Bifurcating Time Series Models for Cell Lineage Data

by

Jin Zhou

(Under the direction of I. V. Basawa)

Abstract

This dissertation studies bifurcating time series models. Our motivation comes from cell lineage data, in which each individual in a generation gives rise to two individuals in the next generation. For general bifurcating autoregressive models, asymptotic normality of least squares estimators of model parameters is established. An application to integer-valued autoregression is given. For the first-order bifurcating autoregressive process with exponential innovations, exact and asymptotic distributions of the maximum likelihood estimator of the autoregressive parameter are derived. Limit distributions for stationary, critical and explosive cases are unified via a single pivot using a random normalization. The pivot is shown to be asymptotically exponential for all values of the autoregressive parameter. Finally, a general class of Markovian non-Gaussian bifurcating models is studied. Examples include bifurcating autoregression, random coefficient autoregression, bivariate exponential, bivariate gamma, and bivariate Poisson models. Quasilikelihood estimation for the model parameters and large-sample properties of the estimates are discussed.

Index words: Cell Lineage Data; Tree-Indexed Data; Bifurcating Autoregressive Models; Least Squares Estimation; Maximum Likelihood Estimation; Quasilikelihood Estimation; Exponential Innovations; Exact Distribution; Limit Distribution; Asymptotic Property; Non-Gaussian Models; Integer-valued Autoregression.

Bifurcating Time Series Models for Cell Lineage Data

by

Jin Zhou

B.S., University of Science and Technology of China, 1997

M.S., University of Science and Technology of China, 2000

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2004

Bifurcating Time Series Models for Cell Lineage Data

by

Jin Zhou

Approved:

Major Professor: I. V. Basawa

Committee: Gauri Datta
Robert Lund
Jaxk Reeves
Anand Vidyashankar

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

Dedication

To My Parents and My Wife

For Their Love and Support

Acknowledgments

I wish to thank my major professor, Dr. I. V. Basawa, for his guidance and support throughout all stages of this dissertation. Were it not for his patient guidance, frequent meetings, and invaluable feedback and suggestions, this work would not have turned out as it did. I am honored to have been his student and I wish him the best.

I also thank my committee members, Dr. Gauri Datta, Dr. Robert Lund, Dr. Jaxk Reeves, and Dr. Anand Vidyashankar, for teaching me wonderful classes, sharing with me their experiences, enhancing this dissertation and encouraging me in my academic development.

Thanks go to Dr. Robert Taylor, now at Clemson University, for his wonderful classes and help in my career development. My thanks are extended to all the faculty members, currently or previously at this program, who educated me, encouraged me and helped me. Special thanks go to Connie Durden, for her timely and professional typing assistance.

The words are frail when I express my appreciation to my parents. Their unconditional love and sacrifice are the most invaluable assets to me.

Last, but not the least, my deepest gratitude and love go to my wife, Se Li. Thank you for your love, support and appreciation of my values. Because of you, my life has become meaningful and beautiful.

Table of Contents

	Page
Acknowledgments	v
List of Tables	viii
Chapter	
1 Introduction	1
2 Literature Review	4
2.1 Cell Lineage Data and Bifurcating Models	4
2.2 Extended BAR Models and Inference	5
2.3 Non-Gaussian Conditional Linear AR(1) Models	10
2.4 Estimating Functions and Quasilikelihood Estimation	11
3 Least Squares Estimation for Bifurcating Autoregressive Processes	13
3.1 Introduction	14
3.2 Least Squares Estimation for BAR(p) Processes	15
3.3 Limit Distributions	16
3.4 Integer-Valued Bifurcating Autoregressive Model	24
3.5 Bifurcating Poisson Model	26
3.6 References	28
4 Maximum Likelihood Estimation for a First-Order Bifurcating Autoregressive Process with Exponential Errors	30
4.1 Introduction	31

4.2 Exact Distribution of the Maximum Likelihood Estimator 33
4.3 Asymptotic Distributions 38
4.4 A Unified Limit Theorem 42
4.5 Simulation Results 46
4.6 References 53
5 Non-Gaussian Bifurcating Models and Quasilikelihood Estimation 54
5.1 Introduction 55
5.2 Specification of the Model: Likelihood and Quasilikelihood Estimation 56
5.3 Examples 57
5.4 Remarks on Asymptotic Properties 60
5.5 Non-Gaussian Conditional Linear Bifurcating Models . . 61
5.6 Concluding Remarks 66
5.7 References 67
6 Future Topics 69
6.1 Bifurcating Random Walk with Drift 69
6.2 Consistency and Asymptotic Normality of QL Estimates 71
6.3 Multiple-Splitting Model 72
Bibliography 75

List of Tables

4.1 Comparison of $\hat{\phi}_{ML}$ and $\hat{\phi}_{LS}$ 48
4.2 Comparison of $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{LS}$ (with $\lambda = 1$) 49
4.3 The proportion of $2T_n > \chi^2_\alpha(2)$ in 1000 simulations 50

Chapter 1

Introduction

Bifurcating models are concerned with modeling data on descendants of an initial individual, in which each individual in a generation gives rise to two individuals in the next generation. Cell lineage data, such as *Escherichia coli* by Powell (1955), EMT6 cells by Collyn d'Hooge *et al.* (1977), are typically of this kind. The most important features of cell lineage data include the bifurcating tree structure and correlation of sister or cousin cells. For biological details of cell lineage data, see Powell (1955, 1956, 1958), Collyn d'Hooge *et al.* (1977), Brooks *et al.* (1980), Hola and Riley (1987), and Staudte *et al.* (1996).

To analyze cell lineage data, Cowan (1984) and Cowan and Staudte (1986) proposed the bifurcating autoregressive (BAR) model. If X_t denotes a measurement on some characteristic of individual t, the first-order BAR model is given by

$$X_t = \phi X_{[t/2]} + \epsilon_t, \qquad t = 2, 3, ...,$$

where $|\phi| < 1$ is assumed for causality of the model, $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of independently and identically distributed (iid) bivariate random vectors with common mean μ, common variance σ^2 and correlation ρ. Here $[u]$ denotes the largest integer less than or equal to u. The motivation for this correlation is that sister cells grow in a similar environment and hence one expects an environmental correlation between sisters.

Staudte *et al.* (1984) proposed an additive model which allows for positive correlation between sister cells but arbitrary correlation between mother and daughter cells. Staudte (1992) extended the BAR model to allow for variable generation means. Huggins and Staudte (1994) introduced variance components models to allow for measurement error and between tree variability. Huggins (1995) derived asymptotic properties of both robust and maximum likelihood estimators. The problem of identifiability of measurement error was discussed by Huggins (1996). A bivariate BAR model was proposed by Bui and Huggins (1998) to analyze bivariate cell data. A random coefficient BAR model was introduced by Bui and Huggins (1999) to allow for environmental factors. Robust inference of BAR models was discussed by Huggins and Marschner (1991), Huggins and Staudte (1994), Huggins (1996), and Bui and Huggins (1998).

Huggins and Basawa (1999) extended BAR models to higher order bifurcating autoregressive and moving average (BARMA) models and fitted models to data from independent trees. A distance model was also introduced by Huggins and Basawa (1999) to account for correlation between cousins. Huggins and Basawa (2000) studied asymptotic properties of maximum likelihood estimators for BAR(p) models.

Most of the work on bifurcating models mentioned above retained the normality assumption. In many applications, the normality assumption may not be realistic. For instance, if X_t denotes the life time of the tth individual, a non-negative random variable, a gamma or an exponential model may be more appropriate. If X_t denotes the number of certain type of genes present, a Poisson model may be considered.

The dissertation is organized as follows. Chapter 2 contains a literature review. In Chapter 3, we discuss general bifurcating autoregressive models, without imposing any specific distributional assumption on errors. Asymptotic normality of least squares estimators of the model parameters is established. An application to bifurcating integer-valued autoregression is given. Chapter 4 introduces a first-order bifurcating autoregressive process with exponential innovations. Exact and asymptotic distributions of the maximum likelihood estimator of the autoregressive parameter are derived. Limit distributions for stationary, critical and explosive cases are unified via a single pivot using a random normalization. The pivot is shown to be asymptotically exponential for all values of the autoregressive parameter. Chapter 5 presents a general class of Markovian non-Gaussian bifurcating models. Examples include bifurcating autoregression, random coefficient autoregression, bivariate exponential, bivariate gamma, and bivariate Poisson models. Quasilikelihood estimation for the model parameters and large-sample properties of the estimates are discussed. In Chapter 6, we discuss several future topics, including bifurcating random walk, asymptotic properties of quasilikelihood estimates and multiple-splitting models.

Chapter 2

Literature Review

2.1 Cell Lineage Data and Bifurcating Models

Cell lineage data consist of measurements on characteristics of the descendants of an initial cell, where each cell in one generation gives rise to two cells in the next generation. This type of data includes *Escherichia coli* data by Powell (1955), EMT6 cells data by Collyn d'Hooge *et al.* (1977), 3T3 cells data by Brooks *et al.* (1980), and epithelial cells data by Hola and Riley (1987). These data are obtained by methods of direct observation, time-lapse photography, or more advanced image analyzers and computer software. Measurements on characteristics of the initial cell and its offspring, such as the cell lifetimes and cell size at division, form a bifurcating tree of dependent data. The objective is to determine the extent to which the characteristic is influenced by environmental and inherited factors. A typical lineage of such data is shown in Figure 1.

The most important feature of cell lineage data lies in its inherent bifurcating structure and the dependence of sister or cousin cells. This feature requires extensions of classical models for statistical analysis of cell lineage data. The bifurcating autoregressive (BAR) model was introduced by Cowan (1984) and developed by Cowan and Staudte (1986) to analyze cell lineage data. Suppose in the cell division of a cell lineage tree, individual t produces daughter cells $2t$ and $2t+1$. Let X_t denote an observation on some characteristic of individual t. The BAR(1) model is given by

$$X_t = \phi X_{[t/2]} + \epsilon_t, \qquad t = 2, 3, ..., \tag{2.1.1}$$

where $|\phi| < 1$, $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of iid bivariate normal random vector with common mean μ, common variance σ^2 and correlation ρ. Here $[u]$ denotes the largest integer less than or equal to u. Another form of BAR(1) model is

$$X_t = \phi_0 + \phi_1 X_{[t/2]} + \epsilon_t, \qquad t = 2, 3, ..., \tag{2.1.2}$$

where the assumptions are the same as in (2.1.1) except that $\{\epsilon_{2t}, \epsilon_{2t+1}\}$ has mean 0. Maximum likelihood estimators are developed and compared via a simulation study in Cowan and Staudte (1986).

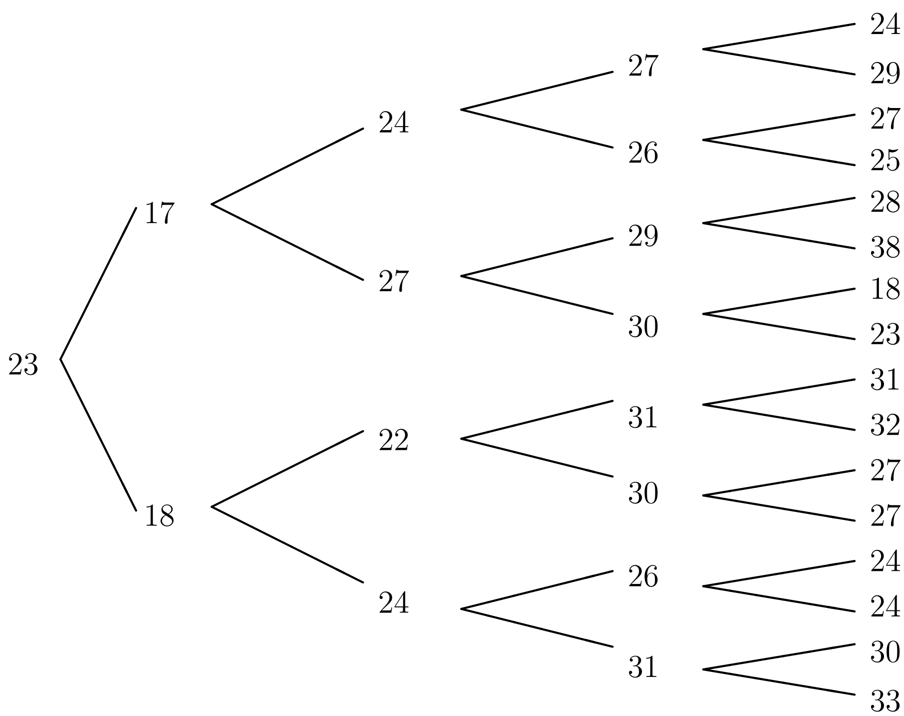


Figure 1: Cell lifetimes of *E. Coli* in minutes (Cowan & Staudte (1986)).

2.2 Extended BAR Models and Inference

Huggins and Basawa (1999) extended the BAR(1) model to the BARMA(p, q) model, which is defined by

$$\phi(b)X_t = \theta(b)\epsilon_t, \quad t = 0, \pm1, \pm2, ..., \tag{2.2.1}$$

where

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - ...\phi_p z^p,$$

$$\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + ... + \theta_q z^q$$

and b denotes the bifurcating operator

$$b^r u_t = u_{[t/2^r]^*} = \begin{cases} u_{[t/2^r]} & , \quad \text{if} \quad t \geq 2^r \\ u_{[\log_2(t/2^r)]+1}, & \text{if} \quad 0 < t < 2^r \\ u_{[t-r]} & , \quad \text{if} \quad t \leq 0 \end{cases}$$

Note that t can be negative in BARMA(p, q) model, where descendants of the initial cell are labeled according to their position in the tree but ancestors of the initial cell are labeled 0, -1, -2, ... In this sense, the BAR(1) model in (2.1.1) can be rewritten as

$$X_t = \phi X_{[t/2]^*} + \epsilon_t, \qquad t = 0, \pm 1, \pm 2, ... \tag{2.2.2}$$

As a special case of BARMA(p, q) model, BAR(p) model is defined by

$$X_t = \phi_1 X_{[t/2]^*} + \phi_2 X_{[t/4]^*} + ... + \phi_p X_{[t/2^p]^*} + \epsilon_t, \qquad t = 0, \pm 1, \pm 2, ... \tag{2.2.3}$$

When the BARMA(p, q) process is causal and invertible in each descendant line in the sense of Brockwell and Davis (1987), it has the stationary solution

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{[t/2^j]^*}, \tag{2.2.4}$$

where

$$\Psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad \text{for} \quad |z| \leq 1.$$

Let $c(s, t) = \max\{v : \text{ for some } r \text{ and } q, v = [t/2^r]^* \text{ and } v = [s/2^q]^*\}$, so $c(s, t)$ represents the most recent common ancestor of in individuals t and s. Also define $g_t(s, t)$ and $g_s(s, t)$ to satisfy $c(s, t) = [t/2^{g_s(s,t)}]$ and $c(s, t) = [t/2^{g_t(s,t)}]$ simultaneously, so that $g_t(s, t)$ and $g_s(s, t)$ represents the number of generations since the most recent common ancestor of t and s. For notational simplicity, write $g_t = g_t(s, t)$ and $g_s = g_s(s, t)$.

To obtain a general form for the covariance between any two individuals, Huggins and Basawa (1999) gave the following lemma.

Lemma 2.1. For any t and s,

$$X_t = \sum_{j=0}^{\infty} \psi_{j+g_t} \epsilon_{[c(s,t)/2^j]^*} + \sum_{j=0}^{g_t-1} \psi_j \epsilon_{[t/2^j]^*}. \tag{2.2.5}$$

According to Lemma 2.1, the covariance between individuals t and s is

$$\text{cov}(X_t, X_s) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+g_t} \psi_{j+g_s} + \rho\sigma^2 \psi_{g_t-1} \psi_{g_s-1} \tag{2.2.6}$$

where $\psi_{-1} = 0$, so that for individuals on the same line of descent, i.e., $\min(g_t, g_s) = 0$, $\text{cov}(X_t, X_s) = \sigma^2 \sum_{j=0}^{\infty} \psi_{j+g_t} \psi_{j+g_s}$.

For the BAR(1) model, the stationary solution is

$$X_t = \sum_{r=0}^{\infty} \psi_j \epsilon_{[t/2^j]^*} = \sum_{j=0}^{\infty} \phi^j \epsilon_{[t/2^j]^*}, \quad t = 0, \pm 1, \pm 2, ...$$

With regard to the covariance structure of BAR(1) model, if individuals s and t are on different lines, i.e., $\min(g_t, g_s) > 0$, then according to (2.2.6),

$$\begin{aligned} \text{cov}(X_t, X_s) &= \sigma^2 \sum_{j=0}^{\infty} \phi^{j+g_t} \phi^{j+g_s} + \rho\sigma^2 \phi^{g_t-1} \phi^{g_s-1} \\ &= \frac{\varphi\sigma^2}{1-\phi^2} \phi^{g_t+g_s-2} \end{aligned} \tag{2.2.7}$$

and hence $\rho(X_t, X_s) = \varphi\phi^{g_t+g_s-2}$, where $\varphi = \phi^2 + (1-\phi^2)\rho$, which is the unconditional correlation between sisters. If s and t are on the same line, i.e., $\min(g_t, g_s) = 0$, then

$$\text{cov}(X_t, X_s) = \frac{\sigma^2}{1-\phi^2} \phi^{g_t+g_s} \tag{2.2.8}$$

and $\rho(X_t, X_s) = \phi^{g_t+g_s}$.

One notable point is that in the analysis of stationary solution and covariance structure, $\{\epsilon_{2t}, \epsilon_{2t+1}\}$ is assumed to be only iid, without any specific distributional assumption. In the following analysis of this section, BAR(p) models are assumed to be Gaussian, which means the error term $\{\epsilon_{2t}, \epsilon_{2t+1}\}$ has a bivariate normal distribution.

Suppose in the BAR(p) model defined in (2.2.3), $\{\epsilon_{2t}, \epsilon_{2t+1}\}$ forms a sequence of iid bivariate normal random vectors with common mean μ, common variance σ^2 and correlation ρ. A bifurcating tree consisting of complete mother-daughter triples (X_t, X_{2t}, X_{2t+1}), $t = 1, 2, ..., n$ is observed. Let $\mathbf{X}_t = (X_{[t/2^j]}, j = 1, 2, ..., p)^T$ denote the vector of the most recent p ancestors of X_{2t} and X_{2t+1}. Then the likelihood of Gaussian BAR(p) model is the product of the conditional densities of (X_{2t}, X_{2t+1}) given $\mathbf{X_t}$, by the Markovian property of BAR(p) model. These conditional distributions are bivariate normal with means $\eta_{2t} = \eta_{2t+1} = \sum_{j=1}^{p} \phi_j X_{[t/2^j]^*}$ and covariance matrix

$$V(\rho, \sigma^2) = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

.

Thus the log-likelihood is

$$\ln L = -\frac{1}{2} n \ln \sigma^4 - \frac{1}{2} n \ln (1 - \rho^2) - \frac{1}{2} \sum_{t=1}^{n} Z_t^T V^{-1} Z_t, \qquad (2.2.9)$$

where $Z_t = (Z_{t1}, Z_{t2})^T = (X_{2t} - \eta_{2t}, X_{2t+1} - \eta_{2t+1})^T$.

Let $\upsilon = (\phi_1, ..., \phi_p, \sigma^2, \rho)^T$ denote the vector of parameters. Define

$$\mu_t(\upsilon) = \begin{bmatrix} -\frac{\partial \eta_{\mathbf{t}}}{\partial \phi} V^{-1} Z_t \\ -\frac{1}{\sigma^2} + \frac{1}{2} Z_t^T V^{-1} \frac{\partial V}{\partial \sigma^2} V^{-1} Z_t \\ -\frac{\rho}{1-\rho^2} + \frac{1}{2} Z_t^T V^{-1} \frac{\partial V}{\partial \rho} V^{-1} Z_t \end{bmatrix},$$

so that the maximum likelihood estimating function is $S_n(\upsilon) = \sum_{t=1}^{n} \mu_t(\upsilon)$.

Huggins and Basawa (2000) gave the following theorem.

Theorem 2.1. If X_t is a BARMA($p, 0$) process, $\mathrm{E}(\epsilon_t^{2(1+\delta)}) < \infty$ for some $\delta > 0$, then there exists a sequence $\hat{\upsilon}_n$ such that $S_n(\hat{\upsilon}_n) = 0$, $\hat{\upsilon}_n \xrightarrow{p} \upsilon$ and $n^{1/2}(\hat{\upsilon}_n - \upsilon) \xrightarrow{d} \mathrm{N}(0, I^{-1}(\upsilon))$.

The information matrix $I(\upsilon)$ used in Theorem 2.1 is given by

$$I(\upsilon) = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix},$$

where A is the $p \times p$ matrix with (i,k)th element $A_{ik} = \frac{2}{(1+\rho)} \sum_{j=0}^{\infty} \psi_j \psi_{j+|i-k|}$ and

$$B = \begin{bmatrix} \frac{1}{\sigma^4} & \frac{-\rho}{\sigma^2(1-\rho^2)} \\ \frac{-\rho}{\sigma^2(1-\rho^2)} & \frac{1+\rho^2}{(1-\rho^2)^2} \end{bmatrix}.$$

With regard to the maximum likelihood estimating function of BAR(1) model, the MLE of ϕ is given by

$$\hat{\phi}_{ML} = \frac{\sum_{t=1}^{n} X_t U_t}{\sum_{t=1}^{n} X_t^2}. \tag{2.2.10}$$

where $U_t = (X_{2t} + X_{2t+1})/2$.

From Theorem 2.1, we have

$$\sqrt{n}(\hat{\phi}_{ML} - \phi) \xrightarrow{d} N(0, \frac{1}{2}(1+\rho)(1-\phi^2)). \tag{2.2.11}$$

The explicit forms of MLEs of σ^2 and ρ are not easy to find, but we can get their asymptotic marginal distributions by Theorem 2.1, which are

$$\sqrt{n}(\hat{\sigma}^2_{ML} - \sigma^2) \xrightarrow{d} N(0, \sigma^4(1+\rho^2)) \tag{2.2.12}$$

and

$$\sqrt{n}(\hat{\rho}_{ML} - \rho) \xrightarrow{d} N(0, (1-\rho^2)^2). \tag{2.2.13}$$

Moreover, $(\hat{\sigma}^2_{ML}, \hat{\rho}_{ML})$ are asymptotically independent of $\hat{\phi}_{ML}$.

Other extensions of BAR models include Staudte (1992), which allowed for non-stationary generation means, and Huggins and Staudte (1994), where a variance component model was proposed to allow for additional sources of variation, namely measurement error and between tree variation.

Huggins and Marschner (1999) proposed a robust estimation procedure for the Cowan-Staudte BAR(1) model. Some conditions for the consistency and asymptotic normality of the robust estimators were given for an estimating function of a general type.

Huggins and Staudte (1994) considered the variance component model and gave asymptotic properties of the robust estimators for a large number of trees. Huggins (1996) derived the asymptotic properties of robust estimators when the data set arises from a single tree.

The derivation of asymptotic properties of estimators for the BARMA(p, q) models and the more complex covariance structure of Huggins and Basawa (1999) remains open.

2.3 Non-Gaussian Conditional Linear AR(1) Models

Let $\{Y_t\}, t = 0, 1, 2, ...$ denote a Markov process. Grunwald *et al.* (2000) have studied non-Gaussian Markov models for which the conditional mean $E(Y_t|Y_{t-1}) = m(Y_{t-1})$ is of the linear form

$$m(Y_{t-1}) = \phi Y_{t-1} + \lambda \tag{2.3.1}$$

Grunwald *et al.* (2000) refer to the model satisfying (2.3.1) as a first-order conditional linear autoregressive (CLAR(1)) model. More than 30 models which were summarized in Grunwald *et al.* (2000) belong to the CLAR(1) models.

To construct CLAR(1) models, several general methods can be used. The innovation method yields the usual autoregressive (AR) model $Y_t = \phi Y_{t-1} + Z_t$ where innovation Z_t has a specified distribution. Alternatively, one could specify a conditional distribution of Y_t given Y_{t-1} to be of a particular form, with mean $m(Y_{t-1})$ given by (2.3.1). A random coefficient model is an extension of the AR model where ϕ is replaced by ϕ_t, an iid sequence of random coefficients such that $E\phi_t = \phi$ and $\{\phi_t\}$ is independent of $\{Z_t\}$. A thinning model is of the form $Y_t = \phi * Y_{t-1} + \lambda$ and the thinning operation denoted by $*$ is defined as

$$\phi * X = \sum_{i=1}^{N(X)} W_i$$

where $N(x)$ is an integer valued random variable and $\{W_i\}$ is a sequence of iid random variables, independent of $N(x)$, such that $E(N(x)W_i|X = x) = \phi x$. Finally, random coefficients combined with thinning can be used to construct CLAR(1) models.

Under mild assumptions, Grunwald *et al.* (2000) derive the stationary mean and stationary variance, using the convergence of geometric series. Furthermore, sufficient but not necessary conditions for the ergodicity of the Markov process $\{Y_t\}$ are given.

The exponentially decaying autocorrelation function (ACF), $\rho_k = \text{corr}(Y_t, Y_{t-k}) = \phi^k \quad (k = 1, 2, ...)$, appears in many special models. Grunwald *et al.* (2000) show that under mild conditions the exponentially decaying ACF is implied by the CLAR(1) model and holds very generally. Thus, the exponentially decaying ACF can be used as a model diagnostic for CLAR(1) structure. Some data sets were analyzed in Grunwald *et al.* (2000) via an approach developed by Tsay (1992) based on bootstrap samples.

2.4 Estimating Functions and Quasilikelihood Estimation

Let $g(x,\theta)$ be a real valued function of the data x and unknown parameter θ. Then $g(x,\theta) = 0$ is referred to as an estimating equation, while $g(x,\theta)$ itself is termed an estimating function.

Godambe (1960) derived the Cramer-Rao type inequality

$$Var(\frac{g(x,\theta)}{E(\frac{\partial g(x,\theta)}{\partial\theta})}) \geq \frac{1}{i(\theta)} \qquad \forall \ \theta,$$

where $g(x,\theta)$ is any unbiased estimating function, i.e., $Eg(x,\theta) = 0$ and $i(\theta)$ is the Fisher's information. The optimal estimating function, in the sense of minimizing $Var(\frac{g(x,\theta)}{E(\frac{\partial g(x,\theta)}{\partial\theta})})$, is $g^* = \frac{\partial \log f(x,\theta)}{\partial\theta}$, the likelihood score function.

When the underlying distribution is unknown, the optimal estimating function (i.e., the likelihood score function), is not known. However, by restricting attention to an appropriate subclass $\mathcal{G}$ of the class of unbiased estimating functions, an optimal estimating function within this subclass can be obtained. Godambe (1985) derived such optimal estimating functions which depend only on the conditional means and variances. These estimating functions are known as quasilikelihood score functions.

Consider a score function $S_n(\theta) = \sum g_k(x_{(k)},\theta)h_{k-1}(x_{(k-1)},\theta)$ where g_k is known, unbiased and h_k is unknown. Here, $x(t)$ denotes $(x_t, x_{t-1}, ..., x_1)$. For the Godambe criterion function $Var(\frac{S_n(\theta)}{ES_n'(\theta)})$ to be minimized, the optimal score function is given by

$$S_n^*(\theta) = \sum g_k(x_{(k)},\theta)E_{k-1}g_k^{'}(E_{k-1}g_k^2)^{-1} \tag{2.4.1}$$

where E_{k-1} denotes conditional expectation given $x_{(k-1)}$ and $g_k^{'} = \frac{\partial g_k}{\partial\theta}$.

Under the multiparameter context, similar optimality criterion can be defined. Let $\{X_t, t \leq T\}$ be a sample of data whose distribution depends on unknown parameter θ of p dimensions. If $\mathcal{G}$ is the class of unbiased, square integrable estimating functions $G_T(X_T, \theta)$ and $\mathcal{H}$ is a subclass of $\mathcal{G}$, then G_T^* is said to be optimal within $\mathcal{H}$ if

$$E(\dot{G}_T^*)'(EG_T^*G_T^{*'})^{-1}E(\dot{G}_T^*) - E(\dot{G}_T)'(EG_TG_T')^{-1}E(\dot{G}_T)$$

is non-negative definite for all $G_T \in \mathcal{H}$, where $\dot{G}_T = ((E\frac{\partial G_{T,i(\theta)}}{\partial \theta_j})), \quad i, j = 1, ..., p$. See Heyde (1997).

If $\hat{\theta}_n$ is a consistent solution of an estimating equation, the asymptotic normality of $\hat{\theta}_n$ can be easily established in the context of independent observations. When the data are dependent, consistency and asymptotic normality are usually derived via martingale limit theory (Hall and Heyde (1980)). See, for instance, Heyde (1997). We shall use the quasilikelihood method to estimate the parameters of non-Gaussian bifurcating models.

Chapter 3

Least Squares Estimation for Bifurcating Autoregressive Processes[1]

[1]J. Zhou and I. V. Basawa. Submitted to *Statistics and Probability Letters*. 2/15/2004.

Abstract

Bifurcating autoregressive processes are used to model each line of descent in a binary tree as a standard AR(p) process, allowing for correlations between nodes which share the same parent. Limit distributions of the least squares estimators of the model parameters for a pth-order bifurcating autoregressive process (BAR(p)) are derived. An application to bifurcating integer-valued autoregression is given. A Poisson bifurcating model is introduced.

Keywords: Cell Lineage Data; Tree-indexed Time Series; Bifurcating Autoregression; Least Squares Estimators; Limit Distributions; Integer-valued Autoregression.

3.1 Introduction

Bifurcating autoregressive models were introduced by Cowan and Staudte (1986) for cell lineage data where each individual in one generation gives rise to two offspring in the next generation. The Cowan-Staudte model views each line of descent as a first-order autoregressive (AR(1)) process with the added complication that the observations on the two sister cells who share the same parent are allowed to be correlated. Staudte et al. (1996) studied data sets in which the observed correlations between cousin cells were significant, thus necessitating higher order models. Huggins and Basawa (1999) proposed bifurcating ARMA(p, q) models to accommodate this extended dependence in the family tree. Huggins and Basawa (2000) discussed maximum likelihood estimation for a Gaussian bifurcating AR(p) process and established the consistency and asymptotic normality of the maximum likelihood estimators of the model parameters. Recently, Basawa and Zhou (2004) introduced non-Gaussian bifurcating autoregressive models and studied some preliminary estimation problems. Zhou and Basawa (2003) have discussed maximum likelihood estimation for an exponential bifurcating AR(1) process. In this paper, we consider the asymptotic properties of the least squares estimators of parameters in a bifurcating AR(p) (BAR(p)) process.

The rest of the paper is organized as follows. The BAR(p) model and the least squares estimators of the model parameters are presented in Section 2. The limit distributions of the least squares estimators are derived in Section 3. Section 4 is concerned with an application to a bifurcating integer-valued AR(1) process. A Poisson bifurcating model is introduced in Section 5.

3.2 Least Squares Estimation for BAR(p) Processes

The pth order bifurcating autoregressive process (BAR(p)) is defined by the equation

$$X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \phi_2 X_{[\frac{t}{4}]} + \cdots + \phi_p X_{[\frac{t}{2^p}]} + \epsilon_t, \tag{3.2.1}$$

where $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of independent identically distributed (i.i.d.) bivariate random variables with $E(\epsilon_{2t}) = E(\epsilon_{2t+1}) = 0$, $Var(\epsilon_{2t}) = Var(\epsilon_{2t+1}) = \sigma^2$, and $Corr(\epsilon_{2t}, \epsilon_{2t+1}) = \rho$. The notation $[u]$ denotes the largest integer less than or equal to u. As in Huggins and Basawa (1999), the bifurcating operator b is defined by

$$b^r u_t = \begin{cases} u_{[\frac{t}{2^r}]^*}, & \text{if} \quad t > 0 \\ u_{t-r}, & \text{if} \quad t < 0 \end{cases}$$

where $[\frac{t}{2^r}]^* = [\frac{t}{2^r}]$ if $(\frac{t}{2^r}) \geq 1$, and $[\frac{t}{2^r}]^* = [\log_2(\frac{t}{2^r})] + 1$, if $(\frac{t}{2^r}) < 1$. This notation implies that the descendants of the initial cell are labeled according to their position in the binary tree and the ancestors of the initial cell are labeled $0, -1, -2, \ldots$. The BAR(p) process in (6.1.1) can then be represented as

$$\phi(b) X_t = \epsilon_t + \phi_0, \tag{3.2.2}$$

where $\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$. We assume that the roots of $\phi(z) = 0$ are greater than 1 in absolute value, so that we can write

$$X_t = \sum_{j=0}^{\infty} (\epsilon_{[\frac{t}{2^j}]^*} + \phi_0) \psi_j \tag{3.2.3}$$

where $\{\psi_j\}$ are the coefficients of z^j in the expansion of $\phi^{-1}(z)$. Moreover, $\sum\limits_{j=0}^{\infty} |\psi_j| < \infty$. The coefficients ψ_j can be determined recursively as in Huggins and Basawa (1999). The autocovariances $Cov(X_t, X_s)$ are determined as discussed in Huggins and Basawa (1999).

In particular, it is seen that

$$\begin{aligned} E(X_t) &= \mu = \phi_0 \sum_{j=0}^{\infty} \psi_j = \phi_0 (1 - \sum_{i=1}^{p} \phi_i)^{-1}, & (3.2.4) \\ Var(X_t) &= \gamma(0) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2, \quad \text{and} \\ Cov(X_t, X_{[\frac{t}{2^k}]^*}) &= \gamma(k) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+k}, \quad k \geq 0. & (3.2.5) \end{aligned}$$

Huggins and Basawa (2000) have discussed the consistency and asymptotic normality of the maximum likelihood estimators of the parameters in a BAR(p) process assuming Gaussian errors. Here, we consider the asymptotic properties of the least squares estimators of $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$, and σ^2 without imposing any specific distributional assumption on $\{\epsilon_t\}$. Let $Y_t = (1, X_{[\frac{t}{2}]}, \ldots, X_{[\frac{t}{2^p}]})$, $t \geq 2^p$. Then the least squares (LS) estimator $\hat{\phi}$ of ϕ based on the observations $\{X_t, t = 2^p, 2^p + 1, \ldots, n\}$ is seen to be

$$\hat{\phi} = (\sum_{t=2^p}^{n} Y_t Y_t')^{-1} \sum_{t=2^p}^{n} Y_t X_t. \tag{3.2.6}$$

Define

$$\hat{\sigma}^2 = \frac{1}{(n - 2^P - p)} \sum_{t=2^p}^{n} (X_t - Y_t' \hat{\phi})^2. \tag{3.2.7}$$

We will derive the limit distributions of $\hat{\phi}$ and $\hat{\sigma}^2$ in the next section. A consistent estimator of ρ is given by

$$\hat{\rho} = \hat{\sigma}^{-2} \Sigma (X_{2t} - Y_{2t}' \hat{\phi})(X_{2t+1} - Y_{2t+1}' \hat{\phi}).$$

3.3 Limit Distributions

Consider the following conditions:

(C.1) All the roots of $\phi(z) = 0$ are greater than 1 in absolute value.

(C.2) $E(\epsilon_t^4) < \infty$, for all t.

Lemma 3.1. Under (C.1), we have, as $n \to \infty$,

(i) $\frac{1}{n} \sum_{t=1}^{n} X_t \xrightarrow{p} \mu$

(ii) $\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)^2 \xrightarrow{p} \gamma(0)$

(iii) $\frac{1}{n}\sum_{t=1}^{n}(X_t-\mu)(X_{[\frac{t}{2^k}]^*}-\mu) \xrightarrow{p} \gamma(k)$, for $k \geq 0$, where μ and $\gamma(k)$ are defined in (3.2.4) and (3.2.5), respectively.

Proof: Note that $\{\epsilon_{[\frac{t}{2^j}]^*}\}$, $j = 0,1,2,\ldots$, are i.i.d. random variables with mean 0 and variance σ^2. The results then follow, via (3.2.3), as shown in Huggins and Basawa (2000). Also, see Brockwell and Davis (1987). □

Define $Z_t = (1, X_t, X_{[\frac{t}{2}]}, \ldots, X_{[\frac{t}{2^{p-1}}]})'$, and let $m = \frac{n-1}{2}$ = the number of triplets (X_t, X_{2t}, X_{2t+1}) observed. We then have

Lemma 3.2. Under (C.1)

$$\frac{1}{m}\sum_{t=2^{p-1}}^{m} Z_t Z_t' \xrightarrow{p} A, \quad \text{as} \quad m \to \infty, \tag{3.3.1}$$

where A is a $(p+1)\times(p+1)$ matrix defined by

$$A = \begin{pmatrix} 1 & \mu & \mu & \ldots & \mu \\ \mu & a(0) & a(1) & \ldots & a(p-1) \\ \mu & a(1) & a(0) & \ldots & a(p-2) \\ \vdots & & & & \\ \mu & a(p-1) & a(p-2) & \ldots & a(0) \end{pmatrix}, \tag{3.3.2}$$

with μ defined in (3.2.4), $a(k) = \mu^2 + \gamma(k)$, and $\gamma(k)$ given by (3.2.5).

Proof: The result follows from Lemma 3.1 after noting that

$$\Sigma Z_t Z_t' = \begin{pmatrix} m & \Sigma X_t & \Sigma X_{[\frac{t}{2}]} & \ldots & \Sigma X_{[\frac{t}{2^{p-1}}]} \\ \Sigma X_t & \Sigma X_t^2 & \Sigma X_t X_{[\frac{t}{2}]} & \ldots & \Sigma X_t X_{[\frac{t}{2^{p-1}}]} \\ \Sigma X_{[\frac{t}{2}]} & \Sigma X_{[\frac{t}{2}]} X_t & \Sigma X^2_{[\frac{t}{2}]} & \ldots & \Sigma X_{[\frac{t}{2}]} X_{[\frac{t}{2^{p-1}}]} \\ \vdots & & & & \\ \Sigma X_{[\frac{t}{2^{p-1}}]} & \Sigma X_{[\frac{t}{2^{p-1}}]} X_t & \Sigma X_{[\frac{t}{2^{p-1}}]} X_{[\frac{t}{2}]} & \ldots & \Sigma X^2_{[\frac{t}{2^{p-1}}]} \end{pmatrix}.$$

□

The following version of the martingale central limit theorem will be used in the derivation of the limit distribution of the least-squares estimator.

Lemma 3.3. Let $\{Y_t\}$, $t = 1, 2, \ldots$, be a sequence of zero-mean vector martingale differences satisfying the following conditions:

(a) $E(Y_t Y_t') = \Omega_t$, a positive definite matrix, and $\frac{1}{n}\sum_{t=1}^{n} \Omega_t \to \Omega$, a positive definite matrix.

(b) $E(Y_{it} Y_{jt} Y_{lt} Y_{mt}) < \infty$ for all t, and all i, j, l, m, where Y_{rt} denotes the rth element of the vector Y_t.

(c) $\frac{1}{n}\sum_{t=1}^{n} Y_t Y_t' \xrightarrow{d} \Omega$.

Then, $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Y_t \xrightarrow{d} N(0, \Omega)$.

Proof: See, for instance, Proposition 7.9 in Hamilton (1994). □

Lemma 3.4. Under (C.1) and (C.2), as $m \to \infty$,

$$\frac{1}{\sqrt{m}} \sum_{t=2^{p-1}}^{m} Z_t V_t \xrightarrow{d} N(0, \sigma^2(1+\rho)A),$$

where A is defined in Lemma 3.2, and $V_t = \frac{1}{\sqrt{2}}(\epsilon_{2t} + \epsilon_{2t+1})$.

Proof: Let $\mathcal{F}_t = \sigma\{\epsilon_j : j \leq 2t + 1\}$. It can be verified that $\sum_{t=2^{p-1}}^{m} Z_t V_t$ is a zero-mean martingale with respect to $\mathcal{F}_t$. In order to verify the central limit theorem for martingales, we now check the conditions of Lemma 3.3.

(a) From (3.2.5), we have $E(Z_t Z_t' V_t^2) = E(Z_t Z_t')E(V_t^2) = A\sigma^2(1+\rho)$, where A is defined in Lemma 3.2. It can be verified that A is a positive definite matrix. Hence, condition (a) is satisfied.

(b) $E(V_t^4 Z_{it} Z_{jt} Z_{kt} Z_{lt}) < \infty$, for all i, j, k, l, where Z_{rt} is the rth element of the vector Z_t. Condition (b) holds from Proposition 7.10 of Hamilton (1994) under the assumption (C.2).

(c) $\frac{1}{m}\sum_{t=2^{p-1}}^{m} V_t^2 Z_t Z_t' \xrightarrow{p} \sigma^2(1+\rho)A$. In order to verify (c), consider

$$\begin{aligned}\frac{1}{m}\sum_{t=2^{p-1}}^{m} V_t^2 Z_t Z_t' &= \frac{1}{m}\sum_{t=2^{p-1}}^{m}[V_t^2-\sigma^2(1+\rho)]Z_t Z_t' + \sigma^2(1+\rho)\frac{1}{m}\sum_{t=2^{p-1}}^{m} Z_t Z_t' \\ &= U_{1m}+U_{2m}, \quad \text{say.}\end{aligned}$$

We have $U_{1m} = \frac{1}{m}\sum_{t=2^{p-1}}^{m} W_t$, where $W_t = (V_t^2-\sigma^2(1+\rho))Z_t Z_t'$. For any $(p+1)$-vector λ, we have $\lambda' U_{1m}\lambda = \frac{1}{m}\sum_{t=2^{p-1}}^{m}\lambda' W_t \lambda$. It is easily verified that $E(\lambda' W_t\lambda|\mathcal{F}_{t-1}) = 0$, and $\{\lambda' W_t \lambda\}$ is a stationary martingale difference sequence with $E(\lambda' W_t\lambda)^2 < \infty$ (see (b) above). Consequently, by the law of large numbers for martingales (see Hall and Heyde (1980)) we conclude that $\lambda' U_{1m}\lambda \xrightarrow{p} 0$, and hence $U_{1m} \xrightarrow{p} 0$.

From Lemma 3.2, $\frac{1}{m}\Sigma Z_t Z_t' \xrightarrow{p} A$, and hence $U_{2m} \xrightarrow{p} \sigma^2(1+\rho)A$. Consequently, condition (c) is verified. The desired limit in Lemma 3.4 then follows from Lemma 3.3. □

The limit distribution of $\hat{\phi}$ is given below.

Theorem 3.1. Under (C.1) and (C.2), we have

$$\sqrt{n}(\hat{\phi}-\phi) \xrightarrow{d} N(0, \sigma^2(1+\rho)A^{-1}), \quad \text{as} \quad n\to\infty.$$

Proof: We have

$$\begin{aligned}\sqrt{n}(\hat{\phi}-\phi) &= (\frac{1}{n}\sum_{t=2^p}^{n} Y_t Y_t')^{-1}[\frac{1}{\sqrt{n}}\sum_{t=2^p}^{n} Y_t\epsilon_t] \\ &= (\frac{1}{m}\sum_{t=2^{p-1}}^{m} Z_t Z_t')^{-1}[\frac{1}{\sqrt{m}}\sum_{t=2^{p-1}}^{m} Z_t V_t] + o_p(1)\end{aligned}$$

The result then follows from Lemmas 3.2, 3.4 and Slutsky's theorem. □

The next theorem gives the limit distribution of $\hat{\sigma}^2$.

Theorem 3.2. Under (C.1) and (C.2), we have, as $n\to\infty$,

$$\sqrt{n}(\hat{\sigma}^2-\sigma^2) \xrightarrow{d} N(0, u_4+u_{22}-2\sigma^4),$$

where $u_4 = E(\epsilon_t^4)$ and $u_{22} = E(\epsilon_{2t}^2\epsilon_{2t+1}^2)$.

Proof: We have

$$\begin{aligned}
\sum_{t=2^p}^{n}(X_t - Y_t'\hat{\phi})^2 &= \sum_{t=2^p}^{n}(X_t - Y_t'\phi - Y_t'(\hat{\phi}-\phi))^2 \\
&= \sum_{t=2^p}^{n}\epsilon_t^2 - 2(\hat{\phi}-\phi)'\sum_{t=2^p}^{n}Y_t\epsilon_t + (\hat{\phi}-\phi)'(\sum_{t=2^p}^{n}Y_tY_t')(\hat{\phi}-\phi) \\
&= \sum_{t=2^p}^{n}\epsilon_t^2 - (\hat{\phi}-\phi)'(\sum_{t=2^p}^{n}Y_tY_t')(\hat{\phi}-\phi).
\end{aligned}$$

Hence,

$$\begin{aligned}
\sqrt{n}(\hat{\sigma}^2 - \sigma^2) &\simeq \frac{1}{\sqrt{n}}\sum_{t=2^p}^{n}(\epsilon_t^2 - \sigma^2) - \sqrt{n}(\hat{\phi}-\phi)'(\frac{1}{n}\sum_{t=2^p}^{n}Y_tY_t')(\hat{\phi}-\phi) \\
&= W_{1n} + W_{2n}, \quad \text{say.}
\end{aligned}$$

Note that $W_{2n} \xrightarrow{p} 0$, since $\frac{1}{n}\sum\limits_{t=2^p}^{n} Y_tY_t' \xrightarrow{p} A$, and $\sqrt{n}(\hat{\phi}-\phi) = O_p(1)$.

We have

$$\begin{aligned}
W_{1n} = \frac{1}{\sqrt{n}}\sum_{t=2^p}^{n}(\epsilon_t^2 - \sigma^2) &\simeq \frac{1}{\sqrt{m}}\sum_{t=2^{p-1}}^{m}(\frac{\epsilon_{2t}^2 + \epsilon_{2t+1}^2 - 2\sigma^2}{\sqrt{2}}) \\
&\xrightarrow{d} N(0, u_4 + u_{22} - 2\sigma^4).
\end{aligned}$$

This completes the proof. □

The limit distribution of $\hat{\rho}$ can be obtained in a similar manner which is omitted. We now illustrate Theorem 3.1 by two examples.

Example 1 BAR(1) Model

Consider the model

$$X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \epsilon_t, \quad \phi_0 \neq 0, \quad \text{and} \quad |\phi_1| < 1.$$

The least squares estimators are given by

$$\begin{aligned}
\hat{\phi}_1 &= \frac{\sum\limits_{t=1}^{m} U_t(X_t - \bar{X})}{\sum\limits_{t=1}^{m}(X_t - \bar{X})^2}, \quad \text{where} \quad U_t = \frac{X_{2t} + X_{2t+1}}{2}, \quad \text{and} \quad \bar{X} = \frac{1}{m}\sum_{t=1}^{m} X_t, \\
\hat{\phi}_0 &= \bar{U} - \hat{\phi}_1\bar{X}, \quad \text{where} \quad \bar{U} = \frac{1}{m}\sum_{t=1}^{m} U_t.
\end{aligned}$$

From Theorem 3.1, we have

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2(1+\rho)A^{-1}),$$

where

$$A = \begin{pmatrix} 1 & \phi_0/(1-\phi_1) \\ \phi_0/(1-\phi_1) & \frac{\sigma^2}{1-\phi_1^2} + (\frac{\phi_0}{1-\phi_1})^2 \end{pmatrix}.$$

If $\phi_0 = 0$, we have $\hat{\phi}_1 = \sum_{t=1}^{m} U_t X_t / \sum_{t=1}^{m} X_t^2$, and $A = EX_t^2 = \frac{\sigma^2}{1-\phi_1^2}$. Consequently, we have, for $\phi_0 = 0$,

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \xrightarrow{d} N(0, (1+\rho)(1-\phi_1^2)).$$

Example 2. BAR(2) Model

For the model

$$X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \phi_2 X_{[\frac{t}{4}]} + \epsilon_t,$$

we have under (C.1) and (C.2),

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, \sigma^2(1+\rho)A^{-1}),$$

where

$$A = \begin{pmatrix} 1 & \mu & \mu \\ \mu & a(0) & a(1) \\ \mu & a(1) & a(0) \end{pmatrix}.$$

In particular, when $\phi_0 = 0$, and $\phi = (\phi_1, \phi_2)'$, we have

$$\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{d} N(0, (1+\rho)B),$$

where

$$B = \begin{pmatrix} 1-\phi_2^2 & -\phi_1(1-\phi_2) \\ -\phi_1(1-\phi_2) & 1-\phi_2^2 \end{pmatrix}.$$

Mean-Centered Process

We now consider the mean-centered version of the model in (6.1.1). Model (6.1.1) can be rewritten as

$$X_t - \mu = \phi_1(X_{[\frac{t}{2}]} - \mu) + \phi_2(X_{[\frac{t}{2^2}]} - \mu) + \cdots + \phi_p(X_{[\frac{t}{2^p}]} - \mu) + \epsilon_t, \tag{3.3.3}$$

where $\mu = \phi_0(1 - \sum_{i=1}^{p} \phi_i)^{-1}$.

Define

$$\hat{\mu} = \hat{\phi}_0(1 - \sum_{i=1}^{p} \hat{\phi}_i)^{-1}. \tag{3.3.4}$$

Let $\beta = (\mu, \phi_1, \phi_2, \ldots, \phi_p)'$. We then have

$$(\hat{\beta} - \beta) = D(\hat{\phi} - \phi) + o_p(1), \tag{3.3.5}$$

where $\phi = (\phi_0, \phi_1, \ldots, \phi_p)'$,

$$D = \begin{pmatrix} c & c\mu & c\mu & \ldots & c\mu \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & & & & \\ 0 & 0 & 0 & \ldots & 1 \end{pmatrix}, \tag{3.3.6}$$

and $c = (1 - \sum_{i=1}^{p} \phi_i)^{-1}$. The limit distribution of $\hat{\beta}$ is given next.

Theorem 3.3. Under (C.1) and (C.2), we have

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{p} N(0, \sigma^2(1+\rho)DA^{-1}D'), \quad \text{as} \quad n \to \infty,$$

where A is defined in (3.3.2), and D in (3.3.6).

Proof: The result follows from Theorem 3.1 and (3.3.5). □

Remark: It is easily verified that

$$DA^{-1}D' = \begin{pmatrix} c^2 & \mathbf{0} \\ \mathbf{0} & \Gamma^{-1} \end{pmatrix} \tag{3.3.7}$$

where

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(p-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(p-2) \\ \vdots & & & \\ \gamma(p-1) & \gamma(p-2) & \dots & \gamma(0) \end{pmatrix}. \tag{3.3.8}$$

In order to check (3.3.7), first note that

$$A = \begin{pmatrix} 1 & \mu \boldsymbol{u}' \\ \mu \boldsymbol{u} & \Gamma + \mu^2 \boldsymbol{u}\boldsymbol{u}' \end{pmatrix} = P'\Sigma P,$$

where $\boldsymbol{u} = (1, 1, \dots, 1)'$, is a $(p \times 1)$ unit vector,

$$P = \begin{pmatrix} 1 & \mu \boldsymbol{u}' \\ \mathbf{0} & I \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \Gamma \end{pmatrix}.$$

Also,

$$D = \begin{pmatrix} c & c\mu \boldsymbol{u}' \\ \mathbf{0} & I \end{pmatrix} = QP,$$

where

$$Q = \begin{pmatrix} c & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}.$$

We thus have

$$\begin{aligned} DA^{-1}D' &= (QP)(P'\Sigma P)^{-1}(QP)' \\ &= Q\Sigma^{-1}Q' = \begin{pmatrix} c^2 & \mathbf{0} \\ \mathbf{0} & \Gamma^{-1} \end{pmatrix}. \end{aligned} \tag{3.3.9}$$

Hence, the result in (3.3.7) is verified.

It then follows that

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, c^2\sigma^2(1+\rho)),$$

and

$$\sqrt{n}(\hat{\phi}^* - \phi^*) \xrightarrow{d} N(0, \sigma^2(1+\rho)\Gamma^{-1}),$$

where $\phi^* = (\phi_1, \phi_2, \ldots, \phi_p)'$. Moreover, $\hat{\mu}$ is asymptotically independent of $\hat{\phi}^*$. It can further be noted that

$$A^{-1} = P^{-1}\Sigma^{-1}(P^{-1})' = \begin{pmatrix} 1 + \mu^2 \boldsymbol{u}'\Gamma^{-1}\boldsymbol{u} & -\mu\boldsymbol{u}'\Gamma^{-1} \\ -\mu\Gamma^{-1}\boldsymbol{u} & \Gamma^{-1} \end{pmatrix}. \tag{3.3.10}$$

Example 1 (Continued)

The centered version of the BAR(1) model is

$$X_t - \mu = \phi_1(X_{[\frac{t}{2}]} - \mu) + \epsilon_t, \quad \text{where} \quad \mu = \phi_0(1 - \phi_1)^{-1}.$$

It follows from Theorem 3.3 that

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2(1+\rho)(1-\phi_1)^{-2}),$$

and

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \xrightarrow{d} N(0, (1+\rho)(1-\phi_1^2)).$$

Moreover, $\hat{\mu}$ is asymptotically independent of $\hat{\phi}_1$.

3.4 Integer-Valued Bifurcating Autoregressive Model

In this section, we introduce an extension of the first-order integer-valued autoregression (INAR(1)) (see Al-Osh and Alzaid (1987)) to a binary tree-indexed process and discuss least squares estimation for the model parameters. Consider the process $\{X_t\}$ satisfying the relation:

$$X_t = \phi_1 o X_{[\frac{t}{2}]} + \epsilon_t, \quad 0 < \phi_1 < 1, \tag{3.4.1}$$

where $\phi_1 o X_{[\frac{t}{2}]}$ denotes the binomial thinning operation defined by

$$\phi_1 o X_{[\frac{t}{2}]} = \sum_{i=1}^{X_{[\frac{t}{2}]}} Y_i, \tag{3.4.2}$$

where $\{Y_i\}$, $i = 1, 2, \ldots,$ are i.i.d. Bernoulli random variables with $P(Y_i = 1) = \phi_1$ and $P(Y_i = 0) = 1 - \phi_1$, $0 < \phi_1 < 1$. The error process $\{\epsilon_t\}$ is characterized by the fact that $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$, $t = 1, 2, \ldots$, are i.i.d. integer-valued bivariate random variables with $E(\epsilon_{2t}) = E(\epsilon_{2t+1}) = \phi_0$, $Var(\epsilon_{2t}) = Var(\epsilon_{2t+1}) = \sigma^2$ and $Corr(\epsilon_{2t}, \epsilon_{2t+1}) = \rho$. It is readily verified from (3.4.1) that

$$E(X_t | X_{[\frac{t}{2}]}) = \phi_0 + \phi_1 X_{[\frac{t}{2}]}, \quad \phi_0 > 0 \tag{3.4.3}$$

and

$$Var(X_t | X_{[\frac{t}{2}]}) = \phi_1(1-\phi_1) X_{[\frac{t}{2}]} + \sigma^2. \tag{3.4.4}$$

The conditional least squares (CLS) estimators of ϕ_0 and ϕ_1 are obtained by minimizing $\sum_{t=2}^{n}(X_t - \phi_0 - \phi_1 X_{[\frac{t}{2}]})^2$ with respect to ϕ_0 and ϕ_1, and these are the same as the LS estimators $\hat{\phi}_0$ and $\hat{\phi}_1$ for the BAR(1) model given in Example 1 in Section 3. It can be verified from (3.4.3) and (3.4.4) that the unconditional stationary moments are given by

$$\mu = E(X_t) = \phi_0(1-\phi_1)^{-1}, \tag{3.4.5}$$

and

$$\gamma(0) = Var(X_t) = (\mu\phi_1(1-\phi_1) + \sigma^2)(1-\phi_1^2)^{-1}. \tag{3.4.6}$$

Using basically similar arguments as those for the centered BAR(1) example at the end of Section 3, one can verify that

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{d} N(0, \sigma^2(1+\rho)(1-\phi_1)^{-2}),$$

and

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \xrightarrow{d} N(0, \sigma^2(1+\rho)\gamma^{-1}(0)),$$

where $\gamma(0)$ is given by (3.4.6). Moreover, $\hat{\mu}$ is asymptotically independent of $\hat{\phi}_1$. Even though some of the time series asymptotics used in the previous section are not directly applicable for the model in (3.4.1), one can use the fact that $\{X_t\}$ is an ergodic Markov chain (see Grunwald et al. (2000)) and standard Markov chain asymptotics can then be used to establish the above results. The details are omitted.

3.5 Bifurcating Poisson Model

As an example of the bifurcating INAR(1) model of Section 4, we present here a Poisson bifurcating model, and study some of its properties. Consider the model in (3.4.1) with $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ having a bivariate Poisson distribution defined by

$$P(\epsilon_{2t} = y_1, \epsilon_{2t+1} = y_2) = e^{-(\theta_1+\theta_2+\theta_3)} \sum_{i=0}^{y_1 \wedge y_2} \frac{\theta_1^{y_1-i}\theta_2^{y_2-i}\theta_3^i}{(y_1-i)!(y_2-i)!i!}, \tag{3.5.1}$$

where $y_1 \wedge y_2 = \min(y_1, y_2)$, $\theta_i > 0$, $i = 1, 2, 3$, and $y_j = 0, 1, 2, \ldots, (j = 1, 2)$. The marginal distributions of ϵ_{2t} and ϵ_{2t+1} are then Poisson with means $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$ respectively, and $Cov(\epsilon_{2t}, \epsilon_{2t+1}) = \theta_3$. The joint moment generating function of $(\epsilon_{2t}, \epsilon_{2t+1})$ is seen to be

$$M(t_1, t_2) = \exp[\theta_3(e^{t_1+t_2} - 1) + \theta_1(e^{t_1} - 1) + \theta_2(e^{t_2} - 1)]. \tag{3.5.2}$$

See, for instance, Johnson et al. (1997). We now choose the following parameterization:

$$\theta_1 = \theta_2 = (1-\rho)\phi_0, \quad \text{and} \quad \theta_3 = \rho\phi_0, \quad \text{with} \quad 0 < \rho < 1, \ \phi_0 > 0.$$

We then get $E(\epsilon_{2t}) = E(\epsilon_{2t+1}) = Var(\epsilon_{2t}) = Var(\epsilon_{2t+1}) = \phi_0$, and $Corr(\epsilon_{2t}, \epsilon_{2t+1}) = \rho$.

The conditional distribution of X_t given $X_{[\frac{t}{2}]}$ is obtained from (3.4.1) and (3.5.1), and it is seen to be

$$p(x_t|x_{[\frac{t}{2}]}) = e^{-\phi_0} \sum_{i=0}^{x_t \wedge x_{[\frac{t}{2}]}} \frac{\phi_0^{(x_t-i)}}{(x_t-i)!} \binom{x_{[\frac{t}{2}]}}{i} \phi_1^i (1-\phi_1)^{(x_{[\frac{t}{2}]}-i)}. \tag{3.5.3}$$

We have, from (3.4.3) and (3.4.4),

$$E(X_t|X_{[\frac{t}{2}]}) = \phi_0 + \phi_1 X_{[\frac{t}{2}]},$$

and

$$Var(X_t|X_{[\frac{t}{2}]}) = \phi_1(1-\phi_1)X_{[\frac{t}{2}]} + \phi_0.$$

The conditional least squares estimators of ϕ_0 and ϕ_1 are then obtained as discussed in Section 4.

The likelihood function is given by

$$L_n(\phi_0, \phi_1, \rho) = p(x_1) \prod_{t=1}^{m} p(x_{2t}, x_{2t+1}|x_t),$$

where m is the total number of triplets (x_t, x_{2t}, x_{2t+1}) observed, and $p(x_{2t}, x_{2t+1}|x_t)$ is the conditional distribution of (X_{2t}, X_{2t+1}) given X_t. However, $p(x_{2t}, x_{2t+1}|x_t)$ does not have a simple form. The conditional moment generating function of (X_{2t}, X_{2t+1}) given X_t is given below.

Lemma 5.1. The conditional moment generating function of (X_{2t}, X_{2t+1}) given X_t is

$$M^{(t_1,t_2)}_{(X_{2t},X_{2t+1})|X_t} = [\phi_1 e^{t_1+t_2} + (1-\phi_1)]^{X_t} M_{(\epsilon_{2t},\epsilon_{2t+1})}(t_1, t_2),$$

where $M_{(\epsilon_{2t},\epsilon_{2t+1})}(t_1, t_2)$ is given by (3.5.2).

Proof: We have

$$\begin{aligned} & E[e^{t_1 X_{2t} + t_2 X_{2t+1}}|X_t] \\ = \; & E[e^{t_1 \sum_{i=1}^{X_t} Y_i + t_1 \epsilon_{2t} + t_2 \sum_{i=1}^{X_t} Y_i + t_2 \epsilon_{2t+1}}|X_t] \\ = \; & E[e^{(t_1+t_2)\sum_{i=1}^{X_t} Y_i}|X_t] E(e^{t_1\epsilon_{2t} + t_2\epsilon_{2t+1}}) \\ = \; & [\phi_1 e^{t_1+t_2} + (1-\phi_1)]^{X_t} M_{(\epsilon_{2t},\epsilon_{2t+1})}(t_1, t_2), \end{aligned}$$

since conditional on X_t, $\sum_{i=1}^{X_t} Y_i$ is a binomial random variable with parameters (X_t, ϕ_1). □

Next, we obtain the unconditional joint distribution of (X_{2t}, X_{2t+1}) for the model given by (3.4.1) and (3.5.1). This turns out to be a bivariate Poisson distribution.

Lemma 5.2. The joint distribution of (X_{2t}, X_{2t+1}) is a bivariate Poisson with $E(X_{2t}) = E(X_{2t+1}) = \frac{\phi_0}{1-\phi_1}$, and $Cov(X_{2t}, X_{2t+1}) = (\rho + \frac{\phi_1}{1-\phi_1})\phi_0$.

Proof: The joint moment generating function of (X_{2t}, X_{2t+1}) is given by

$$\begin{aligned} M_{(X_{2t},X_{2t+1})}(t_1, t_2) &= E[M_{(X_{2t},X_{2t+1})}(t_1, t_2)|X_t] \\ &= M_{(\epsilon_{2t},\epsilon_{2t+1})}(t_1, t_2) E[(\phi_1 e^{t_1+t_2} + (1-\phi_1)^{X_t}]. \end{aligned} \tag{3.5.4}$$

Next, note that the marginal distribution of X_t is Poisson with mean $\frac{\phi_0}{1-\phi_1}$. This is seen from representing X_t in (3.4.1) in terms of $\{\epsilon_{[\frac{t}{2^j}]}\}$, $j = 0, 1, \ldots$,

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{[\frac{t}{2^j}]},$$

and noting that $\{\epsilon_{[\frac{t}{2^j}]}\}$, $j = 0, 1, 2, \ldots$, is a sequence of i.i.d. Poisson random variables with mean ϕ_0. Consequently,

$$\begin{aligned} & E[(\phi_1 e^{t_1+t_2} + (1-\phi_1))^{X_t}] \\ = \; & E[e^{sX_t}], \quad \text{where} \quad s = \log(\phi_1 e^{t_1+t_2} + (1-\phi_1)) \\ = \; & \exp[\frac{\phi_0}{1-\phi_1}(e^s - 1)] = \exp[\frac{\phi_0}{1-\phi_1}(\phi_1 e^{t_1+t_2} - \phi_1)]. \end{aligned} \tag{3.5.5}$$

Substituting (3.5.5) in (3.5.4), and simplifying, we get the moment generating function of the bivariate Poisson distribution given in (3.5.2) with

$$\theta_1 = \theta_2 = (1-\rho)\phi_0, \quad \text{and} \quad \theta_3 = (\frac{\phi_1}{1-\phi_1} + \rho)\phi_0.$$

The result in the lemma then follows. □

3.6 References

[1] Basawa, I. V. and B. L. S. Prakasa Rao (1980). *Statistical Inference for Stochastic Processes*, Academic Press, London.

[2] Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Series Analysis* 8, 261-275.

[3] Basawa, I. V. and Zhou, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Prob.* 41A, 55-64.

[4] Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods.* Springer, New York.

[5] Cowan, R. and Staudte, R. G. (1986). The bifurcating autoregression model in cell lineage studies. *Biometrics* 42, 769-783.

[6] Grunwald, G. K., Hyndman, R. J., Tedesco, L. and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Aust. N.Z.J. Stat.* 42, 479-495.

[7] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[8] Hamilton, J. D. (1994). *Time Series Analysis.* Princeton Univ. Press, Princeton, N.J.

[9] Huggins, R. M. and Basawa, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Prob.* 36, 1225-1233.

[10] Huggins, R. M. and Basawa, I. V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N.Z.J. Stat.* 42, 423-432.

[11] Johnson, N. L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions.* Wiley, New York.

[12] Staudte, R. G., Zhang, J., Huggins, R. M. and Cowan, R. (1996). A re-examination of the cell lineage data of E. O. Powell. *Biometrics* 52, 1214-1222.

[13] Zhou, J. and Basawa, I. V. (2003). Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. Technical Report, University of Georgia.

Chapter 4

Maximum Likelihood Estimation for a First-Order Bifurcating Autoregressive Process with Exponential Errors[1]

[1] J. Zhou and I. V. Basawa. Submitted to *Journal of Time Series Analysis*. 12/15/2003.

Abstract

Exact and asymptotic distributions of the maximum likelihood estimator of the autoregressive parameter in a first-order bifurcating autoregressive process with exponential innovations are derived. The limit distributions for the stationary, critical and explosive cases are unified via a single pivot using a random normalization. The pivot is shown to be asymptotically exponential for all values of the autoregressive parameter.

Keywords: Bifurcating Autoregression; Exponential Innovations; Maximum Likelihood; Exact Distribution; Limit Distribution; Non-standard Asymptotics.

4.1 Introduction

Consider the first-order autoregressive process

$$X_t = \phi X_{t-1} + \epsilon_t \tag{4.1.1}$$

where $\{\epsilon_t\}$ is a sequence of independent exponential random errors with mean $\lambda > 0$, and $\phi \geq 0$. Nielsen and Shephard (2003) have derived the exact distribution of the maximum likelihood (ML) estimator $\hat{\phi}_n$ of ϕ, conditioning on some initial value X_0,

$$\hat{\phi}_n = \min_{1 \leq t \leq n} \left(\frac{X_t}{X_{t-1}}\right). \tag{4.1.2}$$

Davis and McCormick have studied the limit distribution of $\hat{\phi}_n$ when $0 \leq \phi < 1$ and $\{\epsilon_t\}$ has a more general class of distributions of which exponential distribution is a special case. Nielsen and Shephard (2003) have also derived the limit distribution of $\hat{\phi}_n$ for the exponential innovations. In particular, they have shown that, for $0 \leq \phi \leq 1$ (non-explosive cases),

$$c_n(\hat{\phi}_n - \phi) \xrightarrow{d} Exp(1) \tag{4.1.3}$$

where $Exp(1)$ denotes an exponential random variable with mean 1, and

$$c_n = \begin{cases} (1-\phi)^{-1} n, & \text{for} \quad 0 \leq \phi < 1 \\ n(n-1)/2, & \text{for} \quad \phi = 1. \end{cases}$$

They further show that

$$\phi^n(\hat{\phi}_n - \phi) = O_p(1), \quad \text{for} \quad \phi > 1. \tag{4.1.4}$$

Nielsen and Shephard (2003) also derive the limit distribution of the likelihood ratio statistic for all values of $\phi \geq 0$. In the derivation of the limit distribution of the likelihood ratio statistic, Nielsen and Shephard (2003) show that, for $\phi > 1$,

$$\frac{Z\phi^n}{\lambda(\phi - 1)}(\hat{\phi}_n - \phi) \xrightarrow{d} Exp(1), \tag{4.1.5}$$

where $Z = X_0 + \sum_{j=1}^{\infty} \phi^{-j}\epsilon_j$.

A careful reading of their proof of Theorem 3 reveals that

$$d_n^{-1}\sum_{t=1}^{n} X_{t-1} \xrightarrow{p} \begin{cases} \frac{\lambda}{1-\phi}, & \text{for} \quad 0 \leq \phi < 1 \\ \lambda, & \text{for} \quad \phi = 1, \end{cases} \tag{4.1.6}$$

where

$$d_n = \begin{cases} n, & 0 \leq \phi < 1 \\ \frac{n(n-1)}{2}, & \phi = 1. \end{cases}$$

We also have

$$\phi^{-n}\sum_{t=1}^{n} X_{t-1} \xrightarrow{\text{a.s.}} \frac{Z}{\phi - 1}, \quad \text{for} \quad \phi > 1. \tag{4.1.7}$$

From (4.1.3), (4.1.5), (4.1.6) and (4.1.7), it readily follows that

$$\lambda^{-1}(\sum_{t=1}^{n} X_{t-1})(\hat{\phi}_n - \phi) \xrightarrow{d} Exp(1), \quad \text{for all} \quad \phi \geq 0. \tag{4.1.8}$$

The pivot

$$T_n = \hat{\lambda}_n^{-1}(\sum_{t=1}^{n} X_{t-1})(\hat{\phi}_n - \phi), \tag{4.1.9}$$

where

$$\hat{\lambda}_n = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\phi}_n X_{t-1}), \tag{4.1.10}$$

can, therefore, be used for constructing asymptotic tests and confidence intervals for ϕ without the prior knowledge as to which of the three regions, viz., $\phi < 1$, $\phi = 1$, $\phi > 1$, the parameter ϕ belongs to. It is to be further noted that

$$2T_n \xrightarrow{d} \chi^2(2), \quad \text{for all} \quad \phi \geq 0, \tag{4.1.11}$$

which unifies all the three cases. The unifying limit result in (4.1.8), though not noted explicitly by Nielsen and Shephard, can nevertheless be deduced from a careful reading of their proof of Theorem 3.

The main purpose of this paper is to derive the exact and asymptotic distributions of the maximum likelihood estimator of ϕ for the first-order bifurcating autoregressive process defined by

$$X_t = \phi X_{[\frac{t}{2}]} + \epsilon_t, \quad t \geq 2, \phi \geq 0, \tag{4.1.12}$$

where $[u]$ denotes the largest integer $\leq u$, and $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$, $t \geq 1$, are independent bivariate exponential random variables. Bifurcating autoregressive processes are used to model data indexed by a binary tree where each individual at any node gives rise to two individuals. The model in (4.1.12) with Gaussian errors was orginally introduced by Cowan and Staudte (1986) in the context of modeling cell lineage data. See also, Huggins and Staudte (1994) and Huggins and Basawa (1999, 2000) for various extensions. Basawa and Zhou (2003) have discussed non-Gaussian bifurcating models with $0 \leq \phi < 1$.

The exact distribution of the maximum likelihood estimator of ϕ in (4.1.12) with bivariate exponential innovations, is derived in Section 2. Section 3 contains the derivation of the limit distributions of the ML estimator in the three cases (i) $\phi < 1$, (ii) $\phi = 1$ and (iii) $\phi > 1$. These results are unified via a single pivot using a random normalizing sequence in Section 4. Some simulation results on the comparison of the maximum likelihood and least squares estimators are reported in Section 5.

4.2 Exact Distribution of the Maximum Likelihood Estimator

Consider the model in (4.1.12) with $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$, $t = 1, 2, \ldots$, independent with bivariate exponential distribution defined by

$$P(\epsilon_{2t} > u_1,\ \epsilon_{2t+1} > u_2) = \exp[-\alpha_1 u_1 - \alpha_2 u_2 - \alpha_{12} \max(u_1, u_2)],\ u_1 \geq 0,\ u_2 \geq 0, \tag{4.2.1}$$

where α_1, α_2 and α_{12} are the model parameters satisfying $\alpha_1 > 0$, $\alpha_2 > 0$ and $\alpha_{12} \geq 0$. See Mardia (1970) and Kotz, et al (2000). The marginal distributions of ϵ_{2t} and ϵ_{2t+1} are exponential with means $(\alpha_1 + \alpha_{12})^{-1}$ and $(\alpha_2 + \alpha_{12})^{-1}$ respectively with correlation between ϵ_{2t} and ϵ_{2t+1} given by $\alpha_{12}(\alpha_1 + \alpha_2 + \alpha_{12})^{-1}$. We have chosen this particular form of the bivariate exponential distribution for its simplicity. It also happens to be the only bivariate exponential distribution which possesses the (bivariate) lack of memory property. We now choose the parameters as follows

$$\alpha_1 = \alpha_2 = \frac{1-\rho}{(1+\rho)\lambda}, \quad \text{and} \quad \alpha_{12} = \frac{2\rho}{(1+\rho)\lambda}, \tag{4.2.2}$$

where $\lambda > 0$ and $0 \leq \rho < 1$. With this parametrization, the marginal distributions of ϵ_{2t} and ϵ_{2t+1} are both exponential with mean λ and correlation ρ. Note also that when $\rho = 0$, the innovations $\{\epsilon_t\}$ in (1.12) will be independent and identically distributed exponential random variables with mean λ, which corresponds to the assumption in the AR(1) model (4.1.1).

In order to derive the likelihood function based on the sample $(X_1, \ldots, X_n)$ we need the bivariate density function corresponding to (4.2.1). It is seen that

$$f_{(\epsilon_{2t},\epsilon_{2t+1})}(u_1, u_2) = \begin{cases} \alpha_2(\alpha_1 + \alpha_{12})\exp[-(\alpha_1 + \alpha_{12})u_1 - \alpha_2 u_2], & 0 \leq u_2 < u_1 \\ \alpha_1(\alpha_2 + \alpha_{12})\exp[-\alpha_1 u_1 - (\alpha_2 + \alpha_{12})u_2], & 0 \leq u_1 < u_2 \\ \alpha_{12}\exp[-(\alpha_1 + \alpha_2 + \alpha_{12})u], & u_1 = u_2 = u. \end{cases} \tag{4.2.3}$$

The likelihood function, conditional on X_1, is then given by

$$\begin{aligned} L_n &= \prod_{t=1}^{m} p(x_{2t}, x_{2t+1}|x_t) \\ &= \prod_{t=1}^{m} f_{(\epsilon_{2t},\epsilon_{2t+1})}(x_{2t} - \phi X_t, \quad x_{2t+1} - \phi X_t), \end{aligned} \tag{4.2.4}$$

where m denotes the number of triplets (x_t, x_{2t}, x_{2t+1}) observed and $p(x_{2t}, x_{2t+1}|x_t)$ denotes the conditional density of (X_{2t}, X_{2t+1}) given $X_t = x_t$. Note that $n = 2m + 1$.

After substituting (4.2.2) and (4.2.3) in (4.2.4) and some simplification, we have

$$L_n = [\prod_{t=1}^{m} g_t(\lambda, \rho, \phi, X_t, X_{2t}, X_{2t+1})]I(\phi \leq \min_{2\leq t\leq n}(\frac{X_t}{X_{[\frac{t}{2}]}})), \tag{4.2.5}$$

where $I(\cdot)$ denotes the indicator function and $g_t(\cdot)$ is an increasing function of ϕ. Let $\hat{\phi}_{ML}$ denote the maximizer of $L_n(\phi)$ with respect to ϕ. Since L_n is an increasing function of ϕ, it follows that

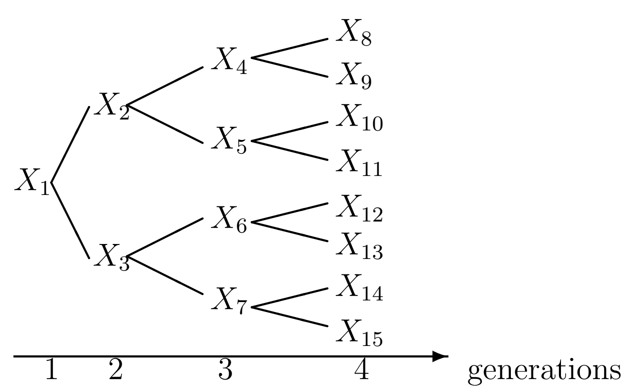
$$\hat{\phi}_{ML} = \min_{2 \le t \le n} (\frac{X_t}{X_{[\frac{t}{2}]}}), \tag{4.2.6}$$

which does not depend on λ and ρ. We will treat λ and ρ as unknown nuisance parameters. Our primary goal in this section is to derive the exact distribution of $\hat{\phi}_{ML}$ in (4.2.6).

Consider

$$\begin{aligned} P(\hat{\phi}_{ML} - \phi > u) &= P(\min_{2 \le t \le n} (\frac{\epsilon_t}{X_{[\frac{t}{2}]}}) > u) \\ &= P(X_t > (\phi + u) X_{[\frac{t}{2}]}, \quad 2 \le t \le n). \end{aligned} \tag{4.2.7}$$

In order to evaluate the multiple integral in (4.2.7), we first arrange the observations $(X_t, t = 1, \ldots, n)$ into k generations, the jth generation consisting of 2^{j-1} observations, $j = 1, 2, \ldots, k$, and $\sum_{j=1}^{k} 2^{j-1} = 2^k - 1 = n$, the total number of observations. Let A_j denote the set of observations in the jth generation. Note that $t \in A_j$ implies $j = [\log_2 t] + 1$. Also, define B_j to be the set of all observations contained in the first j generations. For $n = 15$, the grouping of the observations $\{X_t\}$, $t = 1, \ldots, 15$ into $k = 4$ generations is illustrated below.



The connecting lines indicate the branches in the binary tree. Here, $A_1 = \{X_1\}$, $A_2 = \{X_2, X_3\}$, $A_3 = \{X_4, X_5, X_6, X_7\}$ and $A_4 = \{X_8, \ldots, X_{15}\}$. Also, $B_1 = \{X_1\}$, $B_2 = \{X_1, X_2, X_3\}$, $B_3 = \{X_1, \ldots, X_7\}$ and $B_4 = \{X_1, \ldots, X_{15}\}$. The model in (4.1.12) implies (i) Markovity of observations between successive generations and (ii) conditional independence within each generation given the observations in the previous generation. In the above illustration the Markovity implies, for instance, that the conditional distribution of any observation in the set A_4 given all the previous observations in B_3 depends only on the observation on the same path in A_3. Also, conditional on all the observations in A_3, conditional independence implies that the pairs $(X_8, X_9), \ldots, (X_{14}, X_{15})$ are independent.

From (4.2.7), we have

$$P(\hat{\phi}_{ML} - \phi > u) = E[\prod_{j=2}^{k} W_j], \tag{4.2.8}$$

where $W_j = \prod_{t \in A_j} I(X_t > (\phi + u)X_{[\frac{t}{2}]})$. For any triplet (X_t, X_{2t}, X_{2t+1}), we have, from (4.2.1) and (4.2.2),

$$P(X_{2t} > (\phi + u)x_t, X_{2t+1} > (\phi + u)x_t | X_t = x_t) = \exp(-\frac{2u}{(1+\rho)\lambda} x_t). \tag{4.2.9}$$

Consider the last two generations $(k-1)$th and kth. We have

$$\begin{aligned} E(W_k | B_{k-1}) &= E(W_k | A_{k-1}), \quad \text{by Markovity,} \\ &= \exp(-\frac{2u}{(1+\rho)\lambda} \sum_{t \in A_{k-1}} x_t), \end{aligned} \tag{4.2.10}$$

by conditional independence, and using (4.2.9). In order to proceed further we need the following result which is proved in the appendix.

Lemma 2.1. For any constant b, and any triplet (X_t, X_{2t}, X_{2t+1}), we have

$$E[I(X_{2t} > (\phi + u)x_t, X_{2t+1} > (\phi + u)x_t) \exp(-\frac{2ub}{(1+\rho)\lambda}(X_{2t} + X_{2t+1})) | X_t = x_t]$$

$$= (\frac{1 + (1+2ub)\rho}{(1+2ub)(1+2ub+\rho)}) \exp(-\frac{2ux_t}{(1+\rho)\lambda}(1 + b(2u + 2\phi))). \tag{4.2.11}$$

In what follows the notation $\sum_{t \in A_j} X_t$ is used to denote the sum of all observations in the jth generation.

For the jth generation, for constant b_j, we have

$$\begin{aligned} & E[W_j \exp(-\frac{2ub_j}{(1+\rho)\lambda}\sum_{t\in A_j} X_t)|B_{j-1}] \\ = \; & E[W_j \exp(-\frac{2ub_j}{(1+\rho)\lambda}\sum_{t\in A_j} X_t)|A_{j-1}], \quad \text{by markovity} \\ = \; & a_j \exp(-\frac{2ub_{j-1}}{(1+\rho)\lambda}\sum_{t\in A_{j-1}} X_t), \end{aligned} \tag{4.2.12}$$

where $a_j = (\frac{1+(1+2ub_j)\rho}{(1+2ub_j)(1+2ub_j+\rho)})^{2^{j-2}}$, and $b_{j-1} = 1 + b_j(2\phi + 2u)$. The result in (4.2.12) is obtained from (2.11) and using conditional independence. Recall that 2^{j-2} is the number of observations in the $(j-1)$th generation. The exact distribution of the ML estimator $\hat{\phi}_{ML}$ is given below.

Theorem 2.1. Conditional on X_1, for any $u > 0$, the exact distribution is given by

$$P(\hat{\phi}_{ML} - \phi > u|X_1) = (\prod_{j=2}^{k-1} a_j)\exp(-\frac{2ub_1}{(1+\rho)\lambda}X_1), \tag{4.2.13}$$

where $a_j = (\frac{1+(1+2ub_j)\rho}{(1+2ub_j)(1+2ub_j+\rho)})^{2^{j-2}}$ and $b_j = \sum_{s=0}^{k-j-1}(2\phi+2u)^s$, $j = 1, \dots, k-1$.

Proof: From (4.2.8), we have

$$\begin{aligned} P(\hat{\phi}_{ML} - \phi > u) &= E[\prod_{j=2}^{k} W_j] = E[E(W_k \prod_{j=2}^{k-1} W_j|B_{k-1})] = E[E(W_k|B_{k-1})(\prod_{j=2}^{k-1} W_j)] \\ &= E[\exp(-\frac{2u}{(1+\rho)\lambda}\sum_{t\in A_{k-1}} X_t)(\prod_{j=2}^{k-1} W_j)], \quad \text{from (4.2.10)} \\ &= E[E(\exp(-\frac{2ub_{k-1}}{(1+\rho)\lambda}\sum_{t\in A_{k-1}} X_t)W_{k-1}|B_{k-2})\prod_{j=2}^{k-2} W_j], \quad b_{k-1} = 1, \\ &= a_{k-1}E[\exp(-\frac{2ub_{k-2}}{(1+\rho)\lambda}\sum_{t\in A_{k-2}} X_t)\prod_{j=2}^{k-2} W_j], \quad \text{from (4.2.12)} \\ &= a_{k-1}E[E(\exp(-\frac{2ub_{k-2}}{(1+\rho)\lambda}\sum_{t\in A_{k-2}} X_t)W_{k-2})|B_{k-1})\prod_{j=2}^{k-3} W_j] \\ &= a_{k-1}a_{k-2}E[\exp(-\frac{2ub_{k-3}}{(1+\rho)\lambda}\sum_{t\in A_{k-3}} X_t)\prod_{j=2}^{k-3} W_j], \quad \text{from (4.2.12).} \end{aligned}$$

Continuing this process iteratively and using (4.2.12) repeatedly, we finally obtain

$$\begin{aligned} P(\hat{\phi}_{ML} - \phi > u|X_1) &= a_{k-1}a_{k-2},\ldots,a_3 E[\exp(-\frac{2ub_2}{(1+\rho)\lambda}\sum_{t\in A_2} X_t)W_2] \\ &= a_{k-1}a_{k-2},\ldots,a_3 a_2 \exp(-\frac{2ub_1}{(1+\rho)\lambda}X_1), \quad \text{from (4.2.12).} \end{aligned}$$

The process stops at X_1 since X_1 is assumed fixed. The expression for b_j is obtained by solving the recursive equation $b_{j-1} = 1 + b_j(2\phi + 2u)$, for $j = k,\ k-1, \ldots, 2$, with $b_{k-1} = 1$, to get $b_j = \sum_{s=0}^{k-j-1} (2\phi + 2u)^s$, $j = 1, 2, \ldots, k-1$, and hence, $b_1 = \frac{(2\phi+2u)^{k-1}-1}{(2\phi+2u)-1}$. □

Corollary 2.1. If $\rho = 0$ (i.e. when $\{\epsilon_t\}$ are i.i.d. exponential with mean λ), the exact distribution of $\hat{\phi}_{ML}$, conditional on X_1, is given by (4.2.13) with $a_j = (1 + 2ub_j)^{-2^{j-1}}$ and $b_j = \sum_{s=0}^{k-j-1} (2\phi + 2u)^s$.

Proof: This follows readily from Theorem 2.1, by setting $\rho = 0$ in the expression for a_j. □

Corollary 2.2. If X_1 is a random variable with moment generating function $M_{X_1}(t)$, then the unconditional exact distribution of $\hat{\phi}_{ML}$ is given by

$$P(\hat{\phi}_{ML} - \phi > u) = (\prod_{j=2}^{k-1} a_j) M_{X_1}(-\frac{2ub_1}{(1+\rho)\lambda}). \qquad (4.2.14)$$

Proof: The result follows by taking expectations with respect to X_1 on both sides of (4.2.13) and noting that the a_j's do not depend on X_1. □

4.3 Asymptotic Distributions

In this section we derive the limit distribution of $\hat{\phi}_{ML}$ in the three cases (i) $0 \leq \phi < 1$ (stationary), (ii) $\phi = 1$ (critical) and (iii) $\phi > 1$ (explosive).

Recall that if n is the total number of observations in k generations, we have $n = 2^k - 1$ or $k = \log_2(n+1)$. The limit distributions of $\hat{\phi}_{ML}$ in the three cases are summarized below.

Theorem 3.1. As $k \to \infty$, we have

$$\alpha_k(\hat{\phi}_{ML} - \phi) \xrightarrow{d} Exp(1), \quad \text{for all} \ \ \phi \geq 0, \qquad (4.3.1)$$

where $Exp(1)$ is an exponential random variable with mean 1, and

$$\alpha_k = \begin{cases} \frac{2^k}{(1+\rho)(1-\phi)}, & \text{for} \quad 0 \le \phi < 1 \\ \frac{2^k k}{1+\rho}, & \text{for} \quad \phi = 1 \\ \frac{2^k \phi^{k-1} W}{\lambda(1+\rho)(2\phi-1)}, & \text{for} \quad \phi > 1, \end{cases}$$

W being a positive random variable defined by

$$W = \sum_{j=2}^{\infty} \phi^{-(j-1)} \bar{\epsilon}_j + X_1, \tag{4.3.2}$$

with $\bar{\epsilon}_j = 2^{-(j-1)} \sum_{t \in A_j} \epsilon_t$ = average of ϵ_t's corresponding to the jth generation.

Proof: First, consider the non-explosive cases, $0 \le \phi \le 1$. The exact distribution in (4.2.13) can be rewritten as

$$P(\hat{\phi}_{ML} - \phi > u) = \exp[-\frac{2uX_1A_k}{(1+\rho)\lambda} - \sum_{i=2}^{k-1} 2^{k-i-1} \log Z_i], \tag{4.3.3}$$

where $\log Z_i = \log(1 + 2uA_i) + \log(1 + \frac{2u}{1+\rho}A_i) - \log(1 + \frac{2u\rho}{1+\rho}A_i)$, and $A_i = \sum_{s=0}^{i-2}(2\phi + 2u)^s = \frac{(2\phi+2u)^{i-1}-1}{2\phi+2u-1}$. The main work involved in finding the limit of the right hand side of (4.3.3) after replacing u by $\alpha_k^{-1}x$, for $x > 0$, is to determine the limiting behaviour of $V_k = \sum_{i=2}^{k-1} 2^{k-i-1} \log(1 + 2uA_i)$, as $k \to \infty$ (and hence $n \to \infty$).

Since

$$c - \frac{c^2}{2} < \log(1 + c) < c, \quad c > 0,$$

we have

$$\sum_{i=2}^{k-1} 2^{k-i-1}(2uA_i) - \sum_{i=2}^{k-1} 2^{k-i-1} \frac{(2uA_i)^2}{2} < V_k < \sum_{i=2}^{k-1} 2^{k-i-1}(2uA_i). \tag{4.3.4}$$

(i) For $0 \le \phi < 1$, choose $u = \frac{x}{2^k}$. It can be verified that

$$\sum_{i=2}^{k-1} 2^{k-i-1} \frac{(2uA_i)^2}{2} \to 0, \quad \text{and} \quad \sum_{i=2}^{k-1} 2^{k-i-1}(2uA_i) \to \frac{x}{2(1-\phi)}.$$

Hence, it follows from (4.3.4) that $V_k \to \frac{x}{2(1-\phi)}$. It is then straightforward to see that with $u = \frac{x}{2^k}$,

$$\sum_{i=2}^{k-1} 2^{k-i-1} \log Z_i \to \frac{x}{(1-\phi)(1+\rho)}. \tag{4.3.5}$$

It is also easy to verify that

$$\frac{2uX_1A_k}{(1+\rho)\lambda} \to 0, \quad \text{with} \quad u = \frac{x}{2^k}. \tag{4.3.6}$$

From (4.3.3), (4.3.5) and (4.3.6), it follows that, for $0 \le \phi < 1$, we have the desired result

$$P(2^k(\hat{\phi}_{ML} - \phi) > x) \to \exp(-\frac{x}{(1-\phi)(1+\rho)}).$$

Hence, the result in the theorem follows.

(ii) For $\phi = 1$, choose $u = \frac{x}{2^k k}$. It can then be verified that $V_k \to \frac{x}{2}$, and hence

$$\sum_{i=2}^{k-1} 2^{k-i-1} \log Z_i \to \frac{x}{1+\rho}. \tag{4.3.7}$$

It is seen that

$$\frac{2uX_1A_k}{(1+\rho)\lambda} \to 0, \quad \text{with} \quad u = \frac{x}{2^k k}. \tag{4.3.8}$$

Hence the result for $\phi = 1$ follows from (4.3.3), (4.3.7) and (4.3.8).

(iii) For $\phi > 1$, we will derive the limit distribution more directly. From (1.12), we have

$$X_t = \sum_{j=0}^{[\log_2 t]-1} \phi^j \epsilon_{[\frac{t}{2^j}]} + \phi^{[\log_2 t]} X_1. \tag{4.3.9}$$

Hence

$$\sum_{t \in A_k} X_t = \sum_{j=2}^{k} (2\phi)^{k-j} (\sum_{t \in A_j} \epsilon_t) + (2\phi)^{k-1} X_1 \tag{4.3.10}$$

and

$$(2\phi)^{-(k-1)} \sum_{t \in A_k} X_t = \sum_{j=2}^{k} (2\phi)^{-(j-1)} \sum_{t \in A_j} \epsilon_t + X_1.$$

Since

$$\begin{aligned} \sum_{j=2}^{\infty} E\left|(2\phi)^{-(j-1)} \sum_{t \in A_j} \epsilon_t\right| &= \sum_{j=2}^{\infty} (2\phi)^{-(j-1)} 2^{j-1} \lambda \\ &= \lambda \sum_{j=2}^{\infty} \phi^{-(j-1)} < \infty, \end{aligned}$$

we have

$$(2\phi)^{-(k-1)} \sum_{t \in A_k} X_t \xrightarrow{\text{a.s.}} W, \tag{4.3.11}$$

where W is defined in (4.3.2). Let

$$H_k = \frac{2^k \phi^{k-1}(\hat{\phi}_{ML} - \phi)W}{\lambda(1+\rho)(2\phi - 1)}. \tag{4.3.12}$$

We need to show that $H_k \xrightarrow{d} Exp(1)$, as $k \to \infty$. We will now follow analogous arguments to those used by Nielsen and Shephard (2003), p 343. We have

$$\begin{aligned}
P(H_k > y) &= P(\hat{\phi}_{ML} - \phi > \frac{ya}{((2\phi)^{k-1}W)}), \quad \text{where} \;\; a = \frac{\lambda(1+\rho)(2\phi-1)}{2} \\
&= P(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W}, \quad t = 2, \ldots, n) \\
&= E[\prod_{j=2}^{k} I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W}, \quad t \in A_j)] \\
&= E[E\{\sum_{j=2}^{k} I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W}, \quad t \in A_j)|B_{k-1}\}] \\
&= E[I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W}, \quad t \in B_{k-1}) \exp(-\frac{2ya}{(2\phi)^{k-1}W}(\sum_{t \in A_{k-1}} X_t)\frac{1}{(1+\rho)\lambda})] \\
&\geq E[I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W}, \quad t \in B_{k-1}) \exp(\frac{-2ya}{(2\phi)\lambda(1+\rho)})], \qquad (4.3.13)
\end{aligned}$$

since $\sum_{t \in A_{k-1}} X_t/(2\phi)^{k-2} \leq W$, a.s.. Continuing the process in (4.3.13) iteratively, we have

$$P(H_k > y) \geq \exp[-\frac{2ya}{\lambda(1+\rho)}(\frac{1}{2\phi} + \frac{1}{(2\phi)^2} + \cdots + \frac{1}{(2\phi)^{k-1}})]$$

$$\to \exp(-y), \quad \text{as} \;\; k \to \infty. \tag{4.3.14}$$

Let $\frac{\sum_{t \in A_k} X_t}{(2\phi)^{k-1}} = \eta_k$. Then, by (4.3.11), $\eta_k \xrightarrow{\text{a.s.}} W$. Egorov's theorem (see Lieb and Loss (2001)) then implies that for any $\delta_1, \delta_2 > 0$, there exists a set Ω_1, with $P(\Omega_1) = 1 - \delta_1$ and a k_0 such that, for $w \in \Omega_1$ and $k \geq k_0$, we have $\frac{\eta_k(w)}{W(w)} > 1 - \delta_2$ (i.e. η_k converges to W uniformly on

Ω_1). We have

$$\begin{aligned}
P(H_k > y) &= P(H_k > y, \Omega_1^c) + P(H_k > y, \Omega_1) \\
&\le P(\Omega_1^c) + P(H_k > y, \Omega_1) \\
&= \delta_1 + E\prod_{j=2}^{k}(I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W},\ t \in A_k)I(\Omega_1)) \\
&= \delta_1 + E[(I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W},\quad t \in B_{k-1}))\exp(-\frac{2ya}{(1+\rho)\lambda}\frac{\sum_{t\in A_{k-1}} X_t}{(2\phi)^{k-1}W})I(\Omega_1)] \\
&\le \delta_1 + E[(\prod_{j=k_0}^{k-1} I(\epsilon_t > \frac{yaX_{[\frac{t}{2}]}}{(2\phi)^{k-1}W},\ t \in A_j)\exp(-\frac{2ya}{(1+\rho)\lambda(2\phi)}(1-\delta_2))I(\Omega_1)] \\
&\le \delta_1 \exp(-\frac{2ya(1-\delta)}{(1+\rho)\lambda}(\frac{1}{2\phi} + \frac{1}{(2\phi)^2} + \cdots + \frac{1}{(2\phi)^{k-k_0+1}})) \\
&\to \delta_1 + \exp(-y(1-\delta_2)) \\
&\to \exp(-y), \qquad (4.3.15)
\end{aligned}$$

since δ_1 and δ_2 are arbitrarily small. From (4.3.14) and (4.3.15), the desired result follows. □

4.4 A Unified Limit Theorem

In Theorem 3.1 we had to use different normalizing sequences for the three cases to get the limiting exponential distribution. In this section we will show that the three normalizing sequences α_k can be replaced by a single random normalizing sequence. The result is given below.

Theorem 4.1. As $k \to \infty$,

$$\frac{2}{\lambda(1+\rho)}(\sum_{t\in B_{k-1}} X_t)(\hat{\phi}_{ML} - \phi) \xrightarrow{d} Exp(1), \quad \text{For all} \quad \phi \ge 0. \qquad (4.4.1)$$

Proof: The desired result in (4.4.1) will follow from Theorem 3.1 if we show that

$$2^{-(k-1)} \sum_{t\in B_{k-1}} X_t \xrightarrow{p} (\frac{\lambda}{1-\phi}), \quad \text{for} \quad 0 \le \phi < 1, \qquad (4.4.2)$$

$$(k2^{k-1})^{-1} \sum_{t\in B_{k-1}} X_t \xrightarrow{p} \lambda, \quad \text{for} \quad \phi = 1, \qquad (4.4.3)$$

and

$$(2\phi)^{-(k-1)} \sum_{t \in B_{k-1}} X_t \xrightarrow{\text{a.s.}} \frac{W}{(2\phi-1)}, \quad \text{for} \quad \phi > 1. \tag{4.4.4}$$

The results in (4.4.2)-(4.4.4) then lead to the common normalizing random sequence $2(\lambda(1+\rho))^{-1}\sum_{t\in B_{k-1}} X_t$ for all the three cases, proving (4.4.1). We now proceed to verify (4.4.2)-(4.4.4).

From (4.3.10), we have

$$\begin{aligned}\sum_{t\in B_{k-1}} X_t = \sum_{j=1}^{k-1}\sum_{t\in A_j} X_t &= \sum_{j=1}^{k-1}[\sum_{i=2}^{j}(2\phi)^{j-i}\sum_{t\in A_i}\epsilon_t + (2\phi)^{j-1}X_1] \\ &= \begin{cases} \sum_{j=2}^{k-1}(\frac{1-(2\phi)^{k-j}}{1-2\phi})\sum_{t\in A_j}\epsilon_t + \frac{1-(2\phi)^{k-1}}{1-2\phi}X_1, & \text{for} \quad \phi \neq \frac{1}{2} \\ \sum_{j=2}^{k-1}(k-j)\sum_{t\in A_j}\epsilon_t + (k-1)X_1, & \text{for} \quad \phi = \frac{1}{2}. \end{cases}\end{aligned} \tag{4.4.5}$$

For $0 \le \phi < 1$, one can verify, via (4.4.5), that

$$E(\sum_{t\in B_{k-1}} 2^{-(k-1)}X_t) \to \frac{\lambda}{1-\phi}, \quad \text{and} \quad Var(\frac{\sum_{t\in B_{k-1}} X_t}{2^{k-1}}) \to 0$$

giving the result in (4.4.2). Similarly, (4.4.3) can be verified by checking

$$E(\sum_{t\in B_{k-1}} (k2^{k-1})^{-1}X_t) \to \lambda \quad \text{and} \quad Var(\frac{\sum_{t\in B_{k-1}} X_t}{k2^{k-1}}) \to 0.$$

For $\phi > 1$, we have, from (4.4.5) with $\phi \neq \frac{1}{2}$,

$$\begin{aligned}(2\phi-1)\sum_{t\in B_{k-1}} X_t &= \sum_{j=2}^{k-1}((2\phi)^{k-j}-1)\sum_{t\in A_j}\epsilon_t + ((2\phi)^{k-1}-1)X_1 \\ &= \sum_{t\in A_k} X_t - (\sum_{t\in B_k}\epsilon_t + X_1).\end{aligned} \tag{4.4.6}$$

From (4.3.11), (4.4.6) and noting that $(2\phi)^{-(k-1)}\sum_{t\in B_k}\epsilon_t \xrightarrow{\text{a.s.}} 0$, the result in (4.4) follows. This completes the proof of the Theorem. □

Define the pivot

$$T_n = \frac{2}{\hat{\lambda}(1+\hat{\rho})}(\sum_{t\in B_{k-1}} X_t)(\hat{\phi}_{ML} - \phi) \tag{4.4.7}$$

where $\hat{\lambda}$ and $\hat{\rho}$ are any consistent estimates of λ and ρ. For instance, one can choose

$$\hat{\lambda} = \frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{\phi}_{ML}X_{[\frac{t}{2}]}), \text{ and } \hat{\rho} = \frac{\sum_{t=1}^{m}(X_{2t} - \hat{\phi}_{ML}X_t - \hat{\lambda})(X_{2t+1} - \hat{\phi}_{ML}X_t - \hat{\lambda})}{[(\sum_{t=1}^{m}(X_{2t} - \hat{\phi}_{ML}X_t - \hat{\lambda})^2)(\sum(X_{2t+1} - \hat{\phi}_{ML}X_t - \hat{\lambda})^2)]^{1/2}}. \tag{4.4.8}$$

It then follows that

$$2T_n \xrightarrow{d} \chi^2(2), \quad \text{for all} \quad \phi \geq 0. \tag{4.4.9}$$

The pivot T_n can therefore be used for constructing confidence intervals and tests for ϕ without prior knowledge as to which of the three regions the true value of ϕ belongs to. The result in (4.4.9) thus extends (4.1.11) for the AR(1) model to the bifurcating AR(1) model.

Suppose $\hat{\phi}_{ML}$ is based on observations in the last two generations only, i.e. in A_{k-1} and A_k, and denote the estimate by $\hat{\phi}_{ML}^{(2)}$. We have

$$\hat{\phi}_{ML}^{(2)} = \min_{t \in A_k}(\frac{X_t}{X_{[\frac{t}{2}]}}), \tag{4.4.10}$$

and

$$\begin{aligned} P(\hat{\phi}_{ML}^{(2)} - \phi > u) &= P(X_t > (\phi + u)X_{[\frac{t}{2}]}, \quad t \in A_k) \\ &= E(W_k) = E[E(W_k|B_{k-1})] \\ &= E[\exp(-\frac{2u}{(1+\rho)\lambda}\sum_{t \in A_{k-1}} X_t)], \quad \text{from (4.2.10)}. \end{aligned}$$

If we now choose $u = ((2\sum_{t \in A_{k-1}} X_t)x/(1+\rho)\lambda)^{-1}$, we have

$$P[(\frac{2\sum_{t \in A_{k-1}} X_t}{(1+\rho)\lambda})(\hat{\phi}_{ML}^{(2)} - \phi) > x] = \exp(-x), \quad \text{for all} \quad \phi \geq 0, \tag{4.4.11}$$

and for any k. Thus, the exact distribution of the pivot $(\frac{4\sum_{t \in A_{k-1}} X_t}{(1+\rho)\lambda})(\hat{\phi}_{ML}^{(2)} - \phi)$ is $\chi^2(2)$ for any $\phi \geq 0$. Consequently, the asymptotic distribution of this pivot is also $\chi^2(2)$. The asymptotic relative efficiency of $\hat{\phi}_{ML}^{(2)}$ with respect to $\hat{\phi}_{ML}$ (based on all the generations) is obtained by the limit in probability of the ratio $(\sum_{t \in A_{k-1}} X_t / \sum_{t \in B_{k-1}} X_t)$ of the two corresponding normalizing sequences. It can be verified that

$$2^{-(k-2)}\sum_{t \in A_{k-1}} X_t \xrightarrow{p} \frac{\lambda}{1-\phi}, \quad \text{for} \quad 0 \leq \phi < 1, \tag{4.4.12}$$

$$(k2^{k-2})^{-1} \sum_{t \in A_{k-1}} X_t \xrightarrow{p} \lambda, \quad \text{for} \quad \phi = 1, \tag{4.4.13}$$

and, from (4.3.11), we have

$$(2\phi)^{-(k-2)} \sum_{t \in A_{k-1}} X_t \xrightarrow{\text{a.s.}} W, \quad \text{for} \quad \phi > 1. \tag{4.4.14}$$

Comparing (4.4.12)-(4.4.14) with (4.4.2)-(4.4.4), we can conclude that the relative efficiency of $\hat{\phi}_{ML}^{(2)}$ with respect to $\hat{\phi}_{ML}$ is $\frac{1}{2}$ for $0 \leq \phi \leq 1$, and $1 - (2\phi)^{-1}$ for $\phi > 1$.

Extrapolating this argument, let $\hat{\phi}_{ML}^{(l)}$ denotes the estimate based on the observations in $A_k, A_{k-1}, \ldots, A_{k-l+1}$ (i.e. in the last $l < k$) generations, the asymptotic relative efficiency of $\hat{\phi}_{ML}^{(l)}$, for fixed l and as $k \to \infty$, is seen to be $1 - 2^{-l+1}$ for $0 \leq \phi \leq 1$ and $1 - (2\phi)^{-l+1}$ for $\phi > 1$. Thus, the efficiency increases as we include more generations.

Consider now the special case when $\rho = 0$, i.e. when $\{\epsilon_t\}$ are i.i.d. $Exp(\lambda)$ random variables. The likelihood function (4.2.5) then simplifies to

$$L_n(\lambda, \phi) = \lambda^{-(n-1)} \exp(-\lambda^{-1} \sum_{t=2}^{n} (X_t - \phi X_{[\frac{t}{2}]})) I(\phi \leq \min_{2 \leq t \leq n} (\frac{X_t}{X_{[\frac{t}{2}]}})). \tag{4.4.15}$$

The maximum likelihood estimates of ϕ and λ from (4.4.15) are given by

$$\hat{\phi}_{ML} = \min_{2 \leq t \leq n} (\frac{X_t}{X_{[\frac{t}{2}]}}), \quad \text{and} \quad \hat{\lambda}_{ML} = (n-1)^{-1} \sum_{t=2}^{n} (X_t - \hat{\phi}_{ML} X_{[\frac{t}{2}]}). \tag{4.4.16}$$

The likelihood ratio (LR) statistic for testing $\phi = \phi_0$ is given by

$$Q_n = \frac{L_n(\hat{\lambda}_0, \phi_0)}{L_n(\hat{\lambda}_{ML}, \hat{\phi}_{ML})} = (\frac{\hat{\lambda}_{ML}}{\hat{\lambda}_0})^{n-1}, \tag{4.4.17}$$

where $\hat{\lambda}_0 = (n-1)^{-1} \sum_{t=1}^{n} (X_t - \phi_0 X_{[\frac{t}{2}]})$. Note that

$$Q_n = [1 - \frac{(\hat{\phi}_{ML} - \phi_0) \sum_{t=2}^{n} X_{[\frac{t}{2}]}}{\sum_{t=2}^{n} (X_t - \phi_0 X_{[\frac{t}{2}]})}]^{n-1} = [1 - \frac{2(\sum_{t \in B_{k-1}} X_t)(\hat{\phi}_{ML} - \phi_0)}{\sum_{t=2}^{n} \epsilon_t}]^{n-1}. \tag{4.4.18}$$

From Theorem 4.1, and (4.4.18) it is easy to verify that, for $\phi = \phi_0$, $\phi_0 \geq 0$,

$$\text{(i)} \quad Q_n \xrightarrow{d} \text{Uniform } (0,1),$$

$$\text{(ii)} \quad -\log Q_n \xrightarrow{d} Exp(1),$$

$$\text{and} \qquad \text{(iii)} \quad -2 \log Q_n \xrightarrow{d} \chi^2(2). \tag{4.4.19}$$

The results in (4.4.19) generalize those of Nielsen and Shephard (2003) for the AR(1) model to the bifurcating AR(1) model. For $\rho \neq 0$, the LR statistic does not have a simple form. However, we can always use the simpler statistic T_n in (4.4.7) which corresponds to the Wald statistic.

4.5 Simulation Results

The least-squares (LS) estimators of ϕ and λ for the model in (4.1.12), obtained by minimizing $\sum_{t=1}^{m}(X_{2t} - \phi X_t - \lambda)^2 + \sum_{t=1}^{m}(X_{2t+1} - \phi X_t - \lambda)^2$, are given by

$$\hat{\phi}_{LS} = \frac{\sum_{t=1}^{m} X_t(U_t - \bar{U})}{\sum_{t=1}^{m}(X_t - \bar{X})^2}, \quad \text{and} \quad \hat{\lambda}_{LS} = \bar{U} - \hat{\phi}_{LS}\bar{X}, \tag{4.5.1}$$

where $U_t = \frac{X_{2t}+X_{2t+1}}{2}$, $\bar{U} = \frac{1}{m}\sum_{t=1}^{m} U_t$, and $\bar{X} = \frac{1}{m}\sum_{t=1}^{m} X_t$, and $m =$ number of triplets observed. Recall that $n = 2m + 1$. Denote

$$\hat{\lambda}_{ML} = \bar{U} - \hat{\phi}_{ML}\bar{X} \tag{4.5.2}$$

where $\hat{\phi}_{ML}$ is given by (4.2.6). Note that $\hat{\lambda}_{ML}$ in (4.5.2) is the ML estimate of λ, when $\rho = 0$, and it is not quite the ML estimate when $\rho \neq 0$. However, we shall denote the estimate in (4.5.2) as $\hat{\lambda}_{ML}$ for all $0 \leq \rho < 1$, by an abuse of the notation.

In this section, we first report the results of a simulation study to compare $\hat{\phi}_{ML}$ with $\hat{\phi}_{LS}$, and $\hat{\lambda}_{ML}$ with $\hat{\lambda}_{LS}$. We simulated observations $\{X_t\}$ from the model in (4.1.12) with bivariate exponential errors and parameters $\lambda = 1$, $\rho = 0.5$, $\phi = \{0, 0.5, 0.9, 1, 2\}$, and number of generations $k = \{5, 6, \ldots, 10\}$. The estimates $\hat{\phi}_{ML}$, $\hat{\phi}_{LS}$, $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{LS}$ were computed. The process was replicated 1000 times. The means and mean-squared errors (MSE) of the estimates over 1000 replications were computed. Also, the relative efficiency of the LS estimate with respect to the ML estimate is computed as the ratio MSE(ML)/MSE(LS). The results for the comparison for $\hat{\phi}_{ML}$ and $\hat{\phi}_{LS}$ are given in Table 1. It may be noted, from Table 1, that the MSE's for both $\hat{\phi}_{ML}$ and $\hat{\phi}_{LS}$ descrease as the number of generations k increases. For all k, MSE $(\hat{\phi}_{ML}) <$ MSE $(\hat{\phi}_{LS})$. The relative efficiency of the LS estimate with respect to the ML estimate approaches zero very quickly as k increases. The reason for this behaviour is that $\hat{\phi}_{ML}$ has a much larger rate of convergence than $\hat{\phi}_{LS}$.

Table 2 summarizes the comparison of $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{LS}$. Again, the MSE's of both $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{LS}$ decrease with MSE $(\hat{\lambda}_{ML})$ < MSE $(\hat{\lambda}_{LS})$ for each k. However, the relative efficiency of $\hat{\lambda}_{LS}$ with respect to $\hat{\lambda}_{ML}$ does not approach zero since the rates of convergence for both the estimates are the same.

In order to study the convergence of the pivot $2T_n$ to $\chi^2(2)$-distribution where T_n is given by (4.4.7), we computed $p_n(2T_n > \chi^2_\alpha(2))$ where $p_n(\cdot)$ denotes the proportion out of 1000 values, and $\chi^2_\alpha(2)$ denotes the value such that $P(\text{Chi-square}(2) \leq \chi^2_\alpha(2)) = \alpha$. The results are summarized in Table 3. It is seen that $p_n(2T_n > \chi^2_\alpha(2))$ approaches $1-\alpha$ for all ϕ, as k increases.

ϕ	k	$\hat{\phi}_{ML}$	$\hat{\phi}_{LS}$	$MSE(\hat{\phi}_{ML})$	$MSE(\hat{\phi}_{LS})$	$\frac{MSE(\hat{\phi}_{ML})}{MSE(\hat{\phi}_{LS})}$
0.0	5	0.05104	-0.06442	.005245731	0.09222	0.056881
0.0	6	0.02385	-0.02698	.001134474	0.02905	0.039048
0.0	7	0.01112	-0.00921	.000237704	0.01281	0.018556
0.0	8	0.00578	-0.00671	.000064087	0.00598	0.010717
0.0	9	0.00293	-0.00503	.000016964	0.00286	0.005928
0.0	10	0.00144	-0.00198	.000004222	0.00147	0.002882
0.5	5	0.53033	0.38985	.001890812	0.12790	0.014784
0.5	6	0.51285	0.44987	.000330809	0.03058	0.010818
0.5	7	0.50603	0.47659	.000070483	0.01146	0.006150
0.5	8	0.50296	0.48543	.000016863	0.00517	0.003264
0.5	9	0.50153	0.49166	.000004721	0.00239	0.001972
0.5	10	0.50073	0.49607	.000001101	0.00114	0.000970
0.9	5	0.91843	0.86511	.000708221	0.06374	0.011111
0.9	6	0.90683	0.87502	.000094902	0.01672	0.005677
0.9	7	0.90286	0.88707	.000015997	0.00566	0.002827
0.9	8	0.90125	0.89095	.000003046	0.00229	0.001331
0.9	9	0.90059	0.89469	.000000720	0.00094	0.000770
0.9	10	0.90026	0.89711	.000000140	0.00041	0.000339
1.0	5	1.01622	0.98562	.000548891	0.04435	0.012377
1.0	6	1.00578	0.98674	.000068449	0.01109	0.006171
1.0	7	1.00232	0.99353	.000010565	0.00355	0.002977
1.0	8	1.00097	0.99521	.000001827	0.00139	0.001318
1.0	9	1.00044	0.99721	.000000401	0.00052	0.000769
1.0	10	1.00019	0.99848	.000000072	0.00022	0.000326
2.0	5	2.00504	2.00518	.000052061	0.00195	0.026709
2.0	6	2.00115	2.00139	.000002667	0.00020	0.013418
2.0	7	2.00028	2.00028	.000000165	0.00002	0.007087
2.0	8	2.00007	2.00007	.000000011	0.00000	0.003659
2.0	9	2.00002	2.00002	7.3376E-10	0.00000	0.002168
2.0	10	2.00000	2.00001	4.7961E-11	0.00000	0.001130

Table 4.1: Comparison of $\hat{\phi}_{ML}$ and $\hat{\phi}_{LS}$

ϕ	k	$\hat{\lambda}_{ML}$	$\hat{\lambda}_{LS}$	$MSE(\hat{\lambda}_{ML})$	$MSE(\hat{\lambda}_{LS})$	$\frac{MSE(\hat{\lambda}_{ML})}{MSE(\hat{\lambda}_{LS})}$
0.0	5	0.94884	1.06084	0.049185	0.12984	0.37881
0.0	6	0.97981	1.03294	0.023058	0.05509	0.41853
0.0	7	0.98621	1.00554	0.011283	0.02180	0.51750
0.0	8	0.99402	1.00622	0.005483	0.01076	0.50970
0.0	9	0.99837	1.00640	0.002976	0.00573	0.51940
0.0	10	0.99878	1.00224	0.001472	0.00291	0.50573
0.5	5	0.94851	1.21576	0.049626	0.37730	0.13153
0.5	6	0.98024	1.10859	0.022898	0.13120	0.17453
0.5	7	0.98595	1.04435	0.011278	0.04869	0.23162
0.5	8	0.99407	1.02851	0.005494	0.02403	0.22858
0.5	9	0.99828	1.01816	0.002970	0.01202	0.24708
0.5	10	0.99879	1.00825	0.001473	0.00593	0.24838
0.9	5	0.94828	1.20887	0.049758	0.64264	0.07743
0.9	6	0.98025	1.14837	0.022801	0.31761	0.07179
0.9	7	0.98588	1.07921	0.011293	0.12762	0.08849
0.9	8	0.99408	1.05684	0.005483	0.06116	0.08965
0.9	9	0.99827	1.03572	0.002965	0.02907	0.10199
0.9	10	0.99878	1.02115	0.001476	0.01525	0.09674
1.0	5	0.94825	1.16139	0.049754	0.57351	0.08675
1.0	6	0.98019	1.11982	0.022778	0.28369	0.08029
1.0	7	0.98589	1.06314	0.011291	0.12176	0.09274
1.0	8	0.99406	1.04728	0.005479	0.06128	0.08942
1.0	9	0.99826	1.02979	0.002965	0.02951	0.10047
1.0	10	0.99877	1.01868	0.001476	0.01628	0.09067
2	5	0.94823	1.00640	0.049458	0.24710	0.20016
2	6	0.98006	1.00393	0.022711	0.10194	0.22279
2	7	0.98599	1.00076	0.011241	0.04807	0.23386
2	8	0.99387	1.00283	0.005453	0.02413	0.22597
2	9	0.99823	1.00121	0.002970	0.01106	0.26852
2	10	0.99873	0.99974	0.001472	0.00567	0.25952

Table 4.2: Comparison of $\hat{\lambda}_{ML}$ and $\hat{\lambda}_{LS}$ (with $\lambda = 1$)

ϕ	k	$\hat{\phi}_{ML}$	α						
			.05	.10	.30	.50	.70	.90	.95
0.0	5	0.05104	0.947	0.895	0.687	0.475	0.277	0.102	0.050
0.0	6	0.02385	0.953	0.907	0.695	0.497	0.287	0.084	0.039
0.0	7	0.01112	0.957	0.895	0.693	0.481	0.268	0.081	0.040
0.0	8	0.00578	0.948	0.906	0.706	0.507	0.295	0.095	0.048
0.0	9	0.00293	0.954	0.896	0.693	0.502	0.309	0.102	0.042
0.0	10	0.00144	0.955	0.895	0.716	0.488	0.291	0.091	0.051
0.5	5	0.53033	0.953	0.899	0.699	0.493	0.282	0.095	0.052
0.5	6	0.51285	0.950	0.909	0.699	0.507	0.270	0.079	0.040
0.5	7	0.50603	0.955	0.909	0.702	0.480	0.283	0.085	0.041
0.5	8	0.50296	0.953	0.906	0.709	0.494	0.300	0.083	0.041
0.5	9	0.50153	0.960	0.893	0.701	0.487	0.315	0.112	0.050
0.5	10	0.50073	0.954	0.902	0.689	0.484	0.292	0.088	0.049
0.9	5	0.91843	0.955	0.906	0.696	0.499	0.293	0.098	0.052
0.9	6	0.90683	0.956	0.905	0.710	0.491	0.271	0.082	0.036
0.9	7	0.90286	0.960	0.908	0.696	0.474	0.294	0.090	0.046
0.9	8	0.90125	0.952	0.906	0.706	0.496	0.286	0.084	0.038
0.9	9	0.90059	0.960	0.899	0.702	0.492	0.304	0.111	0.052
0.9	10	0.90026	0.952	0.902	0.694	0.481	0.293	0.094	0.043
1.0	5	1.01622	0.955	0.906	0.702	0.499	0.286	0.098	0.050
1.0	6	1.00578	0.956	0.906	0.711	0.489	0.277	0.079	0.040
1.0	7	1.00232	0.958	0.909	0.697	0.478	0.295	0.088	0.045
1.0	8	1.00097	0.952	0.906	0.710	0.504	0.290	0.087	0.041
1.0	9	1.00044	0.960	0.902	0.709	0.492	0.297	0.113	0.053
1.0	10	1.00019	0.953	0.903	0.693	0.481	0.291	0.100	0.049
2.0	5	2.00504	0.956	0.904	0.707	0.509	0.293	0.095	0.053
2.0	6	2.00115	0.956	0.911	0.704	0.500	0.278	0.087	0.040
2.0	7	2.00028	0.960	0.909	0.681	0.478	0.288	0.087	0.033
2.0	8	2.00007	0.954	0.913	0.707	0.509	0.312	0.103	0.046
2.0	9	2.00002	0.955	0.913	0.716	0.520	0.305	0.112	0.052
2.0	10	2.00000	0.948	0.902	0.698	0.495	0.282	0.106	0.058

Table 4.3: The proportion of $2T_n > \chi^2_\alpha(2)$ in 1000 simulations

Appendix: Proof of Lemma 2.1

From (4.2.3), we have

$$p(x_{2t}, x_{2t+1}|x_t) = f_{(\epsilon_{2t},\epsilon_{2t+1})}(x_{2t} - \phi x_t, x_{2t+1} - \phi x_t)$$
$$= \begin{cases} \exp(\frac{2\phi x_t}{(1+\rho)\lambda})[\frac{1-\rho}{(1+\rho)\lambda^2}\exp(-\frac{x_{2t}}{\lambda} - \frac{(1-\rho)x_{2t+1}}{(1+\rho)\lambda})], & x_{2t+1} < x_{2t},\ \phi \le \min(\frac{x_{2t}}{x_t}, \frac{x_{2t+1}}{x_t}) \\ \exp(\frac{2\phi x_t}{(1+\rho)\lambda})[\frac{1-\rho}{(1+\rho)\lambda^2}\exp(-\frac{(1-\rho)x_{2t}}{(1+\rho)\lambda} - \frac{x_{2t+1}}{\lambda})], & x_{2t} < x_{2t+1},\ \phi \le \min(\frac{x_{2t}}{x_t}, \frac{x_{2t+1}}{x_t}) \\ \exp(\frac{2\phi x_t}{(1+\rho)\lambda})[\frac{2\rho}{(1+\rho)\lambda}\exp(-\frac{2x}{(1+\rho)\lambda})], & x_{2t} = x_{2t+1} = x,\ \phi \le \min(\frac{x_{2t}}{x_t}, \frac{x_{2t+1}}{x_t}). \end{cases} \quad (A.1)$$

We thus have

$$E[I(X_{2t} > (\phi+u)X_t, X_{2t+1} > (\phi+u)X_t)\exp(-\frac{2ub}{(1+\rho)\lambda}(X_{2t} + X_{2t+1}))|X_t = x_t]$$
$$= \exp(\frac{2\phi x_t}{(1+\rho)\lambda})[I_1 + I_2 + I_3], \quad (A.2)$$

where $I_1 =$

$$\int_{(\phi+u)x_t}^{\infty}\int_{(\phi+u)x_t}^{x_{2t}} \exp(-\frac{2ub}{(1+\rho)\lambda}(x_{2t}+x_{2t+1}))(\frac{1-\rho}{(1+\rho)\lambda^2})\exp(-\frac{x_{2t}}{\lambda} - \frac{(1-\rho)x_{2t+1}}{(1+\rho)\lambda})dx_{2t}dx_{2t+1},$$

$I_2 =$

$$\int_{(\phi+u)x_t}^{\infty}\int_{(\phi+u)x_t}^{x_{2t+1}} \exp(-\frac{2ub}{(1+\rho)\lambda})(x_{2t}+x_{2t+1}))(\frac{1-\rho}{(1+\rho)\lambda^2})\exp(-\frac{(1-\rho)x_{2t}}{(1+\rho)\lambda} - \frac{x_{2t+1}}{\lambda})dx_{2t}dx_{2t+1},$$

and

$$I_3 = \int_{(\phi+u)x_t}^{\infty} \exp(-\frac{2ub}{(1+\rho)\lambda}(2x)(\frac{2\rho}{(1+\rho)\lambda})\exp(-\frac{2x}{(1+\rho)\lambda})dx.$$

Note that $I_1 = I_2 =$

$$\int_{(\phi+u)x_t}^{\infty} \exp(-\frac{(2ub+1+\rho)}{(1+\rho)\lambda}x_{2t})\int_{(\phi+u)x_t}^{x_{2t}}(\frac{1-\rho}{(1+\rho)\lambda^2})\exp(-\frac{(2ub+1-\rho)}{(1+\rho)\lambda}x_{2t+1})dx_{2t+1}dx_{2t}$$
$$= \int_{(\phi+u)x_t}^{\infty} \exp(-\frac{(2ub+1+\rho)}{(1+\rho)\lambda}x_{2t})(\frac{1-\rho}{(2ub+1-\rho)\lambda})[\exp(-\frac{(2ub+1-\rho)}{(1+\rho)\lambda}(\phi+u)x_t$$
$$- \exp(-\frac{(2ub+1-\rho)}{(1+\rho)\lambda}x_{2t})]dx_{2t}$$
$$= (\frac{1-\rho}{(2ub+1-\rho)\lambda})\exp(-(\frac{2ub+1-\rho}{(1+\rho)\lambda})(\phi+u)x_t)\int_{(\phi+u)x_t}^{\infty}\exp(-(\frac{2ub+1+\rho}{(1+\rho)\lambda})x_{2t})dx_{2t}$$
$$-(\frac{1-\rho}{(2ub+1-\rho)\lambda})\int_{(\phi+u)x_t}^{\infty}\exp(-\frac{2(2ub+1)}{(1+\rho)\lambda}x_{2t})dx_{2t}$$
$$= (\frac{1-\rho}{2ub+1+\rho})(\frac{1+\rho}{2(2ub+1)})\exp(-\frac{2(2ub+1)}{(1+\rho)\lambda}(\phi+u)x_t). \quad (A.3)$$

Also,

$$\begin{aligned} I_3 &= \int_{(\phi+u)x_t}^{\infty} (\frac{2\rho}{(1+\rho)\lambda}) \exp(-\frac{2(2ub+1)}{(1+\rho)\lambda} x) dx \\ &= (\frac{\rho}{2ub+1}) \exp(-\frac{2(2ub+1)}{(1+\rho)\lambda} (\phi+u)x_t). \end{aligned} \tag{A.4}$$

Substituting (A.3) and (A.4) in (A.2) we have the desired result. □

4.6 References

[1] Basawa, I. V. and Zhou, J. (2003). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Prob.* 41A, 55-64.

[2] Cowan, R. and Staudte, R. G. (1986). The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42, 769-783.

[3] Davis, R. A. and McCormick, W. P. (1989). Estimation for the first-order autoregressive processes with positive or bounded innovations. *Stoch. Proc. App.* 31, 237-250.

[4] Huggins, R. M. and Basawa, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Prob.* 36, 1225-1233.

[5] Huggins, R. M. and Basawa, I. V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Austral. N.Z.J. Statist.* 42, 423-432.

[6] Kotz, S., Balakrishnam, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Vol 1, Models and Applications*, 2nd Edn., John Wiley, New York.

[7] Lieb, E. H. and Loss, M. (2001). *Analysis*, American Mathematical Society, 2nd Edn., Providence, Rhode Island.

[8] Mardia, K. V. (1970). *Families of Bivariate Distributions.* Griffin, London.

[9] Nielsen, B. and Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *J. Time Series Analysis*, 24, 337-344.

Chapter 5

Non-Gaussian Bifurcating Models and Quasilikelihood Estimation[1]

[1] I. V. Basawa and J. Zhou. 2004. *Journal of Applied Probability* 41A, 55-64. First published in Journal of Applied Probability 41A: 55-64 (2004).

Abstract

A general class of Markovian non-Gaussian bifurcating models for cell lineage data is presented. Examples include bifurcating autoregression, random coefficient autoregression, bivariate exponential, bivariate gamma, and bivariate Poisson models. Quasilikelihood estimation for the model parameters and large-sample properties of the estimates are discussed.

Keywords: Tree-Indexed Data; Bifurcating Autoregressive Models; Maximum Likelihood; Quasilikelihood Estimation; Markovian Models.

5.1 Introduction

Binary-splitting or bifurcating models are concerned with modeling data on the descendants of an initial individual, where each individual in one generation gives rise to two offspring in the next generation. Cell lineage data (e.g. Powell (1955)) are typically of this kind. Cowan (1984) and Cowan and Staudte (1986) introduced an important model which regarded each line of descent as an autoregressive process and allowed correlations between sister cells. These models are now known as bifurcating autoregressive (BAR) models. If X_t denotes an observation on some characteristic on individual t, the BAR(1) model is specified by the relation

$$X_t = \phi X_{[\frac{t}{2}]} + \epsilon_t, \quad t = 2, 3, \ldots, \tag{5.1.1}$$

where $[u]$ denotes the largest integer $\leq u$. Cowan and Staudte (1986) assumed that $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ form a sequence of independent and identically distributed bivariate normal random variables with means zero, variances both equal to σ^2, and correlation ρ. Huggins and Basawa (1999, 2000) extended the Cowan-Staudte model to higher order autoregression and moving average (ARMA) models and studied the asymptotic properties of the maximum likelihood estimates of the model parameters for BAR(p) models via the time series methods. Huggins and Basawa (2000), however, retained the normality assumption on the errors.

In many applications, the normality assumption may not be realistic. For instance, if X_t denotes the life time of the tth individual, a non-negative random variable, a gamma or an exponential model may be more appropriate. If X_t denotes the number of certain type of genes present, a Poisson model may be considered. In this paper, we introduce a general class of Markovian bifurcating models which includes the BAR model as a special case. If the model is non-Gaussian, the likelihood function would, in general, be complicated, or not even be known. We therefore, use quasilikelihood methodology for estimation

The goal of this paper is to present an exploratory general modeling strategy with some examples. More rigorous treatment with proofs and theoretical considerations along with data analysis will be pursued elsewhere. Section 2 is concerned with model specification. Some examples are presented in Section 3. An outline of asymptotic properties of the estimates is given in Section 4. Quasilikelihood estimation for specific conditional linear bifurcating models is discussed in Section 5. Section 6 contains comments on some extensions of the model.

5.2 Specification of the Model: Likelihood and Quasilikelihood Estimation

In binary cell lineage trees, individual t gives rise to two daughter cells $2t$ and $2t+1$, upon cell division. Let X_t denote an observation on some characteristic on individual t. Suppose we observe complete mother-daughter triplets (X_t, X_{2t}, X_{2t+1}), $t = 1, 2, \dots, n$. Let $X(t) = (X_{[\frac{t}{2^j}]}, j = 0, 1, 2, \dots)$ denote the vector of observations on the ancestors of individuals $(2t, 2t+1)$. Consider the following assumptions:

(A.1) (Markovity): $f((X_{2t}, X_{2t+1})|X(t)) = f((X_{2t}, X_{2t+1})|X_t)$, where $f(\cdot)$ denotes the appropriate conditional density.

(A.2) (Conditional Independence): Conditional on $(X(t), X(s))$, (X_{2t}, X_{2t+1}) is independent of (X_{2s}, X_{2s+1}), for $t \neq s$.

The likelihood function is then given by

$$L_n = f(x_1) \prod_{t=1}^{n} f((X_{2t}, X_{2t+1})|X_t).$$

Suppose the conditional bivariate density $f((X_{2t}, X_{2t+1})|X_t)$ depends on an unknown vector of parameters β and denote the conditional density by $f_\beta((X_{2t}, X_{2t+1})|X_t)$. Conditional on the initial observation $X_1 = x_1$, the likelihood is then defined as

$$L_n(\beta) = \prod_{t=1}^{n} f_\beta((X_{2t}, X_{2t+1})|X_t). \tag{5.2.1}$$

The conditional bivariate density of (X_{2t}, X_{2t+1}) may not be known precisely. One may then consider models based on only second-order moment assumptions. Consider the assumptions

(B.1) $E(X_{2t}|X_t) = E(X_{2t+1}|X_t) = m_t(X_t; \theta)$,

(B.2) $Var(X_{2t}|X_t) = Var(X_{2t+1}|X_t) = v_t(X_t; \theta, \alpha)$,

(B.3) $Cov((X_{2t}, X_{2t+1})|X_t) = \gamma_t(X_t; \theta, \alpha)$,

where m_t, v_t and γ_t are known functions and (θ, α) are unknown parameters. Denote $\beta = (\theta', \alpha')'$. Let $Z_t(\theta) = \begin{pmatrix} X_{2t} - m_t(X_t; \theta) \\ X_{2t+1} - m_t(X_t; \theta) \end{pmatrix}$, and $V_t(\theta, \alpha) = \begin{pmatrix} v_t(X_t; \theta, \alpha) & \gamma_t(X_t; \theta, \alpha) \\ \gamma_t(X_t; \theta, \alpha) & v_t(X_t; \theta, \alpha) \end{pmatrix}$.
If α is known, the quasilikelihood estimating equation for θ is given by

$$\sum_{t=1}^{n} (\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta, \alpha) Z_t(\theta) = 0. \tag{5.2.2}$$

See Godambe (1985) and Heyde (1997) for a background on quasilikelihood estimation. If α is an unknown nuisance parameter, one typically replaces α in $V_t(\theta, \alpha)$ by a consistent estimate $\tilde{\alpha}_n$, and obtains an approximate quasilikelihood estimate from (5.2.2).

5.3 Examples

We present some examples to illustrate the model.

Example 3.1. Bifurcating Autoregression

Cowan and Staudte (1986) introduced the bifurcating autoregressive model defined by

$$X_{2t} = \phi X_t + \epsilon_{2t}$$

and

$$X_{2t+1} = \phi X_t + \epsilon_{2t+1}, \quad |\phi| < 1, \tag{5.3.1}$$

where $(\epsilon_{2t}, \epsilon_{2t+1})'$, $t = 1, 2, \ldots$, are independent identically distributed bivariate normal vectors with means $(0, 0)'$ and the covariance matrix

$$V(\rho, \sigma^2) = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}. \tag{5.3.2}$$

Here, $m_t = \phi X_t$, $v_t = \sigma^2$ and $\gamma_t = \sigma^2 \rho$. Huggins and Basawa (2000) derived the asymptotic properties of the maximum likelihood estimators of the parameters of a pth order Gaussian bifurcating autoregressive model defined by

$$X_t = \phi_1 X_{[\frac{t}{2}]} + \phi_2 X_{[\frac{t}{4}]} + \cdots + \phi_p X_{[\frac{t}{2^p}]} + \epsilon_t, \tag{5.3.3}$$

where $(\epsilon_{2t}, \epsilon_{2t+1})'$ are i.i.d. bivariate normal with the covariance structure given by (5.3.2).

Example 3.2. Random Coefficient Autoregression

Consider the model

$$X_{2t} = \lambda + (\phi + Y_t) X_t + \epsilon_{2t}$$

and

$$X_{2t+1} = \lambda + (\phi + Y_t) X_t + \epsilon_{2t+1} \tag{5.3.4}$$

where $\{Y_t\}$ are i.i.d. with $E(Y_t) = 0$, $Var(Y_t) = \tau^2$, $\phi^2 + \tau^2 < 1$, $(\epsilon_{2t}, \epsilon_{2t+1})'$ are i.i.d. with $E(\epsilon_{2t}) = E(\epsilon_{2t+1}) = 0$ and covariance matrix given by (5.3.2). It is assumed that $\{Y_t\}$ is independent of $(\epsilon_{2t}, \epsilon_{2t+1})'$. Here, we have

$$m_t = \lambda + \phi X_t, \quad v_t = X_t^2 \tau^2 + \sigma^2,$$

and

$$\gamma_t = X_t^2 \tau^2 + \sigma^2 \rho. \tag{5.3.5}$$

In this example, no specific distributional assumptions are made apart from second-order moment assumptions.

Example 3.3. Bivariate Exponential

Consider the bivariate exponential distribution with the distribution function given by

$$F(u,v) = 1 - e^{-(\alpha_1+\alpha_3)u} - e^{-(\alpha_2+\alpha_3)v} + e^{-\alpha_1 u - \alpha_2 v + \alpha_3 \max(u,v)}, \tag{5.3.6}$$

$$u, v > 0, \quad \alpha_1 > 0, \quad \alpha_2 > 0, \quad \alpha_3 > 0.$$

See Mardia (1970). Here, U and V have marginal exponential distributions with means $(\alpha_1 + \alpha_3)^{-1}$ and $(\alpha_2 + \alpha_3)^{-1}$ respectively and $Corr(U,V) = \frac{\alpha_3}{\alpha_1+\alpha_2+\alpha_3}$. Now, suppose that conditional on X_t, (X_{2t}, X_{2t+1}) has the bivariate exponential density given by (5.3.6) with

$$\alpha_1 = \alpha_2 = m_t^{-1}(X_t) - \alpha_3 = m_t^{-1}(X_t)(\frac{1-\rho}{1+\rho}), \ \text{ and } \ \alpha_3 = 2m_t^{-1}(X_t)(\frac{\rho}{1+\rho}), \ \ 0 \le \rho < 1.$$

If we take $m_t(X_t) = \phi X_t + \lambda$, $\lambda > 0$, we have $v_t = (\phi X_t + \lambda)^2$, and $\gamma_t = (\phi X_t + \lambda)^2 \rho$.

Example 3.4. Bivariate Gamma

Consider the bivariate gamma density

$$p(u,v) = \frac{e^{-(u+v)}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} \int_0^{\min(u,v)} w^{\alpha_3-1}(u-w)^{\alpha_1-1}(v-w)^{\alpha_2-1} e^w dw, \tag{5.3.7}$$

$u, v > 0$ and α_1, α_2, $\alpha_3 > 0$. See Kotz et al. (2000). The marginal distributions of U and V are gamma with parameters $(\alpha_1+\alpha_3, 1)$ and $(\alpha_2+\alpha_3, 1)$ respectively with $E(U) = Var(U) = \alpha_1 + \alpha_3$, $E(V) = Var(V) = \alpha_2 + \alpha_3$, and $Cov(U,V) = \alpha_3$. Suppose now that conditional on X_t, (X_{2t}, X_{2t+1}) has the bivariate gamma density in (5.3.7) with

$$\alpha_1 = \alpha_2 = (1-\rho)m_t(X_t) \ \text{ and } \ \alpha_3 = \rho m_t(X_t), \quad 0 \le \rho < 1.$$

If we take $m_t(X_t) = \phi X_t + \lambda$, $\lambda > 0$, we have $v_t = \phi X_t + \lambda$, and $\gamma_t = (\phi X_t + \lambda)\rho$.

Example 3.5. Bivariate Poisson

Suppose (U,V) have a bivariate Poisson distribution with density

$$p(u,v) = e^{-(\theta_1+\theta_2+\theta_3)} \sum_{i=0}^{\min(u,v)} \frac{\theta_1^{u-i}\theta_2^{v-i}\theta_3^i}{(u-i)!(v-i)!i!}, \tag{5.3.8}$$

$u, v \in (0, 1, 2, \dots)$, $\theta_j > 0$, $j = 1, 2, 3$. See Johnson et al. (1997). The marginal distributions of U and V are Poisson with means $\theta_1 + \theta_3$ and $\theta_2 + \theta_3$ respectively, and $Cov(U, V) = \theta_3$. Conditional on X_t, suppose (X_{2t}, X_{2t+1}) has a bivariate Poisson distribution with $\theta_1 = \theta_2 = (1-\rho)m_t(X_t)$ and $\theta_3 = \rho m_t(X_t)$, $0 \le \rho < 1$.

If $m_t(X_t) = \phi X_t + \lambda$, $\lambda > 0$, we have

$$v_t = \lambda + \phi X_t \quad \text{and} \quad \gamma_t = (\lambda + \phi X_t)\rho.$$

Notice the similarity of expressions for m_t, v_t and γ_t for Examples 3.4 and 3.5.

5.4 Remarks on Asymptotic Properties

If the conditional density $f_\beta(X_{2t}, X_{2t+1}|X_t)$ is known, one may derive the maximum likelihood (ML) estimate $\hat{\beta}_{ML}$ of β as a consistent solution of the likelihood equation

$$\sum_{t=1}^{n} \frac{d \log f_\beta}{d\beta} = 0. \tag{5.4.1}$$

Under appropriate regularity conditions, see, for instance, Billingsley (1961), and Basawa and Prakasa Rao (1980), one can establish the consistency and asymptotic normality of the ML estimate $\hat{\beta}_{ML}$. In particular, we have

$$\sqrt{n}(\hat{\beta}_{ML} - \beta) \xrightarrow{d} N(0, I^{-1}(\beta)), \tag{5.4.2}$$

where $I(\beta) = E((\frac{d \log f_\beta}{d\beta})(\frac{d \log f_\beta}{d\beta})')$, and the expectation is with respect to the stationary distribution. Huggins and Basawa (2000) have established the result in (5.4.2) for the Gaussian model (see Ex. 3.1). The same techniques can be used for the non-Gaussian models.

Let $\hat{\theta}_{QL}$ denote a consistent solution of the (approximate) quasilikelihood equation

$$\sum_{t=1}^{n} (\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta, \tilde{\alpha}_n) Z_t(\theta) = 0, \tag{5.4.3}$$

where $Z_t(\theta)$ and $V_t(\theta, \alpha)$ are as defined in Section 2 and $\tilde{\alpha}_n$ is a $\sqrt{n}$-consistent estimate of α. Under appropriate regularity conditions, see, for instance, Heyde (1997), one can established the result

$$\sqrt{n}(\hat{\theta}_{QL} - \theta) \xrightarrow{d} N(0, A^{-1}(\theta, \alpha)), \tag{5.4.4}$$

where $A(\theta,\alpha) = E[(\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta,\alpha)(\frac{dZ_t(\theta)}{d\theta})]$, and the expectation is with respect to the stationary distribution and hence A does not depend on t.

Even though the ML estimate is asymptotically more efficient than the quasilikelihood estimate in most cases, we may choose to use quasilikelihood estimate when the likelihood function is too unwieldy (as happens in the non-Gaussian examples presented in Section 3) or when only information on conditional second-order moments is available. A class of models for which the quasilikelihood method can be applied readily is discussed in the next section. Note that in the ML method, both the parameters, θ and α are estimated simultaneously, where as in the QL method, we are mainly interested in estimating θ treating α as an unknown nuisance parameter. Even though it is possible to estimate θ and α simultaneously via an extended version of the QL method, we will not address such an extension in this paper.

5.5 Non-Gaussian Conditional Linear Bifurcating Models

Let $\{Y_t\}$, $t = 0, 1, 2, \ldots$, denote a Markov process. Grunwald, et al. (2000) have studied non-Gaussian Markov models for which the conditional mean $E(Y_t|Y_{t-1}) = m(Y_{t-1})$ is of the linear form

$$m(Y_{t-1}) = \phi Y_{t-1} + \lambda. \tag{5.5.1}$$

Grunwald, et al. (2000) refer to the model satisfying (5.1) as a first-order conditional linear autoregressive (CLAR(1)) model. They show that a surprisingly large number of models in the literature belong to the CLAR(1) family. Grunwald, et al. (2000) have established simple sufficient conditions for the ergodicity of the Markov process $\{Y_t\}$ satisfying (5.5.1). If $\mathcal{Y}$ denotes the state space of $\{Y_t\}$, the key conditions for ergodicity are given by

Case 1: $E[|Y_t - m(Y_{t-1})|Y_{t-1} = y] < B$, for all y, and some finite B and $|\phi| < 1$, for $\mathcal{Y} = \Re = (-\infty, \infty)$, and,

Case 2: $0 \le \phi < 1$, for $\mathcal{Y} \subseteq [0, \infty)$.

Note, in particular, that the boundedness condition on $(Y_t - m(Y_{t-1}))$ is not needed for Case 2 (non-negative process $\{Y_t\}$).

These results can readily be extended to the bifurcating models in an obvious way. On any ancestral path $\{X_t, X_{[\frac{t}{2}]}, X_{[\frac{t}{4}]}, \dots\}$ we assume a first-order Markov process. Set $Y_{t-j} = X_{[\frac{t}{2^j}]}$, $j = 0, 1, 2, \dots$, and apply the results of Grunwald, et al. (2000) to establish ergodicity. It will now be assumed that the process $\{X_t\}$ is ergodic along each ancentral path. It will further be assumed that the stationary distribution along each path is the same.

Consider the Markovian bifurcating model $\{X_t\}$ specified by the second-order moment assumptions (B.1) to (B.3) in Section 2. In particular, suppose we further require the mean function m_t to satisfy

$$m_t(X_t; \theta) = E(X_{2t}|X_t) = E(X_{2t+1}|X_t) = \phi X_t + \lambda. \tag{5.5.2}$$

Here, $\theta = (\phi, \lambda)'$. Recall the notation from Section 2

$$Var(X_{2t}|X_t) = Var(X_{2t+1}|X_t) = v_t(\theta, \alpha) \tag{5.5.3}$$

and

$$Cov((X_{2t+1}, X_{2t+1})|X_t) = \gamma_t(\theta, \alpha), \tag{5.5.4}$$

where α is an unknown nuisance parameter. See Section 3 for examples. Suppose our main goal is to estimate θ. We now present a two-step method of estimating the nuisance parameter α, which in turn, will be needed for estimating θ.

Step 1. Find the conditional least squares (CLS) estimate of θ by minimizing

$$Q_1 = \sum_{t=1}^{n} (X_{2t} - \phi X_t - \lambda)^2 + \sum_{t=1}^{n} (X_{2t+1} - \phi X_t - \lambda)^2. \tag{5.5.5}$$

We have

$$\tilde{\phi}_0 = \frac{\sum_{t=1}^{n} X_t(U_t - \bar{U})}{\sum_{t=1}^{n} (X_t - \bar{X})^2}, \quad \text{and} \quad \tilde{\lambda}_0 = \bar{U} - \tilde{\phi}_0 \bar{X}, \tag{5.5.6}$$

where $U_t = \frac{X_{2t} + X_{2t+1}}{2}$, $\bar{U} = \frac{1}{n} \sum_{t=1}^{n} U_t$ and $\bar{X} = \frac{1}{n} \sum_{t=1}^{n} X_t$. Let $\tilde{\theta}_0 = (\tilde{\phi}_0, \tilde{\lambda}_0)'$.

Step 2. Find the (approximate) conditional least squares estimate of α by minimizing

$$Q_2 = \sum_{t=1}^{n}((X_{2t} - \tilde{\lambda}_0 - \tilde{\phi}_0 X_t)^2 - v_t(\tilde{\theta}_0, \alpha))^2 + \sum_{t=1}^{n}((X_{2t+1} - \tilde{\lambda}_0 - \tilde{\phi}_0 X_t)^2 - v_t(\tilde{\theta}_0, \alpha))^2$$

$$+ \sum_{t=1}^{n}((X_{2t} - \tilde{\lambda}_0 - \tilde{\phi}_0 X_t)(X_{2t+1} - \tilde{\lambda}_0 - \tilde{\phi}_0 X_t) - \gamma_t(\tilde{\theta}_0, \alpha))^2. \quad (5.5.7)$$

Let $\tilde{\alpha}_0$ denote the estimate of α so obtained. It can be shown, via the techniques similar to the ones used by Klimko and Nelson (1978) that $\tilde{\alpha}_0$ is a $\sqrt{n}$-consistent estimate. An approximate quasilikelihood estimate $\hat{\theta}_{QL}$ of θ is then obtained by solving equation (5.2.2), with α replaced by $\tilde{\alpha}_0$. The limit distribution of $\hat{\theta}_{QL}$ is given by (5.4.4). A consistent estimate of the quasi-information matrix $A(\theta, \alpha)$ appearing in (5.4.4) is given by

$$\hat{A}_n = \frac{1}{n}\sum_{t=1}^{n}[(\frac{dZ_t(\theta)}{d\theta})'_{\tilde{\theta}_0} V_t^{-1}(\tilde{\theta}_0, \tilde{\alpha}_0)(\frac{dZ_t(\theta)}{d\theta})_{\tilde{\theta}_0}]. \quad (5.5.8)$$

We now return to the estimation problem for the examples discussed in Section 3.

For any model satisfying (5.5.2) to (5.5.4) with $\lambda = 0$ (for simplicity), it can be verified that (5.2.2) leads to the equation

$$\hat{\phi}_{QL} = \frac{\sum_{t=1}^{n}(v_t + \gamma_t)^{-1} X_t U_t}{\sum_{t=1}^{n}(v_t + \gamma_t)^{-1} X_t^2}, \quad (5.5.9)$$

where $U_t = \frac{1}{2}(X_{2t} + X_{2t+1})$. Note that $v_t + \gamma_t$ may depend on unknown parameters. From (5.4.4) we have

$$\sqrt{n}(\hat{\phi}_{QL} - \phi) \xrightarrow{d} N(0, \frac{1}{2}(E(\frac{X_t^2}{v_t + \gamma_t}))^{-1}), \quad (5.5.10)$$

where the expectation $E(\cdot)$ is with respect to the stationary distribution. In all the examples discussed below, the right hand side of eqn. (5.5.9) is free from ϕ. Except for Ex 3.2, $\hat{\phi}_{QL}$ is also free from the nuisance parameter α. In Ex 3.2, $\hat{\phi}_{QL}$ depends on the nuisance parameter α which needs to be estimated before using (5.5.9) as an estimate of ϕ.

The result in (5.5.10) can be verified as follows. From (5.5.9) we have

$$\sqrt{n}(\hat{\phi}_{QL} - \phi) = [\frac{1}{n}\sum_{t=1}^{n}(v_t + \gamma_t)^{-1} X_t^2]^{-1}[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} M_t]$$

where

$$M_t = \sum_{t=1}^{n} (v_t + \gamma_t)^{-1} X_t (U_t - \phi X_t).$$

Note that $\{M_t\}$ is a zero-mean martingale-difference sequence with respect to the σ-field generated by $X(t) = \{X_{[\frac{t}{2^j}]}, \quad j = 0, 1, 2, \dots\}$. It is seen that

$$Var(M_t | X(t)) = \frac{1}{2}(v_t + \gamma_t)^{-1} X_t^2.$$

Suppose that

$$\frac{1}{n} \sum_{t=1}^{n} (v_t + \gamma_t)^{-1} X_t^2 \xrightarrow{p} B, \quad 0 < B < \infty.$$

One can identify, via ergodicity, that $B = E[(v_t + \gamma_t)^{-1} X_t^2]$ where the expectation is with respect to the stationary distribution. Under regularity conditions (see Hall and Heyde (1980)), it follows, by martingale central limit theorem, that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} M_t \xrightarrow{d} N(0, \frac{1}{2} B).$$

This result together with Slutsky's theorem finally yields (5.5.10).

Example 3.1 (Contd.)

Consider the model defined by (5.3.1) and (5.3.2) without the assumption of normality of errors.

We have $m_t = \phi X_t$, $|\phi| < 1$, $v_t = \sigma^2$ and $\gamma_t = \sigma^2 \rho$. It is seen that the quasilikelihood equation

$$\sum_{t=0}^{n} (\frac{dZ_t}{d\phi}) V^{-1} Z_t = 0$$

leads to

$$\hat{\phi}_{QL} = \frac{\Sigma X_t U_t}{\Sigma X_t^2} \tag{5.5.11}$$

where $U_t = \frac{1}{2}(X_{2t} + X_{2t+1})$. Here, the quasilikelihood estimate reduces to the simple conditional least squares (CLS) estimate which does not depend on σ^2 and ρ. The CLS estimates of the nuisance parameters σ^2 and $\gamma = \sigma^2 \rho$ are given by

$$\tilde{\sigma}_0^2 = \frac{1}{2n} [\sum_{t=1}^{n} (X_{2t} - \hat{\phi}_{QL} X_t)^2 + \sum_{t=1}^{n} (X_{2t+1} - \hat{\phi}_{QL} X_t)^2],$$

and

$$\tilde{\gamma}_0 = \frac{1}{n}\sum_{t=1}^{n}(X_{2t} - \hat{\phi}_{QL}X_t)(X_{2t+1} - \hat{\phi}_{QL}X_t).$$

Consequently, $\tilde{\rho}_0 = \tilde{\gamma}_0/\tilde{\sigma}_0^2$.

The quasilikelihood information corresponding to ϕ is given by

$$E[(\frac{dZ_t}{d\phi})'V^{-1}(\frac{dZ_t}{d\phi})] = 2\sigma^{-2}(\frac{EX_t^2}{(1+\rho)}) = 2(1+\rho)^{-1}(1-\phi^2)^{-1},$$

since $EX_t^2 = \frac{\sigma^2}{(1-\phi^2)}$. See Huggins and Basawa (1999, 2000).

We thus have

$$\sqrt{n}(\hat{\phi}_{QL} - \phi) \xrightarrow{d} N(0, \frac{1}{2}(1+\rho)(1-\phi^2)). \tag{5.5.12}$$

The asymptotic variance in (5.5.12) coincides with the corresponding asymptotic variance of $\hat{\phi}_{ML}$ for the BAR(1) Gaussian model obtained by Huggins and Basawa (2000). Hence, if the errors are normal, $\hat{\phi}_{QL}$ has the same limit distribution as $\hat{\phi}_{ML}$.

Example 3.2 (Contd.)

For the random coefficient BAR(1) model, m_t, v_t and γ_t are given by (5.3.5). Note that no distributional assumptions on the errors are made. Set $\lambda = 0$ for simplicity. If the nuisance parameters σ^2, τ^2 and ρ are known, the quasilikelihood estimate of ϕ is seen to be

$$\hat{\phi}_{QL} = \frac{\sum_{t=1}^{n} X_t U_t W_t^{-1}}{\sum_{t=1}^{n} X_t^2 W_t^{-1}}, \tag{5.5.13}$$

where $U_t = \frac{1}{2}(X_{2t}+X_{2t+1})$ and $W_t = 2X_t^2\tau^2+\sigma^2(1+\rho)$. If $\tau^2 = 0$, (5.5.13) reduces to (5.5.11) as it should. The quasi-information is given by

$$E[(\frac{dZ_t}{d\phi})'V_t^{-1}\frac{dZ_t}{d\phi}] = 2E(X_t^2W_t^{-1}),$$

where the expectation is with respect to the stationary distribution. Consequently, we have

$$\sqrt{n}(\hat{\phi}_{QL} - \phi) \xrightarrow{d} N(0, \frac{1}{2}(E(X_t^2W_t^{-1}))^{-1}). \tag{5.5.14}$$

The CLS estimates of the nuisance parameters σ^2 and τ^2 are given by

$$\begin{pmatrix} \tilde{\sigma}_0^2 \\ \tilde{\tau}_0^2 \end{pmatrix} = \begin{pmatrix} n & \Sigma X_t^2 \\ \Sigma X_t^2 & \Sigma X_t^4 \end{pmatrix}^{-1} \begin{pmatrix} \Sigma M_t \\ \Sigma M_t X_t^2 \end{pmatrix}, \tag{5.5.15}$$

where $M_t = \frac{1}{2}[(X_{2t} - \tilde{\phi}_0 X_t)^2 + (X_{2t+1} - \tilde{\phi}_0 X_t)^2]$, and $\tilde{\phi}_0 = (\sum_{t=1}^n X_t U_t)(\sum_{t=1}^n X_t^2)^{-1}$. The estimates in (5.5.15) are obtained by minimizing

$$\Sigma[(X_{2t} - \tilde{\phi}_0 X_t)^2 - \gamma_t)]^2 + \Sigma[(X_{2t+1} - \tilde{\phi}_0 X_t)^2 - \gamma_t]^2$$

ignoring the covariance term. Finally, a consistent estimate of ρ is

$$\tilde{\rho}_0 = \frac{1}{n}\tilde{\sigma}_0^{-2}\sum_{t=1}^{n}(X_{2t} - \tilde{\phi}_0 X_t)(X_{2t+1} - \tilde{\phi}_0 X_t). \tag{5.5.16}$$

Substituting $\tilde{\sigma}_0^2$, $\tilde{\tau}_0^2$ and $\tilde{\rho}_0$ for σ^2, τ^2 and ρ in (5.5.13) we obtain an approximate quasilikelihood estimate whose limit distribution is again given by (5.5.14). A consistent estimate of $E(X_t^2 W_t^{-1})$ in (5.14) is seen to be $\frac{1}{n}\sum_{t=1}^{n} X_t^2 \tilde{W}_t^{-1}$ where $\tilde{W}_t = 2X_t^2\tilde{\tau}_0^2 + \tilde{\sigma}_0^2(1+\tilde{\rho}_0)$.

Example 3.3 (Contd.)

Recall that $m_t = \phi X_t + \lambda$, $v_t = (\phi X_t + \lambda)^2$ and $\gamma_t = (\phi X_t + \lambda)^2\rho$. Assume $\lambda > 0$ and $0 < \phi < 1$. The quasilikelihood estimating equation for $\theta = (\phi, \lambda)'$ is seen to be

$$\begin{pmatrix} \Sigma(U_t - m_t)X_t m_t^{-2} \\ \Sigma(U_t - m_t)m_t^{-2} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \tag{5.5.17}$$

which is free from the nuisance parameters ρ. From (5.4.4) we have

$$\sqrt{n}(\hat{\theta}_{QL} - \theta) \xrightarrow{d} N(0, A^{-1}(\theta))$$

where

$$A(\theta) = \frac{2}{(1+\rho)}\begin{pmatrix} E(X_t^2 m_t^{-2}) & E(X_t m_t^{-2}) \\ E(X_t m_t^{-2}) & E(m_t^{-2}) \end{pmatrix}, \tag{5.5.18}$$

the expectation being with respect to the stationary distribution.

Estimation for Examples 3.4 and 3.5 can be carried out in a similar manner and we omit the details.

5.6 Concluding Remarks

In this paper, we have considered first order bifurcating models where (X_{2t}, X_{2t+1}) depend only on the immediate parent observation X_t. These models can be extended to higher order models representing dependence on the past p ancestors as in Huggins and Basawa (2000).

Covariates can be introduced by considering a mean function satisfying

$$E(X_{2t}|X_t) = \phi X_t + \lambda + c'_{2t}\gamma - \phi c'_t\gamma,$$

and

$$E(X_{2t+1}|X_t) = \phi X_t + \lambda + c'_{2t+1}\gamma - \phi c'_t\gamma,$$

where c_i is a vector of known covariates associated with individual i, and γ is the regression parameter.

We have assumed in this paper that complete (balanced) data, viz., (X_t, X_{2t}, X_{2t+1}), $t = 1, 2, \ldots, n$, are available. If some data are missing, appropriate modifications of the estimation procedure can be made as indicated by Cowan and Staudte (1986).

Finally, generalized linear models can be used to model m_t. For instance, for a given link function $g(\cdot)$, one may consider the model $g(m_t) = \phi X_t + \lambda$. Conditions for stationarity for such models need to be explored.

5.7 References

[1] Basawa, I. V. and B. L. S. Prakasa Rao (1980). *Statistical Inference for Stochastic Processes*, Academic Press, London.

[2] Billingsley, P. (1961). *Statistical Inference for Markov Processes*, Chicago Univ. Press, Chicago.

[3] Cowan, R. (1984). Statistical concepts in the analysis of cell lineage data. In *Proceedings of the 1983 Workshop on Cell Growth and Division*, p. 18-22, Latrobe Uni., Melbourne.

[4] Cowan, R. and R. G. Staudte (1986). The bifurcating autoregression model in cell lineage studies. *Biometrics* 42, 769-783.

[5] Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* 31, 1208-1212.

[6] Godambe, V. P. (1985). The foundations of finite sample estimation for stochastic processes. *Biometrika* 72, 419-428.

[7] Grunwald, G. K., R. J. Hyndman, L. Tedesco and R. L. Tweedie (2000). Non-Gaussian conditional AR(1) models. *Aust. N. Z. J. Stat.* 42, 479-495.

[8] Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[9] Heyde, C. C. (1997). *Quasilikelihood and Its Applications.* Springer, New York.

[10] Huggins, R. M. and I. V. Basawa (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Prob.* 36, 1225-1233.

[11] Huggins, R. M. and I. V. Basawa (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* 42, 423-432.

[12] Johnson, N. L., S. Kotz and N. Balakrishnan (1997). *Discrete Multivariate Distributions*, Wiley, New York.

[13] Klimko, L. A. and P. I. Nelson (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* 6, 629-642.

[14] Kotz, S., N. Balakrishnan and N. L. Johnson (2000). *Continuous Multivariate Distributions, Vol 1: Models and Applications*, 2nd Ed., Wiley, New York.

[15] Mardia, K. V. (1970). *Families of Bivariate Distributions.* Griffin, London.

[16] Powell, E. O. (1955). Some features of the generation times of individual bacteria. *Biometrika* 42, 16-44.

Chapter 6

Future Topics

There are several topics that we are interested in and intend to develop in the future.

6.1 Bifurcating Random Walk with Drift

In Chapter 3 we discussed the least squares (LS) estimation of BAR(p) model

$$X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \phi_2 X_{[\frac{t}{4}]} + \cdots + \phi_p X_{[\frac{t}{2^p}]} + \epsilon_t, \tag{6.1.1}$$

where the roots of $\phi(z) = 0$ are assumed to be greater than 1 in absolute value. Specifically, in the BAR(1) model $X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \epsilon_t$, we assume $|\phi_1| \leq 1$. A natural question is whether the asymptotic properties of LS estimators can be extended to the critical case where $\phi_1 = 1$. In this section, we will discuss the least squares estimation of bifurcating random walk with drift, i.e. BAR(1) model with $\phi_0 \neq 0$ and $\phi_1 = 1$.

Consider the bifurcating random walk with drift model

$$X_t = \phi_0 + \phi_1 X_{[\frac{t}{2}]} + \epsilon_t, \quad t = 2, 3, ..., n, \tag{6.1.2}$$

with assumption that X_1 is constant, $\phi_0 \neq 0, \phi_1 = 1$, and $\{(\epsilon_{2t}, \epsilon_{2t+1})\}$ is a sequence of independent identically distributed (i.i.d.) bivariate random variables with $E(\epsilon_{2t}) = E(\epsilon_{2t+1}) = 0$, $Var(\epsilon_{2t}) = Var(\epsilon_{2t+1}) = \sigma^2$, and $Corr(\epsilon_{2t}, \epsilon_{2t+1}) = \rho$. The least squares estimators of ϕ_0 and ϕ_1 are given by

$$\begin{pmatrix} \hat{\phi}_0 \\ \hat{\phi}_1 \end{pmatrix} = \begin{pmatrix} n-1 & \sum_2^n X_{[\frac{t}{2}]} \\ \sum_2^n X_{[\frac{t}{2}]} & \sum_2^n X^2_{[\frac{t}{2}]} \end{pmatrix}^{-1} \begin{pmatrix} \sum_2^n X_t \\ \sum_2^n X_{[\frac{t}{2}]} X_t \end{pmatrix} \tag{6.1.3}$$

or

$$\begin{pmatrix} \hat{\phi}_0 - \phi_0 \\ \hat{\phi}_1 - \phi_1 \end{pmatrix} = \begin{pmatrix} n-1 & \sum_2^n X_{[\frac{t}{2}]} \\ \sum_2^n X_{[\frac{t}{2}]} & \sum_2^n X^2_{[\frac{t}{2}]} \end{pmatrix}^{-1} \begin{pmatrix} \sum_2^n \epsilon_t \\ \sum_2^n X_{[\frac{t}{2}]}\epsilon_t \end{pmatrix} \tag{6.1.4}$$

i.e.

$$\hat{\phi}_1 - \phi_1 = \frac{\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})\epsilon_t}{\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})^2}, \quad \hat{\phi}_0 - \phi_0 = \bar{\epsilon} - (\hat{\phi}_1 - \phi_1)\bar{X}, \tag{6.1.5}$$

where $\bar{X} = \sum_2^n X_{[\frac{t}{2}]}/(n-1)$ and $\bar{\epsilon} = \sum_2^n \epsilon_t/(n-1)$.

Define $W_t = \sum_{j=0}^{[\log_2 t]-1} \epsilon_{[\frac{t}{2^j}]}, \bar{W} = \sum_1^n W_t/n$, and $\bar{W}^* = \sum_2^n W_{[\frac{t}{2}]}/(n-1)$, where $[u]$ denotes the largest integer less than or equal to u. We can write, for example, $\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})^2$, and $\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})\epsilon_t$, as sum of several terms including ϕ_0, W_t and t respectively. Our guess is that the terms including ϕ_0 dominate the other terms with respect to asymptotic properties. We provide the following two conjectures which are the main keys to further work and need to be proved (or disproved) in the future.

Conjecture 1. $\sum_1^n (W_t - \bar{W})^2 = o_p(n)$

Conjecture 2. $\sum_2^n (W_{[\frac{t}{2}]} - \bar{W}^*)\epsilon_t = o_p(\sqrt{n})$

Under Conjecture 1 and Conjecture 2, we have

Proposition 1. $\frac{\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})^2}{n} \xrightarrow{p} 2\phi_0^2$

Proposition 2. $\frac{\sum_2^n (X_{[\frac{t}{2}]} - \bar{X})\epsilon_t}{\sqrt{n}} \xrightarrow{d} N(0, 2\phi_0^2\sigma^2(1+\rho))$

Proposition 3. $\frac{\sqrt{n}(\hat{\phi}_0 - \phi_0)}{\log_2 n} = -\phi_0\sqrt{n}(\hat{\phi}_1 - \phi_1) + o_p(1)$

From Proposition 1, 2 and 3, we have

Theorem 1.

$$\begin{pmatrix} \frac{\sqrt{n}(\hat{\phi}_0 - \phi_0)}{\log_2 n} \\ \sqrt{n}(\hat{\phi}_1 - \phi_1) \end{pmatrix} \xrightarrow{d} N\left(0, \frac{\sigma^2(1+\rho)}{2} \begin{pmatrix} 1 & -\frac{1}{\phi_0} \\ -\frac{1}{\phi_0} & \frac{1}{\phi_0^2} \end{pmatrix}\right). \tag{6.1.6}$$

It is to be noted that the asymptotic covariance matrix in Theorem 1 is singular.

Simulation study can be done to provide some support for the above theorem, but the final result will depend on the verification of the two conjectures.

6.2 Consistency and Asymptotic Normality of QL Estimates

In section 5.2, we get the estimating function

$$S_n^*(\theta) = \sum_{t=1}^{n} (\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta, \alpha) Z_t(\theta). \tag{6.2.1}$$

where $Z_t(\theta)$ and $V_t^{-1}(\theta, \alpha)$ are defined as in section 5.2. If α is an unknown nuisance parameter, one typically replaces α in $V_t(\theta, \alpha)$ by a $\sqrt{n}$-consistent estimate $\tilde{\alpha}_n$.

Consider the class of estimating functions $S_n(\theta) = \sum_{t=1}^{n} W_t(X_t, \theta) Z_t(X_{2t}, X_{2t+1}, X_t, \theta)$, where $E_t(Z_t) = 0$, and $E_t(\frac{dZ_t(\theta)}{d\theta}) \neq 0$. Then the optimum weights $\{W_t^*\}$, according to the Godambe (1985) criterion, are given by

$$W_t^* = [E_t(\frac{dZ_t(\theta)}{d\theta})]' V_t^{-1}(Z_t) = (\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta, \alpha), \tag{6.2.2}$$

where E_t denotes the conditional expectation with respect to $X_{(t)}$. Hence we see that $S_n^*(\theta)$ is the optimal estimating function in the class of $S_n(\theta)$. Let $\hat{\theta}_n$ be a consistent solution of the equation $S_n^*(\theta) = 0$. By Taylor's expansion, we have

$$S_n^*(\hat{\theta}_n) = S_n^*(\theta) + (\hat{\theta}_n - \theta)(\frac{dS_n^*(\theta)}{d\theta})_{\theta_n^*},$$

where θ_n^* lies in the circle $\{\theta : ||\theta - \hat{\theta}_n|| \leq \epsilon\}$. Setting $S_n^*(\hat{\theta}_n) = 0$, we have

$$\sqrt{n}(\hat{\theta}_n - \theta) = -[\frac{1}{n}(\frac{dS_n^*(\theta)}{d\theta})_{\theta_n^*}]^{-1}[\frac{1}{\sqrt{n}} S_n^*(\theta)]. \tag{6.2.3}$$

By the central limit theorem for martingales,

$$\frac{1}{\sqrt{n}} S_n^*(\theta) \xrightarrow{d} N(0, A(\theta)), \tag{6.2.4}$$

where $A(\theta) = E[(\frac{dZ_t(\theta)}{d\theta})' V_t^{-1}(\theta)(\frac{dZ_t(\theta)}{d\theta}))]$ and the expectation is with respect to the stationary distribution and hence $A(\theta)$ does not depend on t. By the law of large numbers for Markov processes,

$$\frac{1}{n}(\frac{dS_n^*(\theta)}{d\theta}) \xrightarrow{p} E[(\frac{dZ_t(\theta)}{d\theta})'V_t^{-1}(\theta)(\frac{dZ_t(\theta)}{d\theta})] = A(\theta), \tag{6.2.5}$$

Assuming $\frac{1}{n}[\frac{dS_n^*(\theta)}{d\theta} - (\frac{dS_n^*(\theta)}{d\theta})_{\theta_n^*}] \xrightarrow{p} 0$,we get

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, A^{-1}(\theta)). \tag{6.2.6}$$

The general methodology indicated above can be used to establish the consistency and asymptotic normality of the QL estimates.

6.3 Multiple-Splitting Model

The multiple-splitting model is a generalization of the bifurcating model. Suppose in cell division, each individual produces k daughter cells. Let X_t denote an observation on some characteristic on individual t, then observations on its daughter cells are denoted as $X_{kt}, X_{kt+1}, ..., X_{kt+k-1}$. Here k is a positive integer greater than 2 and is fixed.

We can make the following assumptions:

(A$'$.1) (Markovity): $f((X_{kt}, ..., X_{kt+k-1})|X(t)) = f((X_{kt}, ..., X_{kt+k-1})|X_t)$, where $f(\cdot)$ denotes the appropriate conditional density and $X(t) = (X_{[t/k^j]}, j = 0, 1, 2, \dots)$ denotes the vector of observations on the ancestors of individuals $(kt, ..., kt + k - 1)$.

(A$'$.2)(Conditional Independence): Conditional on $(X(t), X(s))$, $(X_{kt}, ..., X_{kt+k-1})$ is independent of $(X_{ks}, ..., X_{ks+k-1})$, for $t \neq s$.

When the likelihood function is not known precisely, one may consider models based on only second-order moment assumptions. Consider the assumptions:

(B$'$.1) $E(X_{kt+j}|X_t) = m_t(X_t; \theta), \quad j = 0, 1, ..., k - 1,$

(B$'$.2) $Var(X_{kt+j}|X_t) = v_t(X_t; \theta, \alpha) \quad j = 0, 1, ..., k - 1,$

(B$'$.3) $Cov((X_{2t+i}, X_{2t+j})|X_t) = \gamma_t(X_t; \theta, \alpha), \quad i \neq j, 0 \leq i, j \leq k - 1.$

where m_t, v_t and γ_t are known functions and (θ, α) are unknown parameters.

A noticeable point here is that we assume the correlations of daughters of one individual are the same, which is reasonable since daughters live in a similar environment.

Using the same approach as in Chapter 5, we can get the quasilikelihood estimating equation for θ and hence the quasilikelihood estmator. The asymptotic properties will be studied subsequently.

Some examples of multiple-splitting models are as follows.

Example 1 Multiple-Splitting Autoregression

A mutiple-splitting autoregressive (MSAR) model is defined by

$$X_t = \phi X_{[t/k]} + \epsilon_t, t = 1, 2, ..., |\phi| < 1 \qquad (6.3.1)$$

where $(\epsilon_{kt}, \epsilon_{kt+1}, ..., \epsilon_{kt+k-1})\prime$, $t = 1, 2, \ldots,$ are independent identically distributed k-dimensional vectors with $k \times 1$ mean vector $(\mu, ..., \mu)'$ and $k \times k$ covariance matrix V with diagonal elements σ^2 and off-diagonal elements $\varphi\sigma^2$.

Example 2 Multivariate Gamma

Consider the multivariate gamma density

$$p(u_1, u_2, ..., u_k) = \frac{e^{-\sum_{i=1}^k u_i}}{\prod_{i=0}^k \Gamma(\alpha_i)} \int_0^{u_{(1)}} w^{\alpha_0 - 1} \{\prod_{i=0}^k (u_i - w)^{\alpha_i - 1}\} e^{(k-1)w} dw, \qquad (6.3.2)$$

where $u_{(1)} = \min(u_1, u_2, ..., u_k)$, $u_1, u_2, ..., u_k > 0$ and α_0, α_1,..., $\alpha_k > 0$. See Kotz *et al.* (2000). The marginal distribution of $U_i, i = 1, 2, ..., k$, is gamma with parameters $(\alpha_i + \alpha_0, 1)$. $E(U_i) = Var(U_i) = \alpha_i + \alpha_0$ and $Cov(U_i, U_j) = \alpha_0$. Suppose now that conditional on X_t, $(X_{kt}, X_{kt+1}, ..., X_{kt+k-1})$ has the above multivariate gamma density with

$$\alpha_1 = \alpha_2 = ... = \alpha_k = (1 - \varphi) m_t(X_t) \text{ and } \alpha_0 = \varphi m_t(X_t), \quad 0 \leq \varphi < 1.$$

If we take $m_t(X_t) = \phi X_t + \lambda$, $\lambda > 0$, we have $v_t = \phi X_t + \lambda$, and $\gamma_t = (\phi X_t + \lambda)\varphi$.

Example 3 Multivariate Poisson

A version of multivariate Poisson distribution is defined by setting

$$U_i = Y_i + Y \ \ , i = 1, 2, ..., k,$$

where $Y, Y_1, ..., Y_k$ are mutually independent Poisson random variables with means $\theta, \theta_1, ..., \theta_k$, respectively. Clearly, $U_1, U_2, ..., U_k$ marginally have Poisson distributions with means $\theta_1 +$

$\theta, \theta_2 + \theta, ..., \theta_k + \theta$, respectively, and the covariance between U_i and U_j is θ. Suppose now that conditional on X_t, $(X_{kt}, X_{kt+1}, ..., X_{kt+k-1})$ has the multivariate Poisson density with

$$\theta_1 = \theta_2 = ... = \theta_k = (1-\varphi)m_t(X_t) \text{ and } \theta = \varphi m_t(X_t), \quad 0 \leq \varphi < 1.$$

If we take $m_t(X_t) = \phi X_t + \lambda$, $\lambda > 0$, we have $v_t = \phi X_t + \lambda$, and $\gamma_t = (\phi X_t + \lambda)\varphi$.

Until now we assume that k is fixed. A more realistic assumption is that k is a random variable with some specified distribution. This question remains open and will be studied in the future.

Bibliography

[1] Al-Osh, M. A. and Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Series Analysis* 8, 261-275.

[2] Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical Inference for Stochastic Processes*. Academic Press, London.

[3] Basawa, I. V. and Zhou, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Prob.* 41A, 55-64.

[4] Billingsley, P. (1961). *Statistical Inference for Markov Processes*. Chicago Univ. Press, Chicago.

[5] Brockwell, P. J. and Davis, R. A. (1987). *Time Series: Theory and Methods*. Springer-Verlog, New York.

[6] Brooks, R. F., Bennett, D. C., and Smith, J. A. (1980). Mammalian cell cycles need two random transitions. *Cell* 19, 493.

[7] Bui, Q. M. and Huggins, R. M. (1998). Robust inference for the bivariate bifurcating autoregressive model. *Aust. N. Z. J. Stat.* 40(2), 151-163.

[8] Bui, Q. M. and Huggins, R. M. (1999). Inference for the random coefficients bifurcating autoregressive model for cell lineage studies. *J. Statist. Planning & Inf.* 81, 253-262.

[9] Collyn D'Hooge, M. P., Valleron, A. J., and Malaise, E. P. (1977). Time lapse cinematography studies of cell cycle and mitois duration. *Experimental Cell Research* 106, 405-407.

[10] COWAN, R. (1984). Statistical concepts in the analysis of cell lineage data. In *Proceedings of the 1983 Workshop on Cell Growth and Division*, p. 18-22. Latrobe Uni., Melbourne.

[11] COWAN, R. AND STAUDTE, R. G. (1986). The bifurcating autoregression model in cell lineage studies. *Biometrics* 42, 769-783.

[12] DAVIS, R. A. AND MCCORMICK, W. P. (1989). Estimation for the first-order autoregressive processes with positive or bounded innovations. *Stoch. Proc. App.* 31, 237-250.

[13] FAN, J. AND YAO, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

[14] FEIGIN, P. AND TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* 6, 1-14.

[15] FEIGIN, P. AND TWEEDIE, R. L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* 6, 1-14.

[16] GAVER, D. P. AND LEWIS, P. A. (1980). First-order autoregressive Gamma sequences and point processes. *Adv. Appl. Prob.* 12, 726-745.

[17] GODAMBE, V. P. (1985). The foundations of finite sample estimation for stochastic processes. *Biometrika* 72, 419-428.

[18] GRIGORIU, M. (1995). *Applied Non-Gaussian Processes.* Prentice Hall, NJ.

[19] GRUNWALD, G. K., HYNDMAN, R. J., TEDESCO, L. AND TWEEDIE, R. L. (2000). Non-Gaussian conditional AR(1) models. *Aust. N. Z. J. Stat.* 42, 479-495.

[20] HALL, P. AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[21] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton, N.J.

[22] Heyde, C. C. (1997). *Quasilikelihood and Its Applications*. Springer, New York.

[23] Hola, M. and Riley, P. A. (1987). The relative significance of growth rate and interdivision time in the size control of cultured mammalian epithelial cells. *J. Cell Sci.* 88, 73-80.

[24] Huggins, R. M. (1995). A law of large numbers for the bifurcating autoregressive process. *Comm. Statistics. Stochastic Models* 11, 273-278.

[25] Huggins, R. M. (1996). On the identifiability of measurement error in the bifurcating autoregressive model. *Statist. Probab. Lett.* 27, 17-23.

[26] Huggins, R. M. and Basawa, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Prob.* 36, 1225-1233.

[27] Huggins, R. M. and Basawa, I. V. (2000). Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* 42, 423-432.

[28] Huggins, R. M. and Marschner, I. C. (1991). Robust analysis of the bifurcating autoregressive process in cell lineage studies. *Aust. J. Statist.* 33, 209-220.

[29] Huggins, R. M. and Staudte, R. G. (1994). Variance components models for dependent cell populations. *J. Amer. Statist. Assoc.* 89, 19-29.

[30] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. Wiley, New York.

[31] Klimko, L. A. and Nelson, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* 6, 629-642.

[32] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions, Vol 1: Models and Applications*, 2nd Ed.. Wiley, New York.

[33] Lieb, E. H. and Loss, M. (2001). *Analysis*, American Mathematical Society, 2nd Edn., Providence, Rhode Island.

[34] Mardia, K. V. (1970). *Families of Bivariate Distributions.* Griffin, London.

[35] Nielsen, B. and Shephard, N. (2003). Likelihood analysis of a first-order autoregressive model with exponential innovations. *J. Time Series Analysis*, 24, 337-344.

[36] Powell, E. O. (1955). Some features of the generation times of individual bacteria. *Biometrika* 42, 16-44.

[37] Powell, E. O. (1956). An improved culture chamber for the study of living bacteria. *J. R. Micr. Soc.* 75, 235.

[38] Powell, E. O. (1958). An outline of the pattern of bacterial generation times. *J. Gen. Microbiol.* 18, 382-417.

[39] Powell, E. O. and Errington, F. P. (1963). Generation times of individual bacteria: Some corroborative measurements. *J. Gen. Microbio.* 31, 315-327.

[40] SIM, C.H. (1990) First-order autoregressive models for gamma and exponential processes. *J. Appl. Prob.*. 27, 325-332.

[41] Staudte, R. G. (1992). A bifurcating autoregression model for cell lineage data with varying generation means. *J. Theoret. Biol.* 156, 183-195.

[42] Staudte, R. G., Guiguet, M., and Collyne D'Hooghe, M. (1984). Additive models for dependent cell populations. *J. Theoret. Biol.* 109, 127-146.

[43] Staudte, R. G., Zhang, J., Huggins, R.M., and Cowan, R. (1996). A reexamination of the cell lineage data of E. O. Powell. *Biometrics* 52, 1214-1222.

[44] Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach.* Oxford Univ. Press, New York.

[45] Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis. *Applied Statistics* 41, 1-15.

[46] Zhou, J. and Basawa, I. V. (2003). Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. Technical Report, University of Georgia.

[47] Zhou, J. and Basawa, I. V. (2004). Least squares estimation for bifurcating autoregressive processes. Technical Report, University of Georgia.