Statistica Sinica 2(1992), 113-135

A BRIDGE BETWEEN NONLINEAR TIME SERIES MODELS AND NONLINEAR STOCHASTIC DYNAMICAL SYSTEMS: A LOCAL LINEARIZATION APPROACH

Tohru Ozaki

Institute of Statistical Mathematics, Tokyo

Abstract: In the present paper we point out a close relationship between nonlinear time series models and nonlinear stochastic dynamical systems. We introduce a time discretization method of stochastic dynamical systems, which brings us a nonlinear time series model from a nonlinear stochastic dynamical system. Then a maximum likelihood method for the estimation of the time series model is given. We also give an identification procedure of the original continuous time dynamical system. The effectiveness of the procedure is numerically checked using simulated data. Implications of the present method in non-Gaussian time series analysis are discussed.

Key words and phrases: Exponential AR model, amplitude-dependent AR model, identification, stochastic dynamical system, diffusion process, local linearization, Pearson system, exponential family, maximum likelihood method.

1. Introduction

There are many examples of dynamic phenomena in nature which can be regarded as stochastic processes, e.g. ship motion in the sea, brain wave records in physiology, animal populations in ecology. Some of them are considered to be stochastic processes by virtue of their own mechanism. Some of them, such as hydrodynamic phenomena or dynamics of economic activity may not be considered stochastic at the microscopic level, but may be considered as such at the macroscopic level. By treating them as stochastic processes, meaningful results both in theory and applications may be obtained.

For inference of the characteristics of these stochastic processes and for their forecasting and control, time series data, obtained by sampling the process at equally spaced intervals of time, are often used. Although many stochastic phenomena in the world can be considered approximately Gaussian processes, some of them are obviously non-Gaussian. This means, in some cases, conventional linear time series models driven by Gaussian white noise are not really appropri-

ate. For the analysis of these non-Gaussian processes and for their forecasting and control, nonlinear time series models are needed. It is also expected that nonlinear time series models may be useful for inference of the nonlinear structure in the dynamics of stochastic processes. Several nonlinear time series models have been introduced and used for the analysis of such non-Gaussian time series data (Priestley (1988)).

On the other hand, nonlinear, stochastic, dynamical system models have been intoduced for the analysis of the same sort of phenomena, namely, non-Gaussian stochastic processes, in many scientific fields such as physics, electrical engineering, chemistry, biology, genetics and macro-economics. One of the significant advantages of nonlinear time series models over stochastic dynamical system models is, since time series models are statistical models, they can be identified from observed data of finite length. Stochastic dynamical system models, on the other hand, are mathematical models and are difficult to identify from the data. However, being a mathematical model gives the stochastic dynamical system model an advantage over nonlinear time series models in characterizing the phenomena, with powerful analytical tools developed in Markov diffusion theory. In this situation, it must be natural for time series analysts to wonder if a bridge can be made between time series models and stochastic dynamical system models so that we can take advantage of both.

To make a bridge between them a time discretization method is introduced in this paper, which yields computationally stable discrete time dynamical system models. This means we can obtain a stationary time series model from any stationary nonlinear stochastic dynamical system model. This point is essential since the sampling interval is fixed in time series analysis; and because of this we are often faced with undesirable computational instability in nonlinear model simulations. A crude idea of the use of local linearization was first presented in Ozaki (1985a). In the present paper a clearer and more consistent presentation of the method is provided. Then we give a maximum likelihood method for estimation of the discretized time series model. Finally it is shown that the original continuous time stochastic dynamical system model can be identified from the estimated nonlinear time series model. Thus, identification of the original continuous time stochastic dynamical system is available, based on the local linearization relationship between the continuous time model and the estimated time series model. The effectiveness of the procedure is numerically checked using simulated data. Implications of the present method in nonlinear and non-Gaussian time series analysis and its application to various problems are discussed.

2. Local Linearization

Suppose we have a continuous time stochastic dynamical system,

$$\dot{y}(t) = f(y(t)) + n(t), \tag{2.1}$$

where $n(t)$ is a continuous time Gaussian white noise with variance σ^2. The idea for discrete time modelling of (2.1) is as follows: first, we try to approximate (2.1) by a Gaussian process,

$$\dot{y}(t) = K_t y(t) + n(t) \tag{2.2}$$

with some appropriate K_t, which we are going to find later, on each infinitesimal interval $[t, t+\Delta t)$. This idea is reasonable if Δt is sufficiently small, because $y(t)$ of (2.1) is a Markov diffusion process, and the main characteristic of the Markov diffusion process is that it is locally Gaussian, where its mean and variance are specified by the drift and diffusion coefficient of the process (see Goel and Richter-Dyn (1974)). Then from (2.2) we have an autoregressive model,

$$y_{t+\Delta t} = A_t y_t + w_{t+\Delta t} \tag{2.3}$$

for the Gaussian process (2.2) on the interval. Once K_t of (2.2) is obtained, it is easy to give A_t of the corresponding autoregressive model (2.3). The solution $y_t(\tau)$ of (2.2) can be expressed as

$$y_t(\tau) = \int_{-\infty}^{\tau} \exp\{K_t(\tau - u)\} n(u) du,$$

and its autocovariance function is $C_t(\tau) = [\sigma^2/(2K_t)]\exp(K_t\tau)$ where the subscript t is used to denote the interval $[t, t+\Delta t)$. Then A_t of (2.3) is given by $A_t = [C_t(0)]^{-1} C_t(\Delta t) = \exp(K_t \Delta t)$. The variance of the discrete time white noise

$$w_{t+\Delta t} = \int_{t}^{t+\Delta t} \exp\{K_t(t + \Delta t - u)\} n(u) du$$

of the model (2.3) is given by $\sigma_w^2 = \sigma^2\{\exp(2K_t\Delta t) - 1\}/(2K_t)$. This yields the following autoregressive model for (2.2) on the interval $[t, t+\Delta t)$,

$$y_{t+\Delta t} = \exp(K_t \Delta t) y_t + \sqrt{\frac{\exp(2K_t\Delta t) - 1}{2K_t}}\, n_{t+\Delta t} \tag{2.4}$$

where $n_{t+\Delta t}$ is a discrete time Gaussian white noise with variance σ^2. The problem is how to find K_t of (2.2).

If we have a discrete time dynamical system model,

$$y_{t+\Delta t} = A_t y_t \tag{2.5}$$

with positive A_t, a natural choice of K_t would be $K_t = \{1/(\Delta t)\} \log A_t$, because the discrete time model for a linear dynamical system, $\dot{y}(t) = K_t y(t)$ is given by $y_{t+\Delta t} = \exp(K_t \Delta t \pm i2\pi m) y_t$, where m is an integer. Therefore, the problem of finding a "good K_t" is equivalent to the problem of finding a "good A_t" of (2.5) as an approximate discrete time dynamical system for the continuous time dynamical system, $\dot{y} = f(y)$. For the criterion of the goodness of A_t, the following requirements are reasonable:
(1) The model is consistent, i.e.

$$\frac{y_{t+\Delta t} - y_t}{\Delta t} \to f(y_t) \quad for \quad \Delta t \to 0.$$

(2) The trajectory of y_t coincides with that of $y(t)$ at the discrete time points t, $t + \Delta t$, $t + 2\Delta t$, $t + 3\Delta t$, ... at least for linear $f(y)$.
(3) The discrete time model preserves the qualitative characteristics of the continuous time model.

The third condition is important since in the nonlinear case the dynamical system $\dot{y} = f(y)$ has many structural variations. Unfortunately, all of the discretization schemes, such as the Euler, Heun and Runge-Kutta methods, well known in the field of numerical analysis, do not satisfy the second and third conditions (Henrici (1962)). For example, by applying the Runge-Kutta method to $\dot{y} = -y^3$ we have

$$y_{t+\Delta t} = y_t + \frac{\Delta t}{6}(k_1 + 2k_2 + 2k_3 + k_4) = p_{80}(y_t) y_t$$

where $k_1 = -y_t^3$, $k_2 = -(y_t + \frac{\Delta t}{2} k_1)^3$, $k_3 = -(y_t + \frac{\Delta t}{2} k_2)^3$, $k_4 = -(y_t + \Delta t k_3)^3$ and $p_{80}(y_t)$ is an 80-th order polynomial of y_t. The polynomial function is known to go to $\pm\infty$ for $|y_t| \to \infty$. Therefore y_t explodes to infinity if it starts from some large initial value, while the trajectory of $\dot{y} = -y^3$ converges to zero regardless of its starting value.

A scheme which satisfies the above conditions is obtained by a simple assumption, i.e. the Jacobian of the linear dynamical system $\dot{y} = K_t y$ of (2.2) over each interval $[t, t + \Delta t)$ is given by the Jacobian $J_t = \{\partial f(y)/\partial y\}$ at $y = y_t$ of the original dynamical system. From this assumption we have, for $t \le s < t + \Delta t$, $\ddot{y}(s) = J_t \dot{y}(s)$. If this is integrated on $[t, t + \tau)(0 \le \tau < \Delta t)$ we have $\dot{y}(t + \tau) = \exp(J_t \tau)\dot{y}(t) = \exp(J_t \tau) f(y(t))$. By integrating this again with respect to τ on $[0, \Delta t)$ we have $y(t + \Delta t) = y(t) + J_t^{-1}\{\exp(J_t \Delta t) - 1\} f(y(t))$. This equation gives the value of y at time point $t + \Delta t$ as a function of y at time point t. This results in the following relation between y_t and $y_{t+\Delta t}$, where

$t = \Delta t, 2\Delta t, 3\Delta t, \ldots$, and $J(y_t) \neq 0$,

$$y_{t+\Delta t} = y_t + J(y_t)^{-1}[\exp\{J(y_t)\Delta t\} - 1]f(y_t). \tag{2.6}$$

For y_t, where $J(y_t) = 0$, we have $y_{t+\Delta t} = y_t + \Delta t f(y_t)$. It is easily checked that the model (2.6) satisfies the above three conditions (Ozaki (1985a)).

The model (2.6) has been used by engineers for the simulations of deterministic nonlinear differential equations (Smith (1977)). The idea of approximating a deterministic nonlinear system by a piecewise linear differential equation has been common in numerical analysis and system theory. For example, approximating the nonlinear system $\dot{y} = f(y)$ by a linear system $\dot{y} = J_t y$, on each interval $[t, t + \Delta t)$, has been used in the Extended Kalman filter. Even though this results in a non-explosive and stable scheme, $y_{t+\Delta t} = \exp(J_t \Delta t) y_t$, it is not a consistent scheme. $(y_{t+\Delta t} - y_t)/\Delta t$ of the scheme tends to $J_t y_t$ for $\Delta t \to 0$ instead of to $f(y_t)$. Only the approach of Smith (1977) based on piecewise linear approximation can produce stable and consistent scheme. It is surprising that the above mentioned natural conditions (2) and (3) were overlooked, and not much attention was paid to a very simple and natural scheme (2.6) by numerical analysts.

From (2.4), with (2.6), we have the following discrete time model of (2.2) for the approximation of the model (2.1) at y_t, where $y_t \neq 0$ and $J(y_t) \neq 0$,

$$y_{t+\Delta t} = A(y_t) y_t + B(y_t) n_{t+\Delta t} \tag{2.7}$$

and where $n_{t+\Delta t}$ is a discrete time Gaussian white noise of variance σ^2,

$$A(y_t) = \exp\{K(y_t)\Delta t\}, \tag{2.8}$$

$$B(y_t) = \sqrt{\frac{\exp\{2K(y_t)\Delta t\} - 1}{2K(y_t)}}, \tag{2.9}$$

$$K(y_t) = \frac{1}{\Delta t} \log[1 + J(y_t)^{-1}\{\exp\{J(y_t)\Delta t\} - 1\} f(y_t)/y_t]$$

and $J(y_t) = [\partial f(y)/\partial y]_{y=y_t}$. Note that if Δt is sufficiently small, then $1 + J(y_t)^{-1}[\exp\{J(y_t)\Delta t\} - 1] f(y_t)/y_t > 0$ and $K(y_t)$ is well defined, and so are $A(y_t)$ and $B(y_t)$ for any $f(y)$ which satisfies $f(0) = 0$. We call the model (2.7) a *locally linearized time series model.* Since the model is a kind of Markov chain defined on a continuous state space, we can also call the model a *locally linearized Markov chain model.*

We note that if we expand $\exp(\cdot)$ in (2.8) and (2.9) with respect to Δt and approximate them up to the order Δt, it follows that

$$A(y_t) = 1 + \Delta t f(y_t) \tag{2.10}$$

$$B(y_t) = \sqrt{\Delta t}. \tag{2.11}$$

The model (2.7), with $A(y_t)$ of (2.10) and $B(y_t)$ of (2.11) is known by probabilists as a consistent discrete time model for the Markov diffusion process $y(t)$ defined by (2.1) (Maruyama (1983), Gikhman and Skorohod (1965)); and it is commonly used for the simulation of Markov diffusion processes. However, when $f(y_t)$ is a nonlinear polynomial function of y_t, the function $A(y_t)$ of (2.10) is also a polynomial function of y_t, and the Markov chain process defined by (2.10) is divergent for a fixed Δt. As we can imagine from (2.11), $B(y_t)$ of (2.9) is almost constant while $A(y_t)$ of (2.8) is quite different from a constant function. We shall see this in some examples in the next section.

It can easily be checked if a locally linearized Markov chain model (2.7) is computationally stable by observing the behaviour of the function $A(y_t)$ for large y_t; it is stable if $A(y_t) < 1$ for large $|y_t|$ and unstable if $A(y_t) > 1$ for large $|y_t|$. Sufficient conditions for $f(y)$ to produce an ergodic locally linearized Markov chain are given in Ozaki (1985a).

The model (2.7) was defined on the region of y_t where $y_t J(y_t) \neq 0$. The remaining problem is how to define the discrete time dynamical system at the region (which is measure zero though) where $y_t = 0$ or $y_t = \eta$ but $J(\eta) = 0$.
If a) $y_t = 0$ and the following limit A_0, is finite,

$$\lim_{y_t \to 0}[1 + J(y_t)^{-1}\{\exp\{J(y_t)\Delta t\} - 1\}f(y_t)/y_t] = A_0 < \infty,$$

then it will be natural to define $y_{t+\Delta t}$ at $y_t = 0$ by $y_{t+\Delta t} = \sqrt{\Delta t}\, n_{t+\Delta t}$.
If b) $y_t = 0$, $A_0 = \infty$ and $J(0) \neq 0$, we can define

$$y_{t+\Delta t} = J(0)^{-1}[\exp\{J(0)\Delta t\} - 1]f(0) + \sqrt{\Delta t}\, n_{t+\Delta t}.$$

If c) $y_t = 0$, $A_0 = \infty$ and $J(0) = 0$, we can define

$$y_{t+\Delta t} = \Delta t f(0) + \sqrt{\Delta t}\, n_{t+\Delta t}.$$

If d) $J(\eta) = 0$ for some $\eta \neq 0$, we have

$$\lim_{y \to \eta}[1 + J(y)^{-1}\{\exp\{J(y)\Delta t\} - 1\}f(y)/y] = 1 + \Delta t f(\eta)/\eta.$$

Then it will be natural to define $y_{t+\Delta t}$ by

$$y_{t+\Delta t} = \eta + \Delta t f(\eta) + \sqrt{\Delta t}\, n_{t+\Delta t}.$$

Therefore we can define the discrete time dynamical system for the region of y_t where $y_t J(y_t) = 0$,

$$y_{t+\Delta t} = \begin{cases} 0 + \sqrt{\Delta t}\, n_{t+\Delta t} & \text{for the case a)} \\ J(0)^{-1}[\exp\{J(0)\Delta t\} - 1]f(0) + \sqrt{\Delta t}\, n_{t+\Delta t} & \text{for the case b)} \\ \Delta t f(0) + \sqrt{\Delta t}\, n_{t+\Delta t} & \text{for the case c)} \\ \eta + \Delta t f(\eta) + \sqrt{\Delta t}\, n_{t+\Delta t} & \text{for the case d).} \end{cases} \tag{2.12}$$

The goodness of the discretization scheme of the stochastic dynamical systems can be measured by speed of convergence of the approximate process to the true process in terms of Δt. We can easily see the optimality of the present scheme (2.7) by direct application of the Markov semi-group expansion technique due to Milshtein (1974).

3. Some Examples

Condiser some examples of the local linearization.

Example 1. $\dot{y} = -y^3 + n(t)$

From this we have the locally linearized model (2.7) with

$$A(y_t) = \frac{2}{3} + \frac{1}{3}\exp(-3y_t^2\Delta t) \quad \text{and} \quad B(y_t) \approx \sqrt{\Delta t}.$$

The function $A(y_t)$ is a smooth function of y_t and it tends to a constant, namely, 2/3, for $|y_t| \to \infty$. This kind of models is what Ozaki (1981a, b) and Haggan and Ozaki (1981) considered in the statistical analysis of some nonlinear time series.

Example 2. $\dot{y} = 2y - y^3 + n(t)$

The function $f(y) = 2y - y^3$ has three zero points, $h = 0$, $\eta^+ = \sqrt{2}$ and $\eta^- = -\sqrt{2}$. They are the singular points of the system. If an initial value y_0 of the system is one of the three singular points, then $y(t)$ stays at y_0 for any $t > 0$ when the white noise input $n(t)$ is suppressed. The functions $A(y_t)$ and $B(y_t)$ of the locally linearized time series model are

$$A(y_t) = \frac{-2y_t^2}{2 - 3y_t^2} + \frac{2 - y_t^2}{2 - 3y_t^2}\exp\{(2 - 3y_t^2)\Delta t\} \quad \text{and} \quad B(y_t) \approx \sqrt{\Delta t}.$$

$A(y_t)$ is a smooth function of y_t and tends to 2/3 for $|y_t| \to \infty$. Simulation data of the locally linearized model is shown in Figure 3.1, where $\Delta t = 0.1$ and $\sigma^2 = 1$. The figure shows that y_t fluctuates around one of the singular points and sometime moves from a stable singular point to another depending on the white noise input. This behaviour is the same as that of the amplitude-dependent autoregressive models given in Ozaki (1980, 1981a).

Example 3. $\dot{y} = -\sqrt{2}\,\alpha \tanh(\sqrt{2}\,y) + n(t)$

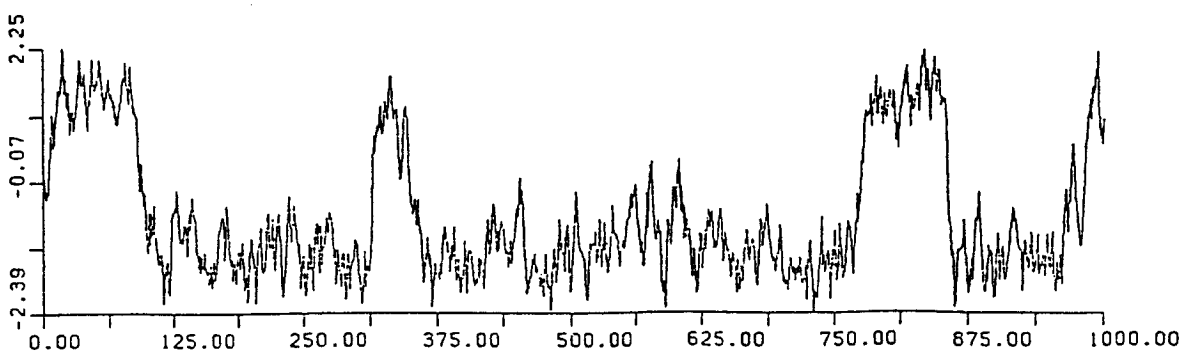

Figure 3.1. Simulated data of Example 2.

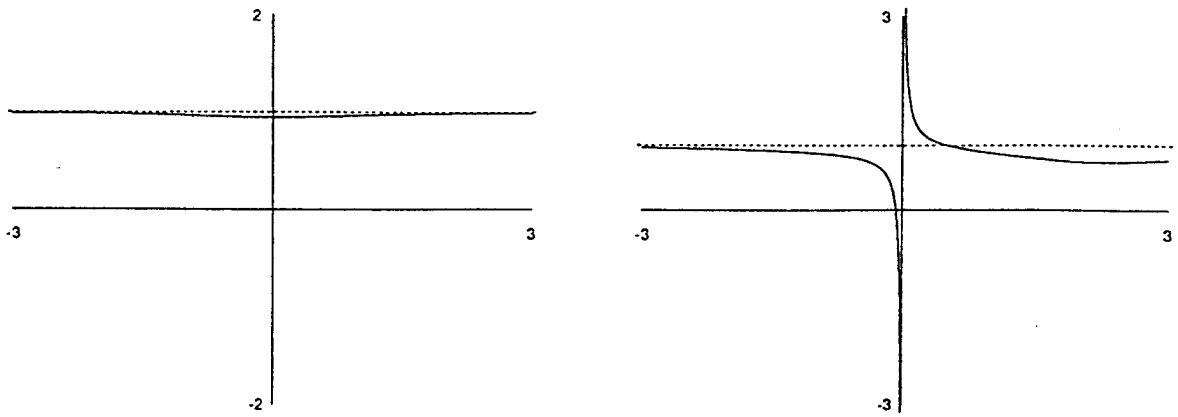

Figure 3.2. $A(y_t)$'s of Examples 3 and 4.
(a) $A(y_t)$ of Example 3.
(b) $A(y_t)$ of Example 4.

The process $x(t)$ obtained from the solution $y(t)$ of this model by a memory-less transformation, $x(t) = \sinh\{\sqrt{2}\, y(t)\}$ is known as a diffusion process whose marginal distribution is Cauchy for $\alpha = 0.5$ (Wong (1963)). A figure of the function $A(y_t)$ is shown in Figure 3.2a where $\Delta t = 0.06$. $B(y_t)$ is almost constant as in the previous examples, while $A(y_t)$ is less than 1 for small $|y_t|$ and tends to 1 for $|y_t| \to \infty$. That means the movement of y_t becomes almost like a Brownian motion when y_t goes far away from the origin. This explains why the process has a marginal distribution with heavy tails.

Example 4. $\dot{y} = \dfrac{\alpha\beta}{\sqrt{2\beta}} - \dfrac{\exp(\sqrt{2\beta}\, y)}{\sqrt{2\beta}} + n(t)$

This is an example with a dynamical system $\dot{y} = f(y)$ with $f(0) \neq 0$. The process $x(t)$ obtained from $y(t)$ by the memory-less transformation, $x(t) = \exp\{\sqrt{2\beta}\, y(t)\}$, is known as a Gamma-distributed diffusion process. A figure of the function $A(y_t)$ for the model is in Figure 3.2b where we used $\Delta t = 0.1$, $\alpha = 2$ and $\beta = 1$. Here the function $A(y_t)$ is defined only for $y_t \neq 0$. Since

$$A_0 = \lim_{y_t \to 0}[1 + J(y_t)^{-1}\{\exp\{J(y_t)\Delta t\} - 1\}f(y_t)/y_t] = \infty$$

and $J(0) \neq 0$ it is in the exceptional case of b) of (2.12) at $y_t = 0$ and the dynamics at $y_t = 0$ is defined by

$$y_{t+\Delta t} = J(0)^{-1}[\exp\{J(0)\Delta t\} - 1]f(0) + \sqrt{\Delta t}\, n_{t+\Delta t}.$$

4. Estimation and Identification

In this section the identification of a continuous time stochastic dynamical system model,

$$\dot{y} = f(y|\boldsymbol{a}) + n(t), \tag{4.1}$$

is considered, where $n(t)$ is a Gaussian white noise of variance σ^2. To identify the model we first try to estimate, from the observation data $y_1, y_2, \ldots, y_N$, the following locally linearized model,

$$\begin{aligned} y_{t+1} &= A(y_t|\boldsymbol{a})y_t + B(y_t|\boldsymbol{a})n_{t+1} \\ A(y_t|\boldsymbol{a}) &= \exp\{K(y_t|\boldsymbol{a})\Delta\} \\ B(y_t|\boldsymbol{a}) &= \sqrt{\frac{\exp\{2K(y_t|\boldsymbol{a})\Delta\} - 1}{2K(y_t|\boldsymbol{a})}} \\ K(y_t|\boldsymbol{a}) &= \frac{1}{\Delta}\log\{1 + J_t^{-1}(e^{J_t\Delta} - 1)f(y_t|\boldsymbol{a})/y_t\} \\ J_t &= \left(\frac{\partial f(y|\boldsymbol{a})}{\partial y}\right)_{y=y_t} \end{aligned} \tag{4.2}$$

where $\boldsymbol{a} = (a_1, a_2, \ldots, a_k)'$ is some parameter vector of $f(y|\boldsymbol{a})$. The data are generated from (2.1) at equally spaced time points, $t, t + \Delta, t + 2\Delta, \ldots$. In this section we use Δ, instead of Δt, for the sampling interval to stress that the sampling interval is fixed at the stage of model estimation and identification. Of course, as in the case of linear time series modelling, the sampling interval needs to be small enough to enable the estimation of the essential features of y_t. After estimating the parameters of the model (2.2) we try to identity the original continuous time stochastic dynamical system model (2.1) using the estimated nonlinear time series model.

4.1. Estimation of $\boldsymbol{a}$ and σ^2

The log-likelihood of model (4.2) is

$$\begin{aligned} &\log p(y_1, y_2, \ldots, y_N|\boldsymbol{a}, \sigma^2) \\ &= \log p(y_2, y_3, \ldots, y_N|y_1, \boldsymbol{a}, \sigma^2) + \log p(y_1|\boldsymbol{a}, \sigma^2) \\ &= \log p(n_2, n_3, \ldots, n_N|y_1, \boldsymbol{a}, \sigma^2) + \log\det\{J(\boldsymbol{y}, \boldsymbol{n})\} + \log p(y_1|\boldsymbol{a}, \sigma^2) \end{aligned} \tag{4.3}$$

where $J(\boldsymbol{y},\boldsymbol{n})$ is the Jacobian matrix of the transformation from $(y_2, y_3, \ldots, y_N)$ to $(n_2, n_3, \ldots, n_N)$. Since $\det\{J(\boldsymbol{y},\boldsymbol{n})\} = B(y_1|\boldsymbol{a})B(y_2|\boldsymbol{a})\cdots B(y_{N-1}|\boldsymbol{a})$, (4.3) is given by

$$\begin{aligned} = &- \sum_{t=1}^{N-1} \frac{\|B(y_t|\boldsymbol{a})^{-1}\{y_{t+1} - A(y_t|\boldsymbol{a})y_t\}\|^2}{2\sigma^2} - \frac{N-1}{2}\log\sigma^2 \\ &- \frac{N-1}{2}\log 2\pi + \sum_{t=1}^{N-1}\log\{B(y_t|\boldsymbol{a})\} + \log p(y_1|\boldsymbol{a},\sigma^2). \end{aligned} \tag{4.4}$$

For large N, the last term of the expression is negligibly small compared with the rest and we can ignore it. Since the log-likelihood function satisfies

$$\left[\frac{\partial \log p(y_1, y_2, \ldots, y_N|\boldsymbol{a},\sigma^2)}{\partial\sigma^2}\right]_{\sigma^2=\hat{\sigma}_N^2} = 0$$

the maximum likelihood estimate $\hat{\sigma}_N^2$ satisfies the equation,

$$\sigma^2 = \frac{1}{N-1}\sum_{t=1}^{N-1}\|B(y_t|\boldsymbol{a})^{-1}\{y_{t+1} - A(y_t|\boldsymbol{a})y_t\}\|^2 \tag{4.5}$$

with $\boldsymbol{a}$ replaced by its estimate. Also, since the first term of (4.4) becomes constant at the maximum point, the maximum log-likelihood is obtained by maximizing

$$-\frac{N-1}{2}\log\sigma^2 + \sum_{t=1}^{N-1}\log\{B(y_t|\boldsymbol{a})\} \tag{4.6}$$

with respeect to $\boldsymbol{a}$, where σ^2 is given by (4.5) and is a function of $\boldsymbol{a}$. For the maximization of the likelihood (4.6) we use some nonlinear optimization procedure (see Fletcher and Powell (1963) or Ishiguro and Akaike (1989)). In the optimization procedure, the standard error of the estimates is easily obtained numerically from the inverse of the Hessian matrix at the maximum point of the log-likelihood (4.6).

The maximum likelihood estimates $\hat{\boldsymbol{h}}_{N,\Delta} = (\hat{\boldsymbol{a}}, \hat{\sigma}_N^2)'$ thus obtained are known to be consistent and asymptotically normally distributed under some regularity conditions (Billingsley (1961)). However, this is under the assumption that the data is generated from the discrete time Markov chain model (4.2) with the true parameter $\boldsymbol{h}_0 = (\boldsymbol{a},\sigma^2)'$. Since the data is generated from the continuous time model (4.1), the maximum likelihood estimate tends not to $\boldsymbol{h}_{d,\Delta}(=\boldsymbol{h}_0)$ for $N \to \infty$ but to some value $\boldsymbol{h}_{c,\Delta}$ which is characterized by the property,

$$E_c[\log p(\boldsymbol{y},\boldsymbol{h}) - \log p(\boldsymbol{y},\boldsymbol{h}_{c,\Delta})] < 0$$

for any $\boldsymbol{h} \neq \boldsymbol{h}_{c,\Delta}$ (Huber (1967)), where E_c means the expectation with respect to the probability measure defined by the continuous time process (4.1), and $p(\boldsymbol{y}, \boldsymbol{h})$ is a likelihood function of $\boldsymbol{h}$ for given data $\boldsymbol{y}$, $\boldsymbol{h}_{c,\Delta}$ is not necessarily equal to $\boldsymbol{h}_{d,\Delta}$. However the difference between them tends to zero, i.e. $|\boldsymbol{h}_{c,\Delta} - \boldsymbol{h}_{d,\Delta}| \to 0$, for $\Delta \to 0$, because as is seen from the definition of the locally linearized model, the process defined by (4.2) weakly converges, for $\Delta \to 0$, to the process defined by (4.1) (see Gikhman and Skorohod (1965), ch.IX). This means the above procedure is consistent in the sense that $\hat{\boldsymbol{h}}_{N,\Delta}$ converges to $\boldsymbol{h}_0$ for $\Delta \to 0$ and $N\Delta \to \infty$. It must be noted, however, that both N and Δ are fixed in real applications.

4.2. Identifying $f(y|\boldsymbol{a})$ from $A(y|\hat{\boldsymbol{a}})$

We note that the estimated parameters $\hat{\boldsymbol{a}}$ and $\hat{\sigma}_N^2$ are for the model (4.2) and not for the original contiuous time model (4.1). What is required next is to identify the continuous time model (4.1) from the estimated model. It will be natural and reasonable to employ the model,

$$\dot{y} = f(y|\hat{\boldsymbol{a}}) + n(t) \tag{4.7}$$

as an identified model for (4.1) with the estimated parameters $\hat{\boldsymbol{a}}$ and $\hat{\sigma}_N^2$ for the model,

$$y_{t+1} = A(y_t|\hat{\boldsymbol{a}})y_t + B(y_t|\hat{\boldsymbol{a}})n_{t+1}. \tag{4.8}$$

This is because the model (4.8) is derived from the model (4.7) by applying the local linearization scheme given in Section 2. However $f(y|\hat{\boldsymbol{a}})$ of (4.7) is not the only function which leads to (4.8) by applying the local linearization.

Knowing $f(y|\hat{\boldsymbol{a}})$ we can obtain $A(y|\hat{\boldsymbol{a}})$ uniquely; however from $A(y|\hat{\boldsymbol{a}})$ we cannot obtain $f(y|\hat{\boldsymbol{a}})$ unless we know the functional form of $f(y|\boldsymbol{a})$. This is because an approximation is involved when we derive $A(y|\boldsymbol{a})$ from $f(y|\boldsymbol{a})$ by the local linearization. In light of the derivation of $A(y|\boldsymbol{a})$ from $f(y|\boldsymbol{a})$ in Section 2 we know that any function $f(y)$ which satisfies the relation,

$$A(y|\hat{\boldsymbol{a}}) = 1 + \left(\frac{\partial f(y)}{\partial y}\right)^{-1}\left[\exp\left\{\frac{\partial f(y)}{\partial y}\Delta\right\} - 1\right]\frac{f(y)}{y} \tag{4.9}$$

yields the model (4.8) by applying local linearization to the continous time model, $\dot{y} = f(y) + n(t)$. Now (4.9) is equivalent to

$$f(y) = \{A(y|\hat{\boldsymbol{a}}) - 1\}\frac{\left\{\dfrac{\partial f(y)}{\partial y}\right\}y}{\exp\left\{\dfrac{\partial f(y)}{\partial y}\Delta\right\} - 1}. \tag{4.10}$$

Unfortunately the function $f(y)$ is not given explicitly by $A(y|\hat{\boldsymbol{a}})$ because of the Jacobian in the right hand side of (4.10). However, $f(y)$ is given approximately,

including $A(y|\hat{a})$ explicitly, by

$$f(y) \approx \frac{1}{\Delta}\{A(y|\hat{a}) - 1\}y, \tag{4.11}$$

which is obtained from (4.10) by employing the approximation,

$$\exp\left\{\frac{\partial f(y)}{\partial y}\Delta\right\} \approx 1 + \frac{\partial f(y)}{\partial y}\Delta.$$

One method to improve the approximate function $f(y)$ of (4.11) is to modify the function using information about its deviation from the relation (4.10), i.e. we can adjust the old function $f^{(k)}(y)$, at each point $y(= y_i, y_i+\Delta y, y_i+2\Delta y, \ldots)$ on any finite interval of y, into a new function $f^{(k+1)}(y)$ by

$$f^{(k+1)}(y) = f^{(k)}(y) - \Delta s\left[f^{(k)}(y) - \{A(y|\hat{a}) - 1\}\frac{\left\{\dfrac{\partial f^{(k)}(y)}{\partial y}\right\}y}{\exp\left\{\dfrac{\partial f^{(k)}(y)}{\partial y}\Delta\right\} - 1}\right] \tag{4.12}$$

where Δs is some small number. (4.12) means that if the old value $f^{(k)}(y)$ is too large or too small to satisfy (4.10), the new value $f^{(k+1)}(y)$ is set by reducing or increasing $f^{(k)}(y)$ proportionally. We note that the iterative procedure (4.12) is equivalent to the numerical solution for the following partial differential equation,

$$\frac{\partial f(y,s)}{\partial s} = -f(y,s) + \{A(y|\hat{a}) - 1\}\frac{\left\{\dfrac{\partial f(y,s)}{\partial y}\right\}y}{\exp\left\{\dfrac{\partial f(y,s)}{\partial y}\Delta\right\} - 1}. \tag{4.13}$$

From (4.13) we know that as $s \to \infty$, $\partial f(y,s)/\partial s \to 0$ and $f(y,\infty)$ satisfies the local linearity relation,

$$f(y,\infty) = \{A(y|\hat{a}) - 1\}\frac{\left\{\dfrac{\partial f(y,\infty)}{\partial y}\right\}y}{\exp\left\{\dfrac{\partial f(y,\infty)}{\partial y}\Delta\right\} - 1}.$$

The Jacobian $\partial f^{(k)}(y,s)/\partial y$ in the iterative procedure (4.12) can be obtained by a numerical differencing method, e.g. $\{f^{(k)}(y_{i+1}) - f^{(k)}(y_i)\}/\Delta y$. If Δs is sufficiently small compared with Δy the iterative numerical procedure for the partial differential equation is known to work properly.

As briefly mentioned earlier, the function thus obtained is not the same as the original function $f(y|\hat{a})$ since an approximation comes in through the local

linearity relation (4.10) and the goodness of the approximation depends on how the data were sampled. So, the function obtained by the above procedure depends on the sampling interval, and we write it as $f_\Delta(y|\hat{a})$. In real data analysis we cannot change Δ of $f_\Delta(y|\hat{a})$ since the sampling interval of the data is fixed and accordingly $A(y|\hat{a})$ is fixed. However it can be seen how fast $f_{\Delta t}(y|a)$ approaches $f(y|a)$ as $\Delta t \to 0$ by recalculating $A(y|a)$ from $f(y|a)$ by using (2.8) for each different Δt and by employing the above numerical procedure for each $A(y|a)$. Figure 4.1 shows $f(y|a)$ and $f_{\Delta t}(y|a)$ of the function in Example 2 for different Δt's.

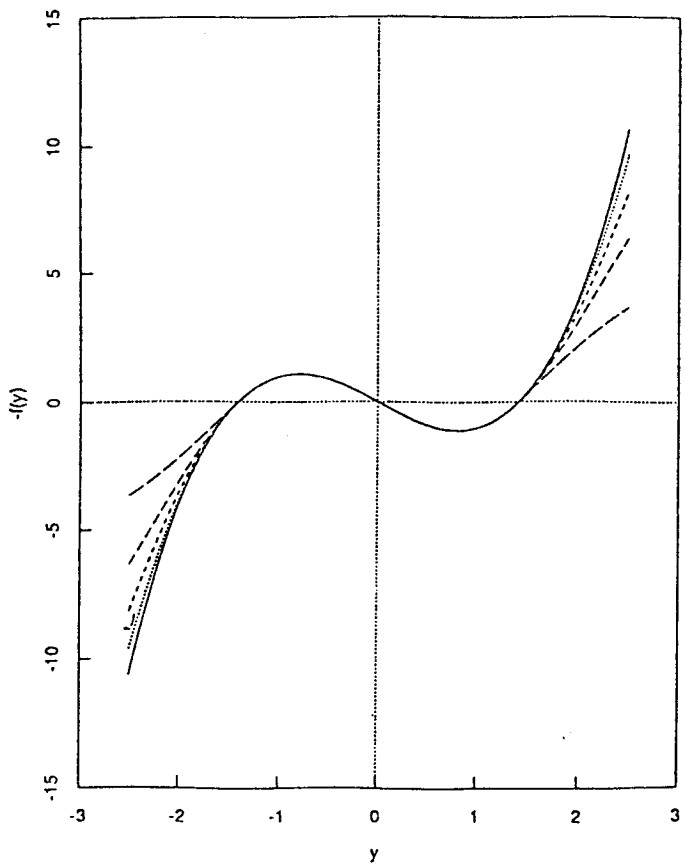

Figure 4.1 True function $f(y|a)$ and approximate functions $f_{\Delta t}(y|a)$ for Δt = 0.2, 0.1, 0.06 and 0.03.

———— : true $f(y|a)$
..... : $\Delta t = 0.03$
- - - - - - - - - - : $\Delta t = 0.06$
– – – – – – – – : $\Delta t = 0.1$
— — — — — : $\Delta t = 0.2$

4.3. Utilizing time series models

From the previous discussion we saw that we can obtain the function $f_{\Delta t}(y|\hat{a})$ numerically from $A(y|\hat{a})$ of the estimated model $y_{t+1} = A(y_t|\hat{a})y_t + B(y_t|\hat{a})n_{t+1}$. This suggests an interesting new method to identify the continuous time model utilizing some parametric nonlinear time series models. Since $B(y_t|\hat{a})$ is related to $A(y_t|\hat{a})$ by the local linearization relations (2.8) and (2.9), it follows that

$$B(y_t|\hat{a}) = \sqrt{\frac{\Delta\{A(y_t|\hat{a})^2 - 1\}}{2\log A(y_t|\hat{a})}}.$$

Then, by using some $A(y_t|\phi)$ parameterized in the discrete time domain with some parameter vector $\phi = (\phi_1, \phi_2, \ldots, \phi_k)'$, we can think of using some nonlinear time series model of the following form,

$$y_{t+1} = A(y_t|\phi)y_t + B(y_t|\phi)n_{t+1} \tag{4.14}$$

where

$$B(y_t|\phi) = \sqrt{\frac{\Delta\{A(y_t|\phi)^2 - 1\}}{2\log A(y_t|\phi)}}. \tag{4.15}$$

Suppose we obtained the estimates $\hat{\phi}$ and $\hat{\sigma}^2$ for the model (4.14). Then from $A(y|\hat{\phi})$, by using the numerical procedure in Section 4.2, we can obtain a figure of the function $f_\Delta(y|\hat{\phi})$ for any finite interval of y. If there is some general parametric representation $A(y_t|\phi)$ for the model (4.14) this procedure provides a useful method of guessing the function figure when we have no idea of the true parametric function form of the original continuous time model.

For the parameterization of $A(y_t|\phi)$, the examples of $A(y_t)$ in Section 3 provide important and useful information. Common features in the first three examples in Section 3 are that $A(y_t)$ is a continuous function of y_t and tends to a constant for $|y_t| \to \infty$. One example of such a parameterization is

$$A(y_t|\phi) = \phi_1 + (\phi_2 + \phi_3 y_t + \cdots + \phi_k y_t^{k-2})\exp(-cy_t^2). \tag{4.16}$$

Another example is

$$A(y_t|\phi) = \begin{cases} \phi_1 + \phi_2 y_t + \cdots + \phi_k y_t^{k-1} & \text{for } |y_t| < T \\ \phi_1 + \phi_2 T + \cdots + \phi_k T^{k-1} & \text{for } |y_t| \geq T \end{cases} \tag{4.17}$$

where T is some threshold value. With these $A(y_t|\phi)$ we can have time series models (4.14) with (4.15).

For this nonlinear time series modelling approach to be valid and work properly it is necessary that the function $A(y_t)$ which corresponds to the true model be finite for the region $\min y_t < y_t < \max y_t$. Therefore the true $f(y)$ needs to be zero at the origin. Further, the sampling interval needs to be sufficiently small so that $A(y_t)$ is positive and $B(y_t)$ is well defined. As for the examples in Section 3, the $A(y_t)$'s of the first three examples are appropriate. However there is some difficulty in Example 4 since the $f(y)$ of the example is not zero at $y = 0$. A special treatment, such as interpolation of the function $f(y)$, is needed in the region near $y = 0$ where $A(y)$ can be negative.

For the parameter estimation of the model (4.14) the same maximum likelihood method is used as the one given in Section 4.1. The computation involved can be drastically reduced if we employ the approximation $\exp\{K(y_t|\hat{\phi})\Delta\} \approx 1 + K(y_t|\hat{\phi})\Delta$. With this approximation, $B(y_t|\hat{a}) \approx \sqrt{\Delta}$ and the model (4.14) becomes

$$y_{t+1} = A(y_t|\phi)y_t + \sqrt{\Delta}\, n_{t+1}. \tag{4.18}$$

If $A(y_t|\phi)$ of the model (4.18) is linearly parameterized with ϕ, then the least squares estimate of ϕ, which is asymptotically equivalent to the maximum likelihood estimate, can be obtained by simply solving the linear normal equation. For example, model (4.18), using parameterization (4.16) with a fixed c, is equivalent to the Extended Exponential AR model of Ozaki (1985a) and the least squares estimate of $\phi = (\phi_1, \phi_2, \ldots, \phi_k)'$ can be obtained by solving the linear equation,

$$X'Y = (X'X)\phi$$

with respect to ϕ where $Y = (y_2, y_3, \ldots, y_N)'$ and

$$X = \begin{bmatrix} y_1 & y_1\exp(-cy_1^2) & \cdots & y_1^{k-1}\exp(-cy_1^2) \\ y_2 & y_2\exp(-cy_2^2) & \cdots & y_2^{k-1}\exp(-cy_2^2) \\ \cdots & \cdots & \cdots & \cdots \\ y_{N-1} & y_{N-1}\exp(-cy_{N-1}^2) & \cdots & y_{N-1}^{k-1}\exp(-cy_{N-1}^2) \end{bmatrix}$$

c needs to be fixed in this simple estimation method. However it does not cause much computational complexity because c is positive and less than $-\log\varepsilon/\max y_t^2$ where ε is a small number such that $\exp(-cy^2) < \varepsilon$ and is negligible for all $y > \max y_t$. So we have only to find ϕ which minimizes the residual variance for several different c's where $0 < c < -\log\varepsilon/\max y_t^2$.

Model (4.18) using parameterization (4.16) with $k = 2$ is equivalent to a first order Exponential AR(Exp AR) model (see Priestley (1981)). The model (4.18) with the parameterization (4.17) is equivalent to the NonLinear Threshold AR(NLTAR) model of Ozaki (1981a). We call a model (4.18) with these

parameterizations an *amplitude-dependent autoregressive model.* This type of parameterization has been considered in nonlinear time series data analysis by Ozaki and Oda (1978) for ship rolling data and by Haggan and Ozaki (1981) for Canadian lynx data.

Thus, we have obtained a practical method to get a continuous time stochastic dynamical system model from a discrete time nonlinear time series model. An important implication of this is that, even though we have no idea of the functional form of continuous time dynamical system model, for given time series data, we can still try to fit a general nonlinear time series model parameterized in the discrete time domain. The model, although it is parametric in the discrete time domain, is nonparametric in the continuous time domain. Having fitted the nonlinear time series model, the fitted model can be transformed into the continuous time domain, by the numerical procedure given in Section 4.2. This will give some idea of the appropriate type of functional form for the continuous time stochastic dynamical system model to be used for the data. Based on this information, we can introduce a new model parameterized in the continuous time domain and estimate it by the maximum likelihood method given in Section 4.1. We can also use the powerful analytic tools for the continuous time stochastic dynamical system model to obtain the probabilistic characteristics of the time series. The bridge brought by the local linearization relation (4.10) is open not only for ExpAR models but also for any nonlinear time series models of (4.14) or (4.18) type with $A(y_t|\phi)$ smooth in terms of y_t.

4.4. Numerical Examples

The estimation and identification methods described above may be understood better with some numerical examples. We applied the maximum likelihood method to the simulated data of Example 2 in Section 3. The data is shown in Figure 3.2, where the variance of the white noise is 1 and $\Delta = 0.1$. The number of data points is 1000. We used the third order polynomial model,

$$\dot{y} = a_1 y + a_2 y^2 + a_3 y^3 + n(t) \tag{4.19}$$

and obtained the following estimates,

$$\hat{a}_1 = 2.0312, \quad \hat{a}_2 = -0.3899 \times 10^{-1}, \quad \hat{a}_3 = -1.0930 \quad \text{and} \quad \hat{\sigma}^2 = 0.8043.$$

The obtained function, $f_p(y|\hat{a}) = \hat{a}_1 y + \hat{a}_2 y^2 + \hat{a}_3 y^3$ is quite similar to the true function, $f(y) = 2y - y^3$. This may be partly because we estimated the coefficients using the polynomial function which includes the true function as a special case.

Usually, in real data analysis, the true function is not known. The resulting estimates may then depend on the parametric model used and in some case the

resulting estimated function may not be close to the true function. This will be the main reason why nonparametric methods are considered in nonlinear time series analysis. To see how the misspecification of the functional form affects the results, we applied the maximum likelihood method to another set of simulated data generated by a model of Example 3, whose nonlinear function is not a polynomial but the following transcendental function,

$$f_T(y|\alpha) = -\sqrt{2}\,\alpha \tanh(\sqrt{2}\,y), \tag{4.20}$$

where $\alpha = 0.5$, the noise variance is $\sigma^2 = 1$, $N = 4000$ and $\Delta = 0.06$. The first 1000 points of the data are shown in Figure 4.2. To this data we first fit the model

$$\dot{y} = -\sqrt{2}\,\alpha \tanh(\sqrt{2}\,y) + n(t).$$

The estimated parameters are $\hat{\alpha} = 0.5682$ and $\hat{\sigma}^2 = 1.0162$. The figures of the functions $f_T(y|a)$ and $f_T(y|\hat{\alpha})$, i.e. (4.20) with the true α and (4.20) with the estimated $\hat{\alpha}$ respectively, are shown in Figure 4.3.

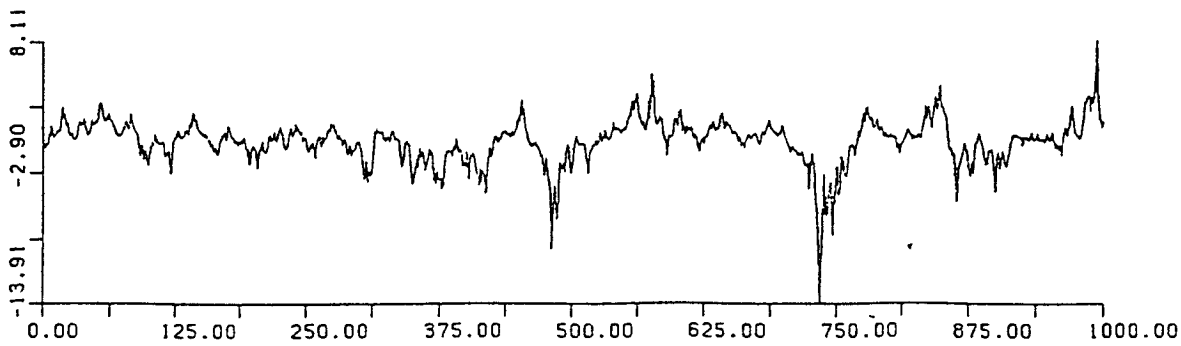

Figure 4.2. The first 1000 data of the simulated data of Example 3.

To see the effect of model misspecification in the parametric approach, we also fit to the data a model with a third order polynomial function, i.e. the model (4.19). The obtained maximum likelihood estimates of the parameters are

$$\hat{a}_1 = -0.8548, \quad \hat{a}_2 = 0.9022 \times 10^{-2}, \quad \hat{a}_3 = 0.4731 \times 10^{-1} \quad \text{and} \quad \hat{\sigma}^2 = 1.0722.$$

The estimated function, $f_p(y|\hat{a}) = \hat{a}_1 y + \hat{a}_2 y^2 + \hat{a}_3 y^3$ and the true function $f_T(y|\alpha)$ are not very different for the region $-3 < y < 3$. However they are very different for large $|y|$ outside the region. This is in a sense natural because the polynomial functions tend to go to $\pm\infty$ for large $|y|$. Thus polynomial functions are useful only to approximate functions within a finite region and are not appropriate to approximate functions which go to a constant as $|y| \to \infty$.

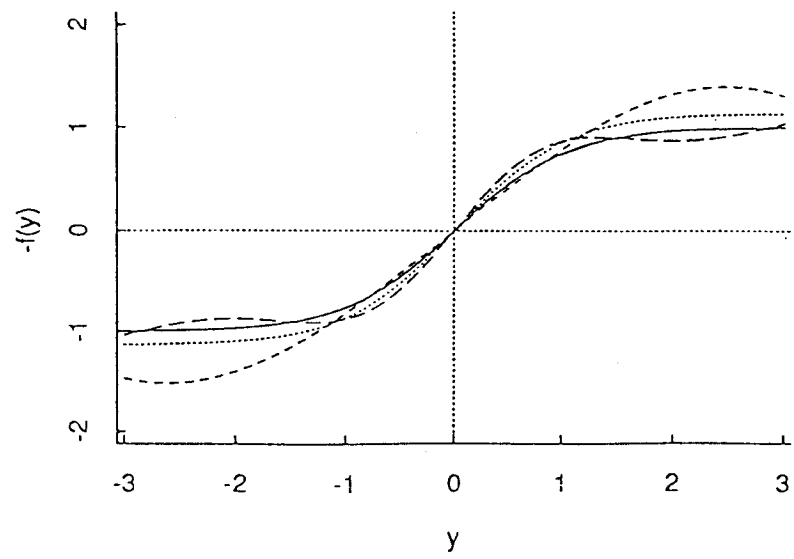

Figure 4.3. True function, $f_T(y|\alpha)$, of (4.21) and estimated functions, $f_T(y|\hat{\alpha})$, $f_p(y|\hat{\boldsymbol{a}})$ and $f_\Delta(y|\hat{\phi})$.

———————— : true function $f_T(y|\alpha)$.
..... : estimated function $f_T(y|\hat{\alpha})$.
- - - - - - - - - - : estimated polynomial function $f_p(y|\hat{\boldsymbol{a}})$.
– – – – – – – – – : identified function $f_\Delta(y|\hat{\phi})$ from the estimated function $A(y_t|\hat{\phi})$.

The above consideration suggests another method of identifying the function $f(y)$, i.e. the use of time series models. We used the model,

$$y_{t+1} = A(y_t|\phi)y_t + B(y_t|\phi)n_{t+1}$$

$$B(y_t|\phi) = \sqrt{\frac{\Delta\{A(y_t|\phi)^2 - 1\}}{2\log A(y_t|\phi)}}$$

$$A(y_t|\phi) = \phi_1 + \phi_2 \exp(-cy_t^2)$$

and the obtained maximum likelihood estimates are

$$\hat{\phi}_1 = 9798, \quad \hat{\phi}_2 = -0.5598 \times 10^{-1}, \quad \hat{c} = -0.5577 \quad \text{and} \quad \hat{\sigma}^2 = 1.0150.$$

The function $f_\Delta(y|\hat{\phi})$ obtained from $A(y|\hat{\phi})$ using the numerical procedure given in Section 4.2 is shown in Figure 4.3. The function approximates the true function $f_T(y|\alpha)$ well for the region $-3 < y < 3$. However, as we might expect the function $f_\Delta(y|\hat{\phi})$ tends to some linear function outside the region.

It is very clear from the figures in Figure 4.3 that the parameterization in discrete time by an amplitude-dependent AR model (4.6) with (4.16) or (4.17)

is preferable to the parameterization by a polynomial function in the continuous time model (4.1) when the function $f(y)$ approaches a linear function or a constant as $|y| \to \infty$.

When we apply the present estimation method to real data, we have to specify Δ in the model (4.2). Since the data is already taken we know what the sampling interval is. However there is an ambiguity in specifying Δ in the model, i.e. the time unit used in the model to represent the time intervals between data, hours or seconds etc. This ambiguity does not cause any essential problem because, as we can see from (4.11), Δ affects only the scale of the function $f(y|\boldsymbol{a})$ and not its shape.

5. Discussions

The time discretization method shown in the present paper provides us with a bridge connecting a nonlinear time series model and a stochastic dynamical system driven by Gaussian white noise. The bridge makes it possible for time series analysts to use not only stochastic dynamical system but also Markov diffusion process models in modelling non-Gaussian time series data. This is because any Markov diffusion process $x(t)$ has a representation,

$$\begin{aligned} x(t) &= h(y(t)) \\ \dot{y} &= f(y) + n(t) \end{aligned} \tag{5.1}$$

where $h(\cdot)$ is a smooth function which gives a memory-less transformation of $y(t)$. $f(\cdot)$ is a smooth function which defines the nonlinear dynamics of $y(t)$ and $n(t)$ is a Gaussian white noise (Goel and Richter-Dyn (1974), Ozaki (1985a)).

The connection of nonlinear time series models with diffusion processes also gives time series analysts a useful tool in understanding the distributional aspects of nonlinear time series. Ozaki (1985a) showed that for any given density distribution $W(x)$ defined by

$$\frac{dW(x)}{dx} = \frac{c_1(x)}{c_2(x)} W(x) \tag{5.2}$$

there is a diffusion process whose marginal density distribution is $W(x)$. The diffusion process is specified by the following Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}\left[\left\{c_1(x) + \frac{\partial c_2(x)}{\partial x}\right\}p\right] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[2c_2(x)p] \tag{5.3}$$

where $c_1(x)$ and $c_2(x)$ of (5.2) are analytic functions. A family of distributions defined by the system (5.2) includes the exponential family and also the distributions defined by the Pearson system. This means, by using the bridge, we

can generate nonlinear time series with any marginal distribution of the Pearson system or the exponential family.

The $c_1(x)$ and $c_2(x)$ of (5.2) are not necessarily mutually irreducible. Therefore we can have infinitely many different diffusion processes which have the same marginal distributions from (5.2) and (5.3) by multiplying the numerator and denominator of (5.2) by any common factor. Two different types of Gamma-distributed processes are shown in Ozaki (1985a). This implies that knowledge of the non-Gaussian marginal distribution of the given time series data does not give us any information of its nonlinear dynamics.

In Markov process theory it is well known that if the process is of continuous type, i.e. the transition probability from x to $R-(x-\varepsilon, x+\varepsilon)$ goes to zero for $\Delta t \to 0$ for any fixed $\varepsilon > 0$, then the process $x(t)$ is a diffusion process. Since any diffusion process has a representation of the form (5.1), this implies that any continuous type non-Gaussian time series can be transformed into Gaussian white noise by a memory-less transformation and a nonlinear dynamic model. Therefore, what we need in nonlinear time series modelling is a memory-less transformation and a nonlinear dynamic model driven by a Gaussian white noise. By these we can transform any continuous type non-Gaussian time series data into Gaussian white noise. It means that a nonlinear dynamic model driven by a non-Gaussian white noise, which may look more general than the nonlinear dynamic model driven by a Gaussian white noise, is not an essential generalization. However, in real time series analysis, we cannot always assume that the series is of continuous type because the data is sometimes contaminated by shot noise type Markov jump processes. In such situations the nonlinear dynamic model driven by non-Gaussian white noise with a fat tail distribution plays an important role in robustifying the estimation procedure.

Once the bridge is made between time series models and stochastic dynamical system models its application may look almost unlimited. However, for the time series modelling approach to be useful in many application fields, the method needs to be extended in several directions. One direction is to a stochastic dynamical system model with an extra observed input process $r(t)$ such as,

$$\dot{y} = f(y) + n(t) + r(t). \tag{5.4}$$

The model (5.4) is an example of a nonlinear storage model, which is used in inventory control in operations research or in stochastic hydrology. An extension of the local linearization method to this type of model is discussed in Ozaki (1985b).

The method needs to be extended to multi-dimensional systems as well, since dynamics of many interesting and important problems in real applications are multi-dimensional. Fortunately, the local linearization method can be ex-

tended to the multi-dimensional stochastic dynamical systems without any essential change, since the analytic solution of a multi-dimensional linear dynamical system, $\dot{z} = Az$ is $z(t) = \exp(tA)z(0)$, where $\exp(tA)$ is a matrix defined by

$$\exp(tA) = I + \sum_{i=1}^{\infty} \frac{t^i}{i!} A^i.$$

An application of the method to two-dimensional nonlinear stochastic dynamical system models in macro-economics is discussed in Ozaki, T. and Ozaki, V. H. (1989).

A nonlinear random vibration model,

$$\ddot{x} + a(x)\dot{x} + b(x)x = n(t) \tag{5.5}$$

can be regarded as an example of the two-dimensional stochastic dynamical system model,

$$\dot{\boldsymbol{z}} = \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{n}(t) \tag{5.6}$$

where $\boldsymbol{z} = (\dot{x}, x)'$, $\boldsymbol{f}(\boldsymbol{z}) = \{-a(x)\dot{x} - b(x)x, \dot{x}\}'$ and $\boldsymbol{n}(t) = (n(t), 0)'$. Since the variance-covariance matrix of $\boldsymbol{n}(t)$ is not of full rank, a special treatment is needed to derive a discrete time model for (5.6) using the local linearization method. The details of this topic can be found in Ozaki (1986, 1989) and Oda, H., Ozaki, T. and Yamanouchi, Y. (1987).

In some applications there are cases where the dimension k of the state variable $\boldsymbol{z}(t)$ is higher than the dimension r of the observed vector time series. A familiar example is a linear Markovian model,

$$\begin{aligned} \boldsymbol{z}_{t+1} &= A\boldsymbol{z}_t + B\boldsymbol{z}_{t+1} \\ \boldsymbol{x}_t &= C\boldsymbol{z}_t \end{aligned} \tag{5.7}$$

where C is not a square matrix but an $r \times k$ rectangular matrix. Since the model (5.7) is known to be equivalent to some ARMA model of AR order k and MA order $k-1$ (Akaike (1974)), there is no difficulty in applying the maximum likelihood method to this situation. However, if the dynamics of the state is nonlinear, the transition matrix A of (5.7) becomes state-dependent and because of this we cannot have a nonlinear ARMA representation for the nonlinear Markovian model. In such a case and also in the case where the observed time series is contaminated by an observation error $\boldsymbol{w}_t$, we need to work on the nonlinear state space representation model,

$$\begin{aligned} \dot{\boldsymbol{z}} &= \boldsymbol{f}(\boldsymbol{z}) + \boldsymbol{n}(t) \\ \boldsymbol{x}_t &= C\boldsymbol{z}(t) + \boldsymbol{w}_t. \end{aligned} \tag{5.8}$$

A maximum likelihood method for the model (5.8) is discussed in Ozaki (1990b) (see also Ozaki (1990c)) using the locally linearized nonlinear filtering method introduced for the model (Ozaki (1990a)).

References

Akaike, H. (1974). Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Statist. Math.* **26**, 363-387.

Billingsley, P. (1961). *Statistical Inference for Markov Processes.* Holt, New York.

Gikhman, I. I. and Skorohod, A. V. (1965). *Introduction to Theory of Random Processes* (translated by Scripta Technica, Inc.). W. B. Saunders Company.

Goel, N. S. and Richter-Dyn, N. (1974). *Stochastic Models in Biology.* Academic Press, New York.

Fletcher, R. and Powell, M. J. D. (1963). A rapidly convergent descent method for minimization. *Computer J.* **6**, 163-168.

Haggan, V. and Ozaki, T. (1981). Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model. *Biometrika* **68**, 189-196.

Henrici, P. (1962). *Discrete Variable Methods in Ordinary Differential Equations.* John Wiley, New York.

Huber, P. J. (1967). The behavior of the maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp. Math. Statist. Probab.* **1**, 221-233.

Ishiguro, M. and Akaike, H. (1989). DALL: Davidon's algorithm for log likelihood maximization — A FORTRAN subroutine for statistical model builders. Computer Science Monographs, No.25. Institute of Statistical Mathematics, Tokyo.

Maruyama, T. (1983). Stochastic theory of population genetics. *Bull. Math. Biology* **45**, 521-554.

Milshtein (1974). Approximate integration of stochastic differential equations. *Theory Probab. Appl.* **23**, 396-401.

Oda, H., Ozaki, T. and Yamanouch, Y. (1987). A nonlinear system identification in the analysis of offshore structure dynamics in random sea. In *Nonlinear Stochastic Engineering System* (Proceeding of IUTAM Symposium, Insbruck/Iglis, Austria, June 21-26, 1987), 87-100, Springer-Verlag.

Ozaki, T. and Oda, H. (1978). Nonlinear time series model identification by Akaike's information criterion. In *Information and Systems* (Edited by B. Dubuisson), 83-91, Pergamon Press.

Ozaki, T. (1980). Nonlinear time series models for nonlinear random vibrations. *J. Appl. Probab.* **17**, 84-93.

Ozaki, T. (1981a). Nonlinear threshold autoregressive models for nonlinear random vibrations. *J. Appl. Probab.* **18**, 443-451.

Ozaki, T. (1981b). Nonlinear phenomena and time series models. Invited paper, 43rd Session of the International Statistical Institute, Buenos Aires, Argentina. *Bull. Internat. Statist. Inst.* **49**, 1193-1210.

Ozaki, T. (1985a). Nonlinear time series models and dynamical systems. In *Handbook of Statistics*, **5** (Edited by E. J. Hannan, et al), 25-83, North Holland.

Ozaki, T. (1985b). Statistical identification of storage models with application to stochastic hydrology. *Water Resources Bull.* **21**, 663-675.

Ozaki, T. (1986). Local Gaussian modelling of stochastic dynamical systems in the analysis of nonlinear random vibrations. In *Essays on Time Series and Allied Processes, Festschrift in honour of Prof. E. J. Hannan.* Applied Probability Trust, 241-255.

Ozaki, T. (1989). Statistical identification of nonlinear random vibration systems. *J. Appl. Mech.* **56**, 186-191.

Ozaki, T. (1990a). A local linearization approach to nonlinear filtering. Research Memo. No.378. Institute of Statistical Mathematics, Tokyo.

Ozaki, T. (1990b). Application of the local linearization filter to the identification of nonlinear stochastic dynamical systems. Research Memo. No.379. Institute of Statistical Mathematics, Tokyo.

Ozaki, T. (1990c). Identification of nonlinearities and non-Gaussianities in time series. To appear in the proceeding of 1990 Summer International Interdisciplinary Workshop on Time Series Analysis at the IMA, July 2-27, 1990.

Ozaki, T. and Ozaki, V. H. (1989). Statistical identification of nonlinear dynamics in macroeconomics using nonlinear time series models. In *Statistical Analysis and Forecasting of Economic Structural Change* (Edited by P. Hackl), 345-365. Springer-Verlag.

Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol 1 and Vol 2. Academic Press, London.

Priestley, M. B. (1988). *Non-linear and Non-stationary Time Series Analysis.* Academic Press, London.

Smith, J. (1977). *Mathematical Modeling and Digital Simulation for Engineers and Scientists.* John Wiley, New York.

Wong, E. (1963). The construction of a class of stationary Markoff process. *Proc. Amer. Math. Soc. Symp. Appl. Math.* **16**, 264-276.

Institute of Statistical Mathematics, 4-6-7 Minami Azabu, Minatoku, Tokyo, Japan.
and
Department of Statistics, Stanford University, Stanford, CA 94305, U.S.A.

(Received January 1990; accepted June 1991)