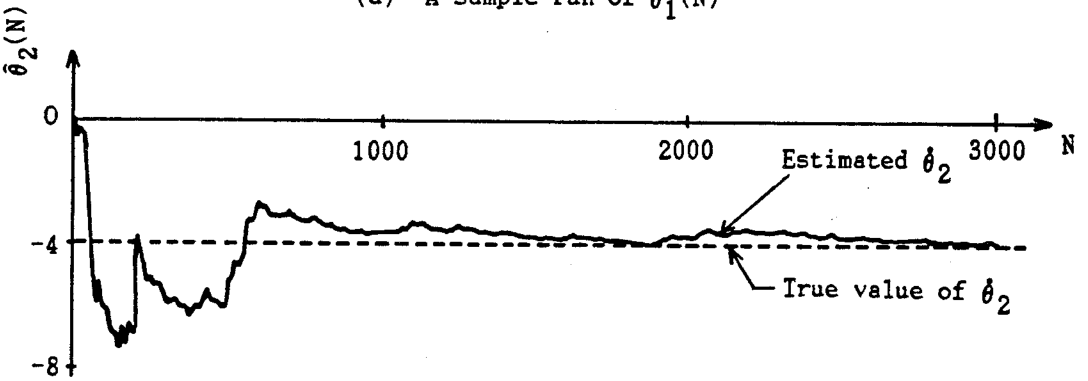


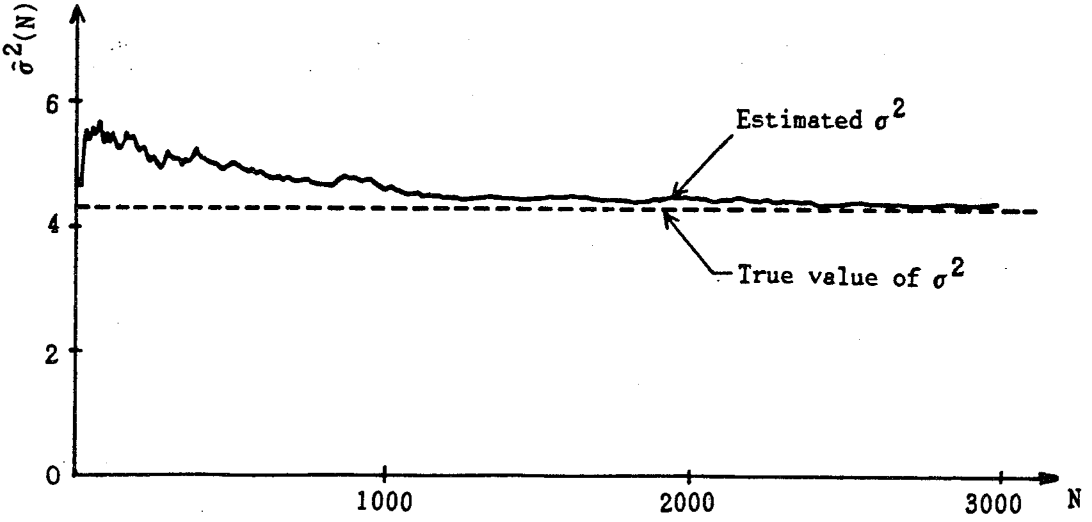
Fig 5.3 Convergence of $\hat{\theta}(N)$ obtained by 100 Monte-Carlo trials



(a) A sample run of $\hat{\theta}_1(N)$

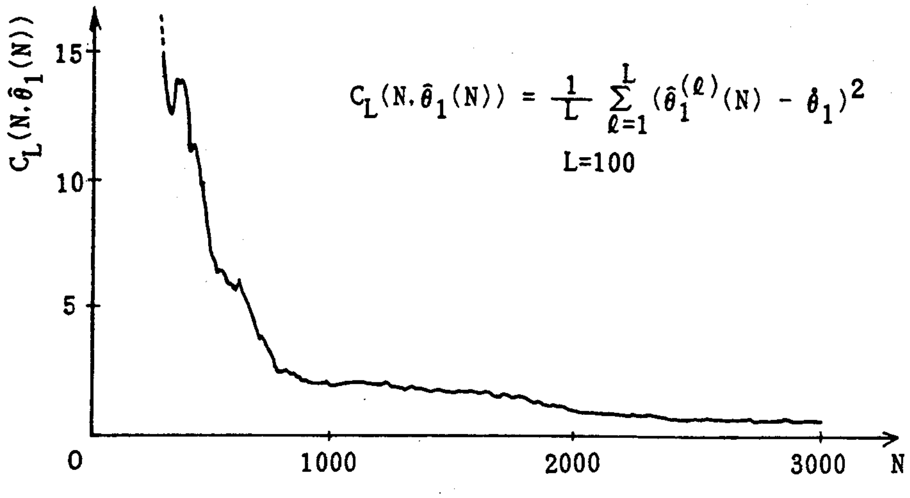


(b) A sample run of $\hat{\theta}_2(N)$

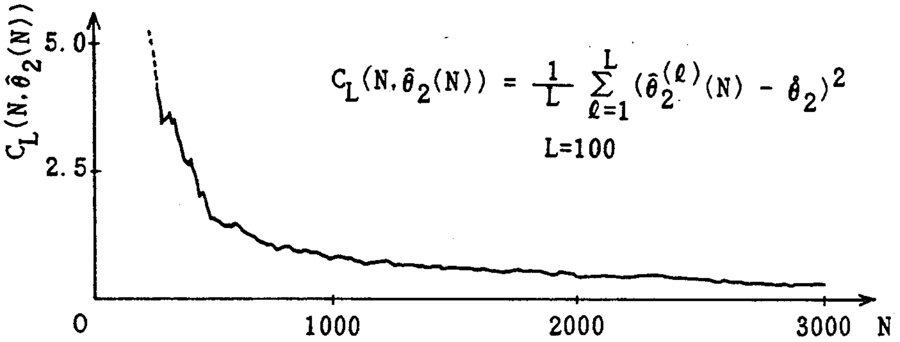


(c) A sample run of $\hat{\sigma}^2(N)$

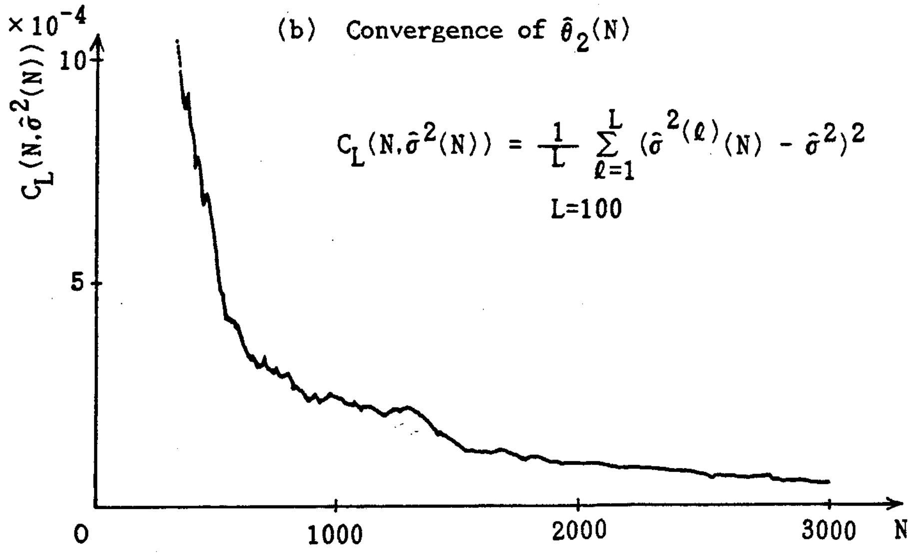
Fig. 5.4 Typical sample runs of estimates of $\hat{\theta}_1(N)$, $\hat{\theta}_2(N)$ and $\hat{\sigma}^2(N)$



(a) Convergence of $\hat{\theta}_1(N)$

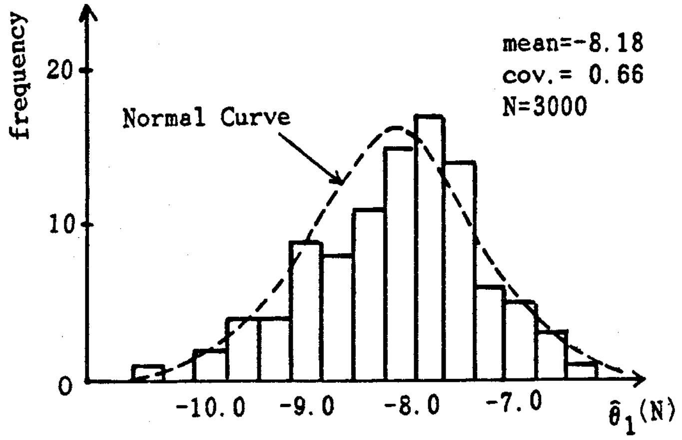


(b) Convergence of $\hat{\theta}_2(N)$

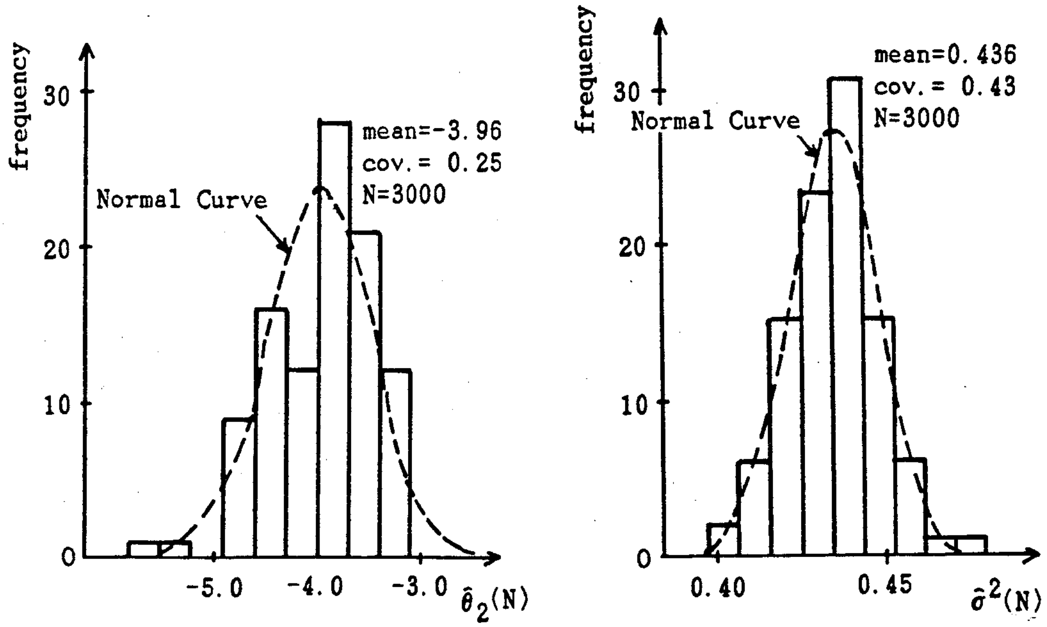


(c) Convergence of $\hat{\sigma}^2(N)$

Fig. 5.5 Convergence behaviours of $\hat{\theta}_1(N)$, $\hat{\theta}_2(N)$ and $\sigma^2(N)$ obtained by 100 Monte-Carlo trials



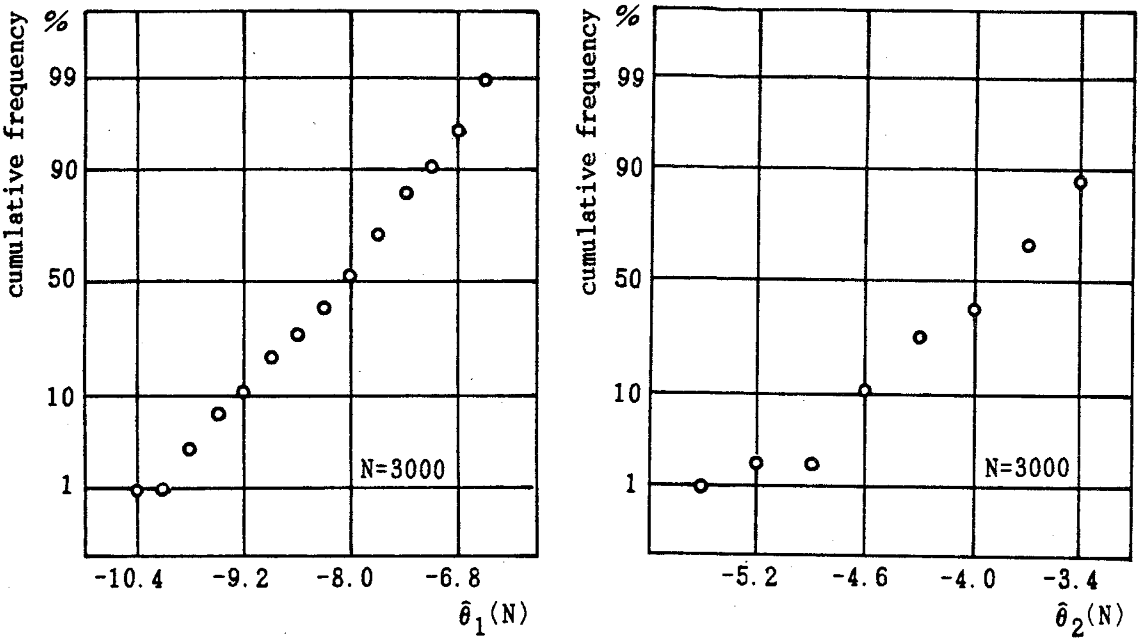
(a) Histogram of $\hat{\theta}_1(N)$



(b) Histogram of $\hat{\theta}_2(N)$

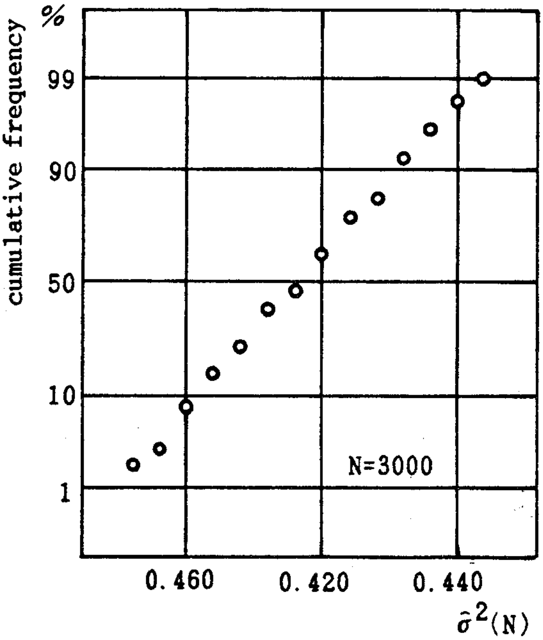
(c) Histogram of $\hat{\sigma}^2(N)$

Fig. 5.6 Histograms of the 100 sample runs of estimates $\hat{\theta}_1(N)$, $\hat{\theta}_2(N)$ and $\hat{\sigma}^2(N)$ at N=3000, where fitted normal curves are also depicted



(a) Cumulative frequency curve of $\hat{\theta}_1(N)$

(b) Cumulative frequency curve of $\hat{\theta}_2(N)$



(c) Cumulative frequency curve of $\hat{\sigma}^2(N)$

Fig 5.7 Cumulative frequency curves of $\hat{\theta}_1(N)$, $\hat{\theta}_2(N)$ and $\hat{\sigma}^2(N)$

Table 5.1 Chi-square test for the check of normality

(a) Chi-square test for $\hat{\theta}_{1N}(N)$ (N=3000)

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above -6.5	1	1.92	0.441
-6.8 ~ -6.5	3	2.54	0.083
-7.1 ~ -6.8	5	4.72	0.017
-7.4 ~ -7.1	6	7.67	0.364
-7.7 ~ -7.4	14	10.91	0.875
-8.0 ~ -7.7	17	13.53	0.890
-8.3 ~ -8.0	15	14.67	0.007
-8.6 ~ -8.3	11	13.89	0.601
-8.9 ~ -8.6	8	11.48	1.055
-9.2 ~ -8.9	9	8.11	0.098
-9.5 ~ -9.2	4	5.40	0.363
-9.8 ~ -9.5	4	2.88	0.436
Below~ -9.8	3	2.28	0.227
Total	100	100.00	5.457
$\chi^2_{0.95}(10) = 18.31 > 5.457$			

(b) Chi-square test for $\hat{\theta}_{2N}(N)$ (N=3000)

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above -3.4	12	13.14	0.099
-3.7 ~ -3.4	21	17.14	0.936
-4.0 ~ -3.7	28	23.14	1.068
-4.3 ~ -4.0	12	21.98	4.531
-4.6 ~ -4.3	16	14.80	0.097
-4.9 ~ -4.6	9	7.02	0.558
Below -4.9	2	3.01	0.339
Total	100	100.23	7.628
$\chi^2_{0.95}(4) = 9.49 > 7.628$			

(c) Chi-square test for $\hat{\sigma}^2(N)$ (N=3000)

Interval	Observed frequency f_i	Theoretical frequency F_i	$(f_i-F_i)^2/F_i$
Above 0.460	2	2.12	0.0068
0.452 ~ 0.460	6	6.75	0.0833
0.444 ~ 0.452	15	16.61	0.1561
0.436 ~ 0.444	31	25.74	1.0749
0.428 ~ 0.436	23	25.22	0.1954
0.420 ~ 0.428	15	15.50	0.0160
0.412 ~ 0.420	6	6.25	0.0100
Below 0.412	2	1.83	0.0158
Total	100	100.02	1.5583
$\chi^2_{0.95}(5) = 11.07 > 1.5583$			

where theoretical frequency F_i denotes the area under the normal curve in the interval while the observed frequency f_i is the actual number of observations which fall in the interval.

5.6 Discussions

In this chapter, the identification procedure for a class of non-stationary nonlinear systems has been derived. It has been also proved mathematically that proposed estimators of unknown system parameters have salient features such as the consistency and the asymptotic normality.

The key assumption is that the system model is, in this chapter, linear with respect to unknown parameters. Estimators of unknown parameters are obtained by using the maximum likelihood concept. Using the martingale convergence theorem and the central limit theorem for the sum of dependent random variables, the consistency of estimators and their asymptotic normality have been proved respectively. It should be emphasized that in case of the single unknown parameter, the consistency of the estimator holds without stable conditions in the sense that the second moment of the output process is bounded,which is usually required to prove the consistency of estimators.

Appendix 5.A Proof of Lemma 5.2

Since both $F_k(y_k)$ and $\Gamma(k)$ are $\mathcal{F}_k$-measurable,

$$(A.1)\quad E\{h(N) \mid \mathcal{F}_N\} = E\{\sum_{k=0}^{N-1} \Gamma(k)F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}$$

$$+\Gamma(N)F_N(y_N)\tilde{B}_{N+1}B_{N+1}v_{N+1} \mid \mathcal{F}_N\}$$

$$= \sum_{k=0}^{N-1} \Gamma(k)F_k(y_k)\tilde{B}_{k+1}B_{k+1}v_{k+1}$$

$$= h(N-1).$$

Then $\{h(N), \mathcal{F}_N\}$ is a martingale. We shall prove that $h(N)$ is bounded in L^2. From (A.1), we obtain

$$(A.2)\quad E\{h^2(N)\} = \sum_{j=0}^{N} E\{h^2(j) - h^2(j-1)\}$$

$$= \sum_{j=0}^{N} E\{E\{h^2(j) - h^2(j-1) \mid \mathcal{F}_j\}\}$$

$$= \sum_{j=1}^{N} E\{E\{h^2(j) - h^2(j-1) \mid \mathcal{F}_j\}\} + E\{h^2(0)\}$$

$$= \sum_{j=1}^{N} E\{E\{(\Gamma(j)F_j(y_j)\tilde{B}_{j+1}^2B_{j+1}v_{j+1})^2$$

$$+ 2(\Gamma(j)F_j(y_j)\tilde{B}_{j+1}^2B_{j+1}v_{j+1})h(j-1) \mid \mathcal{F}_j\}\} + E\{h^2(0)\}$$

$$= \sum_{j=1}^{N} E\{E\{(\Gamma(j)F_j(y_j)\tilde{B}_{j+1}B_{j+1}v_{j+1})^2 \mid \mathcal{F}_j\}\} + E\{h^2(0)\}.$$

Here, from the condition (C-1), it follows that

$$(A.3)\quad E\{h^2(0)\} = E\{(\Gamma(0)F_0(y_0)\tilde{B}_1B_1v_1)^2\}$$

$$= \sigma^2\,\Gamma(0)F_0(y_0)\tilde{B}_1B_1B_1'\tilde{B}_1'\;F_0'(y_0)\,\Gamma(0)$$

$$\leq c\ (\text{const.})$$

and

$$(A.4)\quad E\{(\Gamma(j)F_j(y_j)\tilde{B}_{j+1}B_{j+1}v_{j+1})^2 \mid \mathscr{F}_j\}$$

$$= \Gamma(j)F_j(y_j)\tilde{B}_{j+1}B_{j+1}E\{v_{j+1}v'_{j+1} \mid \mathscr{F}_j\}B'_{j+1}\tilde{B}_{j+1}F'_j(y_j)\,\Gamma(j)$$

$$= \sigma^2\,\Gamma(j)F_j(y_j)\tilde{B}_{j+1}B_{j+1}B'_{j+1}\tilde{B}_{j+1}F'_j(y_j)\,\Gamma(j)$$

$$= \sigma^2\,\Gamma(j)F_j(y_j)F'_j(y_j)\,\Gamma(j)$$

$$= \sigma^2\,\Gamma(j)(\,\Gamma^{-1}(j) - \Gamma^{-1}(j-1)),$$

where the condition (C-1), (5.27) and

$$(A.5)\quad F_j(y_j)\tilde{B}_{j+1}B_{j+1}B'_{j+1}\tilde{B}_{j+1}F'_j(y_j) = F_j(y_j)F'_j(y_j)$$

have been used. Since

$$(A.6)\quad 0 \leq \Gamma^{-1}(j-1)(\,\Gamma(j) - \Gamma(j-1))^2 = \Gamma(j-1) - \Gamma(j) - (\,\Gamma(j) - \Gamma^2(j)\,\Gamma(j-1)),$$

it follows that

$$(A.7)\quad \Gamma^2(j)(\,\Gamma^{-1}(j) - \Gamma^{-1}(j-1)) = \Gamma(j) - \Gamma^2(j)\,\Gamma^{-1}(j-1)$$

$$= (\,\Gamma(j-1) - \Gamma(j)) - \Gamma^{-1}(j-1)(\,\Gamma(j) - \Gamma(j-1))^2$$

$$\leq \Gamma(j-1) - \Gamma(j).$$

Hence, from (A.2) to (A.7), it can be concluded that

$$\text{(A.8)} \quad E\{h^2(N)\} \leq \sigma^2 \sum_{j=1}^{N} E\{\Gamma(j-1) - \Gamma(j)\} + c$$

$$= \sigma^2(\Gamma(0) - E\{\Gamma(N)\}) + c$$

$$\leq \sigma^2 \Gamma(0) + c.$$

Therefore, from Theorem 2.1, h(N) converges to the random variable h with probability one.

Appendix 5.B <u>Proof of Lemma 5.3</u>

From the condition (C-1), we have

$$\text{(B.1)} \qquad \| \tilde{\phi}_k(y_k, \theta) \| \leq c_1 \| y_k \| + c.$$

Then, using (B.1) and (5.1), y_k can be evaluated as follows:

$$\begin{aligned}
\text{(B.2)} \qquad \| y_k \| &\leq \| \tilde{\phi}_{k-1}(y_{k-1}, \theta) \| + \| B_k \| \| v_k \| \\
&\leq c_1 \| y_{k-1} \| + c(1 + \| v_k \|) \\
&\leq c_1 \{ \| \tilde{\phi}_{k-2}(y_{k-2}, \theta) \| + c(1 + \| v_{k-1} \|)\} + c(1 + \| v_k \|) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\leq c_1^k \| y_0 \| + c(c_1^{k-1} \| v_1 \| + \cdots + c_1 \| v_{k-1} \| + \| v_k \| \\
&\qquad\qquad + c_1^{k-1} + \cdots + c_1 + 1) \\
&\leq c_1^k \| y_0 \| + c(\frac{1}{1-c_1} + \sum_{i=1}^{k} c_1^{k-i} \| v_i \|),
\end{aligned}$$

where the relation

$$\sum_{i=0}^{k} c_1^i \leq \sum_{i=0}^{\infty} c_1^i = \frac{1}{1-c_1} \qquad (0 < c_1 < 1)$$

has been used. Hence, from (B.2) and the elementary inequality $(a+b+c)^4 \leq 27(a^4+b^4+c^4)$, we have

$$\begin{aligned}
\text{(B.3)} \qquad E\{ \| y_k \|^4 \} &\leq E\{(c_1^k \| y_0 \| + c(\frac{1}{1-c_1} + \sum_{i=1}^{k} c_1^{k-i} \| v_i \|))^4\} \\
&\leq 27[c_1^{4k} \| y_0 \|^4 + (\frac{c}{1-c_1})^4 + c^4 E\{ \sum_{i=1}^{k} c_1^{k-i} \| v_i \|)^4\}].
\end{aligned}$$

Furthermore, using the condition (C-1), it follows that

$$(\mathrm{B.4})\qquad E\{(\sum_{i=1}^{k} c_1^{k-i} \| v_i \|)\} = \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sum_{i_3=1}^{k} \sum_{i_4=1}^{k} c_1^{4k-i_1-i_2-i_3-i_4}$$

$$\times E\{ \| v_{i_1} \| \| v_{i_2} \| \| v_{i_3} \| \| v_{i_4} \| \}$$

$$\leq \sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \sum_{i_3=1}^{k} \sum_{i_4=1}^{k} c_1^{4k-i_1-i_2-i_3-i_4} [\prod_{j=1}^{4} E\{E\{ \| v_{i_j} \|^4 \mid \mathcal{F}_{i_j-1}\}\}^{1/4}$$

$$\leq c(\sum_{i=1}^{K} c_1^{k-i})^4$$

$$\leq c(\frac{1}{1-c_1})^4 < +\infty .$$

Hence, we can conclude (5.45) from (B.3) and (B.4).

CHAPTER 6 STRUCTURE DETERMINATION OF NONSTATIONART NONLINEAR SYSTEMS

6.1 Introductory Remarks

In this chapter, motivated by the requirement of establishing a mathematical model reflecting nonlinear and nonstationary properties of geophysical data, a method is presented for modeling a class of nonstationary nonlinear systems with unknown system structures. The underlying system is assumed to be described by a nonlinear time-varying stochastic difference equation.

The principal line of attack to determine the system structure is to minimize the upper bound of the entropy associated with errors of both input noises and unknown parameters, where unknown system parameters are estimated by using the maximum likelihood concept.

In Sectoin 6.2, a model for nonlinear time-varying systems is presented and the criterion functions for the structure determination and parameter identification are introduced. For the system model given in Section 6.3, which is linear with respect to unknown parameters, asymptotic properties of the determined system structure and estimated unknown parameters are investigated theoretically in Section 6.4 and numerically in Section 6.5.

In Section 6.6, the proposed method for the structure determination and parameter identification is also tried to apply to several real earthquake data.

6.2 Criterion Function for Structure Determination

It is of interest to consider that obseved discrete data is obtained as the solution of the following stochastic difference equation:

$$(6.1)\quad \begin{cases} y_k = f_s(y_{k-1}, y_{k-2}, \cdots, y_{k-n}; k; \mathring{\theta}) + b_k v_k \\ y_0 = y_{-1} = \cdots = y_{-n+1} = 0 \end{cases}$$

where $\{b_k\}$ is the known positive coefficient with the bounded value for k= 1, 2, ⋯, N; $\mathring{\theta}$ is the vector representing the unknown parameters whose dimension is dependent on both the system order n and the structure of $f_s(\cdot)$; the input $\{v_k\}$ is the sequence of scalar unobservable random variables and satisfies the following condition:

(C-1) The sequence of random variables $\{v_k\}$ satisfies

$$\begin{cases} E\{v_k \mid \mathscr{F}_{k-1}\} = 0 & \text{w.p. } 1 \\ E\{v_k^2 \mid \mathscr{F}_{k-1}\} = \sigma^2 & \text{w.p. } 1 \\ E\{v_k^4 \mid \mathscr{F}_{k-1}\} \leq c & \text{w.p. } 1 \end{cases}$$

where $\mathscr{F}_{k-1}$ is the minimum σ-algebra generated by $\{v_1, \cdots, v_{k-1}\}$.

The purpose of this chapter is to determine the structure of $f_s(\cdot)$ and to estimate the unknown parameters $\mathring{\theta}$ and σ^2 from the observation data $Y_N \triangleq \{y_1, \cdots, y_N\}$. Naturally, it is impossible to find a pair of $(f_s, \mathring{\theta})$ only from the information included in the observation data Y_N. Therefore, the class of $(f, \theta(f))$, where $(f_s, \mathring{\theta})$ is being looked for, is restricted within a certain range of f and $\theta(f)$. We denote this set of pairs $(f, \theta(f))$ by M, i.e.

$$(6.2)\quad M = \{(f, \theta(f)) \mid f \in F,\ \theta(f) \in D_f\} \qquad (D_f \subset R^m;\ m = \text{dim. } \theta(f))$$

where F and D_f are, respectively, the finite set of functions and the bounded domain of parameters associated with f which satisfies $|\theta_i(f)| \leq c$ for i=1, ⋯, m. An element $(f,\theta(f))$ of M represents that the output data $\{y_k\}$ is assumed to be described by

$$(6.3) \quad \begin{cases} y_k = f(y_{k-1},\cdots,y_{k-p};k;\theta(f)) + b_k v_k(f,\theta(f)) \\ \quad y_0 = y_{-1} = \cdots = y_{-p+1} = 0, \end{cases}$$

where p is the model order and $v_k(f,\theta(f))$ is the modeling error due to the form of the function f and the parameter $\theta(f)$. In the following, we use the notation

$$(6.4) \quad D_{f_1} \subset D_{f_2}$$

whose meaning is that, by setting some elements of $\theta(f_2)$ are zero, we obtain $f_1(z_1,\cdots,z_{p_1};k;\theta(f_1)) = f_2(z_1,\cdots,z_{p_2};k;\theta(f_2))$ for any $z_i(i=1,\cdots,p_2;p_1 \leq p_2)$, k and $\theta(f_1)$ $(\in D_{f_1})$.

We introduce the criterion function for determining the structure f_s, derived from evaluating the upper bound of the entropy associated with the errors $\xi_k(f,\theta(f))$ and $\zeta(f,\theta(f))$ defined below. $\xi_k(f,\theta(f))$ is given by

$$(6.5) \quad \xi_k(f,\theta(f)) \triangleq v_k - v_k(f,\theta(f)).$$

The i-th element of $\zeta(f,\theta(f))$ is defined by

$$(6.6a) \quad \zeta_i(f,\theta(f)) \triangleq \mathring{\theta}_j - \theta_i(f)$$

when there exists $\mathring{\theta}_j$ corresponding to $\theta_i(f)$, and

$$(6.6b) \quad \zeta_i(f,\theta(f)) \triangleq - \theta_i(f)$$

when there exists no $\mathring{\theta}_j$ corresponding to $\theta_i(f)$.

The entropy associated with $\{\xi_k(f,\theta(f));k=1,\cdots,N\}$ and $\zeta(f,\theta(f))$ is given by

(6.7) $$H(\xi_1,\cdots,\xi_N;\ \frac{N}{m}\zeta) \triangleq -E\{\log p(\xi_1,\cdots,\xi_N;\ \frac{N}{m}\zeta)\}$$

where $p(\cdot\,;\cdot)$ is the joint probability density function; N is the number of the observation data; m is the dimension of $\zeta(f,\theta(f))$; and $\xi_k(f,\theta(f))$ and $\zeta(f,\theta(f))$ are abbreviated as ξ_k and ζ respectively. Since the number of $\{\xi_k\}$ is N whereas that of ζ is m, contributions of $\{\xi_k\}$ and ζ to the entropy H are evaluated as the order of O(N) and O(m) respectively. Hence, in order to equalize their contributions to the entropy, ζ is multiplied by N/m.

Using Bayes' rule, (6.7) can be represented by

(6.8) $$H(\xi_1,\cdots,\xi_N;\ \frac{N}{m}\zeta) = \sum_{i=1}^{N} H(\xi_i \mid \xi_1,\xi_2,\cdots,\xi_{i-1},\frac{N}{m}\zeta) + H(\frac{N}{m}\zeta)$$

where $H(\cdot \mid *)$ is the conditional entropy defined by using the conditional probability density function $p(\cdot \mid *)$, i.e.,

(6.9) $$H(\xi_i \mid \xi_1,\xi_2,\cdots,\xi_{i-1},\frac{N}{m}\zeta) \triangleq -E\{\log p(\xi_i \mid \xi_1,\cdots,\xi_{i-1},\frac{N}{m}\zeta)\}.$$

Using the properties of entropy (see e.g. [G4]), we obtain

(6.10a) $$H(\xi_i \mid \xi_1,\xi_2,\cdots,\xi_{i-1},\ \frac{N}{m}\zeta) \leq \frac{1}{2}\log 2\pi e + \frac{1}{2}\log[\text{var.}\{\xi_i\}],$$

(6.10b) $$H(\frac{N}{m}\zeta) \leq \frac{1}{2}\,m\log 2\pi e + \frac{1}{2}\,m\log[(\frac{N}{m})^2\text{det.}(\text{var.}\{\zeta\})].$$

We assume that $\{\xi_i\}$ defined by (6.5) is the weak stationary process with $E\{\xi_i\}=0$ for $i=1,2,\cdots,N$. Then, it follows that

(6.11) $$\text{var.}\{\xi_i\} = \sigma^2(f,\theta(f)) - \sigma^2,$$

where $\sigma^2(f,\theta(f)) \triangleq E\{v_k^2(f,\theta(f))\}$. Therefore, since $\text{var.}\{\zeta\} \leq cI$ because of

$|\theta_i| \leq c$ $(i=1,\cdots,m)$, the entropy H given by (6.7) can be evaluated as

$$(6.12)\quad H(\xi_1,\cdots,\xi_N;\ \frac{N}{m}\zeta) \leq \frac{N}{2}\log \sigma^2(f,\theta(f)) + m\log N$$
$$+\frac{1}{2}\{(N+m)\log 2\pi e + m^2\log c - 2m\log m\},$$

where the inequality $\log\{\sigma^2(f,\theta(f))-\sigma^2\} \leq \log\sigma^2(f,\theta(f))$ has been used. For a fixed N, the minimization of the R.H.S. of (6.12) coincides with that of

$$\frac{N}{2}\log\sigma^2(f,\theta(f)) + m\log N + \frac{1}{2}m\log c + \frac{1}{2}m\log 2\pi e - m\log m,$$

and the last three terms are negligible for a large value of N. Hence, the minimization of the R.H.S. of (6.12) is equivalent to minimizing

$$\frac{N}{2}\log\sigma^2(f,\theta(f)) + m\log N.$$

Furthermore, since $\sigma^2(f,\theta(f))$ is unknown, we have to replace $\sigma^2(f,\theta(f))$ with its estimate $\hat{\sigma}_N^2(f,\hat{\theta}_N(f))$. Hence, the following criterion function for the structure determination is introduced:

$$(6.13)\quad \ell_s(N,f,\hat{\theta}_N(f)) \triangleq \frac{N}{2}\log\hat{\sigma}_N^2(f,\hat{\theta}_N(f)) + m\log N.$$

The unknown parameter $\theta(f)$ and the variance of $v_k(f,\theta(f))$ are identified here by minimizing the following criterion function,

$$(6.14)\quad \ell_p(f,\theta,\sigma^2) \triangleq N\log\sigma^2 + \frac{1}{\sigma^2}\sum_{k=1}^{N}\frac{1}{b_k^2}\{y_k - f_p(y_{k-1},\cdots,y_{k-p};k;\theta)\}^2.$$

If $v_k(f,\theta(f))$ is independent and identically distributed Gaussian sequence, the criterion function (6.14) coincides with the maximum likelihood identification method. Differentiating (6.14) with respect to σ^2 and setting $\partial \ell_p / \partial \sigma^2 = 0$, we have

$$(6.15) \qquad \hat{\theta}_N(f) = \arg.\{ \min_{\theta(f) \in D_f} L(N, f, \theta(f))\}$$

$$(6.16) \qquad \hat{\sigma}_N^2(f, \hat{\theta}_N(f)) = L(N, f, \hat{\theta}_N(f))$$

where

$$(6.17) \qquad L(N, f, \theta(f)) \triangleq \frac{1}{N} \sum_{k=1}^{N} \frac{1}{b_k^2} \{ y_k - f(y_{k-1}, \cdots, y_{k-p}; k; \theta(f))\}^2.$$

In order to obtain the estimate $\hat{\sigma}_N^2(f, \hat{\theta}_N(f))$ and $\hat{\theta}_N(f)$ concretely, we have to use some nonlinear optimization technique because of the nonlinearity of $L(N, f, \theta(f))$ with respect to $\theta(f)$. For the detailed aspect of nonlinear optimization techniques, see, e.g. [P2].

Then, the system structure is estimated as f giving the minimum value of (6.13). In the sequel, the estimated structure is represented by $\hat{f}_N$. Closely related criterion functions were derived by the Bayesian argument in Refs. [A5], [S8] and the shortest data description concept in Ref. [R1]. It should be emphasized that the proposed criterion function could be derived without assuming the stationary or ergodic properties of the output process $\{y_k\}$ and without assuming the linearity of the system. The representative criterion functions for structure determination are listed in Table 6.1.

Table 6.1 Comparison with Representative Criterion Functions

	AIC (Akaike[A2])	BIC (Akaike[A5] and Schwarz[S8])	Hannan and Quinn's criterion[H1]	Rissanen's criterion[R1],[R2]	Proposed criterion
basic concept for derivation of criterion function	ninimum Kullback-Leibler Information associated with estimated and true distribution	Bayesian extension of the maximum likelihood estimation	modefied AIC (or BIC) with lesser possibility of over- and under-estimation	minimum estimated number of bits describing observed data	minimum upper bound of estimation error entropy
criterion function	-2log(maximum likelihood)+2m	-2log(maximum likelihood)+mlog N	Nlog $\hat{\sigma}_N^2$ +2mc ×log log N (c:constant such that c≧1)	-log(maximum likelihood)+mlog N	$\frac{N}{2}$ log $\hat{\sigma}_N^2$ + mlog N
assumed model or assumed property of observation data	linear stationary models (AR, ARMA models, etc.)	independent and identically distributed observation data	AR models	nonlinear time-invariant models	nonlinear time-varying models

where m denotes the number of unknown parameters and $\hat{\sigma}_N^2$ denotes the estimate of the variance of the input sequence

6.3 Structure Determination for Linear Models in Parameters

First, for convenience of discussion, consider the nonlinear function $f_s(\cdot)$ in (6.1) as

$$(6.18)\quad f_s(y_{k-1}, y_{k-2}, \cdots, y_{k-n}; k; \hat{\theta}) = \sum_{i=1}^{n} f_s^{(i)}(y_{k-1}, \cdots, y_{k-i}; k; \hat{\theta}).$$

For instance, if $f_s(\cdot)$ is given by

$$f_s(y_{k-1}, \cdots, y_{k-n}; k; \hat{\theta}) = g_1(y_{k-1}; \hat{\theta}) + g_2(y_{k-1}, y_{k-3}; \hat{\theta}) + g_3(y_{k-2}, y_{k-n}; \hat{\theta}),$$

then $f_s^{(i)}(\cdot)$ can be given by

$$f_s^{(i)}(y_{k-1}, \cdots, y_{k-i}; k; \hat{\theta}) = \begin{cases} g_1(y_{k-1}; \hat{\theta}) & \text{for } i=1 \\ g_2(y_{k-1}, y_{k-3}; \hat{\theta}) & \text{for } i=3 \\ g_3(y_{k-2}, y_{k-n}; \hat{\theta}) & \text{for } i=n \\ 0 & \text{otherwise.} \end{cases}$$

In the case where $f_s(\cdot)$ is not decomposed in any sense, $f_s^{(i)}(\cdot)$ is given by

$$f_s^{(i)}(y_{k-1}, \cdots, y_{k-i}; k; \hat{\theta}) = \begin{cases} 0 & \text{for } i=1,2,\cdots,n-1 \\ f_s(y_{k-1}, \cdots, y_{k-n}; \hat{\theta}) & \text{for } i=n. \end{cases}$$

Furthermore, we assume that each nonlinear function $f_s^{(i)}(\cdot)$ can be expanded into

$$(6.19)\quad f_s^{(i)}(y_{k-1}, \cdots, y_{k-i}; k; \hat{\theta}) = \sum_{j=1}^{\ell_i} \hat{\theta}_{ij} f_{s,ij}(y_{k-1}, \cdots, y_{k-i}; k),$$

where $\mathring{\theta}_{ij}$ is the unknown constant parameter, which is the component of

(6.20) $\mathring{\theta} \triangleq [\ \mathring{\theta}_{11} \cdots \mathring{\theta}_{1\ell_1} ; \cdots ; \mathring{\theta}_{n1} \cdots \mathring{\theta}_{n\ell_n}]'$.

Substituting (6.18) and (6.19) into (6.1), it follows that

(6.21) $y_k = \sum_{i=1}^{n} \sum_{j=1}^{\ell_i} \mathring{\theta}_{ij} f_{s,ij}(y_{k-1}, \cdots, y_{k-i} ; k) + b_k v_k$.

For the true system $(f_s, \mathring{\theta})$ described by (6.21), the model set M is assumed to be $\{(f, \theta(f)) \mid f \in F,\ \theta(f) \in D_f\}$ where F is given by

(6.22) $F = \{f \mid f(z_1, \cdots, z_p ; k ; \theta(f)) = \sum_{i=1}^{p} \sum_{j=1}^{q_i} \theta_{ij}(f) f_{p,ij}(z_1, \cdots, z_i ; k)\}$

and $\theta_{ij}(f)$ is the element of

(6.23) $\theta(f) \triangleq [\ \theta_{11}(f) \cdots \theta_{1q_1}(f) ; \cdots ; \theta_{p1}(f) \cdots \theta_{pq_p}(f)]'$.

Hence, the criterion function $\ell_s(\cdot)$ for the structure determination of the system (6.21) is given by

$$\ell_s(N, f, \hat{\theta}_N(f)) \triangleq \frac{N}{2} \log \hat{\sigma}_N^2(f, \hat{\theta}_N(f)) + m \log N.$$

where

(6.24) $m = q_1 + q_2 + \cdots + q_p$.

In the case where the set of functions F is given by (6.22), the function $f \in F$ is linear with respect to $\theta(f)$. Hence, estimates of $\theta(f)$ and $\sigma^2(f, \theta(f))$ are respectively given by

(6.25) $\hat{\theta}_N(f) = Q_N^{-1}(f) s_N(f)$

$$(6.26) \qquad \hat{\sigma}_N^2(f,\hat{\theta}_N(f)) = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{b_k^2}\{y_k - f(y_{k-1},\cdots,y_{k-p};k;\hat{\theta}_N(f))\}^2$$

where

$$(6.27) \qquad s_N(f) \triangleq \sum_{k=1}^{N} \frac{1}{b_k} F_{k-1}(f)y_k$$

$$(6.28) \qquad Q_N(f) \triangleq \sum_{k=1}^{N} F_{k-1}(f)F'_{k-1}(f)$$

$$(6.29) \qquad F_{k-1}(f) \triangleq \frac{1}{b_k}[f_{p,11}(y_{k-1};k) \cdots f_{p,1q_1}(y_{k-1};k); \cdots ;f_{p,p1}(y_{k-1},\cdots,y_{k-p};k) \cdots f_{p,pq_p}(y_{k-1},\cdots,y_{k-p};k)]'.$$

In order to avoid numerical difficulties, we set $\hat{\theta}_N(f)$ at an arbitrary point in D_f, if $\hat{\theta}_N(f) \in D_f$ or (6.25) is not well defined because of the singularity of $Q_N(f)$.

6.4 Asymptotic Properties of Estimators for Linear Models in Parameters

We restrict our discussion to nonlinear system models whose unknown parameters are included linearly in $f_s(\cdot)$. First, we present the theorem concerned with the consistency of $\hat{\theta}_N(f)$ and $\hat{\sigma}_N^2(f, \hat{\theta}_N(f))$.

[Theorem 6.1] Assume that in addition to the condition (C-1), the following conditions hold:

(C-2) $(f_s, \mathring{\theta})$ is stable in the sense that the output process $\{y_k\}$ satisfies $E\{y_k^4\} \leq c$ (const.) for $k=1, 2, \cdots$;

(C-3) $(f_s, \mathring{\theta}) \in M$ and $D_{f_s} \subset D_f$;

(C-4) $\{f_{p,ij}(z_1, \cdots, z_i; k)\}$ satisfies the growth condition, i.e., for any real number z_m $(m=1, 2, \cdots, i)$,

$$| f_{p,ij}(z_1, \cdots, z_i; k) | \leq c(\sum_{m=1}^{i} z_m^2)^{1/2} + c \quad (i=1, \cdots, p; j=1, \cdots, q_i);$$

and

$$\text{(C-5)} \quad \liminf_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} E\{F_{k-1}(f) F_{k-1}'(f) \mid \mathscr{F}_{k-2}\} \geq cI > 0 \qquad \text{w.p. } 1.$$

Then,

$$\text{(6.30)} \quad \hat{\theta}_N(f) \to \mathring{\theta}(f) \qquad \text{w.p. } 1 \quad \text{as} \quad N \to \infty$$

$$\text{(6.31)} \quad \hat{\sigma}_N^2(f, \hat{\theta}_N(f)) \to \sigma^2 \qquad \text{w.p. } 1 \quad \text{as} \quad N \to \infty,$$

where

$$\text{(6.32)} \quad \mathring{\theta}(f) = [\underbrace{\mathring{\theta}_{11} \cdots \mathring{\theta}_{1\ell_1} 0 \cdots 0}_{q_1}; \cdots ; \underbrace{\mathring{\theta}_{n1} \cdots \mathring{\theta}_{n\ell_n} 0 \cdots 0}_{q_n}; \underbrace{0 \cdots 0}_{q_{n+1}+ \cdots +q_p}]'.$$

The proof is given in Appendix 6.A. Setting $f(\cdot)$ as $f_s(\cdot)$ in Theorem 6.1, we have immediately Corollary 6.1.

[Corollary 6.1] Assume that (C-1), (C-2) hold and that, instead of (C-3), (C-4) and (C-5) in Theorem 6.1, the following conditions hold:

(C-3)' $(f_s, \mathring{\theta}) \in M$;

(C-4)' $\{f_{s,ij}(z_1, \cdots, z_i; k)\}$ satisfies the growth condition, i.e., for any real number z_m $(m=1, 2, \cdots, i)$,

$$| f_{s,ij}(z_1, \cdots, z_i; k) | \leq c(\sum_{m=1}^{i} z_m^2)^{1/2} + c \quad (i=1, \cdots, n; j=1, \cdots, \ell_i);$$

and

(C-5)' $\liminf_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N} E\{F_{k-1}(f_s) F'_{k-1}(f_s) \mid \mathcal{F}_{k-2}\} \geq cI > 0$ w.p. 1.

Then,

(6.33) $\hat{\theta}_N(f_s) \to \mathring{\theta}$ w.p. 1 as $N \to \infty$

(6.34) $\hat{\sigma}_N^2(f_s, \hat{\theta}_N(f_s)) \to \sigma^2$ w.p. 1 as $N \to \infty$.

Regarding the asymptotic property of the criteron function for the structure determination, we have the following theorem.

[Theorem 6.2] Assume that (C-1), (C-2), (C-3)', (C-4) and (C-5)' hold. Then, for a sufficiently large N, we have

(6.35) $\ell_s(N, f_s, \hat{\theta}_N(f_s)) \leq \ell_s(N, f, \hat{\theta}_N(f))$ w.p. 1

where f is any element of F given by (6.22).

The proof of Theorem 6.2 is given in Appendix 6.B. Theorem 6.2 suggests that the determined system structure $\hat{f}_N$ becomes asymptotically an element of

(6.36) $G_s = \{f \mid \lim_{N \to \infty} | \ell_s(N, f, \hat{\theta}_N(f)) - \ell_s(N, f_s, \hat{\theta}_N(f_s)) | = 0,$ w.p. 1$\}$.

Note that the number of elements of G_s consists not necessarily of one but usually of more than 2 elements. Restricting domains of unknown parameters D_f and D_{f_s} as $D_f \subset D_{f_s}$ or $D_{f_s} \subset D_f$, the following two theorems are obtained.

[Theorem 6.3] Assume that the same conditions as those in Theorem 6.2 hold except for (C-3)'. Instead of (C-3)', assume that

(C-3)" $(f_s, \mathring{\theta}) \in M$ and $D_f \subset D_{f_s}$.

Then, for a sufficiently large N, the following inequality holds:

$$(6.37) \quad \ell_s(N, f_s, \hat{\theta}_N(f_s)) < \ell_s(N, f, \hat{\theta}_N(f)) \qquad \text{w.p. } 1,$$

where f is any element of F given by (6.22) except for f_s.

[Theorem 6.4] Assume that the same conditions as those in Theorem 6.1 hold. Then, for any element of F given by (6.22) except for f_s, we have

$$(6.38) \quad \ell_s(N, f_s, \hat{\theta}_N(f_s)) < \ell_s(N, f, \hat{\theta}_N(f)) \qquad \text{in prob.}$$

for a sufficiently large N.

Proofs of Theorems 6.3 and 6.4 are given in Appendices 6.C and 6.D respectively. From Theorems 6.3 and 6.4, the following corollary is immediately obtained.

[Corollary 6.2] Assume that the same conditions as those in Theorem 6.1 hold except for (C-3). Further assume that

(C-6) $(f_s, \mathring{\theta}) \in M$ and $D_{f_s} \subset D_f$ or $D_f \subset D_{f_s}$.

Then, for any element of F given by (6.22) except for f_s, we have

$$(6.39) \quad \ell_s(N, f_s, \hat{\theta}_N(f_s)) < \ell_s(N, f, \hat{\theta}_N(f)) \qquad \text{in prob.}$$

for a sufficiently large N.

Corollary 6.2 suggests that if the model set M is given by

$$(6.40) \qquad M = \{(f,\theta(f)) \mid f \in \{f_1,\cdots,f_K\},\ \theta(f_i) \in D_{f_i} : D_{f_1} \subset D_{f_2} \subset \cdots \subset D_{f_K}\},$$

the determined system structure $\hat{f}_N$ is weakly consistent, i.e.,

$$(6.41) \qquad \hat{f}_N \rightarrow f_s \qquad \text{in prob.} \qquad \text{as} \quad N \rightarrow \infty.$$

Furthermore, setting the model set M the same as (6.40), and selecting f_i when $D_{f_i} \subset D_{f_j}$ and $\ell_s(N,f_i,\hat{\theta}_N(f_i))=\ell_s(N,f_j,\hat{\theta}_N(f_j))$, then Theorems 6.2 and 6.3 suggest

$$(6.42) \qquad \hat{f}_N \rightarrow f_s \qquad \text{w.p. } 1 \qquad \text{as} \quad N \rightarrow \infty.$$

We note that if the nonlinear function $f_s(\cdot)$ can be represented by

$$(6.43) \qquad f_s(z_1,\cdots,z_n;k;\mathring{\theta}) = a_s(z_1,\cdots,z_n;k;\mathring{\theta})(z_n \ \cdots \ z_1)' + \phi_s(z_1,\cdots,z_n;k;\mathring{\theta}),$$

$$(a_s(\cdot): \text{n-dim. row vector})$$

the condition (C-2) is replaced by

$$(\text{C-2})' \quad \begin{cases} \max\limits_{i} | \lambda_i\{A_s(z_1,\cdots,z_n;k;\mathring{\theta})\} | < 1 & (i=1,2,\cdots,n) \\ | \phi_s(z_1,\cdots,z_n;k;\mathring{\theta}) | < c, & \end{cases}$$

where $\lambda_i\{*\}$ is the i-th eigenvalue of "*" and $A_s(\cdot)$ is the n-dimensional square matrix defined by

$$(6.44) \qquad A_s(z_1,\cdots,z_n;k;\mathring{\theta}) = \left[\begin{array}{cccc} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ \hdashline \multicolumn{4}{c}{a_s(z_1,\cdots,z_n;k;\mathring{\theta})} \end{array} \right].$$

The derivation of (C-2)' from (C-2) is shown in Appendix 6.E.

6.5 Digital Simulation Studies

In simulation studies, we use

$$(6.45)\quad \begin{cases} y_k = f_s(y_{k-1}, y_{k-2}; \mathring{\theta}) + bv_k \\ \quad y_0 = y_{-1} = 0 \end{cases}$$

where $\{v_k\}$ is the independent Gaussian noise of $N(0, \sigma^2)$, and $f_s(\cdot)$ is given by

$$(6.46)\quad f_s(y_{k-1}, y_{k-2}; \mathring{\theta}) = \mathring{\theta}_{11} y_{k-1} + \mathring{\theta}_{21} g(y_{k-2}) y_{k-2}$$

and

$$(6.47)\quad g(y_{k-2}) = \begin{cases} 1 + \gamma y_{k-2}^2 & \text{for } y_{k-2}^2 \leq \mu \\ 1 + \gamma\mu \quad (>0) & \text{for } y_{k-2}^2 > \mu . \end{cases}$$

From (6.46), it follows that

$$(6.48)\quad f_s(y_{k-1}, y_{k-2}; \mathring{\theta}) = a_s(y_{k-2}, \mathring{\theta})(y_{k-2} \quad y_{k-1})'$$

where

$$(6.49)\quad a_s(y_{k-2}, \mathring{\theta}) \triangleq (\mathring{\theta}_{21} g(y_{k-2}) \quad \mathring{\theta}_{11}).$$

Hence, if all eigenvalues λ_i (i=1,2) of A_s defined by

$$(6.50)\quad A_s(y_{k-2}, \mathring{\theta}) \triangleq \begin{bmatrix} 0 & 1 \\ \mathring{\theta}_{21} g(y_{k-2}) & \mathring{\theta}_{11} \end{bmatrix}$$

are $|\lambda_i| < 1$, the condition (C-2)' is satisfied. Since the characteristic equation of (6.50) is given by

$$\lambda^2 - \mathring{\theta}_{11}\lambda - \mathring{\theta}_{21}g(y_{k-2}) = 0, \tag{6.51}$$

it is easily verified that $|\lambda_i| < 1$ (i=1, 2), if

$$\begin{cases} \mathring{\theta}_{21}g(z) > -1 & \text{for all } z \\ \mathring{\theta}_{11} + \mathring{\theta}_{21}g(z) < 1 & \text{for all } z \\ \mathring{\theta}_{11} - \mathring{\theta}_{21}g(z) > -1 & \text{for all } z\ . \end{cases} \tag{6.52}$$

It is obvious that the input noise sequence $\{v_k\}$ satisfies the condition (C-1). For the system model (6.45), $F_{k-1}(f_s)$ defined by (6.29) is given by

$$F_{k-1}(f_s) = \frac{1}{b}\,[y_{k-1} \quad g(y_{k-2})y_{k-2}]'. \tag{6.53}$$

Then, it follows that

$$\frac{1}{N}\sum_{k=1}^{N} E\{F_{k-1}(f_s)F'_{k-1}(f_s) \mid \mathcal{F}_{k-2}\} \tag{6.54}$$

$$= \frac{1}{b^2}\,\frac{1}{N}\sum_{k=1}^{N}\begin{bmatrix} f_s^2(y_{k-2}, y_{k-3};\mathring{\theta})+b^2\sigma^2 & y_{k-2}f_s(y_{k-2}, y_{k-3};\mathring{\theta})g(y_{k-2}) \\ y_{k-2}f_s(y_{k-2}, y_{k-3};\mathring{\theta})g(y_{k-2}) & g^2(y_{k-2})y_{k-2}^2 \end{bmatrix}.$$

Hence, for an arbitrary vector $x \triangleq [x_1 \quad x_2]'$ $(\neq 0)$, we have

$$x'\{\frac{1}{N}\sum_{k=1}^{N} E\{F_{k-1}(f_s)F'_{k-1}(f_s) \mid \mathcal{F}_{k-2}\}\}x \tag{6.55}$$

$$= \frac{1}{b^2}\,\frac{1}{N}\sum_{k=1}^{N}\{f_s(y_{k-2}, y_{k-3};\mathring{\theta})x_1 + g(y_{k-2})y_{k-2}x_2\}^2 + \sigma^2 x_1^2.$$

Since it is easily verified that

$$\liminf_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} y_{k-1}^2 \geq b^2\sigma^2 > 0 \qquad \text{w.p. 1,} \tag{6.56}$$

it follows that

$$(6.57)\qquad \liminf_{N\to\infty} x'\{\frac{1}{N}\sum_{k=1}^{N} E\{F_{k-1}(f_s)F'_{k-1}(f_s) \mid \mathcal{F}_{k-2}\}\}x$$

$$\geq \begin{cases} \sigma^2 x_1^2 & \text{if } x_1 \neq 0 \\ \liminf_{N\to\infty} \frac{1}{b^2}\frac{1}{N}\sum_{k=1}^{N} g^2(y_{k-2})y_{k-2}^2 x_2^2 \geq \sigma^2 x_2^2 & \text{if } x_1 = 0, \end{cases}$$

which implies that the condition (C-5)' is satisfied.

Along the procedure shown in Fig. 6.1, the system structure determination is performed by the following steps:

(1) *Collection of observed data.* Obtain the observed data $Y_N \triangleq \{y_1, y_2, \cdots, y_N\}$ up to the preassigned step N from the true system at hand. In digital simulation studies, the true values of unknown parameters were set as $\mathring{\theta}_{11}=1.0$, $\mathring{\theta}_{21}=-0.2$ and $\sigma^2=1.0$, and known coefficients were b=0.1, $\gamma=0.5$ and $\mu=4.0$. With these values of coefficients, the condition given by (6.52) is satisfied. The observation data Y_N is obtained by simulating (6.45) on a digital computer. A typical sample run of the observation process $\{y_k\}$ is depicted in Fig. 6.2.

(2) *Assignment of model set* M. In digital simulation studies, we concentrate our attention on the order determination only for simplicity of discussion. Therefore, the model set M is restricted to

$$(6.58)\qquad M = \{(f_p, \theta(p)) \mid \theta(p) \in D_p;\ p=1,2,\cdots,K\}$$

where

$$(6.59)\qquad f_p(z_1,\cdots,z_p;\theta(p)) \triangleq \sum_{i=1}^{p} \theta_{i1}(p)f_{i1}(z_i)$$

$$(6.60)\qquad f_{i1}(z_i) \triangleq z_i \quad (i=1,2,\cdots,p-1)$$

(6.61) $$f_{p1}(z_p) \triangleq g(z_p)z_p.$$

(6.62) $$\theta(p) \triangleq [\ \theta_{11}(p) \ \cdots \ \theta_{p1}(p)]'.$$

Since we are free to choose the maximum order of K, throughout simulation experiments, K was set as 5. From (6.46), (6.59), (6.60) and (6.61), conditions (C-3)', (C-4) and (C-4)' are clearly satisfied when the set D_p (p=1, 2, ⋯, K) is taken to be sufficiently large. Hence, all conditions of Corollary 6.1 and Theorem 6.2 are satisfied.

(3) *Comparison of* ℓ_o *(N,p).* Using the observed data Y_N, compute (6.25) and (6.26) for p=1, 2, ⋯, K. For the system model given by (6.58), the estimated $\theta(p)$ and $\sigma^2(p) \triangleq \sigma^2(f_p, \theta(p))$ are given by

(6.63) $$\hat{\theta}_N(p) = Q_N^{-1}(p)s_N(p)$$

(6.64) $$\hat{\sigma}_N^2(p) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{b^2} \{y_k - f_p(y_{k-1}, \cdots, y_{k-p}; \hat{\theta}_N(p))\}^2$$

where

(6.65) $$\begin{cases} \hat{\theta}_N(p) \triangleq [\ \hat{\theta}_{11}(p) \ \cdots \ \hat{\theta}_{p1}(p)]' \\ s_N(p) \triangleq \frac{1}{b} \sum_{k=1}^{N} F_{k-1}(p) y_k \\ Q_N(p) \triangleq \sum_{k=1}^{N} F_{k-1}(p) F_{k-1}'(p). \\ F_{k-1}(p) \triangleq \frac{1}{b} [f_{11}(p) \ f_{21}(p) \ \cdots \ f_{p1}(p)]'. \end{cases}$$

Using the estimate $\hat{\sigma}_N^2(p)$ obtained by (6.64), the criterion function for order determination is given by

$$(6.66) \quad \ell_o(N,p) \triangleq \ell_s(N, f_p, \hat{\theta}_N(p)) = \frac{N}{2} \log \hat{\sigma}_N^2(p) + p \log N.$$

Figure 6.3 plots $\ell_o(N,p)$-function at every 500 steps.

(4) *Determination of system order.* Accept p as the system order for which the criterion function $\ell_o(N,p)$ takes its minimum value, and determine the system order to be p. Then, adopt $\hat{\theta}_N(p)$ as estimates of unknown system parameters. From Fig. 6.3, we can conclude that $\ell_o(N,2)$ takes its minimum value for $N \geq 1000$, and that the system order is determined to be p=2. Values of $\{\hat{\theta}_{i1}(p); i=1, \cdots, p\}$, $\hat{\sigma}_N^2(p)$ and $\ell_o(N,p)$ at N=3000 are listed for p= 1, 2, ⋯, 5 in Table 6.2. From Table 6.2 and Fig. 6.3, we may fairly say that the parameter identification is well achieved as well as the determination of the system order.

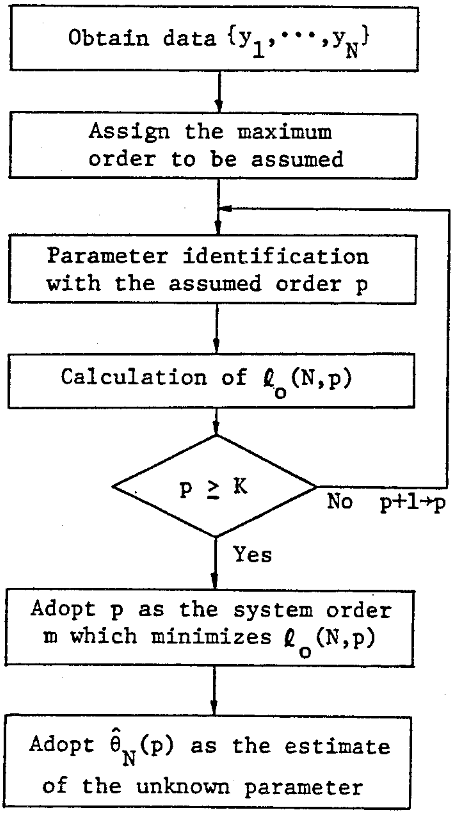


Fig. 6.1 Illustration of the procedure for system order determination and parameter identification

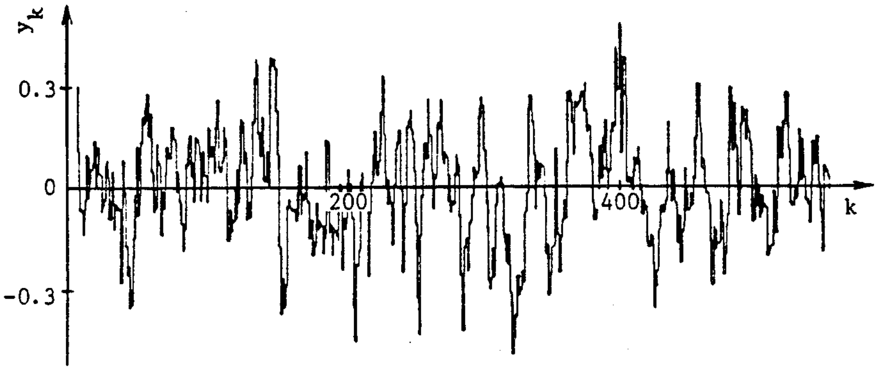


Fig. 6.2 A typical sample run of the observation process $\{y_k\}$

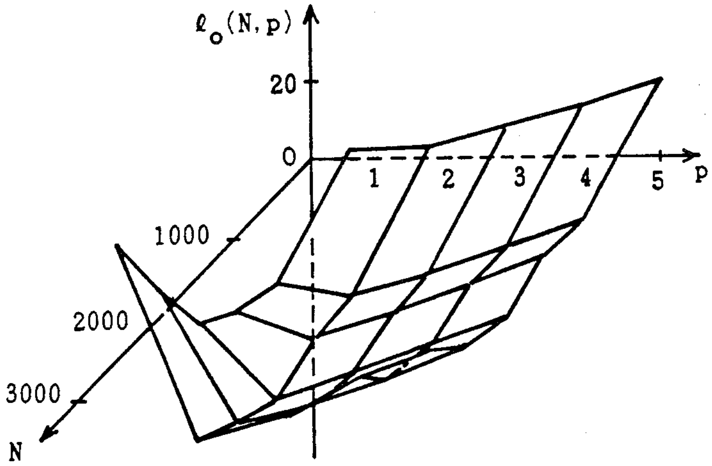


Fig. 6.3 $\ell_o(N,p)$-function at every 500 steps.

Table 6.2 Numerical comparisons of the estimated parameters with the assumed order p ($1 \leq p \leq 5$) at N=3000.

Order p	$\hat{\theta}_{11}(p)$	$\hat{\theta}_{21}(p)$	$\hat{\theta}_{31}(p)$	$\hat{\theta}_{41}(p)$	$\hat{\theta}_{51}(p)$	$\hat{\sigma}_N^2(p)$	$\ell_o(N,p)$
1	0.793	——	——	——	——	1.021	40.04
2	0.992	-0.183	——	——	——	0.982	-10.65
3	0.990	-0.175	-0.185	——	——	0.981	-4.67
4	0.990	-0.174	-0.020	-0.001	——	0.981	3.24
5	0.990	-0.174	-0.021	0.004	-0.005	0.981	11.21

6.6 Application to Geophysical Data

In this section, an important aspect of the application of the present method to real earthquake data is described. As a typical example, a sample path of the earthquake data depicted in Fig. 6.4 was taken into account which was already presented in Ref. [T2]. According to Ref. [T2], Fig. 6.4 is the acceleration record (accelerogram) of the earthquake at San Fernando, California, U.S.A. in 1971 which was observed at the Pacoima Dam near the epicenter. The outstanding features of the earthquake data are that (i) the data exhibits a rise to the maximum level and then a fall to the background noise level in a short period of time, and (ii) it also exhibits sporadically large peak values. The feature (i) mentioned above shows that the data may be a sample run of a nonstationary stochastic process and the feature (ii) shows the nonlinearity of the mechanism which generates the earthquake. Therefore, we assume that the acceleration record of the earthquake shown in Fig. 6.4 was generated by the following input/output relation:

$$(6.67)\quad \begin{cases} y_k = \sum_{i=1}^{n-1} \hat{\theta}_{i1} y_{k-i} + g(y_{k-n};\hat{\theta}) + b_k v_k \\ \\ y_0 = y_{-1} = \cdots = y_{-n+1} = 0. \end{cases}$$

The nonlinear function $g(y_{k-n};\hat{\theta})$ in (6.67) is assumed to be expanded into the following orthogonal series:

$$(6.68)\quad g(y_{k-n};\hat{\theta}) = \sum_{j=1}^{\ell} \hat{\theta}_{nj} H_{j-1}(y_{k-n}) \exp\{-\frac{1}{2}\, y_{k-n}^2\}$$

where $\{H_j(\cdot);j=0,1,\cdots,\ell-1\}$ are the Hermite polynomials, and $\{\hat{\theta}_{i1};i=1,2,\cdots,n-1\}$ and $\{\hat{\theta}_{nj};j=1,\cdots,\ell\}$ are constants and need to be identified. Considering that the envelope of the data loosely represents the values of $\{b_k\}$, the values of b_k in (6.67) are preassigned by fitting the cubic spline curve to several representative points of $\{ | y_k | ; k=1, \cdots, N\}$. A result of the cubic spline fitting is shown in Fig. 6.5.

Since the number of terms ℓ of the series (6.68) can not be preassigned, we decide ℓ as well as the system order n from the observation data. Hence as the model set, we use $M=\{(f, \theta(p,q)) | f \in F, \theta(p,q) \in D_f\}$, where

$$(6.69) \quad F = \{ f | f(z_1, z_2, \cdots, z_p; \theta(f)) = \sum_{i=1}^{p-1} \theta_{i1}(f) z_i + \sum_{j=1}^{q} \theta_{pj}(f) H_{j-1}(z_p) \exp\{- \frac{1}{2} z_p^2\} ; \; p=1, 2, \cdots, K \text{ and } q=1, 2, \cdots, L\}$$

and the maximum system order to be assumed and the maximum number of expansion were set as K=7 and L=7, respectively.

The criterion function for the determination n and ℓ is given by

$$(6.70) \quad \ell_s(N, p, q) \triangleq \ell_s(N, f, \hat{\theta}_N(p, q)) = \frac{N}{2} \log \hat{\sigma}_N^2(p, q) + (p + q - 1) \log N$$

where $\hat{\sigma}_N^2(p,q)$ is the estimated variance of modeling error with the assumption that the system order is p and the terms of series expansion of $g(\cdot)$ is q. $\hat{\sigma}_N^2(p,q)$ and the estimate $\hat{\theta}_N(p,q)$ are given respectively by

$$(6.71) \quad \hat{\sigma}_N^2(p, q) = \frac{1}{N} \sum_{k=0}^{N} \{y_k - f(y_{k-1}, \cdots, y_{k-n}; \hat{\theta}_N(p, q))\}^2$$

$$(6.72) \quad \hat{\theta}_N(p, q) = Q_N^{-1}(p, q) s_N(p, q)$$

where

$$
(6.73)\quad
\begin{cases}
\hat{\theta}_N(p,q) = [\ \hat{\theta}_{11}(p,q) \cdots \hat{\theta}_{p-1,1}(p,q); \\
\qquad\qquad\qquad \hat{\theta}_{p1}(p,q) \quad \hat{\theta}_{p2}(p,q) \cdots \hat{\theta}_{pq}(p,q)]' \\
Q_N(p,q) = \sum_{k=1}^{N} F_{k-1}(p,q)F'_{k-1}(p,q) \\
s_N(p,q) = \sum_{k=1}^{N} \frac{1}{b_k} F_{k-1}(p,q)y_k \\
f(y_{k-1},\cdots,y_{k-n};\hat{\theta}_N(p,q)) = \sum_{i=1}^{p-1} \hat{\theta}_{i1}(p,q)f_{i1}(y_{k-i}) \\
\qquad\qquad\qquad + \sum_{j=1}^{q} \hat{\theta}_{pj}(p,q)f_{pj}(y_{k-p}) \\
F_{k-1}(p,q) = \frac{1}{b_k} [f_{11}(y_{k-1}) \cdots f_{p-1,1}(y_{k-p+1}); \\
\qquad\qquad\qquad f_{p1}(y_{k-p}) \cdots f_{pq}(y_{k-p})]' \\
f_{i1}(y_{k-i}) = y_{k-i} \qquad\qquad \text{for } i=1,2,\cdots,p-1 \\
f_{pj}(y_{k-p}) = H_{j-1}(y_{k-p})\exp\{-\frac{1}{2} y_{k-p}^2\} \qquad \text{for } j=1,2,\cdots,\ q.
\end{cases}
$$

A similar procedure to that in Section 6.5 is applicable to decide n and ℓ in the system model (6.82). Figure 6.6 shows results of the computation of the criterion function $\ell_s(N,p,q)$ at N=1700. From Fig. 6.6, we may conclude that n=3 and ℓ=2. The identified values of the unknown parameter $\hat{\theta}$ are [1.31 -0.42 0.02 -0.24]'. The modeled system with the identified parameter values was simulated on a digital computer to examine the success of the system modeling and its result is depicted in Fig. 6.7.

Using the same system model structure as given by (6.69), we tried to model two other real earthquake data. One of the real data used here is the acceleration record of the earthquake observed at El Centro, which occurred at Imperial Valley, California, U.S.A. in 1940 and the another one is of the San Francisco Earthquake in 1957, which was observed at Oakland. They are depicted respectively Figs. 6.8(a) and (b). The system structure and unknown parameters are identified by using the same procedure as that for the San Fernando Earthquake. Figures 6.9 and 6.10 show respectively the fitted spline curve of b_k and the values of the criterion function $\ell_s(N,p,q)$ at N=1800. The simulated output data with identified system models are shown in Fig. 6.11.

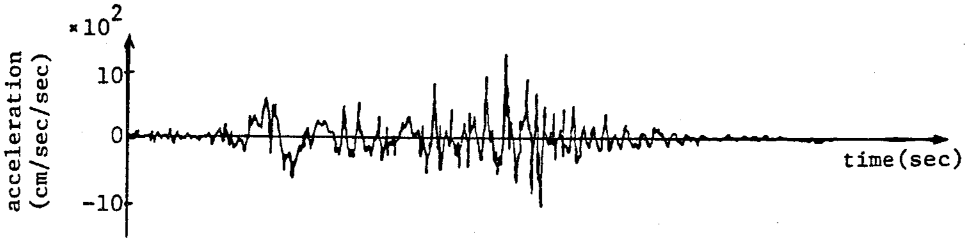


Fig. 6.4 Typical earthquake data-Accelerogram of the San Fernando Earthquake recorded at Pacoima dam, California, U.S.A. on Feb. 9, 1971 (given in Ref. [T2])

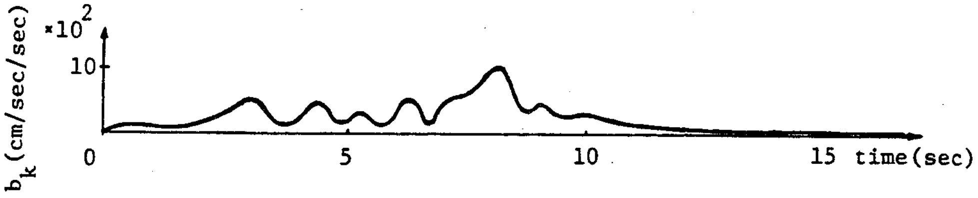


Fig. 6.5 Cubic spline fit of b_k (San Fernando Earthquake)

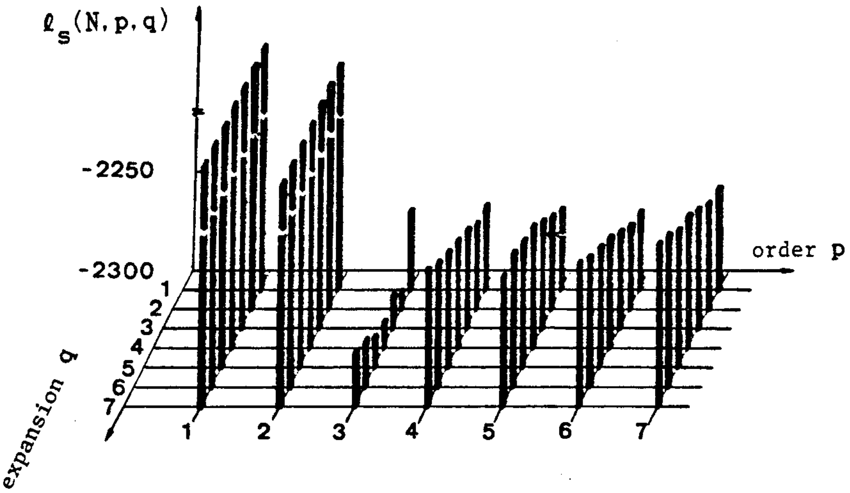


Fig. 6.6 Comparison of $\ell_s(N,p,q)$ at N=1700 (San Fernando Earthquake)

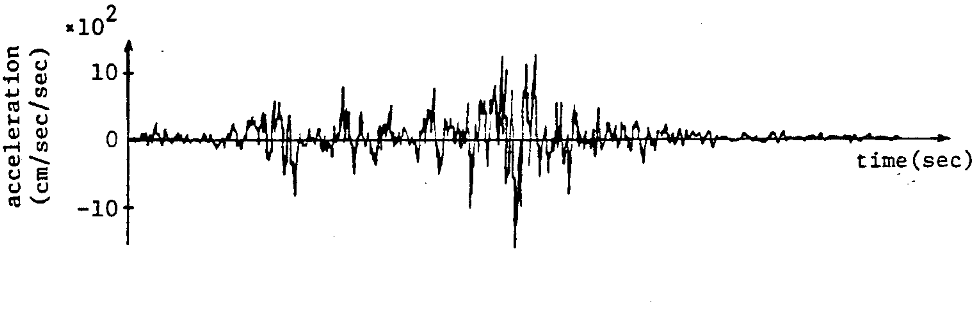
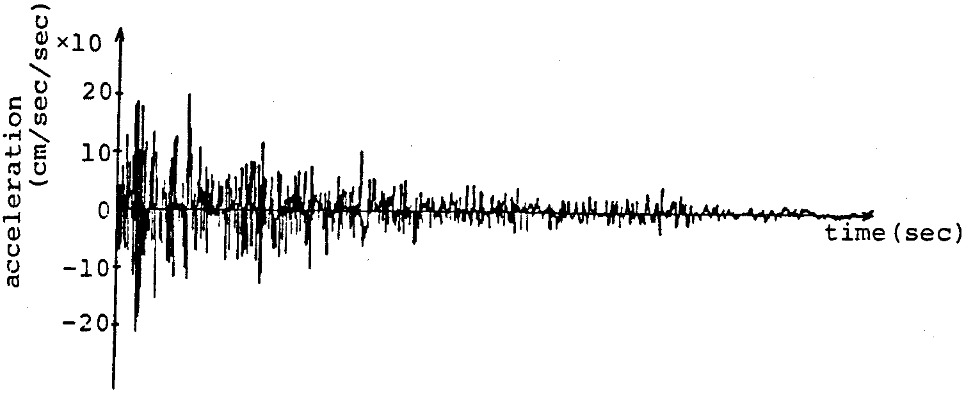
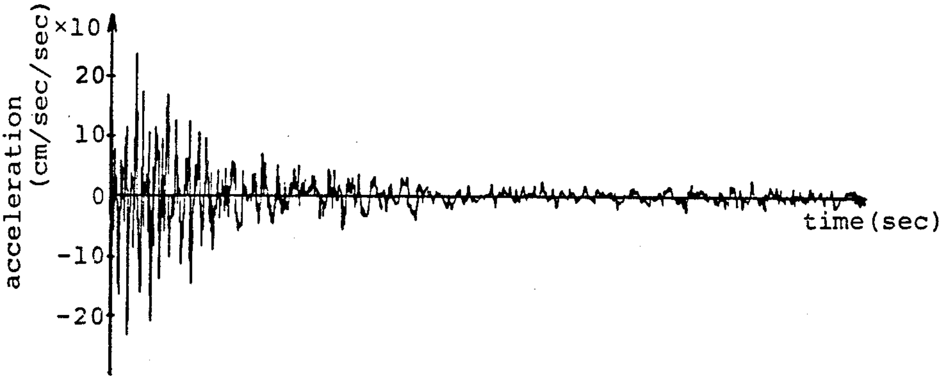


Fig 6.7 Accelerogram simulation of San Fernando Earthquake from the modeled system with identified parameters

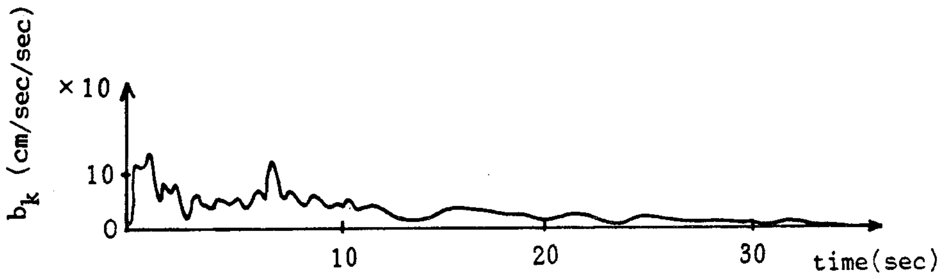


(a) Accelerogram of the Imperial Valley Earthquake

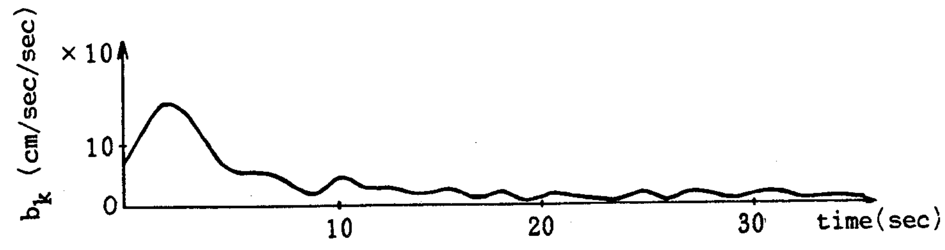


(b) Accelerogram of the San Francisco Earthquake

Fig 6.8 Accelerograms of the Imperial Valley Earthquake at El Centro, U.S.A. in 1940 and San Francisco Earthquake at Oakland, U.S.A. in 1957

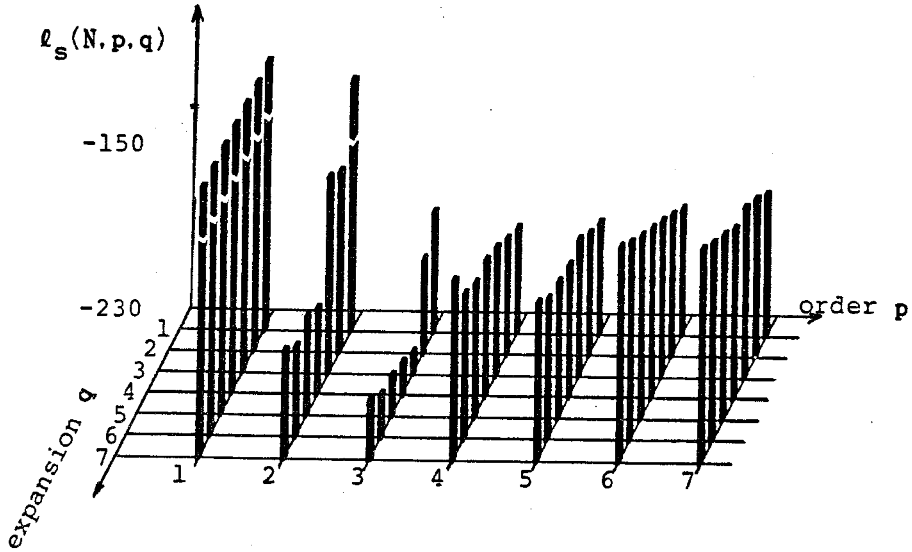


(a) The Imperial Valley Earthquake

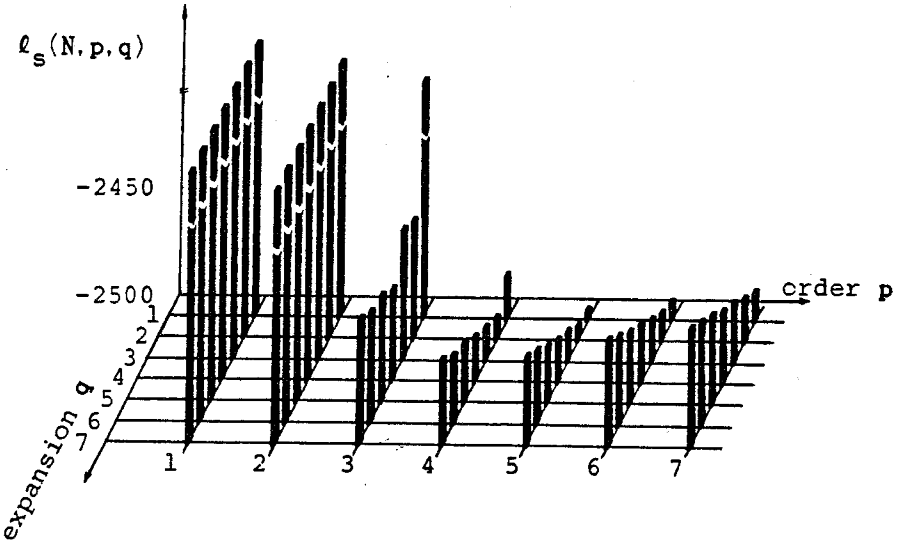


(b) The San Francisco Earthquake

Fig. 6.9 Cubic spline fits of b_k

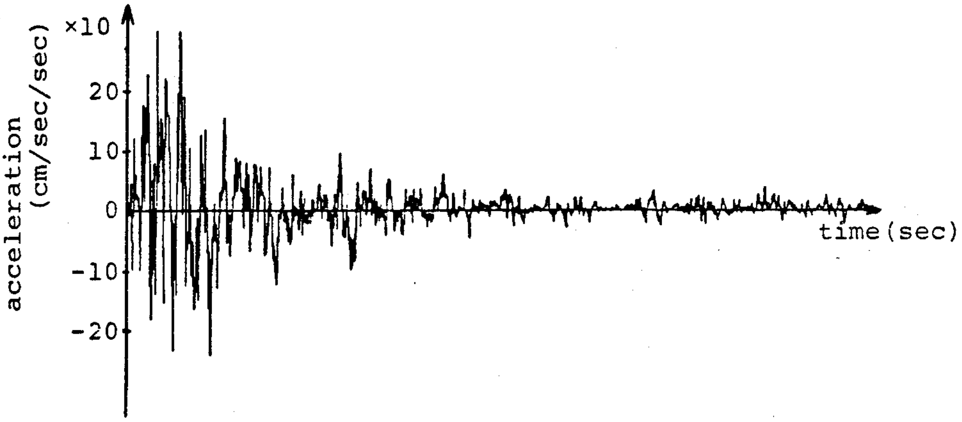


(a) The Imperial Valley Earthquake

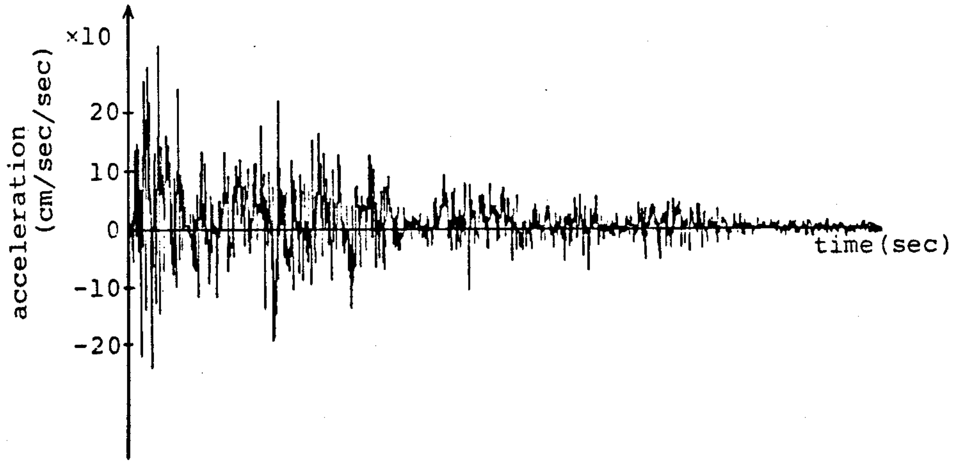


(b) The San Francisco Earthquake

Fig 6.10 Comparisons of $\ell_s(N,p,q)$ at N=1800



(a) The Imperial Valley Earthquake



(b) The San Francisco Earthquake

Fig. 6.11 Accelerogram simulations from the modeled systems with identified parameters

6.7 Discussions

The method for modeling a class of nonstationary nonlinear systems has been presented, where the underlying system model shows a high possibility of fitting real data, such as given in Section 6.6.

The criterion function for the structure determination of the system has been derived from evaluating the upper bound of the error entropy associated with both the input noise and unknown parameters. It has also been shown theoretically that, for a class of nonstationary nonlinear models like those given in Section 6.3, the determined system structure has salient asymptotic features and the estimated unknown parameters have the consistency in some stochastic sense.

The proposed criterion function for structure determination of system models has been applied to the modeling of the real data, and good performance has been achieved.

Appendix 6.A Proof of Theorem 6.1

First, we shall show the consistency property of $\hat{\theta}_N(f)$. The estimation error $\tilde{\theta}_N(f)$ is defined by

(A.1) $$\tilde{\theta}_N(f) \triangleq \mathring{\theta}(f) - \hat{\theta}_N(f).$$

Then, from (6.25), it follows that

(A.2) $$\tilde{\theta}_N(f) = \mathring{\theta}(f) - Q_N^{-1}(f)s_N(f).$$

Furthermore, noting that

(A.3) $$f_{s,ij}(z_1, \cdots, z_i ; k) = f_{p,ij}(z_1, \cdots, z_i ; k) \quad \text{for } i=1,2,\cdots,n;\ j=1,2,\cdots,\ell_i$$

because of $D_{f_s} \subset D_f$, we have from (6.21) and (6.27) that

(A.4) $$s_N(f) = Q_N(f)\mathring{\theta}(f) + \sum_{k=1}^{N} F_{k-1}(f)v_k.$$

Hence, substituting (A.4) into (A.2), we have

(A.5) $$\tilde{\theta}_N(f) = -Q_N^{-1}(f) \sum_{k=1}^{N} F_{k-1}(f)v_k.$$

Define

(A.6) $$q_N \triangleq \sum_{k=1}^{N} x'F_{k-1}(f)v_k$$

and

(A.7) $$r_N \triangleq \sum_{k=1}^{N} x'[F_{k-1}(f)F_{k-1}'(f) - E\{F_{k-1}(f)F_{k-1}'(f) \mid \mathcal{F}_{k-2}\}]x$$

where x is an arbitrary vector such that $x \neq 0$. It is easily verified that, with conditions (C-1), (C-2) and (C-4), $\{q_N, \mathcal{F}_{N-1}\}$ and $\{r_N, \mathcal{F}_{N-2}\}$ are martingales satisfying

(A.8) $E\{q_N^2\} \leq cN,$

(A.9) $E\{r_N^2\} \leq cN.$

Therefore, from Theorem 2.2, it follows that

(A.10) $\frac{1}{N} \sum_{k=1}^{N} F_{k-1}(f)v_k \to 0 \quad \text{w.p. } 1 \quad \text{as } N \to \infty$

(A.11) $\frac{1}{N} \sum_{k=1}^{N} [F_{k-1}(f)F'_{k-1}(f) - E\{F_{k-1}(f)F'_{k-1}(f) \mid \mathcal{F}_{k-2}\}] \to 0 \quad \text{w.p. } 1$

$\text{as } N \to \infty.$

Using the elementary inequality

$$\inf_{n>N} (a_n - b_n) \leq \inf_{n>N} a_n - \inf_{n>N} b_n,$$

and (C-5), we have from (A.11) that

(A.12) $\lim_{N\to\infty} \sup \left[\frac{1}{N} Q_N(f)\right]^{-1} < cI \quad \text{w.p. } 1.$

Hence, combining (A.10) with (A.12), the R.H.S. of (A.5) can be evaluated by

(A.13) $Q_N^{-1}(f) \sum_{k=1}^{N} F_{k-1}(f)v_k = \left[\frac{1}{N} Q_N(f)\right]^{-1} \frac{1}{N} \sum_{k=1}^{N} F_{k-1}(f)v_k \to 0 \quad \text{w.p. } 1$

$\text{as } N \to \infty,$

which implies (6.30).

Next, we shall show the consistency property of $\hat{\sigma}_N^2(f,\hat{\theta}_N(f))$. Using (6.21), (6.22), (6.26), (6.29) and (A.3), we have

$$(\mathrm{A.14}) \quad \hat{\sigma}_N^2(f,\hat{\theta}_N(f)) = \tilde{\theta}_N'(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)F_{k-1}'(f)\}\tilde{\theta}_N(f)$$

$$+ 2\tilde{\theta}_N'(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)v_k\} + \frac{1}{N}\sum_{k=1}^{N}v_k^2.$$

The first term of the R.H.S. of (A.14) can be rewritten by using (6.28) and (A.5) as

$$(\mathrm{A.15}) \quad \tilde{\theta}_N'(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)F_{k-1}'(f)\}\tilde{\theta}_N(f)$$

$$= \{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)v_k\}'NQ_N^{-1}(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)v_k\}.$$

Hence, it can be shown from (A.10) and (A.12) that

$$(\mathrm{A.16}) \quad \tilde{\theta}_N(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)F_{k-1}'(f)\}\tilde{\theta}_N(f) \rightarrow 0 \quad \text{w.p. 1} \quad \text{as} \quad N \rightarrow \infty.$$

From (A.5), (A.10) and (A.12), the second term of the R.H.S. of (A.14) is

$$(\mathrm{A.17}) \quad \tilde{\theta}_N'(f)\{\frac{1}{N}\sum_{k=1}^{N}F_{k-1}(f)v_k\} \rightarrow 0 \quad \text{w.p. 1} \quad \text{as} \quad N \rightarrow \infty.$$

Furthermore, it is easily shown by using Theorem 2.3, that

$$(\mathrm{A.18}) \quad \frac{1}{N}\sum_{k=1}^{N}v_k^2 \rightarrow \sigma^2 \quad \text{w.p. 1} \quad \text{as} \quad N \rightarrow \infty.$$

Hence, from (A.16), (A.17) and (A.18), we can conclude (6.31).

Appendix 6.B Proof of Theorem 6.2

Define $\Delta_N(f,f_s)$ by

$$\Delta_N(f,f_s) \triangleq \ell_s(N,f,\hat{\theta}_N(f)) - \ell_s(N,f_s,\hat{\theta}_N(f_s)) \tag{B.1}$$

$$= \frac{N}{2}\log\frac{\hat{\sigma}_N^2(f,\hat{\theta}_N(f))}{\hat{\sigma}_N^2(f_s,\hat{\theta}_N(f_s))} + (m_f - m_s)\log N,$$

where $m_f \triangleq \dim.\ \hat{\theta}_N(f)$ and $m_s \triangleq \dim.\ \hat{\theta}_N(f_s)$. From (6.1) and the definition of $\hat{\sigma}_N^2(f,\hat{\theta}_N(f))$ given by (6.26), it follows that

$$\begin{aligned}\hat{\sigma}_N^2(f,\hat{\theta}_N(f)) = {}& \frac{1}{N}\sum_{k=1}^{N}\frac{1}{b_k^2}\{f_s(y_{k-1},\cdots,y_{k-n};k;\mathring{\theta}) \\ &\qquad - f(y_{k-1},\cdots,y_{k-p};k;\hat{\theta}_N(f))\}^2 \\ &+ \frac{2}{N}\sum_{k=1}^{N}\frac{1}{b_k}\{f_s(y_{k-1},\cdots,y_{k-n};k;\mathring{\theta}) \\ &\qquad - f(y_{k-1},\cdots,y_{k-p};k;\hat{\theta}_N(f))\}v_k \\ &+ \frac{1}{N}\sum_{k=1}^{N}v_k^2.\end{aligned} \tag{B.2}$$

Since $f_s(\cdot)$ and $f(\cdot)$ can be represented by

$$\begin{cases} f_s(y_{k-1},\cdots,y_{k-n};k;\mathring{\theta}) = \mathring{\theta}'F_{k-1}(f_s) \\ f(y_{k-1},\cdots,y_{k-p};k;\hat{\theta}_N(f)) = \hat{\theta}_N'(f)F_{k-1}(f), \end{cases} \tag{B.3}$$

it is easily verified by Theorem 2.2 that

$$\begin{cases} \frac{1}{N}\sum_{k=1}^{N}\frac{1}{b_k}f_s(y_{k-1},\cdots,y_{k-n};k;\mathring{\theta})v_k \to 0 \quad \text{w.p. } 1 \quad \text{as} \quad N \to \infty \\ \frac{1}{N}\sum_{k=1}^{N}\frac{1}{b_k}f(y_{k-1},\cdots,y_{k-p};k;\hat{\theta}_N(f))v_k \to 0 \quad \text{w.p. } 1 \quad \text{as} \quad N \to \infty. \end{cases} \tag{B.4}$$

Hence, from (A.18) and (B.4) and for a sufficiently large N,

$$\text{(B.5)} \qquad \hat{\sigma}_N^2(f, \hat{\theta}_N(f)) \geq \sigma^2 \qquad \text{w.p. 1.}$$

Recalling that $\hat{\sigma}_N^2(f_s, \hat{\theta}_N(f_s)) \to \sigma^2$ w.p. 1 from Corollary 6.1, (B.5) suggests

$$\text{(B.6)} \qquad \frac{\Delta_N(f, f_s)}{N} = \frac{1}{2} \log \frac{\hat{\sigma}_N^2(f, \hat{\theta}_N(f))}{\hat{\sigma}_N^2(f_s, \hat{\theta}_N(f_s))} + (m_f - m_s) \frac{\log N}{N}$$

$$\geq 0 \qquad \text{w.p. 1}$$

for a sufficiently large N, which implies (6.35).

Appendix 6.C <u>Proof of Theorem 6.3</u>

From the condition (C-3)", we have

$$\text{(C.1)}\quad f_{p,ij}(z_1,\cdots,z_i;k) = f_{s,ij}(z_1,\cdots,z_i;k)$$

for $i=1,2,\cdots,p$ and $j=1,2,\cdots,q_i$ because of $D_f \subset D_{f_s}$. Hence

$$\text{(C.2)}\quad \hat{\sigma}_N^2(f,\hat{\theta}_N(f)) = \tilde{\theta}_N'(f_s,f)\{\frac{1}{N}\sum_{k=1}^{N} F_{k-1}(f_s)F_{k-1}'(f_s)\}\tilde{\theta}_N(f_s,f)$$
$$+ 2\tilde{\theta}_N'(f_s,f)\{\frac{1}{N}\sum_{k=1}^{N} F_{k-1}(f_s)v_k\} + \frac{1}{N}\sum_{k=1}^{N} v_k^2,$$

where

$$\text{(C.3)}\quad \tilde{\theta}_N(f_s,f) \triangleq [\tilde{\theta}_{11,N}(f_s,f) \cdots \tilde{\theta}_{1\ell_1,N}(f_s,f);$$
$$\cdots ; \tilde{\theta}_{n1,N}(f_s,f) \cdots \tilde{\theta}_{1\ell_n,N}(f_s,f)]'$$

and

$$\text{(C.4)}\quad \tilde{\theta}_{ij,N}(f_s,f) \triangleq \begin{cases} \mathring{\theta}_{ij} - \hat{\theta}_{ij,N}(f) & (i=1,2,\cdots,p; j=1,2,\cdots,q_i) \\ \mathring{\theta}_{ij} & (i=1,2,\cdots,p; j=q_i+1,\cdots,\ell_i \text{ and} \\ & \quad i=p+1,\cdots,n;\ j=1,\cdots,\ell_i). \end{cases}$$

The first term of the R.H.S. of (C.2) can be evaluated as

$$\text{(C.5)}\quad \liminf_{N\to\infty} \tilde{\theta}_N'(f_s,f)\{\frac{1}{N}\sum_{k=1}^{N} F_{k-1}(f_s)F_{k-1}'(f_s)\}\tilde{\theta}_N(f_s,f)$$
$$\geq c \liminf_{N\to\infty} \{\sum_{i=1}^{p}\sum_{j=1}^{q_i} \tilde{\theta}_{ij,N}(f_s,f))^2\}$$
$$+ \sum_{i=1}^{p}\sum_{j=q_i+1}^{\ell_i} \mathring{\theta}_{ij}^2 + \sum_{i=p+1}^{n}\sum_{j=1}^{\ell_i} \mathring{\theta}_{ij}^2 > 0 \quad \text{w.p. 1},$$

where the inequality (A.12) has been used. Using the same procedure as that in (B.4), we have

$$(C.6) \qquad \tilde{\theta}_N'(f_s, f)\frac{1}{N}\sum_{k=1}^{N} F_{k-1}(f_s)v_k \to 0 \qquad \text{w.p. } 1 \qquad \text{as} \quad N \to \infty.$$

Therefore, from (C.5), (C.6) and (A.18), for a sufficiently large N, we can conclude that

$$(C.7) \qquad \hat{\sigma}_N^2\,(f, \hat{\theta}_N(f)) > \sigma^2 \qquad \text{w.p. } 1$$

and

$$(C.8) \qquad \frac{\Delta_N(p, n)}{N} > 0 \qquad \text{w.p. } 1$$

for a sufficiently large N, which gives (6.37).

Appendix 6.D **Proof of Theorem 6.4**

Since

$$(D.1)\qquad f_{p,ij}(z_1,\cdots,z_i;k) = f_{s,ij}(z_1,\cdots,z_i;k) \qquad \text{for } i=1,\cdots,n;j=1,\cdots,\ell_i$$

because of $D_{f_s} \subset D_f$, using the Taylor series expansion, we have

$$\begin{aligned}(D.2)\quad \hat{\sigma}_N^2(f_s,\hat{\theta}_N(f_s)) &= \hat{\sigma}_N^2(f,\hat{\theta}_N(f)) + \frac{\partial}{\partial\theta(f)}[\hat{\sigma}_N^2(f,\hat{\theta}_N(f))]'(\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f)) \\ &\quad + \frac{1}{2}(\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f))'\frac{\partial}{\partial\theta(f)}(\frac{\partial}{\partial\theta(f)}\hat{\sigma}_N^2(f,\hat{\theta}_N(f))) \\ &\quad \times (\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f)) + o(\|\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f)\|^2),\end{aligned}$$

where $\hat{\theta}_N(f_s,f)$ is defined by

$$\begin{aligned}(D.3)\quad \hat{\theta}_N(f_s,f) \triangleq {} & \underbrace{[\hat{\theta}_{11,N}(f_s)\ \cdots\ \hat{\theta}_{1\ell_i,N}(f_s)\ 0\ \cdots\ 0}_{q_i};\ \cdots \\ & \underbrace{;\hat{\theta}_{n1,N}(f_s)\ \cdots\ \hat{\theta}_{n\ell_n,N}(f_s)\ 0\ \cdots\ 0}_{q_n};\ \underbrace{0\ \cdots\ 0}_{q_{n+1}+\ \cdots\ +q_p}]'.\end{aligned}$$

Using (6.25) and (6.28), the derivatives of $\hat{\sigma}_N^2(f,\hat{\theta}_N(f))$ can be calculated as

$$(D.4)\qquad \frac{\partial}{\partial\theta(f)}\hat{\sigma}_N^2(f,\theta(f)) = -\frac{2}{N}[s_N(f) - Q_N(f)\theta(f)],$$

$$(D.5)\qquad \frac{\partial}{\partial\theta(f)}(\frac{\partial}{\partial\theta(f)}\hat{\sigma}_N^2(f,\theta(f)) = \frac{2}{N}Q_N(f),$$

$$(D.6)\qquad \frac{\partial^m}{\partial\theta(f)^m}\hat{\sigma}_N^2(f,\theta(f)) = 0 \qquad \text{for } m = 3,4,\cdots.$$

Hence, substituting (D.4), (D.5) and (D.6) into (D.2), it follows that

$$(D.7)\qquad \hat{\sigma}_N^2(f_s,\hat{\theta}_N(f_s)) = \hat{\sigma}_N^2(f,\hat{\theta}_N(f)) + \frac{1}{N}\,(\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f))'Q_N(f)(\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f)).$$

Then, substituting (D.7) into (B.1), the following equation is obtained:

$$(D.8)\qquad \Delta_N(f_s,f) = -\frac{1}{2\hat{\sigma}_N^2(f,\hat{\theta}_N(f))}\,h_N + (m_f - m_s)\log N + o(\frac{1}{\hat{\sigma}_N^2(f,\hat{\theta}_N(f))}\,h_N)$$

where

$$(D.9)\qquad h_N \triangleq (\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f))'Q_N(f)(\hat{\theta}_N(f_s,f) - \hat{\theta}_N(f)).$$

and $\log(1+x) = x + o(x)$ has been used. Setting as $f=f_s$ in (6.25) and (A.4), we have

$$(D.10)\qquad \hat{\theta}_N(f_s) = \mathring{\theta} + Q_N^{-1}(f_s)\sum_{k=1}^{N} F_{k-1}(f_s)v_k.$$

Hence, substituting (6.25) and (D.10) into (D.9), it follows that

$$(D.11)\qquad h_N = (\sum_{k=1}^{N} F_{k-1}(f)v_k)'Q_N^{-1}(f)(\sum_{k=1}^{N} F_{k-1}(f)v_k) - (\sum_{k=1}^{N} F_{k-1}(f_s)v_k)'Q_N^{-1}(f_s)(\sum_{k=1}^{N} F_{k-1}(f_s)v_k).$$

Then, we have

$$(D.12)\qquad |h_N| \le \| NQ^{-1}(f)\|\ \|\frac{1}{\sqrt{N}}\sum_{k=1}^{N} F_{k-1}(f)v_k\|^2 + \| NQ^{-1}(f_s)\|\ \|\frac{1}{\sqrt{N}}\sum_{k=1}^{N} F_{k-1}(f_s)v_k\|^2.$$

From (C-1), (C-2), (C-4) and (6.29), since

(D.13) $$E\{ \| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} F_{k-1}(f) v_k \|^2 \} \le c,$$

and

(D.14) $$E\{ \| \frac{1}{\sqrt{N}} \sum_{k=1}^{N} F_{k-1}(f_s) v_k \|^2 \} \le c$$

we obtain

(D.15) $$\frac{h_N}{\log N} \to 0 \qquad \text{in prob.} \qquad \text{as } N \to \infty.$$

From Theorem 6.1, $\hat{\sigma}_N^2(f, \hat{\theta}_N(f)) \to \sigma^2$ w.p. 1 as $N \to \infty$. Then

(D.16) $$\hat{\sigma}_N^2(f, \hat{\theta}_N(f)) \to \sigma^2 \qquad \text{in prob.} \qquad \text{as } N \to \infty.$$

Thus, we can conclude from (D.8), (D.15) and (D.16) that

(D.17) $$\frac{\Delta_N(f, f_s)}{\log N} \to m_f - m_s > 0 \qquad \text{in prob.} \qquad \text{as } N \to \infty.$$

Appendix 6.E Replacement of (C-2) by (C-2)'

Let define the state vector $x_k \triangleq [x_{1k} \cdots x_{nk}]'$ as

$$x_{ik} \triangleq y_{k-n+i} \qquad \text{for } i=1,2,\cdots,n. \tag{E.1}$$

Then, the state space representation of (6.1) is given by

$$x_k = A_s(x_{k-1};k;\mathring{\theta})x_{k-1} + \varphi_s(x_{k-1}) + \eta_k v_k \tag{E.2}$$

where

$$\left\{\begin{array}{l} \varphi_s(x_{k-1}) = [\underbrace{0 \cdots 0}_{n-1} \ \phi_s(x_{k-1})]' \\ \\ \eta_k = [\underbrace{0 \cdots 0}_{n-1} \ b_k]'. \end{array}\right. \tag{E.3}$$

Hence, from (E.3), it follows that

$$\| x_k \| \leq \| A_s(x_{k-1};k;\mathring{\theta}) \| \ \| x_{k-1} \| + \| \varphi_s(x_{k-1}) \| + | b_k | \ | v_k | . \tag{E.4}$$

Since the condition (C-2)' implies that

$$\prod_{i=1}^{k+1} \| A_s(x_{i-1};i;\mathring{\theta}) \| \leq c \, c_1^{k+1} \qquad (c_1 < 1) \tag{E.5}$$

$$\| \varphi_s(x_{k-1}) \| \leq c, \tag{E.6}$$

we have from (E.4) and (E.6) that

$$\| x_k \| \leq \prod_{i=1}^{k+1} \| A_s(x_{k-i};k-i+1;\mathring{\theta}) \| \ \| x_{-1} \| + \sum_{j=1}^{k} \left(\prod_{i=1}^{j} \| A_s(x_{k-i};k-i+1;\mathring{\theta}) \| \right) \| \varphi_s(x_{k-1-j}) \| \tag{E.7}$$

$$+ \sum_{j=1}^{k} \Big(\prod_{i=1}^{j} \| A_s(x_{k-i};k-i+1;\theta) \| \Big) | b_{k-j} | \; | v_{k-j} |$$

$$\leq c\, c_1^{k+1} \| x_{-1} \| + c^2 \sum_{j=1}^{k} c_1^j + c^2 \sum_{j=1}^{k} c_1^j | v_{k-j} |$$

$$\leq c\, c_1^{k+1} \| x_{-1} \| + \frac{c^2 c_1}{1 - c_1} + c^2 \sum_{j=1}^{k} c_1^j | v_{k-j} |$$

where the inequalities $| b_k | < c$ and $\sum_{j=1}^{k} c_1^i < c_1/(1-c_1)$ have been used. Hence, from (E.7) and the elementary inequality $(a+b+c+d)^4 \leq 64(a^4+b^4+c^4+d^4)$, it follows that

$$\text{(E.8)} \quad E\{ \| x_k \|^4 \} \leq E\{[c\, c_1^{k+1} \| x_{-1} \| + \frac{c^2 c_1}{1 - c_1} + c^2 \sum_{j=1}^{k} c_1^j | v_{k-j} |]^4\}$$

$$\leq 64c^4 \| x_{-1} \|^4 c_1^{4(k+1)} + \Big(\frac{c^2 c_1}{1 - c_1} \Big)^4$$

$$+ c^8 \Big(\sum_{j=1}^{k} c_1^j | v_{k-j} | \Big)^4.$$

Furthermore, by using the condition (C-1), it follows that

$$\text{(E.9)} \quad E\{(\sum_{j=1}^{k} c_1^j | v_{k-j} |)^4\} = E\{(\sum_{j=1}^{k-1} c_1^{k-j} | v_j |)^4\}$$

$$= \sum_{i_1=1}^{k-1} \sum_{i_2=1}^{k-1} \sum_{i_3=1}^{k-1} \sum_{i_4=1}^{k-1} c_1^{4k-i_1-i_2-i_3-i_4} E\{ | v_{i_1} | \, | v_{i_2} | \, | v_{i_3} | \, | v_{i_4} | \}$$

$$\leq \sum_{i_1=1}^{k-1} \sum_{i_2=1}^{k-1} \sum_{i_3=1}^{k-1} \sum_{i_4=1}^{k-1} c_1^{4k-i_1-i_2-i_3-i_4} [\prod_{j=1}^{4} E\{E\{ | v_{i_j} |^4 | \mathcal{F}_{i_j-1}\}\}]^{1/4}$$

$$\leq c\Big(\sum_{i=0}^{k-1} c_1^{k-i} \Big)^4$$

$$\leq c\Big(\frac{c_1}{1 - c_1} \Big)^4 < c.$$

Therefore, from (E.8) and (E.9), we can conclude that

(E.10) $E\{\| x_k \|^4\} < c$

which implies $E\{y_k^4\} < c$.

CHAPTER 7 CONCLUSIONS

7.1 Concluding Remarks

The modeling problem for nonstationary and nonlinear systems has been considered, where system models are described by stochastic differential equations or stochastic difference equations driven by the white noise.

The key assumption on the whole aspect of this dissertation is that time-varying functions and nonlinear functions in system models are well approximated by finite series of known functions with unknown constant coefficients.

In Chapter 3, augmenting the state variables by unknown parameters and using an approximated nonlinear filter, unknown parameters of a nonstationary system model has been identified. The unknown system order has been also determined by using the likelihood-ratio function associated with Bayesian hypothesis test.

In Chapter 4, the parameter identification method has been developed for a class of stationary but nonlinear models. The system model in this chapter was given by the nonlinear MA model whose nonlinear MA terms are described by a set of Hermite polynomials. Using the fact that unknown parameters in the nonlinear MA model are uniquely described as functions of second and third moments of the output process, the estimators of unknown parameters have been obtained by the moment method. The proposed nonlinear MA model has extended to a nonlinear ARMA model in order to handle the data whose serial dependence is much longer than the single time unit.

From more practical viewpoints that the most of all physical systems has both some nonlinearities and nonstationarity, the parameter identification for a class of nonstationary nonlinear system has been considered in Chapter 5, where the system was assumed to be described by a nonlinear difference equation with time-varying coefficients. Unknown parameters were estimated by the maximum likelihood concept. It should be emphasized that if there exists only one system parameter to be identified, the consistency of the estimator has proved without the assumption of bounded-input bounded-output stability which are usually required.

Restricting the system model to the single-input single-output one, the structure determination problem for a nonstationary nonlinear model has been considered in Chapter 6. The key notion for deriving a criterion function for structure determination is to evaluate the upper bounded of the entropy associated with the estimation errors of input noise and unknown parameters, where the contributions of the input noise and unknown parameters to the values of the entropy are adjusted to be equal. The proposed criterion function has been applied for modeling actual earthquake data.

7.2 Discussions

The proposed models in this dissertation are essentially linear with respect to unknown constant parameters, whereas the system models are themselves nonlinear or nonstationary. Hence the estimators of unknown parameters are rather easily obtained and it has become possible to investigate mathematically their asymptotic properties. In case of nonlinear models with respect to unknown parameters, the identification problem usually becomes a nonlinear optimization problem. Although there are many numerical techniques which can exploit for the identification of nonlinear-in-the-parameters models, it is usually very hard to guarantee theoretically the asymptotic properties such as consistency, asymptotic normality, etc. However, for some classes of stationary nonlinear models, the difficulty mentioned above may be overcomed by utilizing the information from statistical moments higher than third order, since one of the outstanding feature of nonlinear system lies in the non-Gaussian statistics of the output process.

Structure determination for a general class of nonstationary nonlinear systems is another very difficult problem which is still unsolved. For instance, the system structure may be changed with time evolution when the dominant time-varying system parameters vary between some non-zero values to zeros, and this fact leads to the structure determination method based on the large number of observation data meaningless. Since the structure determination problem is closely related with the stochastic realization of systems, we should investigate the fundamental properties of nonstationary nonlinear systems by using the system theoretic approach.

List of References

* The reference with an asterisk
is the work of the author

[A1] K.J. Åstörom and P. Eykoff, "System identification-a survey," Automatica, vol.7, pp.123-162, 1971.

[A2] H. Akaike, "Information Theory and an extension of the maximum likelihood principle," in Proc. 2nd International Symposium on Information Theory, eds. B.N. Petrov and Càski, Akadèmia Kiado, pp.267-281, Budapest, 1973.

[A3] H. Akaike, "A new look at the statistical model: identification," IEEE Trans. Automat. Contr., vol.AC-19, no.6, pp.716-823, 1974.

[A4] H. Akaike and et al., "A time series analysis and control program package (1)," Computer Science Monographs, Institute of Statistical Mathematics, Mar., 1975

[A5] H. Akaike, "On entropy maximization principle," in Application of Statics (ed. R. Krishnaiah), pp. 27-41, North-holland, Amsterdom, 1977.

[A6] K.J. Åström and T. Söderström, "Uniqueness of the maximum likelihood estimtes of the parameters of ARMA model," IEEE Trans. Automat. Contr., vol.AC-19, no.6, pp.769-773, 1974.

[A7] T.W. Anderson, "On asymptotic distributions of estimates of parameters of stochastic difference equation," Ann. Math. Stat., vol.30, pp.676-687, 1959.

[B1] E.E.P. Box and G.M. Jenkins, Statistical Models for Forecasting and Control, Revised edition, San Francisco, Calif.:Holden-Day, 1976.

[B2] B.M. Brown, "Martingale central limit theorems," Ann. Math. Statist., vol.42, no.1, pp.59-66, 1971.

[C1] J.C. Chow, "On estimating the orders of an autoregressive moving-average processes with uncertain observations," IEEE Trans. Autom. Contr., vol.AC-17, no.5, pp.707-709, 1972.

[C2] Y.S. Chow, "On the L_p-convergence for $n^{-1/p}S_n$, $0<p<2$," Ann. Statist. vol.42, no.1, pp.393-394, 1971.

[D1] J.L. Doob, Stochastic Processes, New York: John-Willy, 1953.

[E1] P. Eykhoff, System Identification, New York: John-Willy, 1974.

[G1] G.C. Goodwin and R.L. Pyne, Dynamic System Identification, Experimental Design and Data Analysis, New York: Academic Press, 1977.

[G2] R. Gwidorzi, "Cannonical structures in identification of multivariable sytems," Automatica, vol. 11, pp.361-374, 1975.

[G3] R. Gran, "System identificaton using appropriate nonlinear filters," in Proc. 3rd Symp. on Nonlinear Estimation And Its Applications, San Diego, Calif., 1972.

[G4] R.G. Gallager, Information Theory and Reliable Communication, San Francisco, Calif.: John-Willy, 1968.

[G5] D. Gabor, W.P.L. Wilby and R. Woodcock, "A universal nonlinear filter predictor and simulator which optimize itself by a learning process," IEE proc., vol.108B, pp.422-433.

[H1] E.J. Hannan and B.G. Quinn, "The determination of the order of an autoregression," J.R. Statist. Soc. B, vol.41, no. 2, pp.190-195, 1979.

[I1] Ivakhnenko, "Heuristic self-organization in problems of engineering cybernetics," Automatica, vol.1, pp.207-219, 1970.

[J1] A.H. Jazwinski, Stochastic Processes and Filtering Theory, New York: Academic Press, 1970.

[K1] R.E. Kalman, P.L. Falb and M.A. Arbib, Topics in Mathematical System Theory, New york: McGraw-Hill, 1969.

[K2] F. Kozin and F. Nakajima, "The order determination problem for linear time-varying AR model," IEEE Trans. Automat. Contr., vol.AC-25, no.2, pp.250-257, 1980.

[K3] T. Kailath, "A general likelihood-ratio formula for random signals in Gaussian noise," IEEE Trans. Information Theory, vol.IT-15, pp.350-361, 1969.

[K4] S. Kodama and S. Suda, Matrix Theory for Control Enginners, Tokyo: SICE, 1978 (in Japanese).

[K5] K. Knopp, Infinite Sequence and Series, New York: Dover, 1956.

[K6] R.Z. Khazminskii, Stability of Systems of Differential Equations under Random Disturbances of Their Parameters, Moscow: Nauka, 1969.

[K7] J. Kiefer and Wolfowitz, "Stochastic estimation of the maximum of a regression function," Ann. Math. Statist., vol.23, pp.462-466, 1952.

[L1] E. Lukacs, Stochastic Convergence, 2nd edition, New York: Academic Press, 1975.

[L2] L. Ljung, "Asymptotic behavior of the extended Kalman filter as a parameter estimtion for linear systems," IEEE Trans. Automat. Contr., vol.AC-19, no.6, pp.36-50, 1979.

[L3] M. Loeve, Probability Theory, 3rd edition, Princeton, N.J.: D. Van Nostrand, 1963.

[L4] T. Lee, Consitent Estimators for a Class of Nonstationary Models, Ph.D. Thesis, Polytechnique Institute of New York, 1976.

[M1] R.K. Mehra, "On-line identification of linear dynamic sytems with applications to Kalman filtering," IEEE Trans. Automat. Contr., vol.AC-16, pp.12-21, 1971.

[M2] D.Q. Mayne, "Parameterization and identification of linear multivariable systems," in Stability of Stochastic Systems, Lecture Note in Mathematics, vol.294, Berlin, Splinger-verlag, 1972.

[N1] J. Neveu, Mathematical Fundation of the Calculas of Probability, San Francisco: Holden-Day, 1975.

[N2] A.N. Netravali and R.P.De Figueiredo, "On the identification of Nonliner dynamical sytems," IEEE Trans. Automat. Contr., vol.AC-16, no.1, pp.28-36, 1971.

[N3] F. Nakajima and F. Kozin, "A characterization of consistent estimators," IEEE Trans. Automat. Contr., vol.AC-24, no.5, pp.758-765, 1979.

[O1] T. Ozaki and H. Oda, "Nonlinear time series model identification by Akaike's information criterion," in Proc. of IFAC Workshop on Informtion and Sytems, Compiegue, France, 1977.

[P1] A.S. Poznyak, "Estimating the parameters of autoregression processes by the method of least squares," Int. J. Systems Science, vol.11, pp. 577-588, 1980.

[P2] E. Polak, Computational Methods in Optimization: A Unified Approach, New York: Academic Press, 1971.

[R1] J. Rissanen, "Consistent order-estimate of autoregressive processes by shortest description of data," in Analysis and Optimization of Stochastic Systems, Univ. Oxford, 1978.

[R2] J. Rissanen, "Modeling by shortest data description," Automatica, vol.14, pp.465-471, 1978.

[R3] A. Renji, Fundations of Probability, San Francisco: Academic Press, 1970.

[R4] B. Rosen, "On the central limit Theorem for sums of dependent random variables," Z. Wahrscheinlichkeitstheorie verw, Geb. 7, pp. 48-82, 1967.

[R5] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain, "Psychological Rev., vol. 65, pp. 386-408, 1958.

[R6] H. Robbins and S. Monro, "A stochastic approximation method," Ann. Math. Statist., vol. 22, pp. 400-407, 1951.

[R7] P.M. Robinson, "The estimatin of a nonlinear moving average model," Stochastic Processes and Their Applications, vol. 5, pp. 81-90, 1977.

[S1] R. Shibata, "Selection of the order of an autoregressive model by Akaike s information criterion," Biometrika, vol. 63, no. 1, pp. 117-126, 1976.

[S2] R. Shibata, "Asymptotically efficient selection of the order of the model for estimating parameters of a linear process," Ann. Statist., vol. 8, pp. 147-164, 1980.

[S3] R. Shibata, "An optimal selection of regression variable," Biometrika, vol. 68, no. 1, pp. 45-54, 1981.

[S4] F.C. Schweppe, Uncertain Dynamical Systems, New York: Prentice-Hall, 1973.

[S5] Y. Sunahara, "On a method of signal detection by using the nonlinear estimation theory," in Proc. 3rd Symposium on Nonlinear Estimation Theory and Its Applications, San Diego, Calif., pp. 216-223, 1972.

[S6] Y. Sunahara, A. Ohsumi and M. Imamura, "A method of parameter identification for linear ditributed parameter systems," Automatica, vol. 12, pp. 245-256, 1976.

[S7] A.P. Sage and J.L. Melsa, Estimation Theory with Applications to Communication and Control, New York: McGraw-Hill, 1971.

[S8] G. Schwarz, "Estimating the dimension of a Model," Annals of statistics, vol. 6, no. 2, pp. 461-464, 1978.

[S9]* Y. Sunahara, A. Ohsumi and T. Fukuda, "On modeling a random data via a class of nonlinear moving-average models," Trans. SICE, vol. 18, no. 3, pp. 224-231, 1982 (in Japanese).

[S10]* Y. Sunahara, A. Ohsumi and T. Fukuda, "Parameter identification for a class of nonlinear ARMA models," in Proc. of 12th JAACE Symp. on Stochastic Sytems, pp. 35-38, Kanazawa, Oct. 21-23, 1980.

[S11]* Y. Sunahara and T. Fukuda, "Parameter identification for non stationary nonlinear systems," Bulletin of Informatics and Cybernetics, vol. 20, no. 1～2, p. 11-24, 1982.

[S12]* Y. Sunahara, T. Fukuda and H. Jusa, "On a modeling a class of nonstationary nonlinear systems," Trans. SICE, vol. 19, no. 10, 1983 (in Japanese).

[S13]* Y. Sunahara , A. Ohsumi and T. Fukuda, "On modeling of linear stochastic systems from noisy data," Trans. SICE, vol. 14, no. 5, 1978 (in Japanese).

[S14]* Y. Sunahara, T. Fukuda and H. Jusa, "Order determination of nonlinear models by using minimum error entropy principle," in Proc. of 14th JAACE Symp. on Stochastic Systems, Kobe, pp. 20-22, Oct. 20-22, 1982.

[S15]* Y. Sunahara, T. Fukuda and H. Jusa, "A method for order determination of nonstationary nonlinear system models," System and Control, vol. 29, no. 4, 1985 (in Japanese).

[S16]* Y. Sunahara and T. Fukuda, "A modeling method for a class of nonstationary and nonlinear systems," in Proc. of 7th IFAC/IFORS Symp.

on Identification and System Parameter Estimation, England, July 3-7, 1985.

[T1] E. Tse and H.L. Weinert, "Structure determination and parameter identification from multivariable stochastic linear systems," IEEE Trans. Automat. Contr., vol.AC-20, pp.603-613, 1975.

[T2] M.D. Trifunac, A.G. Brady and D.E. Hudson, Strong Motion Earthquake Accelerograms-Digitized and Plotted Data, vol.II, Part C, A Report of Earthquake Engineering Research Laboratory, California Institute of Technology, Calif., 1973.

[V1] J. Vallis, "On-line identification of multivariable linear systems of unknown structure from input output data," in Proc. IFAC Symp. on Identification and Process Parameter Estimation, Paper no. 1.5, Prague, 1970

[V2] H.L. Van Trees, Detection, Estimation and Modulation Theory, Part I, New York: John Willy, 1968.

[W1] C. M. Woodside, "Estimation of the order of linear systems," Automatica, vol.7, pp.727-733, 1971.

[W2] N. Wiener, Nonlinear Problems in Random Theory, New York: John-Willy, 1958.