NONLINEAR TIME SERIES ANALYSIS FOR DYNAMICAL SYSTEMS OF CATASTROPHE TYPE

Loren Cobb
Department of Biometry
Medical University of South Carolina
Charleston, SC 29425

Shelemyahu Zacks
Department of Mathematical Sciences
State University of New York
Binghamton, NY 13901

1 Introduction

This chapter is concerned with a class of nonlinear time series models inspired by catastrophe theory, a branch of differential topology. The nonlinear models of catastrophe theory offer interesting modes of behavior that are not found in the usual linear models of time series analysis. For example, the cusp catastrophe model exhibits a phenomenon called "bistability," which means that the state variable has two attracting equilibrium points, separated by a repelling equilibrium point. Figure 1.1 shows a bistable time series generated by a stochastic cusp catastrophe model. The two attracting equilibria are at ± 1 (indicated by dotted lines), while the repelling equilibrium is at zero.

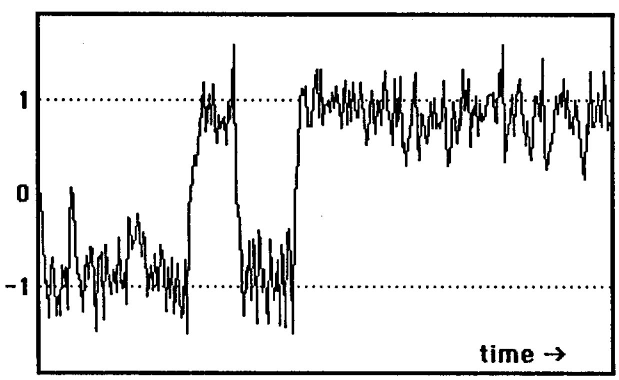

Figure 1.1 A time series generated by a bistable stochastic catastrophe model.

Note the tendency of the process to remain for a while within the domain of attraction of one attracting equilibrium, before finally moving past zero and into the domain of attraction of the other.

Catastrophe models share with most nonlinear models the characteristic that their qualitative behavior (*e.g.* the number of attracting equilibria) depends on their parameters. Thus slowly changing one parameter of a catastrophe model can result in the sudden appearance or disappearance of pairs of equilibria. Characterizing these qualitative modes of behavior is, in fact, the subject matter of catastrophe theory.

The models of catastrophe theory are not statistical models, but we present a method for deriving statistical time series models from the canonical catastrophes. The resulting models would most naturally take the form

$$X_{t+1} = \theta_0 + \theta_1 X_t + \theta_2 X_t^2 + \theta_3 X_t^3 + \ldots + \theta_d X_t^d + U_t, \qquad (1.1)$$

were it not the fact that such models are, in general, neither stationary nor ergodic. This is a major problem for the statistical analysis of such models, since most methods assume that their models are both stationary and ergodic. As Tjøstheim (1986) has remarked, the task of finding nonlinear models satisfying these assumptions is far from trivial! In Section 2 of this chapter we show that the source of the difficulty for models of the form (1.1) lies in the existence of a periodic orbit in the deterministic version of these models, and possibly uncontrolled oscillations. In Section 3 we present a modification of the model that is both stationary and ergodic, while still retaining the essential characteristics of the canonical catastrophe model from which it was ultimately derived. We consider in some detail a special case of the modified cubic time series model, and analyze the maximum likelihood estimators of its parameters. This article is designed to show that the array of research problems that is opened up when nonlinear time series analysis is seen from the perspective of dynamical systems theory is vast, and that straightforward applications of results which are available in the statistical literature is not always possible.

2 Polynomial Time Series Models

2.1 Nonlinear Dynamical Systems

In this section we describe the varieties of behavior exhibited by the class of first-order time series models that are nonlinear in their state variables, namely models of the general form

$$X_{t+1} = f(\theta, X_t) + U_t, \qquad (2.1.1)$$

where f is a function th

independent and identic

in its first argument, the

model." In order to avo

we shall call the mode

systems." For example,

$X_{t+1} = a + bX_t + c$

is a linear model for a n

$X_{t+1} = e^{-at}X_t + U$

is a nonlinear model fo

former kind, not the la

behavior of the determ

reserved for Section 3.

The subclass of n

variable is of central im

exhibited by the entire

theory, which use the po

of nonlinear models.

classification of deter

discrete time such as (2

The major topic

as seen in catastrophe

discrete time in time se

2.2 First-Ord

A *first-order* dyn

rate of change of the s

previous values:

$\Delta x_t = f(x_t)\Delta t,$

where Δ is the forward

$\Delta x_t = x_{t+\Delta t} - x_t.$

where f is a function that is nonlinear in its second argument, and $\{U_t\}$ is a sequence of independent and identically distributed normal random variables. Note that if f is linear in its first argument, the parameter vector θ, then in statistical terminology it is a "linear model." In order to avoid confusion in the usage of the terms "linear" and "nonlinear", we shall call the models discussed here "time series models for nonlinear dynamical systems." For example,

$$X_{t+1} = a + bX_t + cX_t^2 + U_t \tag{2.1.2}$$

is a linear model for a nonlinear dynamical system, whereas

$$X_{t+1} = e^{-at}X_t + U_t \tag{2.1.3}$$

is a nonlinear model for a linear dynamical system. Our interest lies with models of the former kind, not the latter. Further, we shall restrict our attention in this section to the behavior of the deterministic parts of these models. The stochastic behavior will be reserved for Section 3.

The subclass of models of the form (2.1.1) in which f is a polynomial in the state variable is of central importance in the description of the varieties of behavior that can be exhibited by the entire class. This importance originates in the theorems of catastrophe theory, which use the polynomial models as the canonical members of equivalence classes of nonlinear models. However, the subject matter of catastrophe theory is the classification of deterministic systems in continuous time, not stochastic systems in discrete time such as (2.1.1).

The major topic of this section is the classification of nonlinear dynamical systems as seen in catastrophe theory, and the additional considerations implied by the use of discrete time in time series models.

2.2 First-Order Stochastic Dynamical Systems

A *first-order* dynamical system in discrete-time is characterized by the fact that the rate of change of the state variable is dependent only on the current value, and not upon previous values:

$$\Delta x_t = f(x_t)\Delta t, \tag{2.2.1}$$

where Δ is the forward difference operator (with interval Δt) defined by

$$\Delta x_t = x_{t+\Delta t} - x_t. \tag{2.2.2}$$

When this dependence is modified by the presence of additive stochastic noise, as in

$$\Delta X_t = f(X_t)\Delta t + U_t, \quad U_t \sim N(0, \sigma^2 \Delta t), \tag{2.2.3}$$

then we have a model which is an autoregressive nonlinear dynamical system. The connection between (2.2.3) and the AR(1) model (as it is known in the time series analysis literature) is evident if we let $\Delta t = 1$ and $f(x) = \mu + \theta x$, which yields:

$$X_{t+1} = \mu + (1+\theta)X_t + U_t. \tag{2.2.4}$$

Both equations (2.2.3) and (2.2.4) are examples of *stochastic difference equations.*

A first-order (non-stochastic) dynamical system in continuous time is characterized by the differential equation:

$$\dot{x}_t = f(x_t), \tag{2.2.5}$$

where $\dot{x}$ is the time derivative of x. The corresponding *stochastic differential equation* is:

$$dX_t = f(X_t)dt + \sigma dW_t, \tag{2.2.6}$$

where W_t is the standard Wiener Process (Liptser & Shiryayev, 1977, pp. 82-88). The close relationship between this system and the discrete-time system (2.2.3) becomes apparent when we identify U_t with $\sigma \Delta W_t$: the continuous-time version is just a discrete-time system with an infinitesimal Δt (see Stroyan & Bayod, 1986, for a rigorous derivation of stochastic differential equations from infinitesimal difference equations).

The theory of statistical estimation and inference for the continuous-time nonlinear systems (2.2.6) is quite well developed (Liptser & Shiryayev, 1977, 1978) in comparison to the discrete-time nonlinear systems (2.2.3). The goal of this chapter is to show that statistical estimation and inference for discrete-time nonlinear systems is both feasible and practical, provided care is taken in the initial specification of the statistical model. We restrict this exposition, however, to the special case of cubic models.

2.3 Attractors and Repellors in First-Order Systems

The class of nonlinear systems presents a dramatically greater variety of behavior than the class of linear systems. Consider first the equilibrium structure. The equilibrium points of (2.2.2) and (2.2.5) are defined as the points x such that $\Delta x = 0$ or $\dot{x} = 0$, respectively, *i.e.*

$$\{ x: f(x) = 0 \}. \tag{2.3.1}$$

Clearly, if the system is linear and nontrivial then there is at most one equilibrium point. A nonlinear system, by contrast, has as many equilibria as there are roots of $f(x) = 0$.

An equilibrium point is said to be *repelling* if the state variable eventually departs from any arbitrarily small neighborhood of the point, never to return. Conversely, an equilibrium point is said to be *attracting* if the state variable approaches the point asymptotically as $t \to \infty$. For example, the nonlinear dynamical system defined by

$$\dot{x} = \theta(a-x)(b-x)(c-x), \quad \theta < 0. \tag{2.3.2}$$

has three equilibria, two attracting and one repelling, as depicted in Figure 2.1:

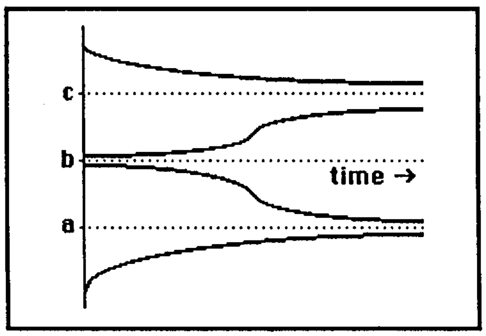

Figure 2.1 Trajectories of the system defined by (2.3.2). The equilibria at a and c are attracting, but the equilibrium at b is repelling.

The *domain of attraction* for an attractor is the set of points in the state space that are attracted to it. Thus the interval $(-\infty, b)$ is the domain of attraction for a in the above figure, while (b, ∞) is the domain of attraction for c.

Now consider the discrete-time dynamical system

$$\Delta x = -\theta x^3 \Delta t, \quad \theta > 0. \tag{2.3.3}$$

This system clearly has an attracting equilibrium at $x = 0$, but what is its domain of attraction? From Figure 2.2 we see that there is an interval (-A, A) within which the system moves asymptotically towards the origin. However, given an initial position outside this interval the system begins oscillating about the origin with an ever-increasing amplitude. We shall call such destructive oscillations an *explosion.*

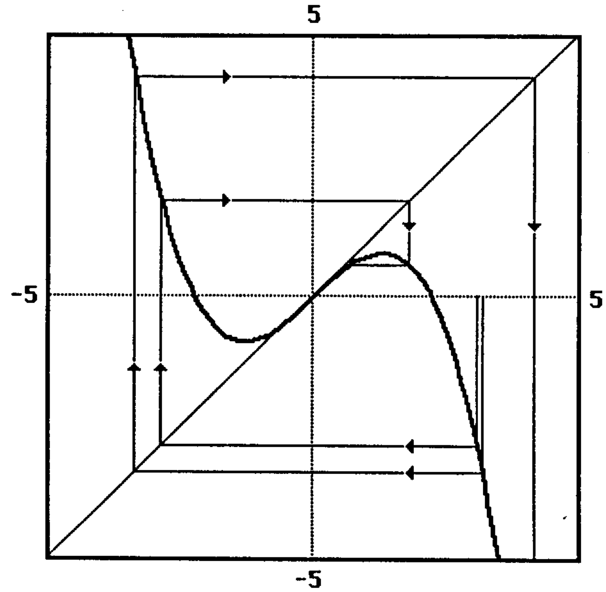

Figure 2.2 A "cobweb" graph for two trajectories of the system defined by Equation (2.3.3), from two initial positions: 3.1 and 3.2, with $\theta = 0.2$ and $\Delta t = 1$. The set $\{\pm\sqrt{10}\}$ is a period-2 repellor, while the origin is an attractor.

The boundary of the domain of attraction for the origin in the system (2.3.3) depends upon θ. It consists of the set of points A such that if $x_t = A$ then $x_{t+\Delta t} = -A$. Equation (2.3.3) is easily solved for these points:

$$A = \pm(2/\theta)^{1/2}\,\Delta t. \qquad (2.3.4)$$

Notice that the set $\{-A, A\}$ constitutes a *periodic orbit* (of period 2) for this system. It is easy to see that this orbit is repelling, since any deviation from the orbit will cause the system either (a) to move towards the attractor at the origin, or (b) to begin an explosion. A periodic orbit that is attracting is also known as a *limit cycle.*

Repelling (or attracting) periodic orbits are not found in linear dynamical systems. The periodic behavior of the linear second-order system

$$x_{t+1} = x_t - \theta x_{t-1}$$

is neither attracting no
system results only in th
different amplitude and
form of behavior that i
systems.

These two examp
dynamical systems whic

(1) *Multip*
(2) *Attrac*

To describe the
systems, it will be nece
follow in part the survey
before, we are concern

dx = f(x)dt *or*

The collection of solut
ψ:S$\rightarrow$S such that $\psi_s(x_t$

$\psi_s(\psi_t) = \psi_{s+t}$.

For systems in discrete
on multiples of Δt. N
equilibria of the system

An *orbit* passing
discrete-time case, the
periodic if $\psi_\tau(x_t) = x_t$ f
orbit of a discrete-time
for all $0 < m < n$.

A point x is *non-*
$\in$ Z if discrete-time), th
the *non-wandering set*
equilibria as well as pe

$$x_{t+1} = x_t - \theta x_{t-1}, \tag{2.3.5}$$

is neither attracting nor repelling, since any deviation from a periodic orbit of this system results only in the establishment of a new periodic orbit of the same frequency but different amplitude and phase. Thus the periodic orbits of nonlinear systems constitute a form of behavior that is qualitatively different from anything seen in linear dynamical systems.

These two examples illustrate the two most important characteristics of nonlinear dynamical systems which are missing in linear dynamical systems:

(1) *Multiple attracting and repelling equilibrium points.*
(2) *Attracting and repelling periodic orbits.*

To describe the entire range of behavior of first-order nonlinear dynamical systems, it will be necessary to introduce some additional terminology. In these notes we follow in part the survey by Zeeman (1982). Suppose now that the state space S is $\Re^n$. As before, we are concerned with either differential or difference systems:

$$dx = f(x)dt \quad or \quad \Delta x = f(x)\Delta t. \tag{2.3.6}$$

The collection of solutions $\{x_t\}$ for $dx = f(x)dt$ constitutes a *flow*, which is a function ψ:S→S such that $\psi_s(x_t) = x_{s+t}$, or, stated a little more elegantly,

$$\psi_s(\psi_t) = \psi_{s+t}. \tag{2.3.7}$$

For systems in discrete time ψ is often called a *map* instead of a flow, and is defined only on multiples of Δt. Note that $\psi_{\Delta t}(x_t) = x_t + f(x_t)\Delta t$. The fixed points of $\psi_{\Delta t}$ are the equilibria of the system.

An *orbit* passing through a point x is the curve described by $\psi_t(x)$, for all t. In the discrete-time case, the orbit is the set of discrete points encountered by $\psi_t(x)$. An orbit is periodic if $\psi_\tau(x_t) = x_t$ for some $0 < \tau < \infty$ and all points x in the orbit at any t. A *period-n orbit* of a discrete-time system consists of points x such that $\psi_{n\Delta t}(x) = x$, but $\psi_{m\Delta t}(x) \neq x$ for all $0 < m < n$.

A point x is *non-wandering* if for every neighborhood N of x and all $t \in \Re$ (all $t/\Delta t \in Z$ if discrete-time), there is an $s > t$ such that $N \supseteq \psi_s(N)$. The set of all such points is the *non-wandering set* Ω. The non-wandering set of a nonlinear system may include equilibria as well as periodic orbits and non-periodic orbits.

A subset Λ of the state space is said to be *attracting* if there is a closed neighborhood N of Λ such that $\psi_t(N) \subset N$ for all $t > 0$, and $\bigcap_{t>0} \psi_t(N) = \Lambda$. Four kinds of attractors are known, in general:

i. Point attractors (equilibria).
ii. Periodic attractors (limit cycles).
iii. Toroidal attractors (quasi-periodic orbits).
iv. Strange attractors (chaotic orbits).

Strange attractors are particularly important. A flow or map which contains a strange attractor typically exhibits an extreme sensitivity to initial conditions, such that orbits that begin close together rapidly diverge. This gives an unpredictable character to the motion, which explains why it is called "chaos." *All of the nonlinear discrete-time catastrophe models that we consider here are capable of exhibiting chaos.*

The cubic dynamical system

$$\dot{x} = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 \tag{2.3.8}$$

may have from one to three equilibria; the number depends upon the parameter vector θ. In the case of three equilibria, as in Figure 2.1, there are typically two possibilities for the non-wandering set: an attractor surrounded by two repellors, or a repellor surrounded by two attractors.

The discrete-time case,

$$\Delta x = (\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3)\Delta t, \tag{2.3.9}$$

has a much more complicated non-wandering set. In addition to three equilibria, located at the roots of the cubic polynomial, the non-wandering set can contain periodic, toroidal, and chaotic orbits. An example of a chaotic orbit is shown in Figure 2.3, on the next page. Note its characteristic pseudo-random trajectory.

Recall that the cubic dynamical systems (2.3.8) and (2.3.9) have one or three equilibria depending upon the parameters of the model. In general the structure of the non-wandering set of a nonlinear dynamical is very dependent upon the parameters. As a parameter is smoothly varied, the non-wandering set may undergo sudden changes in composition, *e.g.* from one to two attracting equilibria, or from a toroidal attractor to a strange attractor, *etc.* A *catastrophe* (in the broadest sense) is a transition in which an attractor disappears from the non-wandering set. The cubic dynamical systems considered here are among the simplest nonlinear models which can exhibit catastrophes.

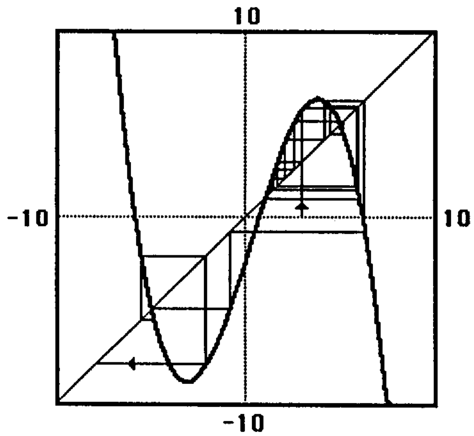

Figure 2.3 A cobweb graph that exhibits chaotic motion. The system is governed by $\Delta x = -0.09(x-5)(x-1)(x+5)\Delta t$, with an initial position at $x = 3$, and $\Delta t = 1$.

A sufficient condition for chaos is the existence of a period-3 orbit (Li & Yorke, 1975). The boundary of the chaotic domain for "bimodal" maps (a category which includes the polynomial systems considered here) has been analyzed exhaustively by MacKay & Tresser (1985).

A cubic dynamical system of the form (2.3.8) with $\theta_3 < 0$ may or may not have any period-2 orbits. If it does, then one period-2 orbit is a repellor which encloses all of the equilibria of the system. Lemma 2.1 (below) gives sufficient conditions for the existence of such a period-two orbit. We provide the proof in complete detail because of the importance of the existence of this orbit for the censored model presented in Section 3.

Lemma 2.1: Let $x_{t+1} = \psi_1(x_t)$, where

$$\psi_1(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3, \quad \alpha_3 < 0. \tag{2.3.10}$$

If this system has at least one attracting equilibrium point, then it also has at least one period-2 orbit, and the largest such orbit encloses all fixed points of ψ_1.

Proof of Lemma 2.1: Let $\psi_2(x) = \psi_1(\psi_1(x))$. Thus ψ_2 is a ninth degree polynomial, say $\psi_2(x) = \beta_0 + \beta_1 x + \ldots + \beta_9 x^9$. Notice that $\beta_9 = \alpha_3{}^4 > 0$, and that all the fixed points of ψ_1 are also fixed points of ψ_2. Obviously, $\psi_1(x) - x = 0$ has at most three real roots. We distinguish between two cases:

Case 1: $\psi_1(x) - x = 0$ has one real root of multiplicity 1;

Case 2: $\psi_1(x) - x = 0$ has three real roots (not necessarily distinct).

Case 1: If r is the unique attracting fixed point of ψ_1, then $|\psi_1'(r)| < 1$. Indeed, as is easy to check, if $\psi_1'(r) < -1$ then r is not an attractor. Moreover, $\psi_1'(r) < 0$; otherwise, since $\psi_1(x) \to \infty$ as $x \to \infty$, the assumptions of Case 1 are violated. By simple differentiation, we find that $\psi_2'(x) = \psi_1'(x)\psi_1'(\psi_1(x))$. Hence $\psi_2'(r) = [\psi_1'(r)]^2$, and $0 < \psi_2'(r) < 1$. Furthermore, $\psi_2'(x)$ is a continuous function, such that $\psi_2'(x) = 9\alpha_3{}^4x^8 + O(x^7)$ as $x \to \infty$. Hence, $\exists s > r$ such that $\psi_2'(x) > 1$ for all $x > s$. Let t be the point such that $\psi_2(t) = t$. Thus t is a fixed point of ψ_2 which is not a fixed point of ψ_1. Consequently $\{t, \psi_2(t)\}$ is a period-2 orbit of the system defined by (2.3.10).

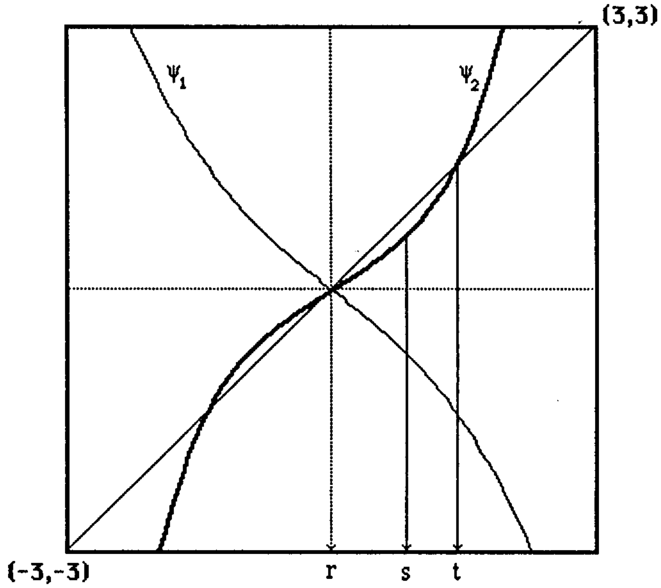

Figure 2.4 The functions ψ_1 and ψ_2 are graphed here in thin and thick lines, respectively, with $\psi_1(x) = -0.8x + 0.1x^3$. The origin is the fixed point of ψ_1. Note that $\psi_1'(0) = -0.8$, and that ψ_2 has three fixed points. This illustrates Case 1 of Lemma 2.1.

Case 2: Let the

2.5). There must exist a

line labelled **A** in Figu

$\psi_2(s) = \psi_1(\psi_1(s)) = \psi_1$

point, say t, such that t <

similar argument using

labelled **D**. The points

points of ψ_1.

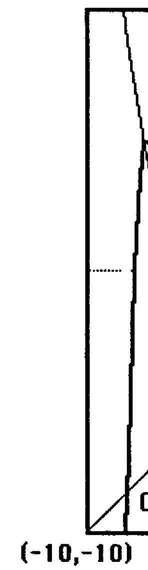

Figure 2.

thick lines

The period

(End of Proof)

Case 2: Let the fixed points of ψ_1 be r_1, r_2 and r_3, with $r_1 \le r_2 \le r_3$ (see Figure 2.5). There must exist a point, say s, such that $s < r_1$ and $\psi_1(s) = r_3$, because $\alpha_3 < 0$. The line labelled **A** in Figure 2.5 extends from (s,r_3) to (r_3,r_3). But $\psi_2(s) = r_3$ also, since $\psi_2(s) = \psi_1(\psi_1(s)) = \psi_1(r_3) = r_3$. Now the fact that $\beta_9 > 0$ implies that there must exist a point, say t, such that $t < s$ and $\psi_2(t) = t$. The point $(t,\psi_2(t))$ is labelled **C** in the figure. A similar argument using the line labelled **B** in the figure shows the existence of the point labelled **D**. The points **C** and **D** constitute a period-2 orbit, within which lie all the fixed points of ψ_1.

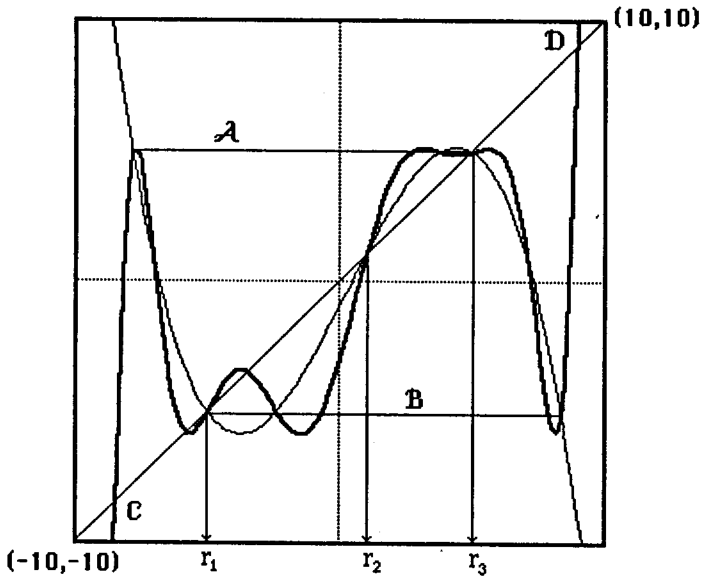

Figure 2.5 The functions ψ_1 and ψ_2 are graphed in thin and thick lines, respectively, with $\psi_1(x) = x - 0.04(x-5)(x-1)(x+5)$. The period-2 orbit is {C,D}. This illustrates Case 2.

(End of Proof)

3 Nonlinear Time Series Analysis

3.1 Introduction

Three main questions are connected with the topic of nonlinear time series analysis. One is related to the choice of model, the second with the estimation of the parameters of the chosen model, and the third with assessing the statistical properties of the estimators. The question of which nonlinear model to choose depends on the type of data under consideration, and on the characteristics of the model. In the present paper, we focus attention on cubic models of order one, with or without covariates. This class of models is derived from the catastrophe models presented in Section 1. The general form of the cubic time series (CTS) model without covariates is:

$$\Delta X_t = \theta_0 + \theta_1 X_t + \theta_2 X_t^2 + \theta_3 X_t^3 + \sigma U_t, \quad (t = 0, 1, \ldots), \tag{3.1.1}$$

where $\{U_t: t=0,1,\ldots\}$ is a stationary stochastic process with mean zero for all t and a specified correlation function. The general form of the CTS with k covariates is similar to (3.1.1), with

$$\theta_i = \alpha_i + \sum_{j=1}^{k} \beta_{ij} Y_{tj}, \quad (i = 0, 1, 2, 3), \tag{3.1.2}$$

where $\{Y_{t\cdot}\}$ is a given sequence of (non-random) k-dimensional covariates.

Unfortunately, models (3.1.1) and (3.1.2) are ill-behaved for the purposes of time series analysis, since they are not, in general, stationary. The nonstationarity is due to the finiteness of the domain of attraction—outside the domain of attraction the system exhibits destructively increasing oscillations. There are several possible ways to overcome this problem: (1) the cubic polynomial can be approximated with a piece-wise linear function, or (2) the model can be modified so that it is polynomial on a finite interval surrounding its equilibria, but changes to a different form outside this interval. We have chosen the latter alternative, so as to preserve the close connection between these models and the polynomial models of catastrophe theory. The former alternative is also viable, as in the piecewise linear SETAR models of Tong (1980, 1983), but these models lack the close connection to catastrophe theory that we are seeking.

In this section we present a class of "censored" polynomial models for time series analysis. These models are ergodic, and manage to preserve almost all of the desirable features of discrete time catastrophe models. We devote the remainder of this section to a thorough analysis of a special one-parameter model with no covariates which, it is hoped, faithfully reflects the essential difficulties that will be found in the more general models.

3.2 Propertie

Consider the CTS

$X_0 \equiv 0,$

$\Delta X_t = -\omega X_t^3 + \sigma$

with $\omega > 0$, $\sigma > 0$, and {

the deterministic case (σ

2.1 there is also a perio

(−A,A) are attracted to t

$\pm\infty$. Thus the origin is a

Let g(x,t) denote

(3.2.1) we obtain

$g(x,1) = \varphi(x/\sigma)/$

$g(x,t) = \int_{-\infty}^{\infty} \varphi([x-y-$

where $\varphi(z)$ denotes the

induction on t, that g(-

symmetric around the ori

The variance of X_t

$\mathrm{Var}[X_0] = 0,$

$\Delta\mathrm{Var}[X_t] = \sigma^2 - 2$

This shows that $\mathrm{Var}[X_t]$

when t=1 we have $X_1 \sim$

$\Delta\mathrm{Var}[X_1] = \sigma^2 - ($

This difference is greater

$\sqrt{2/\omega}$ is small, and wit

domain of attraction (-

following lemma shows

3.2 Properties of a Special Cubic Time Series Model

Consider the CTS given by

$$X_0 \equiv 0,$$

$$\Delta X_t = -\omega X_t^3 + \sigma U_t, \quad (t = 0, 1, 2, \ldots), \tag{3.2.1}$$

with $\omega > 0$, $\sigma > 0$, and $\{U_t\}$ an *i.i.d.* sequence of standard normal random variables. In the deterministic case (σ=0) there is one attracting equilibrium, at the origin. By Lemma 2.1 there is also a period-2 orbit at $\{-A,A\}$, where $A = \sqrt{2/\omega}$. All points in the interval $(-A,A)$ are attracted to the origin, while all points outside of $(-A,A)$ are repelled towards $\pm\infty$. Thus the origin is an attractor, while the periodic orbit is a repellor.

Let $g(x,t)$ denote the PDF of X_t. From the Markovian properties of the CTS (3.2.1) we obtain

$$g(x,1) = \varphi(x/\sigma)/\sigma,$$
$$g(x,t) = \int_{-\infty}^{\infty} \varphi([x-y+\omega y^3]/\sigma)g(y,t-1)dy, \quad (t \geq 2), \tag{3.2.2}$$

where $\varphi(z)$ denotes the PDF of the standard normal distribution. One can prove, by induction on t, that $g(-x,t) = g(x,t)$, for all $t \geq 1$, therefore the distribution of X_t is symmetric around the origin. Hence, all the odd moments of X_t are zero for all $t \geq 0$.

The variance of X_t satisfies the difference equation:

$$\mathrm{Var}[X_0] = 0,$$

$$\Delta \mathrm{Var}[X_t] = \sigma^2 - 2\omega E[X_t^4] + \omega^2 E[X_t^6], \quad (t=1,2,\ldots). \tag{3.2.3}$$

This shows that $\mathrm{Var}[X_t]$ may grow very fast with t, if ω is sufficiently large. Indeed, when t=1 we have $X_1 \sim N(0,\sigma^2)$ and

$$\Delta \mathrm{Var}[X_1] = \sigma^2 - 6\omega\sigma^4 + 15\omega^2\sigma^6. \tag{3.2.4}$$

This difference is greater than σ^2 whenever $\omega > 6/15\sigma^2$. Moreover, if ω is large then $A = \sqrt{2/\omega}$ is small, and with high probability the stochastic process will rapidly exit the domain of attraction $(-A,A)$, and begin destructive oscillations towards $\pm\infty$. The following lemma shows that the time required for the CTS (3.2.1) to exit the domain of attraction is stochastically smaller than a random variable having a geometric distribution.

Lemma 3.1: Let X_t satisfy (3.2.1), and let τ_A be the first exit time from the domain of attraction (A,-A), where $A = \sqrt{2/\omega}$. Thus

$$\tau_A = \text{least } t \geq 1 \text{ such that } |X_t| \geq A. \tag{3.2.5}$$

Then

$$P\{\tau_A > t\} < [\,\Phi(2^{3/2}/\sigma\sqrt{\omega})\,]^t, \quad (t \geq 1), \tag{3.2.6}$$

where $\Phi(z)$ is the standard normal integral.

Proof of Lemma 3.1: Define the "defective" cumulative distribution function

$$H(x,t) = P\{X_t \leq x, \tau_A \geq t\}, \quad (t \geq 1). \tag{3.2.7}$$

Notice that $H(x,1) = \Phi(x/\sigma)$. Because the CTS has the Markov property,

$$H(x,t) = \int_{-A}^{A} \Phi(\,[x-y+\omega y^3]/\sigma\,)h(y,t\text{-}1)dy, \quad (t \geq 2), \tag{3.2.8}$$

where $h(x,t) = \partial H(x,t)/\partial x$. In particular, for all $t \geq 2$,

$$H(A,t) = \int_{-A}^{A} \Phi(\,[A-y+\omega y^3]/\sigma\,)h(y,t\text{-}1)dy. \tag{3.2.9}$$

It is easy to verify that

$$\sup_{-A \leq y \leq A} \{\, A-y+\omega y^3 \,\} = 2A. \tag{3.2.10}$$

Hence, from the monotonicity of $\Phi(z)$, we obtain

$$H(A,t) \leq \Phi(2A/\sigma)\int_{-A}^{A} h(y,t\text{-}1)dy \tag{3.2.11}$$

$$\leq \Phi(2A/\sigma)\,[H(A,t-1) - H(-A,t-1)].$$

Notice that, from defin

P{ τ_A > t } = H(

< H(A

≤ Φ(2A

≤ [Φ(2

≤ [Φ(2

≤ [Φ(2

= [Φ(2

Thus the first ex

having a geometric dis

P{ τ^* = τ } = [1

It follows that the expe

E[τ_A] ≤ 1/[1–

If σ√ω is large then

Conversely, if σ√ω is s

In a similar man

of attraction, the time

than the geometric rand

3.3 A Censo

In order to avoi

model, we now consid

namely

Y_0 ≡ 0,

Y_{t+1} = [Y_t – ω

Notice that, from definition (3.2.7),

$$P\{\tau_A > t\} = H(A,t) - H(-A,t) \tag{3.2.12}$$

$$< H(A,t)$$

$$\leq \Phi(2A/\sigma)\, P\{\tau_A > t-1\}$$

$$\leq [\Phi(2A/\sigma)]^{t-1}\, P\{\tau_A > 1\}$$

$$\leq [\Phi(2A/\sigma)]^{t-1}\, [2\Phi(A/\sigma)-1]$$

$$\leq [\Phi(2A/\sigma)]^{t}$$

$$= [\Phi(2^{3/2}/\sigma\sqrt{\omega})]^{t}.$$

(End of Proof)

Thus the first exit time τ_A is stochastically smaller than a random variable τ^* having a geometric distribution:

$$P\{\tau^* = \tau\} = [1-\Phi(2^{3/2}/\sigma\sqrt{\omega})][\Phi(2^{3/2}/\sigma\sqrt{\omega})]^{t-1}, \quad (t>0). \tag{3.2.13}$$

It follows that the expected first exit time from the domain of attraction satisfies

$$E[\tau_A] \leq 1/[1-\Phi(2^{3/2}/\sigma\sqrt{\omega})]. \tag{3.2.14}$$

If $\sigma\sqrt{\omega}$ is large then this expectation is small and the first exit will occur rapidly. Conversely, if $\sigma\sqrt{\omega}$ is small then the expected time to the first exit is very large.

In a similar manner we can show that, once the process (3.5) is outside the domain of attraction, the time until the first entrance back into (-A,A) is stochastically greater than the geometric random variable having a mean of $1/[1/2 - \Phi(-2^{3/2}/\sigma\sqrt{\omega})]$.

3.3 A Censored Cubic Time Series Model

In order to avoid the undesirable behavior of the unmodified cubic time series model, we now consider a version which is "censored" within the domain of attraction, namely

$$Y_0 \equiv 0,$$

$$Y_{t+1} = [\,Y_t - \omega Y_t^3 + \sigma U_t\,]_{-A}^{+A}, \quad (t = 0, 1, 2, \ldots), \tag{3.3.1}$$

where $A = \sqrt{2/\omega}$, $\omega > 0, \sigma \geq 0$, and

$$[x]_{-A}^{+A} = \begin{cases} A, & x > A \\ x, & -A \leq x \leq A. \\ -A, & x < -A \end{cases} \qquad (3.3.2)$$

Like the uncensored model (3.2.1), the censored model is Markov, but unlike (3.2.1) it is ergodic, with positive recurrent states. Oscillations of the censored CTS (3.3.1) can occur at the points $\pm A$, which are the boundaries of the domain of attraction. Moreover, if $\tau^*(A)$ denotes the number of trials needed to return to the domain of attraction, given that $Y_t = \pm A$, then $\tau^*(A)$ has a geometric distribution

$$P\{\ \tau^*(A) = k \mid Y_t = \pm A\ \} = \Psi(A)(1-\Psi(A))^{k-1}, \quad (k = 1, 2, \ldots), \qquad (3.3.3)$$

regardless of t, where

$$\Psi(A) = \Phi(2^{3/2}/\sigma\sqrt{\omega}) - 1/2. \qquad (3.3.4)$$

The transition cumulative distribution function for the censored process has two mass points, at $x = \pm A$. It is given by

$$p(x, y; \omega, \sigma) = P\{\ Y_{t+1} \leq x \mid Y_t = y\ \} \qquad (3.3.5)$$

$$= \begin{cases} 0 & \text{if } x < -A, \\ \Phi(\,[x-y+\omega y^3]/\sigma\,) & \text{if } -A \leq x \leq A, \\ 1 & \text{if } x \geq A. \end{cases}$$

Furthermore, the cumulative distribution function of Y_t, $G(x,t;\ \omega,\sigma)$, can be determined recursively, according to these equations:

$$G(x,1;\ \omega,\sigma) = \begin{cases} 0 & \text{if } x < -A, \\ \Phi(x/\sigma) & \text{if } -A \leq x < A \\ 1 & \text{if } A \leq x, \end{cases} \qquad (3.3.6)$$

$$G(x,t;\ \omega,\sigma) = \mathbf{I}\{x \geq A\} + \mathbf{I}\{-A \leq x < A\}\ G(-A,t-1;\omega,\sigma)\ \Phi([x-A]/\sigma)$$

$$+ \int_{-A}^{A} g(y,t-1;\omega,\sigma)\Phi([x-y+\omega y^3]/\sigma)dy + (1-G(A,t-1;\omega,\sigma))\Phi([x+A]/\sigma)\ ], \qquad (3.3.7)$$

where $I\{\cdot\}$ is an indicator function and $g(y,t;\,\omega,\sigma) = \partial G(y,t;\,\omega,\sigma)/\partial y$.

The stationary distribution of Y_t, *i.e.* $G^*(x) = \lim_{t\to\infty} G(x,t)$, can be approximated by the stationary distribution of a discrete Markov Chain obtained by partitioning (-A,A) into N subintervals, and including two additional states corresponding to the points -A and A.

3.4 Maximum Likelihood Estimators for the Censored Model

In this section, we present an algorithm for the determination of the maximum likelihood estimators of the parameters (ω,σ) of the censored cubic time series model (3.3.1).

Let $y^{[T]} = [y_1, \ldots, y_T]$ be a vector of T observed consecutive sample values of the stochastic process Y_t, where Y_t is governed by (3.3.1). Define the random variables

$$m_T = \min\{y_1, \ldots, y_T\},$$

$$M_T = \max\{y_1, \ldots, y_T\},$$

$$\omega^*_T = \min\{\, 2/m_T^{\,2},\, 2/M_T^{\,2}\, \}.$$

Since $|Y_t| \le (2/\omega)^{1/2}$ for all t with probability one,

$$\omega \le \omega^*_T. \tag{3.4.1}$$

Thus, the likelihood function of (ω,σ), given the observed vector $y^{[T]}$, satisfies $L(\omega,\sigma;\, y^{[T]}) = 0$, for all $\omega > \omega^*_T$. For $\omega \in (0,\omega^*_T)$, the likelihood function can be written as

$$L(\omega,\sigma;\, y^{[T]}) = \sigma^{-T}\exp[\, -\sum_{t=0}^{T-1} (y_{t+1}-y_t+\omega y_t^{\,3})^2/2\sigma^2\,], \tag{3.4.2}$$

with $Y_0 \equiv 0$. The value of the likelihood function at ω^* can be determined as follows. Define

$$K_1 = \{t:\, y_t = -(2/\omega^*_T)^{1/2} \text{ and } t \in \{0,1, \ldots ,T\}\ \},$$

$$K_2 = \{t:\, y_t = +(2/\omega^*_T)^{1/2} \text{ and } t \in \{0,1,\ldots,T\}\ \},$$

$$K_3 = (\neg K_1) \cap (\neg K_2).$$

Let $J_i = |K_i|$ be the cardinality of K_i. Notice that $J_i \geq 0$ for i = 1 and 2, but $J_1+J_2 \geq 1$. Both J_1 and J_2 are positive if $m_T = -M_T$. Thus the likelihood at ω^*_T is

$$L(\omega^*,\sigma; y^{[T]}) = \prod_{t\in K_1}\Phi(\, U(\omega^*_T,t-1)/\sigma\,) \prod_{t\in K_2}\Phi(\, V(\omega^*_T,t-1)/\sigma\,)\ \sigma^{-J_3} \exp[\, -Q(\omega^*_T)\,], \tag{3.4.3}$$

where

$$U(\omega,t) = -(2/\omega)^{1/2} - y_t + \omega y_t^{\,3},$$

$$V(\omega,t) = +(2/\omega)^{1/2} + y_t - \omega y_t^{\,3}, \quad \text{and}$$

$$Q(\omega) = \sum_{t\in K_3} (y_{t+1}-y_t+\omega y_t^{\,3})^2.$$

Notice that if K_i is empty, then the corresponding product in (3.4.3) is equal to 1 by definition.

The log-likelihood function $\lambda(\omega,\sigma; y^{[T]})$ is, for $\omega < \omega^*_T$,

$$\lambda(\omega,\sigma; y^{[T]}) = -T \log \sigma - \sum_{t=0}^{T-1} (y_{t+1} - y_t + \omega y_t^{\,3})^2/2\sigma^2. \tag{3.4.4}$$

At $\omega = \omega^*_T$ we have

$$\lambda(\omega^*_T,\sigma; y^{[T]}) = \sum_{t\in K_1} \log \Phi(\, U(\omega^*_T,t-1)/\sigma\,) + \sum_{t\in K_2} \log \Phi(\, V(\omega^*_T,t-1)/\sigma\,)$$

$$- J_3 \log \sigma - Q(\omega^*_T)/2\sigma^2. \tag{3.4.6}$$

Note that $U(\omega^*_T,t-1) \leq 0$ for all $t\in K_1$, and $V(\omega^*_T,t-1) \leq 0$ for all $t\in K_2$, since $|y_t| \leq (2/\omega)^{1/2}$ for all t.

The *Maximum Likelihood Estimator* (MLE) of (ω,σ) is the point in $[0,\omega^*_T] \times (0,\infty)$ at which $\lambda(\omega,\sigma; y^{[T]})$ attains its supremum. In order to determine the MLE, we find first a point (ω,σ) for which $\lambda(\omega,\sigma; y^{[T]})$ attains its supremum over the subset $(0,\omega^*_T) \times (0,\infty)$, *i.e.* we first ignore the possibility that $\omega = \omega^*_T$. It is easy to check that the *unique* point maximizing (3.4.4) is

$$\hat{\omega}_T = \min\{\, \omega^*_T, -\sum_{t=0}^{T-1} y_t^3 \Delta y_t \,/\, \sum_{t=0}^{T-1} y_t^6 \,\} \quad \text{if } \textstyle\sum_t y_t^3 \Delta y_t < 0,$$

$$= 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{if } \textstyle\sum_t y_t^3 \Delta y_t \geq 0,$$

$$\hat{\sigma}_T = \Big[\sum_{t=0}^{T-1} (y_{t+1} - y_t + \omega_T y_t^3)^2 / T \,\Big]^{1/2}. \tag{3.4.7}$$

The maximal likelihood

$$\lambda_1 = \lambda(\hat{\omega}_T, \hat{\sigma}_T; y^{[T]}) \tag{3.4.8}$$

is then compared with the boundary likelihood

$$\lambda_2 = \sup_{0<\sigma<\infty} \lambda(\omega^*_T, \sigma; y^{[T]}). \tag{3.4.9}$$

If $\lambda_1 \geq \lambda_2$ then the MLE is $(\hat{\omega}_T, \hat{\sigma}_T)$, otherwise the MLE is (ω^*_T, σ^*_T), where σ^*_T is the value of σ for which λ_2 is attained in (3.4.9).

Furthermore, due to the continuity of the distributions $G_t(y)$, for $|y|<A$,

$$P\{\, (J_1>1) \cup (J_2>1) \,\} = P_{\omega,\sigma}\{\omega^* = \omega\}. \tag{3.4.10}$$

Hence, for all $y^{[T]}$ such that the event $\{J_1>1\} \cup \{J_2>1\}$ occurs,

$$\lambda(\omega,\sigma; y^{[T]}) = 1 \;\; \text{if } \omega = \omega^*_T,$$

$$= 0 \;\; \text{otherwise.}$$

We show now that, if $J_1+J_2 < N$, there exists a unique point σ^*_T, in $(0,\infty)$, for which $\lambda(\omega^*_T,\sigma; y^{[T]})$ attains its supremum. Indeed, partially differentiating (3.4.6) with respect to σ, and equating the derivative to zero, we find that σ^*_T is the root of the equation

$$\sigma^2 = Q(\omega^*_T)/J_3 + (\sigma/J_3)\,\Big\{ \sum_{t\in K_1} |U(\omega^*_T,t-1)|\, \phi(\, U(\omega^*_T,t-1)/\sigma \,) \,/\, \Phi(\, U(\omega^*,t-1)/\sigma \,)$$

$$+ \sum_{t\in K_2} |V(\omega^*_T,t-1)|\, \phi(\, V(\omega^*_T,t-1)/\sigma \,) \,/\, \Phi(\, V(\omega^*,t-1)/\sigma \,) \Big\}. \tag{3.4.11}$$

We see immediately that $(\sigma^*_T)^2 > Q(\omega^*_T)/J_3$. Thus, when $J_3 = 0$, there is no finite ML estimator of σ. Moreover, if $H^*(\sigma)$ denotes the r.h.s. of (3.4.11), we see that $H^*(\sigma)$ is continuous, that $H(0) = Q(\omega^*_T)/J_3$, and that

$$H^*(\sigma) \leq Q(\omega^*_T)/J_3 + \sigma(J_1+J_2) / \{J_3(\pi\omega^*_T)^{1/2}\,\Phi(-(2/\sigma\omega^*_T)^{1/2})\}. \tag{3.4.12}$$

As $\sigma \to \infty$ the increase of the r.h.s. of (3.4.12) is approximately linear, while the l.h.s. of (3.4.11) is quadratic. Hence there exists a unique root, σ^*_T, of (3.4.11).

3.5 Properties of the MLE for the Censored Model

We have seen in Section 3.3 that, with probability one, almost all realizations of the censored model (3.3.1) attain the boundaries $\pm A$ infinitely often. Thus sooner or later, with probability one, $J_1+J_2 > 1$, and $\omega^*_T = \omega'$, where ω' is the true value of ω. **Thus the MLE of ω is strongly consistent,** *i.e.*

$$\lim_{T\to\infty} P_{\omega',\sigma}\{\, MLE_T(\omega) = \omega' \,\} = 1, \tag{3.5.1}$$

where $MLE_T(\omega)$ denotes the MLE of ω, given $y^{[T]}$. The likelihood function of σ, under ω^*_T, *i.e.* $L(\omega^*_T,\sigma; y^{[T]})$, satisfies the regularity conditions for consistency and asymptotic normality of $MLE_T(\sigma)$, as can be checked from (Basawa & Rao, 1980, pp. 122-125). The asymptotic distribution of $MLE_T(\omega)$ is, however, not normal.

3.6 Discussion

It is noteworthy that the our maximum likelihood estimator (3.4.7) is, in many cases, identical to the least-squares estimator

$$LSE_T(\omega) = -\sum_{t=0}^{T-1} y_t^3 \Delta y_t \Big/ \sum_{t=0}^{T-1} y_t^6, \tag{3.6.1}$$

and also closely resembles the maximum likelihood estimator for the stochastic differential equation (recall that $\Delta t = 1$)

$$dy_t = -\omega y_t^3 dt + \sigma dw_t, \tag{3.6.2}$$

namely

$$MLE_T(\omega) = -\int_0^T y_t^3$$

We have overcon
largest period-2 orbit, ar
simple special case of a
distribution of the MLE
because of the non-ergoc
have shown that the asyn
normal distribution and a

In the present pap
the cusp catastrophe mo
Zacks, 1985). Extendin
stochastic systems is,
elementary catastrophe
deterministic systems th
The problem is not th
equations—that is easily
deterministic part of th
that one-dimensional d
wandering sets that cont
$\Delta x = a+bx+cx^2$ has a ch
theorems of catastroph
relation on discrete-time
theory to motivate the s
on the basis that the fam
bifurcations of its isol
theory. Elementary c
periodic and chaotic orb

In summary, this
time series models is su
as the fundamental clas
that all the isolated eq
periodic orbit. Third,
orbit, we obtain a statio
parameters this model tl

namely

$$MLE_T(\omega) = -\int_0^T y_t^3 dy_t \, / \int_0^T y_t^6 dt. \tag{3.6.3}$$

We have overcome the nonstationarity problem by censoring the model at its largest period-2 orbit, and studied the properties of maximum likelihood estimators in a simple special case of a censored model. We have not discussed here the asymptotic distribution of the MLE of ω. It is clear that the classical theory does not apply here because of the non-ergodicity in the non-censored case. In the censored case simulations have shown that the asymptotic distribution of the MLE of ω is, apparently, a mixture of a normal distribution and a singular mass centered at ω.

In the present paper we have considered only cubic polynomial models, inspired by the cusp catastrophe model. More general polynomial models are available (see Cobb & Zacks, 1985). Extending the results and methods of catastrophe theory to discrete time stochastic systems is, however, quite problematic. The classification theorems of elementary catastrophe theory (Poston & Stewart, 1978) apply only to continuous time deterministic systems that have non-wandering sets which contain only discrete points. The problem is not the distinction between stochastic and deterministic differential equations—that is easily dealt with by restating the theory in terms of perturbations of the deterministic part of the stochastic differential equation. The fundamental problem is that one-dimensional discrete-time nonlinear systems do not, in general, have non-wandering sets that contain exclusively discrete points. Even the simple quadratic model $\Delta x = a+bx+cx^2$ has a chaotic domain within its parameter space. Thus one cannot use the theorems of catastrophe theory to claim the existence of an exhaustive equivalence relation on discrete-time models of low codimension. However, one can use catastrophe theory to motivate the selection of polynomial models for nonlinear time series analysis, on the basis that the family of polynomial time series models does at least exhibit generic bifurcations of its isolated equilibria—this much we can recover from catastrophe theory. Elementary catastrophe theory is silent on the question of bifurcations to periodic and chaotic orbits, which requires further mathematical research.

In summary, this is our approach: First, we claim that the family of polynomial time series models is sufficiently rich in its range of qualitative behavior to justify its use as the fundamental class of models for nonlinear time series analysis. Second, we show that all the isolated equilibria of a cubic polynomial model lie within an enclosing periodic orbit. Third, we show that by censoring the model at the enclosing periodic orbit, we obtain a stationary and ergodic process. Fourth, we derive estimators for the parameters this model that converge to the correct values with probability one.

4 References

Basawa, I.V. & Rao, B.L.S. (1980) **Statistical Inference for Stochastic Processes.** New York: Academic Press.

Cobb, Loren, and Zacks, Shelemyahu (1985) "Applications of catastrophe theory for statistical modeling in the biosciences," **Journal of the American Statistical Association, 78,** 124-130.

Li, T.Y. and Yorke, J.A. (1975) "Period three implies chaos," **American Mathematical Monthly, 82,** 985-992.

Liptser R.S. & Shiryayev, A.N. (1977) **Statistics of Random Processes I: General Theory.** New York: Springer-Verlag.

Liptser R.S. & Shiryayev, A.N. (1978) **Statistics of Random Processes II: Applications.** New York: Springer-Verlag.

MacKay, Robert S. and Tresser, C. (1985) "Boundary of chaos for bimodal maps of the interval," preprint, Institute of Mathematics, University of Warwick, Coventry, England.

Poston, Timothy, and Stewart, Ian N. (1978) **Catastrophe Theory and its Applications.** London: Pitman.

Stroyan, K.D., and Bayod, J.M. (1986) **Foundations of Infinitesimal Stochastic Analysis.** Amsterdam: North-Holland.

Thompson, J.M.T. (1982) **Instabilities and Catastrophes in Science and Engineering.** New York: Wiley.

Tjøstheim, Dag (1986) "Estimation in nonlinear time series models," **Stochastic Processes and their Applications, 21,** 251-273.

Tong, Howell, & Lim, K.S. (1980) "Threshold autoregression, limit cycles, and cyclical data (with discussion)," **Journal of the Royal Statistical Society, Series B, 42,** 245-292.

Tong, Howell (1983) **Threshold Models in Nonlinear Time Series Analysis.** Lecture Notes in Statistics #21. New York: Springer-Verlag.

Zeeman, E.C. (1982) "Bifurcation and catastrophe theory," **Contemporary Mathematics, 9,** 207-272.

NONLINEAR PROCESSING WIT

F.J. Bugnon and R.R. Moh
Department of Electrical
Oregon State University
Corvallis, OR 97331 (US

1. INTRODUCTION

Signal processing te
that are random in natur
The classical way to sum
moments (mean, variance
perfectly be described t
to keep the whole origin
process raw measurements
computers.

In this chapter, nor
analogous form to the st
order sequences are buil
are to be processed inst
equation, a generating f
measurement sequences to

The direction-findin
vantages of the Mth-orde
generated from Mth-order
problem, and it produce
in the number of sensor

2. NONLINEAR SEQUENCE

2.1 Mth-Order Sign

A vector r of random
case, at time t_k. Let

$r_{i,M}(t) - r_{i,M-1}(t)$