18.1 INTRODUCTION

It is more than ten years since the excellent time series book by Priestley was published (Priestley, 1981). That book, which ended by describing some interesting nonlinear time series models developed in late 1970s, was further supplemented by his later work (Priestley, 1988). As mentioned there, most nonlinear models of the late 1970s such as exponential AR models, threshold AR models and bilinear models, were introduced to answer such pragmatic concerns as (a) Can the model give a better prediction than linear models?; (b) Can we estimate the model from the data?; and (c) Can we produce a similar series to the original data by simulating the estimated model? A model which cannot be estimated from data, or one which cannot reproduce a similar series by simulation is certainly not interesting to statisticians although it may still interest some mathematicians. Nonlinear models which cannot give a better prediction performance than linear models are also unattractive and meaningless for applied time series analysts. Considering these three criteria, quite a lot of progress has been made in nonlinear time series modelling over the last ten years, and several books on nonlinear time series models have been published. However, as far as the exponential AR model is concerned, it must be admitted, to the author's regret as one of the originators, that the model, in spite of its potential, has not been sufficiently exploited either by applied time series analysts or by theoretical time series analysts. On the occasion of celebrating Professor Priestley's 60th birthday, I would like to reconsider and discuss the ExpAR model as a general nonlinear time series model by adding, to the above three criteria, a fourth question, i.e. Can the model characterize the non-Gaussian characteristics of the series?,
which, I hope, may contribute to future developments of nonlinear time series analysis both in theory and practice.

18.2 ExpAR PROCESSES AND MARGINAL DISTRIBUTIONS

In this section we will see how the non-Gaussian marginal distribution of an ExpAR process is characterized by the first order ExpAR model,

$$x_{i+1} = \{\phi_1 + \phi_2 \exp(-\gamma x_i^2)\} x_i + n_{i+1}.$$  \hfill (18.1)

In (18.1), $\gamma$ is some scaling constant and $n_{i+1}$ is Gaussian white noise with variance $\sigma^2$.

18.2.1 Histograms

One of the simplest and most natural ways of checking the non-Gaussianity of time series is to check its histogram. Many nonlinear time series models have been introduced, but it is not known how the parameters of these models affect the shape of the distribution of the series. The ExpAR model is an exceptional example whose parameters clearly explain what kind of marginal distribution the generated time series is going to have.

Figures 18.1 (a)–(c) show the histograms of three sets of time series, each with 80,000 observations, generated from the following three ExpAR models with $\sigma^2 = 1$,

$$x_{i+1} = \{1 - 0.2 \exp(-x_i^2)\} x_i + n_{i+1}$$ \hfill (18.2)

$$x_{i+1} = \{0.8 + 0.2 \exp(-x_i^2)\} x_i + n_{i+1}$$ \hfill (18.3)

$$x_{i+1} = \{0.8 + 0.4 \exp(-x_i^2)\} x_i + n_{i+1}$$ \hfill (18.4)

The three histograms show three different non-Gaussian characteristics of the density functions. The first histogram clearly shows a distribution with long heavy tails like a stable distribution. This type of histogram is generated from any ExpAR model (18.1) with $\phi_1 = 1$ and $-1 < \phi_2 < 0$. The second histogram shows short light tails and a fairly concentrated centre, which is a characteristic of exponential family distributions. This type of histogram is generated from any ExpAR model with $\phi_1 + \phi_2 = 1, 0 < \phi_1 < 1$ and $0 < \phi_2 < 1$. The third histogram shows a bimodal distribution with light and short tails, which is also a characteristic of some distributions in the exponential family. This type of histogram is generated from any ExpAR model with $\phi_1 + \phi_2 > 1, 0 < \phi_1 < 1$ and $0 < \phi_2 < 1$.

These examples of ExpAR models clearly show that we can generate time series with various types of marginal distributions by controlling the two parameters, $\phi_1$ and $\phi_2$, of the ExpAR model (18.1).
ExpAR processes

Figure 18.1. Histograms of the series $x_1, x_2, \ldots, x_{80-000}$: (a) of the model (18.2); (b) of the model (18.3); (c) of the model (18.4).

18.2.2 Exponential family and stochastic dynamical systems

The second and third histograms in the previous section imply that the marginal distribution of an ExpAR process may belong to the exponential family. Actually, 'Exp'AR models and the 'exponential' family of distributions are not only related phonetically. To show this we need to establish, first, the relationship between the exponential family and a class of diffusion processes. It was Wong (1963) who first pointed out a close relationship
between a certain class of distributions and some class of diffusion processes. He pointed out that for any density function \( W(x) \) belonging to the Pearson system,

\[
\frac{dW(x)}{dx} = \frac{f_0 + f_1 x}{g_0 + g_1 x + g_2 x^2} W(x), \tag{18.5}
\]

it is possible to give a diffusion process whose marginal density is \( W(x) \). The result was extended (Ozaki, 1985a) to any density function \( W(x) \) defined by proper analytic functions \( f(x) \) and \( g(x) \) in a distribution system given by,

\[
\frac{dW(x)}{dx} = \frac{f(x)}{g(x)} W(x). \tag{18.6}
\]

The corresponding Markov diffusion process is given (Ozaki, 1985a, 1992) by the following Fokker–Planck diffusion equation,

\[
\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x} \left[ \left\{ f(x) + \frac{\partial g(x)}{\partial x} \right\} p \right] + \frac{1}{2} \frac{\partial^2}{\partial x^2} \left[ 2g(x)p \right]. \tag{18.7}
\]

When \( g(x) \) is a constant and \( f(x) \) is a polynomial, density functions \( W(x) \) defined by (18.6) constitute the exponential family of distributions. On the other hand, the diffusion process in (18.7) has, when \( g(x) \) is a constant \( g_0 \), the stochastic dynamical system representation,

\[
x(t) = f(x(t)) + n(t), \tag{18.8}
\]

where \( n(t) \) is a Gaussian white noise of variance \( 2g_0 \). Thus, from a diffusion process \( x(t) \), defined by (18.7), we have its marginal density function, \( W(x) \), which satisfies \( dW(x)/dx = \{ f(x)/g_0 \} W(x) \). Since a distribution of the exponential family is completely characterized by a finite number of its moments, the marginal distribution of the process \( x(t) \) of (18.8) is also characterized by a finite number of its moments. We can see this in the following example. Let

\[
x(t) = ax(t) + bx(t)^3 + cx(t)^5 + n(t), \tag{18.9}
\]

where \( n(t) \) is a Gaussian white noise of variance \( \sigma^2 \).

The marginal density function \( W(x) \) satisfies the distribution system

\[
\frac{dW(x)}{dx} = \frac{(ax + bx^3 + cx^5)}{\sigma^2} W(x), \tag{18.10}
\]

and the density function is given by

\[
W(x) = W_0 \exp \left( \frac{a x^2 + b x^4 + c x^6}{2 \sigma^2} \right). \tag{18.10}
\]
where $W_0$ is a normalizing constant and $\sigma^2$ is the variance of the white noise $n(t)$.

Since $W(x)$ satisfies the relation

$$
\frac{d}{dx}\{x^{2k-2} W(x)\} = (2k-1)x^{2k-3} W(x) + \left[ \frac{2\{ax^{2k} + bx^{2k+2} + cx^{2k+4}\}}{\sigma^2} \right] W(x),
$$

(18.11)

we have, for $k = 1, 2, \ldots$, the relations,

$$
(2k-1)\mu_{2k-2} + \left( \frac{2}{\sigma^2} \right)(a\mu_{2k} + b\mu_{2k+2} + c\mu_{2k+4}) = 0.
$$

(18.12)

The relations (18.12) can also be derived from the stochastic dynamical system (18.9). From the dynamical system relation we have

$$
\frac{d}{dt}\left( \frac{x^{2k}}{2k} \right) = x^{2k-1}\{ax(t) + bx(t)^3 + cx(t)^5\} + \frac{1}{2}\{(2k-1)x^{2k-2}\}n(t)^2 + x^{2k-1}n(t)
$$

(18.13)

for $k = 1, 2, \ldots$. This is because for any analytic function $F(\cdot)$, it holds that

$$
F(x + \Delta x) = F(x(t)) + F'(x(t))\Delta x(t) + \frac{1}{2}F''(x(t))[(\Delta x(t))^2 + \ldots,
$$

and by Ito's stochastic calculus this implies that

$$
F(x(t + \Delta t)) = F(x(t)) + F'(x(t))\{ax(t) + bx(t)^3 + cx(t)^5\}\Delta t
+ \Delta \omega(t) + \frac{1}{2}F''(x(t))\Delta \omega(t)^2 + \ldots,
$$

(18.14)

where $\omega(t)$ is a Brownian motion process such that

$$
\frac{\Delta \omega(t)}{(\Delta t)^{1/2}} \rightarrow n(t) \quad \text{for} \quad \Delta t \rightarrow 0.
$$

(18.13) is obtained from (18.14) by using $F(x) = x^{2k}/2k$ and letting $\Delta t \rightarrow 0$. Taking the expectation of both sides of the equation (18.13), we have, for $k = 1, 2, \ldots$,

$$
0 = \frac{d}{dt} E\left[ \frac{x^{2k}}{2k} \right]
= E[ax(t)^{2k} + bx(t)^{2k+2} + cx(t)^{2k+4}] + E\left[ \left( \frac{2k-1}{2} \right)x^{2k-2}n(t)^2 \right]
+ E[x^{2k-1}n(t)]
= a\mu_{2k} + b\mu_{2k+2} + c\mu_{2k+4} + \frac{(2k-1)\sigma^2}{2} \mu_{2k-2}.
$$

(18.15)

Since the process $x(t)$ is symmetric about $x(t) = 0$, $\mu_0$ and all the odd order
moments are zero. The equations show that $\mu_2$, $\mu_4$ and $\mu_6$ are sufficient statistics since all the higher order moments, $\mu_8$, $\mu_{10}$, $\mu_{12}$, ... may be calculated recursively from relation (18.12) and $\mu_2$, $\mu_4$ and $\mu_6$. Process $x(t)$ of (9) is completely specified by the parameters $a$, $b$, $c$ and $\sigma^2$, the marginal distribution, and so $\mu_2$, $\mu_4$ and $\mu_6$, could also be explicitly written using those parameters. In spite of the importance of this type of distribution in nonlinear and non-Gaussian time series analysis, not much work has been devoted to the study of the exponential family, except for a general treatment (Barndorff-Nielsen, 1978), and even the moment generating function of distribution function (18.10) does not seem to have been discussed in the statistics literature.

18.2.3 Stochastic dynamical systems and ExpAR models

The general nonlinear stochastic dynamical system

$$\dot{x}(t) = f(x(t)|a) + n(t) \quad (18.16)$$

and its discrete time approximate model have been studied by applied probabilists, physicists, geneticists and engineers. The main purpose of those studies has been to simulate model (18.16). It has been widely known in this field that when $f(x)$ is nonlinear, the simulation of the discretized model often ends up in a computational explosion when $\Delta t$ is not small enough. This often happens even though the original continuous time process is known to be stationary. For example, $x(t)$ of

$$\dot{x}(t) = -x^3 + n(t) \quad (18.17)$$

is known to be stationary and has a stationary marginal distribution,

$$W(x) = W_0 \exp \left( \frac{x^4}{4} - \frac{x^2}{2\sigma^2} \right)$$

where $W_0$ is a normalizing constant and $\sigma^2$ is the variance of the white noise $n(t)$. We can easily see this computational problem by simulating (18.17) with large $\sigma^2$ and $\Delta t$. To avoid the computational explosion in simulations people often take a small time interval $\Delta t$ in the discretization.

An inductive use of discretized models of dynamical system (18.16) came from time series analysis in statistics, where researchers are interested in estimating the parameters of the model in order to characterize the non-Gaussian and nonlinear nature of the time series. Here there is a problem because we cannot choose both the time interval $\Delta t$ and the white noise variance $\sigma^2$. Both are fixed by the sampling of the time series and it is often the case that the noise variance is quite large and the sampling interval is not very small. Naturally, special attention was paid to the computational stability of the discretization scheme used to derive the model with fairly
large $\Delta t$ and with very large $\sigma^2$. It was reported in Ozaki and Oda (1978) that when the Euler type discretization scheme is employed, simulations of the estimated model computationally explode very easily, even though the process defined by the model of the Euler type discretization scheme converges, in probability, to the original continuous time process $x(t)$ on the finite time interval $[0, T]$ when $\Delta t \to 0$. In fact, not only the Euler type scheme but also more sophisticated schemes, such as the Heun or Runge–Kutta types, are inappropriate for inductive use, even though all three types are consistent. This is because they all yield polynomial type discrete time models when $f(x)$ in (18.16) is of polynomial type as for example in (18.17) and such discrete time models driven by Gaussian white noise are known to be transient and computationally explosive in a finite time interval with probability one.

Ozaki (1985a, b, 1986, 1990a, b, 1992) has introduced a scheme (called the local linearization scheme) which yields a model which has both computational stability and consistency properties. From a nonlinear stochastic dynamical system (18.16), the local linearization yields the following nonlinear time series model:

$$
x_{i+1} = A(x_i | a) x_i + B(x_i | a) n_i + \Delta t
$$

$$
A(x_i | a) = \exp \{ K(x_i | a) \Delta t \}
$$

$$
B(x_i | a) = \left( \frac{\exp \{ 2K(x_i | a) \Delta t \} - 1}{2K(x_i | a)} \right)^{1/2}
$$

$$
K(x_i | a) = \frac{1}{\Delta t} \log \left\{ 1 + J_i^{-1} (e^{\Delta t} - 1) f(x_i | a) / x_i \right\}
$$

$$
J_i = \left( \frac{\partial f(x_i | a)}{\partial x} \right)_{x=x_i}
$$

(18.18)

Since it holds that

$$
A(x_i | a) x_i = x_i + \Delta t f(x_i | a) + o(\Delta t)
$$

and

$$
B(x_i | a) x_i = (\Delta t + o(\Delta t))^{1/2},
$$

the locally linearized model (18.18) is asymptotically equivalent to the Euler type model and shares the same asymptotic properties with regard to $\Delta t$.

When the original model (18.16) is specified, we can check if the locally linearized model (18.18) is a stationary Markov chain by using Tweedie's (1975) theorem. Ozaki (1985a) gave sufficient conditions for the function $f(x)$ in (18.16) to produce a stationary locally linearized model (18.17). The conditions are not very strict and are satisfied by most interesting nonlinear functions seen in defining non-Gaussian diffusion processes.

An interesting point in the local linearization is that the original continuous time function $f(x)$ can be recovered from the locally linearized discrete time
model by a numerical method (see Ozaki (1985b, 1992)). Of course, since approximation is involved in the discretization, the recovered function is an approximation to the original $f(x)$. As is usual for any function approximation, the accuracy of the approximation depends on how it is approximated—i.e. in the local linearization case, on the time interval $\Delta t$ of the discretization. Naturally the difference between the two functions becomes zero when $\Delta t \to 0$.

Another interesting point in the local linearization is that it yields, from the stochastic dynamical system model (18.17), the following $A(x_i)$.

$$ A(x_i) = \frac{2}{3} + \frac{1}{3} \exp(-3x_i^2\Delta t) $$

Since $B(x_i)$ is almost constant and is given by $B(x_i) \approx (\Delta t)^{1/2}$, the discretized model is almost equivalent to an ExpAR model. From the stochastic dynamical system,

$$ \dot{x} = x - x^3 + n(t) $$

It yields

$$ A(x_i) = \frac{-2x_i^2}{1 - 3x_i^2} + \frac{1 - x_i^2}{1 - 3x_i^2} \exp\{(1 - 3x_i^2)\Delta t\}. $$

The function $A(x_i)$ is a positive valued smooth function and goes to a constant, $2/3$, when $x_i$ goes to infinity, and $1 < A(x_i)$ when $-1 < x_i < 1$. The shape of the function $A(x_i)$ is very similar to the $\phi(\cdot)$ function of the ExpAR model (18.3). Model (18.19) is known to have a bimodal distribution similar to ExpAR model (18.3). Looking at the above examples we can see that the local linearization scheme yields, from the class of general nonlinear stochastic dynamical systems,

$$ \dot{x}(t) = a_1 x(t) + a_2 x(t)^2 + \cdots + a_{2k+1} x(t)^{2k+1} + n(t), \quad (k = 1, 2, \ldots) \tag{18.20} $$

a smooth and positive valued function $A(x_i)$ which approaches the constant values $2k/(2k + 1)$ when $x_i \to \infty$. This type of function is exactly what Ozaki (1985a) considered in his extended version of the ExpAR model, i.e.

$$ x_{t+1} = \phi(x_t) x_t + n_{t+1}, \tag{18.21} $$

where

$$ \phi(x_t) = \phi_1 + \{\phi_2 + \phi_3 x_t + \cdots + \phi_r x_t^{r-2}\} \exp(-\gamma x_t^2). $$

By taking $r$ of the $\phi(\cdot)$ function sufficiently large we can approximate any smooth function $A(x_i)$ which converges to a constant for $x_i \to \infty$. It means that the discrete time process obtained from the diffusion process (18.20) can be very closely approximated using the extended ExpAR model (18.21). It also means that if the $\phi(\cdot)$ function of an extended ExpAR model is given, it may be closely approximated by $A(x_i)$ which may be found from a function of the form

$$ f(x) = a_1 x(t) + a_2 x(t)^2 + \cdots + a_{2k+1} x(t)^{2k+1}, $$
and we can characterize the non-Gaussian behaviour of the series using the nice relationship between the diffusion process defined by (18.20) and the exponential family. Further work in this field, both empirical and theoretical, will certainly be worth doing in the future.

18.3 ARCH MODELS AND ExpAR MODELS

In nonlinear time series analysis, it is sometime argued that nonlinear models driven by homogeneous Gaussian white noise, like ExpAR models, are not appropriate for analysing time series whose marginal distribution has heavy tails. Such time series are found in financial time series data. For the analysis of such time series data, researchers often use models with Gaussian white noise whose variance is dependent on the value of the series. A well-known example is the ARCH model (Engel, 1982). One simple example of an ARCH model is given by

\[ x_{t+1} = \phi x_t + (\alpha + \beta x_t^2)^{1/2} n_{t+1}. \]  

This type of model looks very different from an ExpAR model. However, we have already seen in section 18.2 that ExpAR models can also generate time series having a marginal distribution with heavy tails. Further investigation of these two different types of nonlinear time series model reveals the interesting role of instantaneous transformations in non-Gaussian time series analysis.

We know that the model (18.22) can be obtained by discretizing a continuous time diffusion process model of the type

\[ \dot{x} = -ax + (\alpha + \beta x^2)^{1/2} n(t). \]  

Here we assume the Stratonovich type stochastic calculus. The process \( x(t) \) defined by the model has a stable marginal distribution (see Guegan, 1992). The model (18.23) can be transformed into the form

\[ \dot{y} = \frac{a}{\beta^{1/2}} \tan h(\alpha^{1/2} y) + n(t), \]  

by using the transformation

\[ y = \frac{1}{\beta^{1/2}} \sin h^{-1} \left[ \left( \frac{\beta}{\alpha} \right)^{1/2} x \right]. \]

It is shown in Ozaki (1992) that the model (18.24) yields a discrete time model

\[ x_{t+1} = A(x_t)x_t + B(x_t)n_{t+1}, \]

where \( B(x_t) \) is almost constant and equals \( (\Delta t)^{1/2} \), while \( A(x_t) \) is \( 0 < A(x_t) < 1 \) and is a smooth function of \( x_t \), approaching 1 for \( |x_t| \rightarrow \infty \). This means that the ExpAR model combined with an instantaneous variable transformation,
Non-Gaussian characteristics

\[ y_{t+1} = \left(1 - \pi \exp(-\gamma y_t^2)\right) y_t + n_{t+1} \]
\[ x_{t+1} = \left(\frac{\alpha}{\beta}\right)^{1/2} \sin h(\beta^{1/2} y_{t+1}^{1/2}) \]

is a reasonable model for a heteroscedastic time series if the parameters \(\alpha, \beta, \gamma\)
and \(\pi\) are properly chosen. Thus ExpAR models combined with instantaneous
variable transformations can characterize quite large varieties of non-Gaussian
processes.

18.4 ExpAR\((p)\) MODELS AND NONLINEAR STATE-SPACE
REPRESENTATIONS

We have so far seen the non-Gaussian properties of ExpAR models of lag
order 1. Such models are not appropriate for fitting non-Gaussian oscillatory
time series whose spectrum has peaks. One example of non-Gaussian oscilla-

tory time series is data generated from
\[
\ddot{x}(t) + \alpha \ddot{x}(t) + b(x(t))x(t) = \epsilon(t).
\]

Model (18.25) is known to be a model of hard spring type nonlinear oscilla-
tions if \(b(x) = b_1 + b_2 x^2\) with \(b_2 > 0\) and of soft spring type nonlinear
oscillations if \(b_2 < 0\). \(x(t)\) is known to have the following non-Gaussian marginal
distribution (Caughey, 1963; Ozaki, 1990b),
\[
p(x) = p_0 \exp \left\{ \frac{-2a}{\sigma^2} \int b(\xi) \xi d\xi \right\},
\]

where \(\sigma^2\) is the variance of the Gaussian white noise \(\epsilon(t)\) and \(p_0\) is the
normalizing constant. In linear Gaussian cases, i.e. when \(b_2 = 0\), AR\((p)\) models,
or ARMA\((p,q)\) models, with \(p > 1\) are usually considered for the analysis of
such time series. To analyse non-Gaussian and, at the same time, oscillating
time series, the second order ExpAR\((2)\) model
\[
x_{t+1} = \{\phi_{1,1} + \phi_{1,2} \exp(-\gamma x_t^2)\} x_t + \{\phi_{2,1} + \phi_{2,2} \exp(-\gamma x_t^2)\} x_{t-1} + n_{t+1},
\]
or the \(p\)th lag order ExpAR\((p)\) model
\[
x_{t+1} = \{\phi_{1,1} + \phi_{1,2} \exp(-\gamma x_t^2)\} x_t + \cdots + \{\phi_{p,1} + \phi_{p,2} \exp(-\gamma x_t^2)\} x_{t-p+1} + n_{t+1},
\]
have been introduced (Ozaki & Oda (1978), Haggan & Ozaki (1981)).
Although ExpAR\((p)\) models are useful for explaining many interesting
nonlinear phenomena, they are not, in a sense, general enough; they share
the same defects as the linear AR models, i.e. they are not appropriate for
time series whose spectrum has sharp troughs. It is widely recognized in
applied time series analysis that ARMA\((p,q)\) models are much more efficient
to model such series parsimoniously. Also, it is known that a discretization
of the linear vibration model

\[ \ddot{x}(t) + a\dot{x}(t) + bx(t) = n(t) \]

yields the ARMA(2, 1) model,

\[ x_{t+1} = \phi_1 x_t + \phi_2 x_{t-1} + \theta n_t + n_{t+1} \]

where \( \phi_1, \phi_2 \) and \( \theta \) are given as functions of \( a \) and \( b \) (Pandit and Wu, 1975). Naturally, a generalization of the ExpAR(2) model has been considered as a discrete time version for the nonlinear vibration model (18.25) so that the generalized model includes the ARMA(2, 1) model as a special case. Ozaki (1980) proposed the ExpARMA(2, 1) model

\[ x_{t+1} = \{ \phi_1 + \pi_1 \exp(-\gamma x_t^2) \} x_t + \{ \phi_2 + \pi_2 \exp(-\gamma x_t^2) \} x_{t-1} + \{ \theta_1 + \theta_2 \exp(-\gamma x_t^2) \} n_t + n_{t+1} \]

as a discrete version of the nonlinear vibration model (18.25), and the ExpARMA(p, p - 1) model as a discrete version of the nonlinear stochastic differential equation model

\[ x^{(p)}(t) + a_1(x(t))x^{(p-1)}(t) + \cdots + a_{p-1}(x(t))\dot{x}(t) + a_p(x(t))x(t) = n(t). \] (18.26)

However, the problem of these generalizations is that the maximum likelihood method does not work properly in practice. For the estimation of the parameters of these models we need to use a nonlinear optimization method to maximize the likelihood, as in the linear ARMA cases, but in most cases the Hessian becomes almost singular. This may be because the model has too many parameters, twice as many as in the linear case.

In order to eliminate these numerical problems we need to generalize ExpAR models in a compact and parametrically parsimonious way. A useful idea arises from checking the nonlinear characteristics of the model introduced from the continuous time model (18.25) by the local linearization (Ozaki, 1986 and 1989). The nonlinear random vibration model (18.25) can be written down as the two-dimensional stochastic dynamical system

\[ \dot{z} = f(z|\alpha) + \eta(t), \] (18.27)

where \( \eta(t) = \{ n(t), 0 \}^t \), \( z = (\dot{x}, x)^t \) and \( f(z|\alpha) = \{ -a(x|\alpha)\dot{x}, b(x|\alpha)x \}^t \). The local linearization of (18.27) is obtained in the same way as in the scalar case as follows:

\[ z_{t+\Delta t} = A(z_t|\alpha)z_t + B(z_t|\alpha)\eta_{t+\Delta t} \]

\[ A(z_t|\alpha) = \text{Exp} \{ K(z_t|\alpha)\Delta t \} \]

\[ K(z_t|\alpha) = \frac{1}{\Delta t} \text{Log} \{ 1 + J_t^{-1}(e^{J_t\Delta t} - 1)F(z_t|\alpha) \} \]

\[ J_t = \left( \frac{\partial f(z|\alpha)}{\partial z} \right)_{z=z_t}. \] (18.28)
Non-Gaussian characteristics

$F(z|\alpha)$ is a matrix which satisfies $F(z|\alpha)\dot{z}_t = f(z|\alpha)$, $\text{Exp}(\cdot)$ is the matrix exponential function and $\log(\cdot)$ is the matrix logarithmic function. The elements of the matrix $B(z|\alpha)$ are explicitly given as functions of the eigenvalues of the matrix $K(z|\alpha)$ (see Appendix). In many applications we usually have observation records only of $x(t)$ and the record of $dx/dt$ is not available. Then we have a nonlinear state space representation model

$$z_{t+\Delta t} = A(z|\alpha)z_t + B(z|\alpha)u_{t+\Delta t},$$
$$x_t = H z_t,$$

where $H = (0, 1)$. Each element of the matrix $A(z|\alpha)$ and $B(z|\alpha)$ for the model (18.28) is given as a function of the parameters of the original continuous model (18.25). By checking the elements of the matrix $A(z|\alpha)$ we know (see Figures 4.2 and 4.3 of Ozaki, 1986) that the (1,1)th and (1,2)th elements, $a_{1,1}$ and $a_{1,2}$, of the matrix $A(z|\alpha)$ are smooth functions of $x_n$, while the (2,1)th and (2,2)th elements are almost constant, i.e. $a_{2,1} = \Delta t$ and $a_{2,2} = 1$. This suggests that the state space model

$$z_{t+1} = F z_t + G n_{t+1},$$
$$x_t = H z_t,$$  \hspace{1cm} (18.29)

is a reasonably parsimonious discrete time model for nonlinear random vibrations, where

$$F = \begin{pmatrix} f_1(x) \\ f_2(x) \\ c \\ 1 \end{pmatrix}, \quad G = \begin{pmatrix} \sigma_1 \\ \sigma_2 \end{pmatrix}, \quad H = (0, 1),$$

$$f_1(x) = \phi_1 + \pi_1 \exp(-\gamma x^2_t) \quad \text{and} \quad f_2(x) = \phi_2 + \pi_2 \exp(-\gamma x^2_t).$$

$n_{t+1}$ is a Gaussian white noise of variance $\sigma^2$. The two-dimensional state vector $z_t$ is given by $z_t = (\xi_t, x_t)^T$, where the conceptual unobserved variable $\xi_t$ is introduced to represent the velocity of the vibrations. Thus $x_{t+1}$ of the model (18.29) is given by $x_{t+1} = x_t + c \xi_t$. The parameter $c$ corresponds to the time interval of the time series data, which is fixed when the data is sampled, but we do not know which scaling unit we should use in the model (18.29) in real data analysis. When $\pi_1 = 0$ and $\pi_2 = 0$, obviously the model (18.29) is equivalent to the ARMA(2, 1) model

$$x_{t+1} = \lambda_1 x_t + \lambda_2 x_{t-1} + H(F - \lambda_1 I)G n_t + H G n_{t+1},$$

where $\lambda_1$ and $\lambda_2$ are the characteristic coefficients of the matrix $F$, satisfying

$$F^2 - \lambda_1 F - \lambda_2 I = 0.$$

More generally, the eigenvalues of the matrix $F$ are functions of $x_t$. Then the model yields a van der Pol type nonlinear oscillator if $\phi_1, \phi_2, \pi_1$ and $\pi_2$ are such that the absolute value of the roots of

$$\Lambda^2 - (1 + \phi_1 + \pi_1) \Lambda + \{(\phi_1 + \pi_1) - c(\phi_2 + \pi_2)\} = 0.$$
are less than one, and the absolute value of the roots of
\[ \Lambda^2 - (1 + \phi_1)\Lambda + (\phi_1 - c\phi_2) = 0 \]
are greater than one. One such example is the model (18.29) with the \( F \)-matrix
\[
F = \begin{pmatrix}
0.94 + 0.26 \exp(-x_i^2) & -0.2 \\
0.1 & 1
\end{pmatrix}
\]  
(18.30)

Figure 18.2 shows the typical limit cyclic behaviour obtained by simulating model (18.30) with an initial value \( z_i = (\xi_n, x_i)' = (0.001, 0.001)' \) and with zero input noise.

The idea is immediately extendible to the \( p \)-th order nonlinear stochastic differential equation model (18.26), for which we have the \( p \)-dimensional state space representation model
\[
z_{i+1} = Fz_i + G\eta_{i+1},
\]
x_i = Hz_i,  
(18.31)

where
\[
F = \begin{pmatrix}
f_1(x_i) & \cdots & f_p(x_i) \\
c & 1 & 0 & 0 & \cdots & 0 \\
0 & c & 1 & 0 & \cdots & 0 \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
0 & \cdots & c & 1 
\end{pmatrix},
\]
\[
G = (\sigma_1, \sigma_2, \ldots, \sigma_p),
\]
\[
H = (0, \ldots, 0, 1),
\]
\[
f_i(x_i) = \phi_i + \pi_i \exp(-\gamma x_i^2) \quad \text{for} \quad i = 1, 2, \ldots, p.
\]
The state vector \( z_i \) is given by \( z_i = (\xi_p, \xi_{p-1}, \ldots, \xi_2, x_i)' \), where \( \xi_i (i = 2, 3, \ldots, p) \) stands for the \((i - 1)\)th order differential of \( x(t) \), i.e. \( d^{i-1}x/dt^{i-1} \) at time point \( t \).

---

Figure 18.2. Limit cycles simulated by the model (18.29); (\( F \) is given by (18.30)).
Non-Gaussian characteristics

Estimating such a model as (18.31) from data, \( x_t (t = 1, 2, \ldots, N) \), may look difficult since the state variables \( \xi_p, \xi_{p-1}, \ldots, \xi_2 \) are not observed. However, if we use nonlinear Kalman filtering we can transform the data into prediction error sequences, i.e. Gaussian white noise innovations. This allows us to write down the likelihood of the non-Gaussian distributed series \( x_t \) by using the Gaussian likelihood of the innovations and by maximizing the likelihood we can obtain the maximum likelihood estimates of the model parameters. Asymptotic properties, such as consistency and normality, of the estimates have been proved (Ljung, 1976; Caines, 1978). Effectiveness of the estimation method in practice can be checked using simulation data. Numerical studies of the estimation method will be given in a future paper.

18.5 CONCLUSION

It has been shown that a simple first lag order ExpAR model can generate time series with various types of marginal distributions by varying the two model coefficients. A close relationship between ExpAR models and the exponential family of distributions is pointed out. It is shown that for a given distribution of the exponential family we can have a special type of diffusion process whose diffusion coefficient is constant and whose marginal distribution is equivalent to the given distribution. Any such stationary homogeneous diffusion process with constant diffusion coefficient can be closely approximated by a discrete time nonlinear autoregressive model whose autoregressive coefficient is a smooth function of the process and approaches a constant for \(|x_t| \to \infty\). This means that for a given distribution of the exponential family we can introduce an ExpAR model whose marginal distribution is very close to the given distribution.

For the analysis of time series with a stable marginal distribution, heteroscedastic nonlinear time series models such as ARCH models are used. It has shown that an ExpAR model combined with an instantaneous variable transformation is useful for the modelling of such heteroscedastic time series.

A nonlinear state space model was introduced as an extension of the ExpAR model. The model-set defined by the nonlinear state space representation includes ARMA\((p, p - 1)\) models as a subset. Thus it can be regarded as a natural nonlinear extension of linear Markovian models for time series.

APPENDIX

Matrix \( B(z_t | \Theta) \) in (18.28).

The variance–covariance matrix of the discrete time white noise

\[
\eta_{t+\Delta t} = \int_t^{t+\Delta t} \exp\{K(t + \Delta t - u)\} n(u) du
\]
of the model (18.27) is given by

\[
E \left[ \int_t^{t+\Delta t} \exp \{ K_i(t + \Delta t - u) \} n(u) du \int_t^{t+\Delta t} n(u) \exp \{ K'_i(t + \Delta t - u) \} du \right]
\]

\[
= \int_t^{t+\Delta t} \exp \{ K_i(t + \Delta t - u) \} \Sigma \exp \{ K'_i(t + \Delta t - u) \} du
\]

\[
= \int_t^{t+\Delta t} V_i M V^{-1}_i \Sigma (V^{-1}_i)' M V'_i du
\]

\[
= V_i S V'_i,
\]

where

\[
M = \begin{bmatrix}
\exp \{ \mu_1(t + \Delta t - u) \} & 0 \\
0 & \exp \{ \mu_2(t + \Delta t - u) \}
\end{bmatrix}.
\]

\(V_i\) is a matrix which satisfies

\[
V^{-1}_i K_i V_i = \begin{bmatrix}
\mu_1 & 0 \\
0 & \mu_2
\end{bmatrix}
\]

and is given by

\[
V_i = \begin{bmatrix}
\mu_1 & \mu_2 \\
1 & 1
\end{bmatrix}.
\]

\(\Sigma\) is the variance–covariance matrix of the white noise \(n(t)\) given by

\[
\Sigma = \begin{bmatrix}
\sigma^2 & 0 \\
0 & 0
\end{bmatrix}.
\]

The matrix \(S\) is

\[
S = V^{-1}_i \Sigma (V^{-1}_i)'
\]

\[
= \begin{bmatrix}
e_{11} s & e_{12} s \\
e_{21} s & e_{22} s
\end{bmatrix},
\]

where \(s = \sigma^2/(\mu_1 - \mu_2)^2\), \(e_{11} = \exp(2\mu_1 \Delta t)/2\mu_1\), \(e_{12} = [\exp((\mu_1 + \mu_2) \Delta t) - 1]/(2\mu_2)\), \(e_{21} = e_{12}\) and \(e_{22} = \exp(2\mu_2 \Delta t)/2\mu_2\).

The matrix \(S\) is diagonalized by the unitary matrix \(U_i\),

\[
S = U_i \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix} U_i^*,
\]

where \(\lambda_1\) and \(\lambda_2\) are eigenvalues of the matrix \(S\) and are given as roots of

\[
\Lambda^2 - (e_{11} s + e_{22} s) \Lambda + e_{11} e_{22} s^2 - e_{12} e_{22} s^2 = 0.
\]

The unitary matrix \(U_i\) is given by

\[
U_i = (e_{12}^2 s^2 + (\lambda_1 - e_{11} s)^2)^{-1/2} \begin{pmatrix}
e_{12} s & \lambda_2 - e_{22} s \\
\lambda_1 - e_{11} s & e_{12} s
\end{pmatrix}.
\]
Non-Gaussian characteristics

Using $V_n, U_n, \lambda_1$ and $\lambda_2$ the matrix $B(z|\alpha)$ of (18.28) is given by

$$B(z|\alpha) = V_n U_n \begin{pmatrix} (\lambda_1)^{1/2} & 0 \\ 0 & (\lambda_2)^{1/2} \end{pmatrix}.$$ 

REFERENCES


