

Contemporary review on the bifurcating autoregressive models : Overview and perspectives[†]

S. Y. Hwang¹

¹Department of Statistics, Sookmyung Women's University
Received 5 July 2014, revised 1 August 2014, accepted 13 August 2014

Abstract

Since the bifurcating autoregressive (BAR) model was developed by Cowan and Staudte (1986) to analyze cell lineage data, a lot of research has been directed to BAR and its generalizations. Based mainly on the author's works, this paper is concerned with a contemporary review on the BAR in terms of an overview and perspectives. Specifically, bifurcating structure is extended to multi-cast tree and to branching tree structure. The AR(1) time series model of Cowan and Staudte (1986) is generalized to tree structured random processes. Branching correlations between individuals sharing the same parent are introduced and discussed. Various methods for estimating parameters and related asymptotics are also reviewed. Consequently, the paper aims to give a contemporary overview on the BAR model, providing some perspectives to the future works in this area.

Keywords: Bifurcating autoregression, branching correlation, branching tree, estimating function, multi-cast model.

1. Introduction

Cowan and Staudte (1986) suggested bifurcating autoregressive (BAR) models to analyze binary splitting data where each individual (mother) gives rise to exactly two offspring (daughters) in the next generation. Powell's data, G15 + C data and cell lineage data are typically of this kind (cf. Huggins and Basawa, 1999). Figure 1.1 shows a binary splitting (that is, bifurcating) data consisting of 7 observations. See, for instance, Hwang and Basawa (2009). Let X_i be an observation of individual i . For example, X_i may represent blood pressure, cholesterol level, dose of a drug and protein content of a cell i , etc. In a bifurcating model, i produces $(2i, 2i + 1)$. Cowan and Staudte (1986) introduced the first order BAR(1) model recursively defined by, with the initial observation $X_1 = x_1$,

$$\begin{aligned}X_{2i} &= \theta X_i + \epsilon_{2i} \\ X_{2i+1} &= \theta X_i + \epsilon_{2i+1}\end{aligned}\tag{1.1}$$

where $|\theta| < 1$ and $(\epsilon_{2i}, \epsilon_{2i+1})$ is a sequence of iid bivariate random vectors with common mean zero, common variance σ^2 and $Corr(\epsilon_{2i}, \epsilon_{2i+1}) = \rho$. It is noted that the BAR(1) model views each line of descendants as a standard AR(1) time series.

[†] This research was supported by a grant (2013) from Sookmyung Women's University.

¹ Professor, Department of Statistics, Sookmyung Women's University, Seoul 140-742, Korea.
E-mail: shwang@sm.ac.kr

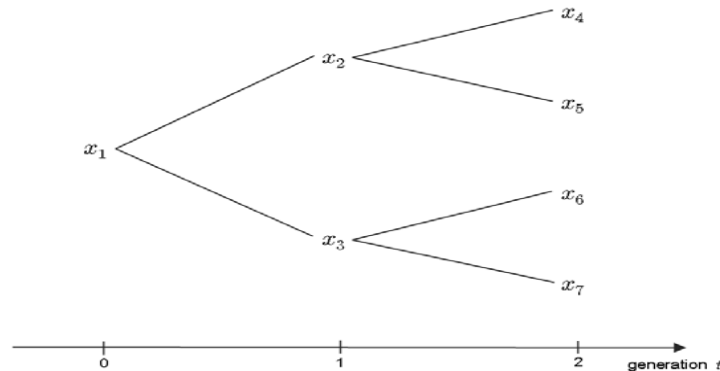


Figure 1.1 A binary splitting (bifurcating) data consisting of 7 observations

Three issues are discussed to generalize BAR(1) time series (1.1). First issue is on the nature of trees (as in Figure 1.1) indexing the data. Specification of the time series models defined on the tree may be the second issue. The correlation (conditionally on their mother) between sisters is referred to as the branching correlation (cf. Hwang, 2011). It is noted in the BAR(1) formulation (1.1) that ρ represents a correlation between sisters, conditionally on their mother. That is, the branching correlation of the BAR(1) models is given by

$$\rho = \text{Corr}(X_{2i}, X_{2i+1} | X_i). \quad (1.2)$$

Mathematical modeling of the branching correlation is discussed as the third issue in generalizing BAR(1) time series. Details of the three issues will be provided next.

Regarding the parameter estimation for BAR(1) and its generalizations, it is reasonable to assume that the exact likelihood is not known. Equivalently, no specific distributional assumptions are made on $(\epsilon_{2i}, \epsilon_{2i+1})$. Then, quasi-maximum likelihood (QML) estimator is obtained by maximizing the pseudo-objective function which is taken as Gaussian bivariate innovations $(\epsilon_{2i}, \epsilon_{2i+1})$. When BAR(1) process X_i is partially specified only through the first and second order conditional moments, a systematic approach for a partially specified model is via the so called quaslikelihood (QL). Due to the wideness of the generalized structures discussed in this paper, it will be appropriate to use a QL estimation. We refer to, for instance, Heyde (1997) and Hwang *et al.* (2014b) for a background on QL and QML in a broader context of stochastic processes.

2. From the bifurcating tree to multicast and branching tree

Many applications require a multicast tree in which each node may have m branches where m may be larger than 2. For instance, in an internet traffic data, a package (signal) may be sent along m links and delay times X over each link may be of interest. One may then extend bifurcating structure to a multicast tree where each individual splits exactly into m -offspring ($m \geq 1$). A tri-casting ($m = 3$) data is illustrated in Figure 2.1. See for instance, Hwang and Choi (2009) and Hwang and Kang (2012).

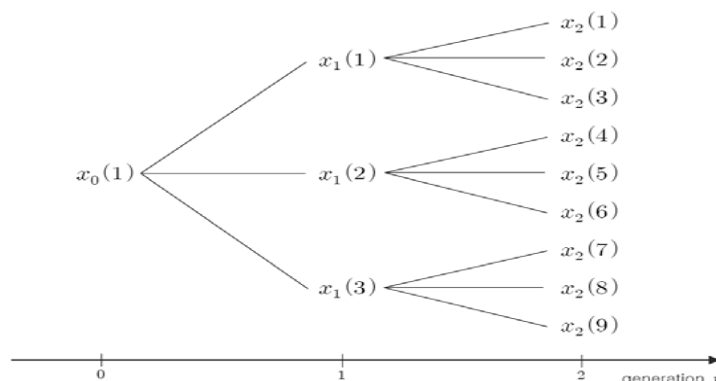


Figure 2.1 A multicast tree : Tri-casting tree ($m=3$)

We consider a multicast autoregressive (MAR, for short) model to accommodate this class of multicast data. Consider the following zero mean process defined by the m -equations

$$\begin{aligned}
 X_{mi-(m-2)} &= \theta X_i + \epsilon_{mi-(m-2)} \\
 &\vdots \\
 X_{mi} &= \theta X_i + \epsilon_{mi} \\
 X_{mi+1} &= \theta X_i + \epsilon_{mi+1}
 \end{aligned} \tag{2.1}$$

where $(\epsilon_{mi-(m-2)}, \dots, \epsilon_{mi}, \epsilon_{mi+1}), i = 1, 2, \dots$ is a sequence of iid m -variate random vectors with mean zero vector and variance-covariance matrix Ψ given by

$$\Psi = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \sigma^2. \tag{2.2}$$

We shall refer to the first-order process defined in (2.1) as the multicast AR(1) ; MAR(1). In particular when $m = 2$, MAR(1) reduces to the BAR(1). To view MAR(1) as a time series along with the generation, we will rewrite $\{X_i\}$ in (2.1) in terms of $\{X_t(j)\}$ for which $X_t(j)$ stands for the j -th observation on the t -th generation. Here, $t = 0, 1, 2, \dots$ and $j = 1, \dots, m^t$. A tri-casting structure ($m = 3$) is illustrated in Figure 2.1. The subscript t in $X_t(j)$ is used for denoting ' t -th generation'. To clarify ancestral path of the particular observation $X_t(j)$, let $X_{t-1}(t(j))$ denote the observation on the immediate parent of $X_t(j)$. In Figure 2.2, note that $x_2(4)$ is the 4-th observation on the second generation ($t = 2$) and in turn gives $x_1(2(4)) = x_1(2)$. The MAR(1) process (2.1) can then be written in terms of $\{X_t(j), t = 0, 1, 2, \dots, j = 1, \dots, m^t\}$ as

$$X_t(j) = \theta X_{t-1}(t(j)) + \epsilon_t(j) \tag{2.3}$$

where

$$\{(\epsilon_t(1 + sm), \epsilon_t(2 + sm), \dots, \epsilon_t(m + sm)); t = 1, 2, \dots, s = 0, \dots, m^{t-1} - 1\}$$

are iid m -variate error vectors with mean zero and variance-covariance matrix Ψ specified in (2.2).

We now consider the branching tree which was investigated by Hwang and Basawa (2009, 2011). In many situations dealing with population biology, epidemiology, physics and chemistry, one may be interested in measuring some characteristics of interest on each individual of a Galton-Watson (G-W) branching process such as life-time of a cell, presence (or absence) of a certain chemical of a protein, radio-activity level and severity of an epidemic, etc. A Branching Markov process (BMP) is a tree-indexed process where the tree-index is a branching process $Z_t, t = 0, 1, 2, \dots$ with Z_t denoting the t -th generation size.

Let $X_t(j), j = 1, 2, \dots, Z_t$ and $t = 0, 1, 2, \dots$, denote observation on the j -th individual in the t -th generation. Figure 2.2 illustrates a branching tree and a sample path of BMP (see Hwang and Basawa, 2009, 2014). We assume that Z_t follows a standard super-critical Galton-Watson (G-W) branching process for which $E(Z_1) = m > 1$ and $Var(Z_1) = \sigma_Z^2 > 0$ where m and σ_Z^2 are the offspring mean and variance respectively. It is well known that there exists a random variable W to which Z_n/m^n converges almost surely as $n \rightarrow \infty$, and $P(W > 0) = 1$. Similar to (2.3), we use the notation $X_{t-1}(t(j))$ to denote the observation on the immediate mother of the $X_t(j)$. Here, the subscript $t - 1$ is used for denoting $(t - 1)$ -th generation. A simple BMP is a branching-AR model defined by

$$X_t(j) = \theta_0 + \theta_1 X_{t-1}(t(j)) + \epsilon_t(j) \tag{2.4}$$

where $\epsilon_t(j), t = 1, 2, \dots$ and $j = 1, 2, \dots$ are iid random variables with mean zero and variance σ^2 . The data for the BMP is given as

$$(z_t, x_t(j)); t = 1, 2, \dots, n; j = 1, 2, \dots, z_n$$

with initial observation $x_0(1)$ on $Z_0 = 1$.

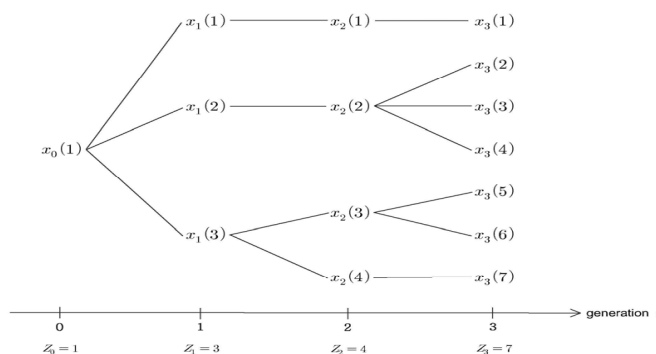


Figure 2.2 A branching tree and a path of the BMP model

It is worth noting that the branching tree reduces to the multicast tree when Z_t is deterministic and is given by $Z_t = m^t$ for all t .

3. The branching correlation between individuals

It is noted both in the BAR(1) model (1.1) and in the MAR(1) formulation (2.1) that ρ represents a correlation between sisters, conditionally on their mother. We refer to such a

conditional correlation ρ between sisters as a branching correlation. In (2.2) for MAR(1), Ψ corresponds to an equi-correlation model between sisters. It can be verified that for fixed $m \geq 3$, the matrix Ψ in (2.2) associated with the equi-correlation model is positive-definite if and only if $-(m - 1)^{-1} < \rho < 1$ (cf. Hwang and Kang, 2012; Hwang, 2011). For instance, a tri-casting case ($m = 3$) requires $-1/2 < \rho < 1$ for a proper covariance matrix. As m increases, one has to essentially excludes negative values of ρ . This is an obvious drawback inherent in Ψ associated with the MAR(1). To circumvent, one needs to propose another useful covariance matrix which does not require any restrictions on ρ . For illustration, consider the case when there is a natural ordering between m -sisters. It will then be useful to consider

$$\Psi_1 = \begin{pmatrix} 1 & \rho & \cdots & \rho^{m-1} \\ \rho & 1 & \cdots & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \cdots & 1 \end{pmatrix} \sigma^2. \tag{3.1}$$

It is noted in Ψ_1 that the branching correlation between two sisters k -position apart is given by ρ^k ($k = 1, 2, \dots, m - 1$) which is exponentially decaying as k increases. The formulation of Ψ_1 is well motivated by standard first order autoregression AR(1) in time series. Another useful formulation is

$$\Psi_2 = \begin{pmatrix} 1 & \rho & \cdots & 0 \\ \rho & 1 & \rho \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \sigma^2. \tag{3.2}$$

Note that Ψ_2 shows no branching correlation beyond 1-position apart and only adjacent sisters assume ρ for branching correlation. The correlation structure specified in Ψ_2 is closely related to that of a first-order moving average MA(1) time series model. In general, instead of the equi-correlation model (2.2), the correlation matrix of the error vectors $(\epsilon_{mi-(m-2)}, \dots, \epsilon_{mi}, \epsilon_{mi+1})$ in the MAR(1) formulation (2.1) is denoted by $R(\rho)$ where $R(\rho)$, as a function of ρ , denotes an appropriate form of $m \times m$ correlation matrix. Accordingly,

$$\Psi = R(\rho) \cdot \sigma^2. \tag{3.3}$$

A rich class of $R(\rho)$ can be obtained from the diverse autocorrelation functions of the ARMA (autoregressive moving average) time series models. It is noted that the branching correlation discussed above for the MAR(1) is homoscedastic. To see this, note that there are exactly m -sisters $X_{mi-(m-2)}, \dots, X_{mi}, X_{mi+1}$ sharing the same mother $X_i, i = 1, 2, \dots$, in (2.1) of MAR(1). Define sister vector S_i as

$$S_i = (X_{mi-(m-2)}, \dots, X_{mi}, X_{mi+1})^T : m \times 1. \tag{3.4}$$

Here ‘ T ’ indicates transpose of a matrix (or a vector). It is noted that the conditional mean vector of S_i is given by $E(S_i|X_i) = (\theta X_i)1_m$ for MAR(1) model in (2.1). Here 1_m is a $m \times 1$ vector of ones. In addition, the conditional variances of each m sister are the same and are given by

$$Var(X_{mi+k}|X_i) = \sigma^2, k = -(m - 2), -(m - 1), \dots, 1. \tag{3.5}$$

Thus, the MAR model (2.1) provides a homoscedastic conditional variance σ^2 . Alternatively, one can develop heteroscedastic branching correlation as follows. We will consider the following heteroscedastic conditional variance h_t which is a quadratic function of the mother observation X_i and such an h_i is called a heteroscedastic branching correlation (HBC). We refer to, for instance, Hwang and Choi (2011) for various HBC. In particular, a quadratic HBC is defined below.

Definition 3.1 Quadratic HBC h_t is defined as, for some non-negative constants β_0, β_1 and β_2

$$h_i = Var(X_{mi+k}|X_i) = \beta_0 + \beta_1 X_i + \beta_2 X_i^2, \quad k = -(m-2), -(m-1), \dots, 1. \quad (3.6)$$

Examples are illustrated below.

Example 3.1 Random coefficient MAR(1)

Consider the process $X_i, i \geq 1$ such that

$$S_i = [\theta_0 + (\theta + \theta_i)X_i]1_m + e_i \quad (3.7)$$

where $e_i = (\epsilon_{mi-(m-2)}, \dots, \epsilon_{mi}, \epsilon_{mi+1})^T$ and θ_i denotes random coefficient of the autoregressive coefficient θ , and θ_i is a sequence of iid random variables with mean zero and variance σ_θ^2 , independently with $e_i, i \geq 1$. It is easy to see that the quadratic HBC is given by

$$h_i = \sigma_\theta^2 X_i^2 + \sigma^2. \quad (3.8)$$

Example 3.2 Binomial thinning MAR(1)

Consider the following integer-valued process $X_i, i \geq 1$ defined by

$$S_i = (\theta \circ X_i)1_m + e_i \quad (3.9)$$

where \circ denotes the binomial thinning operator defined by $\theta \circ X_i = \sum_{j=1}^{X_i} B_j$ where B_j is a sequence of iid Bernoulli random variables with success probability $\theta, 0 < \theta < 1$. Two processes B_j and e_i are assumed to be independent. Here, m -tuple error process e_i is a sequence of iid integer-valued random vectors (e.g., multivariate Poisson vectors (c.f. Hwang and Basawa, 2011) with mean vector $\lambda 1_m, \lambda > 0$ and variance-covariance matrix Ψ in (2.2). See, e.g., Grunwald *et al.* (2000) and Baek *et al.* (2012). Notice that $E(S_i|X_i) = (\theta X_i + \lambda)1_m$ and the HBC can be verified to be

$$h_i = \sigma^2 + \theta(1 - \theta)X_i. \quad (3.10)$$

Consequently, two examples belong to our quadratic class HBC (see (3.8) and (3.10)). For the case of HBC, the heteroscedastic conditional variance covariance matrix $Var(S_i|X_i) = V_i$ is given by, as a function of mother X_i

$$V_i = \begin{pmatrix} 1 & \rho_i & \cdots & \rho_i \\ \rho_i & 1 & \cdots & \rho_i \\ \vdots & \vdots & \ddots & \vdots \\ \rho_i & \rho_i & \cdots & 1 \end{pmatrix} \sigma_i^2. \quad (3.11)$$

For more examples belonging to the quadratic HBC class, refer to Hwang and Kang (2012), Baek *et al.* (2012) and Basawa and Zhou (2004).

4. Various models on the multicast tree and branching tree

It would be useful to illustrate various models defined on the multicast tree. First, recall (2.1) and (2.2). To clarify ancestral path of the particular individual i , let $i(1)$ denote the first ancestor (i.e., mother) of the individual i . It can be shown that

$$i(1) = \left\lceil \frac{i + (m - 2)}{m} \right\rceil, \quad m \geq 2 \tag{4.1}$$

and $i(1) = i - 1$ for $m = 1$ (cf. Hwang and Choi, 2009; Hwang and Kang, 2012). Here $\lceil \cdot \rceil$ denotes the greatest integer function. For the bifurcating case of $m = 2$, $i(1)$ reduces to $\lceil i/2 \rceil$. We reformulate the MAR(1) model defined in (2.1) in terms of a single equation as

$$X_i = \theta X_{i(1)} + \epsilon_i, \quad i \geq 2. \tag{4.2}$$

Example 4.1 MAR(1) with an intercept

A MAR(1) model with an intercept θ_0 is defined by

$$X_i = \theta_0 + \theta_1 X_{i(1)} + \epsilon_i, \quad |\theta_1| < 1. \tag{4.3}$$

Example 4.2 Nonlinear MAR(1)

Consider the model

$$X_i = \mu(X_{i(1)}) + \epsilon_i \tag{4.4}$$

where $\mu(\cdot)$ denotes the conditional mean function, viz.,

$$\mu(X_{i(1)}) = E(X_i | X_{i(1)}). \tag{4.5}$$

In particular, a threshold model is given by

$$\mu(x) = \theta_0 + \theta_1(x - \psi)^+ + \theta_2(x - \psi)^-. \tag{4.6}$$

where ψ stands for a “threshold” constant.

Example 4.3 Conditional exponential family on the multicast tree

Consider the following conditional exponential family defined by

$$p(x_i | x_{i(1)}) = c(x_{i(1)}) \exp[x_i \zeta_i - k(\zeta_i)] \tag{4.7}$$

where the “natural parameter” ζ_i is specified by, with appropriate link function $g(\cdot)$,

$$\zeta_i = g(\mu_i) \text{ and } k(\zeta_i) = \log \int c(x_{i(1)}) \exp(x_i \zeta_i) dx_i < \infty \tag{4.8}$$

where μ_i denote the conditional mean defined in (4.5). Multicast conditional exponential family includes conditional gamma, conditional exponential, conditional Poisson, conditional normal and conditional binomial models as special cases. Refer to, for instance, Hwang and Kang (2012) for more treatment of conditional exponential family on the multicast tree.

Conditional moments of X_i given $X_{i(1)}$ can be obtained via successive differentiations of $k(\zeta_i)$ with respect to the natural parameter ζ_i . In particular, the conditional mean μ_i and conditional variance v_i are given by

$$\mu_i = dk(\zeta_i)/d\zeta_i \text{ and } v_i = d^2k(\zeta_i)/d\zeta_i^2.$$

Example 4.4 Partially specified MAR(1)

We postulate that MAR(1) is specified only by the first two moments, viz., the conditional mean and covariance without requiring the likelihood and/or defining recursive equations. Partially specified MAR(1) model X_i , $i \geq 1$ is such that the conditional distribution of X_i given ancestors depends only on their mother ($X_{i(1)}$), satisfying $E(X_i|X_{i(1)}) = (\theta X_{i(1)})1_m$, $|\theta| < 1$. This class is motivated by the non-Gaussian conditionally linear AR(1) time series of Grunwald *et al.* (2000). This partially specified class is rich enough to include random coefficient MAR, conditionally heteroscedastic MAR and binomial-thinning MAR. See Baek *et al.* (2012) for diverse partially specified models. Details are omitted.

We consider models defined on a branching tree. Recall (2.3) and Figure 2.2. We are willing to investigate the (k -variate) vector process $X_t(j)$.

Example 4.5 Multivariate BMP

Consider the k -variate process $X_t(j)$ defined by

$$X_t(j) = \phi_0 + \Phi X_{t-1}(t(j)) + \epsilon_t(j) \quad (4.9)$$

where $\epsilon_t(j)$, $t = 1, 2, \dots$ and $j = 1, 2, \dots, Z_t$ are iid k variate normal random vectors with mean zero and covariance matrix given by Σ_ϵ of dimension $k \times k$. Here ϕ_0 is a constant vector of size k and Φ denotes $k \times k$ matrix of constants. To illustrate, consider $k = 2$ case with $\phi_0 = 0$. Denoting the i -th component of the vector $X_t(j)$ by $X_{t,i}(j)$, (4.9) reduces to

$$\begin{pmatrix} X_{t,1}(j) \\ X_{t,2}(j) \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} X_{t-1,1}(t(j)) \\ X_{t-1,2}(t(j)) \end{pmatrix} + \begin{pmatrix} \epsilon_{t,1}(j) \\ \epsilon_{t,2}(j) \end{pmatrix}.$$

It is noted that there is no interaction (i.e., feedback) between the two components (or, traits) $X_{t,1}(j)$ and $X_{t,2}(j)$ in $X_t(j)$ if and only if $\phi_{12} = \phi_{21} = 0$.

Example 4.6 Multivariate conditionally heteroscedastic BMP

Let $X_t(j)$ be generated by the equation

$$X_t(j) = H_t(j)^{1/2} \cdot e_t(j) \quad (4.10)$$

where $e_t(j)$, $t = 1, 2, \dots$ and $j = 1, 2, \dots, Z_t$ denote iid random vectors (k variate) with zero mean vector and identity matrix as a covariance matrix. Also, $H_t(j)$ represents conditional covariance matrix, that is, $H_t(j) = \text{Var}(X_t(j)|X_{t-1}(t(j))) : k \times k$, and $H_t(j) = H_t(j)^{1/2} \cdot H_t(j)^{1/2}$ where $H_t(j)^{1/2}$ denotes the (symmetric) half matrix of $H_t(j)$.

Example 4.7 Multivariate partially specified BMP

Consider the multivariate MAR(1) in (4.9) for which $\epsilon_t(j)$ is not necessarily Gaussian. It is noted that

$$E(X_t(j)|F_{t-1}) = \phi_0 + \Phi X_{t-1}(t(j)) \quad (4.11)$$

where F_{t-1} denote the information up to and including $(t - 1)$ -th generation. Here, we do not require the autoregression-specification (4.9) and this class of models is defined only through conditionally-linear-mean specification (4.11), enlarging the class of models under consideration.

Example 4.8 Multivariate thinning BMP

Let $X_{ti}(j)$ denote the i -th component of the vector $X_t(j), i = 1, 2, \dots, k$. Consider the standard binomial-thinning branching integer-valued AR(1) model defined by, for each $i = 1, 2, \dots, k$

$$X_{ti}(j) = \sum_{u=1}^{X_{(t-1)i}(t(j))} \xi_{ui}(j) + Z_{ti}(j) \tag{4.12}$$

where $\xi_{ui}(j)$ is an iid Bernoulli random variables with success probability ϕ_i . One can introduce dependence between the k components of $X_t(j)$ via dependence among the innovations $Z_{ti}(j)$. Let innovation vectors $Z_{ti}(j) = (Z_{t1}(j), \dots, Z_{tk}(j))^T$ be iid k variate Poisson random vectors which are independent of the Bernoulli random variables $\xi_{ui}(j)$. Details are omitted. We refer to Hwang and Basawa (2011) for more details on multivariate BMPs defined on the branching trees.

5. Inferential issues

To estimate parameters involved in BAR(1) and its generalized models, we assume that the exact likelihood is not known. Equivalently, no specific distributional assumptions are made on the innovations. Then, quasi-maximum likelihood (QML) estimator seems to be reasonable to employ and is obtained by maximizing the pseudo-objective function which is taken as Gaussian bivariate innovations $(\epsilon_{2i}, \epsilon_{2i+1})$. In particular when BAR(1) process X_t is partially specified only through the first and second order conditional moments, a systematic approach for a partially specified model is via the so called quaslikelihood (QL). Due to the wideness of the generalized structures, it will be appropriate to discuss further QL estimation. We refer to, for instance, Heyde (1997), Hwang *et al.* (2014b) and Lee (2012) for a background on QL and QML in a broader context.

Quasi likelihood (QL)

Let the conditional variance of X_i given the past information F_{i-1} be denoted by $h_i(\theta) = Var(X_i|F_{i-1})$. Note that $\mu_i(\theta)$ and $h_i(\theta)$ are F_{i-1} -measurable and the parameter vector θ will be suppressed in $\mu_i(\theta)$ and $h_i(\theta)$ for notational simplicity. A quaslikelihood (QL) score (see, e.g., Godambe, 1985) is defined by

$$u_i(\theta) = \left(\frac{\partial \mu_i}{\partial \theta} \right) \cdot h_i^{-1} \cdot (X_i - \mu_i) \tag{5.1}$$

and the QL estimator is obtained by the estimating equation $\sum_{i=1}^n u_i(\theta) = 0$

Example 5.1 QL estimation for BAR(1) model

This example is taken and modified from Hwang and Basawa (2014). Recall the BAR(1) formulation (4.2) given by

$$X_i = \theta X_{i(1)} + \epsilon_i, \quad i \geq 2. \tag{5.2}$$

We generalize (5.2) to the following nonlinear version of BAR(1) defined by

$$X_i = \mu_i + \sqrt{h_i} \cdot e_i \quad (5.3)$$

where the variance of e_i is set to be unity and $\mu_i = \mu(X_{i(1)})$ and $h_i = h(X_{i(1)})$ stand respectively for the conditional mean and variance function defined by

$$\mu_i = E(X_i|X_{i(1)}) \text{ and } h_i = Var(X_i|X_{i(1)}). \quad (5.4)$$

It is noted that μ_i and h_i are $X_{i(1)}$ -measurable. Suppose that the distribution of e_i is not known and then we rely on a quaslikelihood score. Let

$$a_i(\theta) = \begin{pmatrix} X_i - \mu_i(\theta) \\ (X_i - \mu_i(\theta))^2 - h_i(\theta) \end{pmatrix}. \quad (5.5)$$

Denote the (2×2) conditional variance-covariance matrix of $a_i(\theta)$ given the first ancestor $X_{i(1)}$ by $V_i(\theta)$, viz.,

$$V_i(\theta) = E[a_i(\theta)a_i^T(\theta)|X_{i(1)}] = \begin{pmatrix} h_i(\theta) & \mu_{3i}(\theta) \\ \mu_{3i}(\theta) & \mu_{4i}(\theta) - h_i^2(\theta) \end{pmatrix} \quad (5.6)$$

where $\mu_{3i}(\theta) = E[(X_i - \mu_i(\theta))^3|X_{i(1)}]$ and $\mu_{4i}(\theta) = E[(X_i - \mu_i(\theta))^4|X_{i(1)}]$ (cf. Hwang *et al.*, 2014a). The conditional expectation of the derivative matrix of $a_i(\theta)$ is given by

$$E[\partial a_i(\theta)/\partial \theta^T|X_{i(1)}] = \begin{pmatrix} -(\partial \mu_i(\theta)/\partial \theta)^T \\ -(\partial h_i(\theta)/\partial \theta)^T \end{pmatrix} : (2 \times k). \quad (5.7)$$

We now have a quaslikelihood score which is optimum within the Godambe class (cf. Hwang and Basawa, 2011a) is

$$\begin{aligned} QL(\theta) &= \sum_{i=1}^n w_{i-1}^O(\theta) a_i(\theta) = \sum_{i=1}^n (E_{i-1}[\partial a_i(\theta)/\partial \theta^T])^T (E_{i-1}[a_i(\theta)a_i^T(\theta)])^{-1} a_i(\theta) \\ &= - \sum_{i=1}^n \left(\frac{\partial \mu_i(\theta)}{\partial \theta}, \frac{\partial h_i(\theta)}{\partial \theta} \right) V_i^{-1}(\theta) a_i(\theta) \end{aligned} \quad (5.8)$$

where $V_i(\theta)$ is given in (5.6). We then have

$$\sqrt{n}(\hat{\theta}_{QL} - \theta) \xrightarrow{d} N(0, K^{-1}) \quad (5.9)$$

where

$$K = E \left[\left(\frac{\partial \mu_i(\theta)}{\partial \theta}, \frac{\partial h_i(\theta)}{\partial \theta} \right) V_i^{-1}(\theta) \left(\frac{\partial \mu_i(\theta)}{\partial \theta}, \frac{\partial h_i(\theta)}{\partial \theta} \right)^T \right]. \quad (5.10)$$

Quasi maximum likelihood (QML)

It is usually the case that the true likelihood is unknown to researchers, and thus we need to presume a tractable likelihood for the data which is called a pseudo-likelihood (PL). A PL

may be a falsely specified likelihood. A pseudo-maximum likelihood estimator is obtained by maximizing the objective function of pseudo-likelihood score. Often, the pseudo-likelihood is taken via Gaussian errors, standardized t -distributions with unknown degrees of freedom, and generalized error distributions. In particular when Gaussian errors are chosen as the pseudo-likelihood, the resulting estimator is referred to as the quasi maximum likelihood (QML) estimator. It is obvious that the QML-estimator reduces to the maximum likelihood (ML) estimator provided the (unknown) true likelihood is indeed Gaussian. It is interesting to note that even when the true likelihood is different from the Gaussian-likelihood, the QML-estimator continues to be consistent and asymptotically normal under some regularity conditions (cf., Straumann and Mikosch, 2006; Francq and Zakoian, 2013; Hwang *et al.*, 2014b).

Alternative methods of estimation such as conditional least squares, weighted least squares, maximum likelihood, pseudo-likelihood, and quasilielihood can also be employed. Readers may refer to recent reference of Hwang and Basawa (2014) for unifying various estimation methods via martingale estimating functions. For conditional least squares estimation for higher order MAR processes, see Mao (2014) and references therein.

Semi-parametric approach will also be useful to estimate parameters of interest. Various estimating equations may involve nuisance parameters. For instance, the QL estimating function $QL(\theta)$ may involve nuisance parameter, say η which needs to be estimated. Innovation distribution can be treated as a (infinite dimensional) nuisance parameter η . Then innovation distribution may be estimated via suitable kernel density estimation (KDE) based on residuals. Under some regularity conditions, asymptotic optimality of the $QL(\theta)$ continues to hold when replacing innovation density by KDE.

Regarding tests, consider the problem of testing a null hypothesis specified by $H : \theta = \theta_0$. When the likelihood function is known, one can use classical tests, viz., the Rao's score test, Wald test and likelihood ratio test. When the likelihood is not available however, we rely on an appropriate estimating function, for instance, a quasilielihood score $U_n(\theta)$ and QL estimator $\hat{\theta}_n$ obtained from $U_n(\theta) = 0$. The QL score test Q_{1n} for testing $H : \theta = \theta_0$ is defined by

$$Q_{1n} = U_n^T(\theta_0)\xi_n^{-1}(\theta_0)U_n(\theta_0)$$

where $U_n(\theta)$ is the QL score and $\xi_n(\theta)$ is a norming matrix. The Wald-type test statistic based on QL estimator $\hat{\theta}_n$ is defined as

$$Q_{2n} = (\hat{\theta}_n - \theta_0)^T [I(\theta_0)]^{-1} (\hat{\theta}_n - \theta_0) \tag{5.11}$$

where $I(\theta)$ is the QL-information matrix. Unlike the classical three tests, it is not obvious that QL based tests are asymptotically efficient. Readers refer to Hwang *et al.* (2014a) and Hwang and Basawa (2011) for a rigorous discussion on the certain asymptotic optimality of QL based tests Q_{1n} and Q_{2n} .

6. Non-stationary cases

Revisit the BAR(1) formulation

$$X_i = \theta X_{i(1)} + \epsilon_i, \quad i \geq 2. \tag{6.1}$$

For $|\theta| < 1$, it is (path-wise) stationary. Consider the non-stationary case of $|\theta| > 1$. Based on the data X_1, \dots, X_n , the conditional least squares estimator $\hat{\theta}_n$ of θ is given by

$$\hat{\theta}_n = \sum_{t=1}^n X_{t(1)} X_t / \sum_{t=1}^n X_{t(1)}^2. \quad (6.2)$$

Hwang and Choi (2009) derived the limiting distribution of $\hat{\theta}_n$ for MAR(1) model. Due to Theorem 4.1 of Hwang and Choi (2009), we have as n tends to infinity

$$c_n(\hat{\theta}_n - \theta) \xrightarrow{d} U^{-1}V$$

where the norming constant c_n is given by $c_n = 2(2\theta^2 - 1)^{-1}(\sqrt{2}\theta)^n$ and U and V are random variables as described in Lemmas 4.1 and 4.2 of Hwang and Choi (2009). The unit root case of $|\theta| = 1$ is now under investigation. One may well expect the limit distribution of $\hat{\theta}_n$ (when $|\theta| = 1$) is given by a functional on the standard Brownian motion. However, preliminary investigation reveals a different result on $|\theta| = 1$ for the MAR(1) model. This result will be addressed elsewhere.

7. Concluding remarks

This paper provides contemporary reviews on the bifurcating autoregression (BAR) models which has been useful to analyze cell lineage data. Three issues are discussed to generalize the BAR model to the “future” models. Specifically, multi-cast trees and branching tree structures are introduced in place of the bifurcating tree. In stead of the AR(1) time series structure, we accommodate diverse random processes defined on the extended tree structures. Branching correlations between individuals sharing the same parent are highlighted and discussed in details. Focusing on the quasi-likelihood (QL) approach in order to deal with the unspecified likelihood of the data, various methods for estimating parameters and related asymptotics are also reviewed in the context of the estimating function approach. Accordingly, the current paper is able to give a contemporary overview on the BAR model, bringing some perspectives to the future works in this area. The results presented in the paper are mostly adapted from the author’s works in the reference and recent literature on the BAR models.

We have not discussed testing problems, non-parametric issues and the Bayesian approaches. In addition, computing related issues needs to be addressed in this field since the number of data grows exponentially. For instance, the number of observations n_g for the MAR(1) model up to the g th generation is given by

$$n_g = 1 + m + \dots + m^g = (m^{g+1} - 1)/(m - 1), \quad m \neq 1.$$

Consequently, many of the BAR-related problems are still to be resolved. Hoping that this review paper is useful to the future works, we encourage researchers and practitioners to pursue the open problems in this field.

References

- Baek, J. S., Choi, M. S. and Hwang, S. Y. (2012). A broad class of partially specified autoregressions on multicasting data. *Communications in Statistics - Theory and Methods*, **41**, 178-193.
- Basawa, I. V. and Zhou, J. (2004). Non-Gaussian bifurcating models and quasi likelihood estimation. *Journal of Applied Probability*, **41A**, 55-64.
- Cowan R. and Staudte, R. G. (1986). The bifurcating autoregression model in cell lineage studies. *Biometrics*, **42**, 769-783.
- Francq, C. and Zakoian, J. M. (2013). Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of Royal Statistical Society B*, **75**, 345-367.
- Godambe, V. P. (1985). The foundation of finite sample estimation in stochastic processes. *Biometrika*, **72**, 419-428.
- Grunwald G. K., Hyndman, R. J., Tedesco, L. and Tweedie, R. L. (2000). Non-Gaussian conditional linear AR(1) models. *Australian and New Zealand Journal of Statistics*, **42**, 479-495.
- Heyde, C. C. (1997). *Quasi-likelihood and its application*, Springer, New York.
- Huggins, R. M. and Basawa, I. V. (1999). Extensions of the bifurcating autoregressive model for cell lineage data. *Journal of Applied Probability*, **36**, 1225-1233.
- Hwang, S. Y. (2011). An overview on models for tree-indexed time series. *Quantitative Bio-Sciences*, **30**, 9-11.
- Hwang, S. Y. and Basawa, I. V. (2009). Branching Markov processes and related asymptotics. *Journal of Multivariate Analysis*, **100**, 1155-1167.
- Hwang, S. Y. and Basawa, I. V. (2011). Asymptotic optimal inference for multivariate branching-Markov processes via martingale estimating functions and mixed normality. *Journal of Multivariate Analysis*, **102**, 1018-1031.
- Hwang, S. Y. and Basawa, I. V. (2011a), Godambe estimating functions and asymptotic optimal inference. *Statistics & Probability Letters*, **81**, 1121-1127.
- Hwang, S. Y. and Basawa, I. V. (2014). Martingale estimating functions for stochastic processes : A review toward a unifying tool. In *Contemporary Developments in Statistical Theory*, edited by S. Lahiri *et al.*, Springer, Switzerland, 9-28.
- Hwang, S. Y., Basawa, I. V., Choi, M. S. and Lee, S. D. (2014a). Non-ergodic martingale estimating functions and related asymptotics. *Statistics*, **48**, 487-507.
- Hwang, S. Y., Choi, M. S. and Yeo, In-kwon (2014b). Quasilikelihood and quasi maximum likelihood for GARCH-type processes : Estimating function approach. *Journal of the Korean Statistical Society*, DOI :10.1016/j.jkss.2014.01.005.
- Hwang, S. Y. and Choi, M. S. (2009). Modeling and large sample estimation for multi-casting autoregression. *Statistics & Probability Letters*, **79**, 1943-1950.
- Hwang, S. Y. and Choi, M. S. (2011). Preliminary identification of branching heteroscedasticity for tree indexed autoregressive processes. *Communications of the Korean Statistical Society*, **18**, 809-816.
- Hwang, S. Y. and Kang, K. H. (2012). Asymptotics for a class of generalized multicast autoregressive process. *Journal of the Korean Statistical Society*, **41**, 543-554.
- Lee, H. Y. (2012). Property of regression estimators in GEE models for ordinal responses, *Journal of the Korean Data & Information Science Society*, **23**, 209-218.
- Mao, M. (2014). The asymptotic behaviors for least square estimation of multicast autoregressive processes. *Journal of Multivariate Analysis*, **129**, 110-124
- Straumann, D. and Mikosch, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series : A stochastic recurrence equations approach. *Annals of Statistics*, **34**, 2449-2495.