

Preliminary Identification of Branching-Heteroscedasticity for Tree-Indexed Autoregressive Processes

S. Y. Hwang^{1,a}, M. S. Choi^a

^aDepartment of Statistics, Sookmyung Women's University

Abstract

A tree-indexed autoregressive(AR) process is a time series defined on a tree which is generated by a branching process and/or a deterministic splitting mechanism. This short article is concerned with conditional heteroscedastic structure of the tree-indexed AR models. It has been usual in the literature to analyze conditional mean structure (rather than conditional variance) of tree-indexed AR models. This article pursues to identify quadratic conditional heteroscedasticity inherent in various tree-indexed AR models in a unified way, and thus providing some perspectives to the future works in this area. The identical conditional variance of sisters sharing the same mother will be referred to as the branching heteroscedasticity(BH, for short). A quasilielihood but preliminary estimation of the quadratic BH is discussed and relevant limit distributions are derived.

Keywords: Branching heteroscedasticity, quasilielihood, tree-indexed AR.

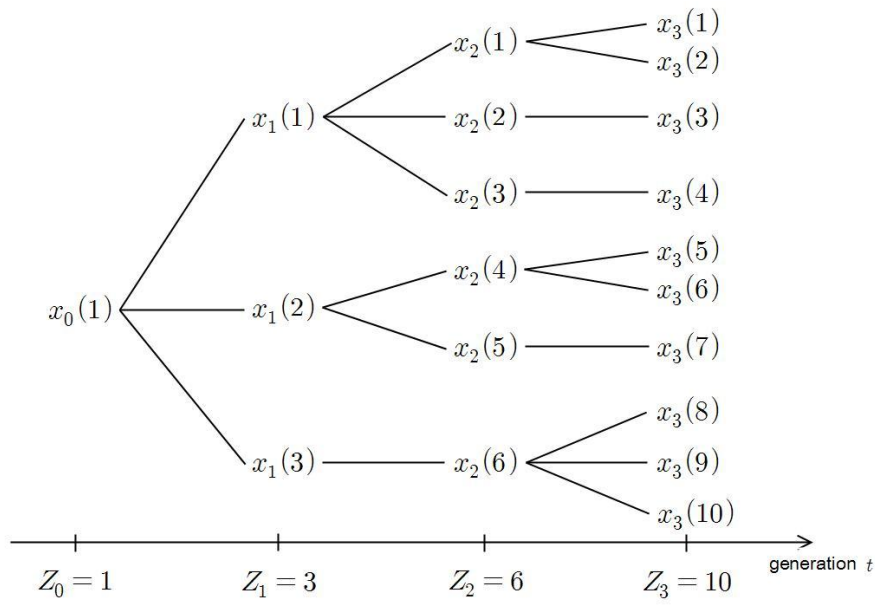
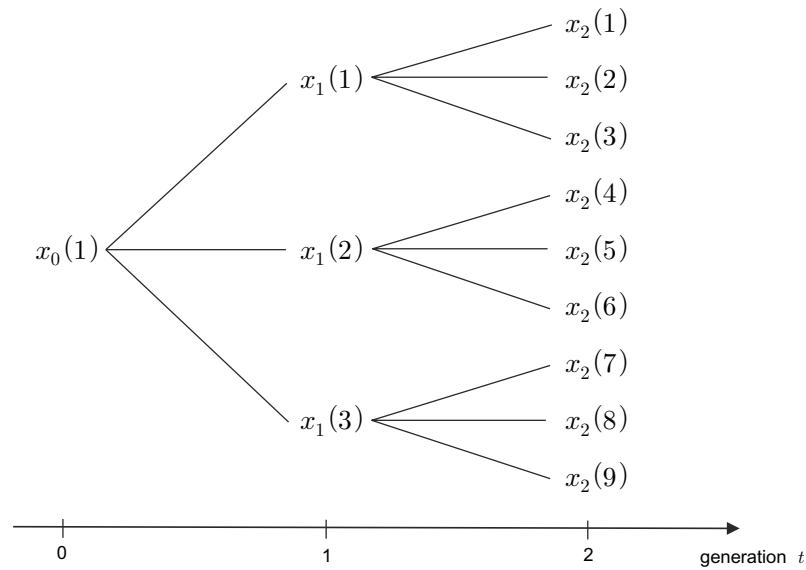
1. Motivation of the Study

We first construct a tree-index on which an AR time series (X) of interest is defined. Following the lines as in Hwang and Basawa (2011), consider the successive generation sizes $\{Z_t\}$ with the initial size $Z_0 = 1$. In particular, the super critical G-W(Galton-Watson) branching process $\{Z_t\}$ is defined by

$$Z_t = \sum_{j=1}^{Z_{t-1}} \eta_{tj}, \quad t = 1, 2, \dots,$$

where $\{\eta_{tj}, t = 1, 2, \dots, j = 1, 2, \dots\}$ is an array of iid non-negative integer-valued random variables with common (offspring) mean $m > 1$ and variance $\sigma_\eta^2 \geq 0$. Let $X_t(j)$ denote the observation on the j^{th} individual in t^{th} generation. In addition, let $X_{t-1}(t(j))$ denote the observation on immediate mother of the j^{th} individual in t^{th} generation. It is noticed that $X_{t-1}(t(j))$ is an observation made in the $(t-1)^{th}$ generation. In Figure 1, note that $x_2(3(1)) = x_2(1)$; $x_2(3(10)) = x_2(6)$. As with *e.g.*, Hwang (2011), one can consider two cases separately according as $\sigma_\eta^2 > 0$ and $\sigma_\eta^2 = 0$. Corresponding to $\sigma_\eta^2 > 0$, Figure 1 illustrates a tree consisting of three generations. It is noted in Figure 1 that there are random number of individuals in each generation. On the other hand, the case of $\sigma_\eta^2 = 0$ (see Figure 2) is referred to as a multi-casting tree where each individual (mother) gives rise to exactly m -offspring (daughters) in the next generation. When $m = 2$, the multi-casting tree reduces to a bifurcating case, *i.e.*, a binary-splitting tree studied by several authors including Cowan and Staudte (1986) and Basawa and Zhou (2004) among others. Most of the research on the multi-casting case of $\sigma_\eta^2 = 0$ has been directed to identification of the conditional mean function of the models. For traditional issues on the

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

Figure 1: A tree with $\sigma_\eta^2 > 0$ Figure 2: A tri-casting tree with $\sigma_\eta^2 = 0$

multi-casting tree such as the estimation of mean parameters and stability of the models, we refer to, for instance, Hwang and Choi (2009), Baek *et al.* (2011), Hwang and Basawa (2011), and Hwang (2011).

This article, however, focuses on the conditional variance function of the tree-indexed AR models, identifying preliminary conditional heteroscedasticity inherent in various tree-indexed AR models in

a unified way. To be more precise, throughout the paper, the identical conditional variance (denoted by h_t) of sisters sharing the same mother will be referred to as the branching-heteroscedasticity(BH, hereafter). It will be assumed that BH is of a quadratic function of the observations and hence BH is random rather than a constant. Quadratic nature of the BH is seen to be satisfied for various tree-indexed AR models, thereby enlarging the class of models under investigation. Due to the wideness of the BH structure, a quasilielihood estimation of BH in a broad context is discussed and relevant limit distribution is derived. We proceed as follows. In Section 2, the BH in tree-indexed AR models is introduced and is illustrated via various examples including standard AR, random coefficient AR and binomial thinning processes. A quasilielihood but preliminary identification for BH is studied and relevant asymptotic distributions are presented in Section 3. Although, for simplicity of presentation, we only confine ourselves to multi-casting cases of $\sigma_\eta^2 = 0$, main arguments can be extended to cover the case of $\sigma_\eta^2 > 0$.

2. Branching Heteroscedasticity(BH) and Illustrative Examples

Cowan and Staudte (1986) introduced a bifurcating-AR(BAR) model $\{X_t, t = 1, 2, \dots\}$ defined recursively by

$$\begin{aligned} X_{2t} &= \theta X_t + \epsilon_{2t} \\ X_{2t+1} &= \theta X_t + \epsilon_{2t+1}, \end{aligned} \quad (2.1)$$

where $\{(\epsilon_{2t}, \epsilon_{2t+1}); t = 1, 2, \dots\}$ is a sequence of iid bivariate random vector with mean vector zero, common variance $\sigma^2 > 0$ and correlation between ϵ_{2t} and ϵ_{2t+1} is given by ρ . It is noted that there are exactly two offspring X_{2t} and X_{2t+1} from the common mother X_t . Extending (2.1) to the multi-casting case of m offspring ($m \geq 3$), Hwang and Choi (2009) proposed the following multi-casting AR(MCAR) model generated by the m equations.

$$\begin{aligned} X_{mt-(m-2)} &= \theta X_t + \epsilon_{mt-(m-2)} \\ &\vdots \\ X_{mt} &= \theta X_t + \epsilon_{mt} \\ X_{mt+1} &= \theta X_t + \epsilon_{mt+1}, \end{aligned} \quad (2.2)$$

where $\{(\epsilon_{mt-(m-2)}, \dots, \epsilon_{mt}, \epsilon_{mt+1}), t = 1, 2, \dots\}$ is a sequence of iid m -variate normal random vectors with mean zero vector and the common variance $\sigma^2 > 0$. The correlation between any of the two among $\epsilon_{mt-(m-2)}, \dots, \epsilon_{mt+1}$ is modelled as ρ . Note that there are exactly m -sisters $X_{mt-(m-2)}, \dots, X_{mt+1}$ sharing the same mother $X_t, t = 1, 2, \dots$. Define sister vector S_t as

$$S_t = (X_{mt-(m-2)}, \dots, X_{mt}, X_{mt+1})^T : m \times 1. \quad (2.3)$$

Here ‘ T ’ indicates transpose of a matrix (or a vector). It is noted that the conditional mean vector of S_t is given by $E(S_t|X_t) = (\theta X_t)1_m$ for MCAR model in (2.2). Here and in the sequel 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

$$\text{Var}(X_{mt+i}|X_t) = \sigma^2, \quad i = -(m-2), -(m-1), \dots, 1. \quad (2.4)$$

Note that the MCAR model (2.1) provides a homoscedastic conditional variance σ^2 . We will consider the following heteroscedastic conditional variance h_t which is a quadratic function of the mother observation X_t and such an h_t is called a BH(branching heteroscedasticity).

Definition 1. BH h_t is defined as, for some non-negative constants β_0, β_1 and β_2 ,

$$h_t = \text{Var}(X_{m+i}|X_t) = \beta_0 + \beta_1 X_t + \beta_2 X_t^2, \quad i = -(m-2), -(m-1), \dots, 1. \quad (2.5)$$

Various examples for BH h_t are illustrated below.

Example 1. [MCAR with a intercept]

In accordance with S_t , define

$$e_t = (\epsilon_{m-(m-2)}, \dots, \epsilon_{mt}, \epsilon_{m+1}), \quad t = 1, 2, \dots$$

Consider the following MCAR model with a intercept θ_0 , as defined by

$$S_t = [\theta_0 + \theta X_t] 1_m + e_t, \quad (2.6)$$

where $\{e_t\}$ are iid vectors with mean zero vector and variance-covariance matrix Σ ,

$$\Sigma = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \sigma^2. \quad (2.7)$$

We note the homoscedastic BH given by $h_t = \sigma^2$.

Example 2. [Random coefficient MCAR; RC-MCAR]

Consider the process $\{X_t, t \geq 1\}$ such that

$$S_t = [\theta_0 + (\theta + \theta_t)X_t] 1_m + e_t, \quad (2.8)$$

where $\{e_t\}$ is defined in Example 1, $\{\theta_t\}$ denotes random coefficient of the autoregressive coefficient θ , and $\{\theta_t\}$ is a sequence of iid random variables with mean zero and variance σ_θ^2 , independently with $\{e_t, t \geq 1\}$. It is easy to see that the BH is given by

$$h_t = \sigma_\theta^2 X_t^2 + \sigma^2. \quad (2.9)$$

Example 3. [Binomial thinning MCAR]

Consider the following integer-valued process $\{X_t, t \geq 1\}$ defined by

$$S_t = (\theta \circ X_t) 1_m + e_t, \quad (2.10)$$

where \circ denotes the binomial thinning operator defined by $\theta \circ X_t = \sum_{i=1}^{X_t} B_i$ where $\{B_i\}$ is a sequence of iid Bernoulli random variables with success probability θ , $0 < \theta < 1$. Two processes $\{B_i\}$ and $\{e_t\}$ are assumed to be independent. Here, m -tuple error process $\{e_t\}$ 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 Σ in (2.7). See, e.g., Grunwald *et al.* (2000) and Baek *et al.* (2011). Notice that $E(S_t|X_t) = (\theta X_t + \lambda) 1_m$ and the BH can be verified to be

$$h_t = \sigma^2 + \theta(1 - \theta)X_t. \quad (2.11)$$

Consequently, Example 1 to Example 3 belong to our quadratic class BH defined in (2.5). For more examples belonging to the quadratic BH class, refer to Baek *et al.* (2011) and Basawa and Zhou (2004).

3. Main Results: Quasikelihood Identification of BH

Let Λ be a collection of multi-casting AR models $\{X_t\}$ satisfying the quadratic BH in (2.5). Note that Λ includes various models as discussed in Section 2. For the model $\{X_t\} \in \Lambda$, introduce the conditional mean (scalar) function $\mu_t(X_t)$ defined by

$$\mu_t(X_t) = E(X_{mt+i}|X_t), \quad i = -(m-2), -(m-1), \dots, 1. \quad (3.1)$$

We then have

$$E(S_t|X_t) = \mu_t(X_t)1_m.$$

The ‘residual’ process $\{r_t\}$ is constructed via

$$r_{mt+i} = X_{mt+i} - \mu_t(X_t), \quad i = -(m-2), -(m-1), \dots, 1, \quad (3.2)$$

and it is noted that

$$h_t = E(r_{mt+i}^2|X_t), \quad i = -(m-2), -(m-1), \dots, 1. \quad (3.3)$$

It then follows from (3.3) that $r_{mt+i}^2 - (\beta_0 + \beta_1 X_t + \beta_2 X_t^2)$ is a martingale difference for each $i = -(m-2), -(m-1), \dots, 1$. Consider the $m \times 1$ vector martingale process

$$R_t(\beta) = (r_{mt-(m-2)}^2 - (\beta_0 + \beta_1 X_t + \beta_2 X_t^2), \dots, r_{mt+1}^2 - (\beta_0 + \beta_1 X_t + \beta_2 X_t^2))^T. \quad (3.4)$$

Here $\beta = (\beta_0, \beta_1, \beta_2)^T$. Note that $E(R_t(\beta)|X_t) = 0$ and define the variance-covariance matrix $V_t(\beta) = \text{Var}(R_t(\beta)|X_t)$ which is a function of X_t . Let the data consist of S_1, \dots, S_n with the starting observation X_1 . Due to Godambe (1985), the quasikelihood estimator of $\beta = (\beta_0, \beta_1, \beta_2)^T$ is obtained from the quasikelihood estimating function

$$Q_n(\beta) = \sum_{t=1}^n E \left[\left(\frac{\partial R_t(\beta)}{\partial \beta^T} \right)^T \middle| X_t \right] V_t^{-1}(\beta) R_t(\beta) : 3 \times 1 \text{ vector}. \quad (3.5)$$

Due to the special structure of the model, $Q_n(\beta)$ is further simplified as

$$Q_n(\beta) = - \sum_{t=1}^n \begin{pmatrix} 1_m^T \\ X_t 1_m^T \\ X_t^2 1_m^T \end{pmatrix} V_t^{-1}(\beta) R_t(\beta). \quad (3.6)$$

A quasikelihood estimator $\hat{\beta}_{QL}$ of β is obtained by solving quasikelihood estimating equation viz., $Q_n(\beta) = 0$. It is noted that $Q_n(\beta)$ is optimal within a certain class of estimating functions in the sense of providing a maximum Godambe information matrix $I(Q_n(\beta))$ given by

$$I(Q_n(\beta)) = E \left(\frac{\partial Q_n(\beta)}{\partial \beta} \right) \left[E(Q_n(\beta) Q_n^T(\beta)) \right]^{-1} \left[E \left(\frac{\partial Q_n(\beta)}{\partial \beta} \right) \right]^T : 3 \times 3. \quad (3.7)$$

It is often the case in practice that $\hat{\beta}_{QL}$ may be obtained using one step solution when the quasikelihood estimating equation $Q_n(\beta) = 0$ is difficult to solve explicitly. For instance, the one-step solution $\tilde{\beta}_{QL}$ of $Q_n(\beta) = 0$ can be obtained via

$$\tilde{\beta}_{QL} = \tilde{\beta} - \left[\frac{\partial Q_n(\tilde{\beta})}{\partial \beta} \right]^{-1} Q_n(\tilde{\beta}), \quad (3.8)$$

where $\tilde{\beta}$ is a preliminary consistent estimator of β .

In particular when sisters are conditionally (on X_t) independent, $\hat{\beta}_{QL}$ can be of a explicit form. Define the conditional (central) fourth order moment $\kappa(X_t)$ as

$$\kappa_{mt+i}(X_t) = E(r_{mt+i}^4 | X_t), \quad i = -(m-2), -(m-1), \dots, 1. \quad (3.9)$$

Then, $V_t(\beta) = \text{Var}(R_t(\beta) | X_t) = E(R_t(\beta)R_t^T(\beta) | X_t)$ reduces to a diagonal matrix of order m with the diagonal elements $\kappa_{mt+i}(X_t)$ in (3.9). Consequently, under conditional independence, the QL estimating function $Q_n(\beta)$ reduces to

$$Q_n(\beta) = - \sum_{t=1}^n \begin{pmatrix} 1_m^T \\ X_t 1_m^T \\ X_t^2 1_m^T \end{pmatrix} \begin{pmatrix} r_{mt-(m-2)}^2 \kappa_{mt-(m-2)}^{-1} - (\beta_0 + \beta_1 X_t + \beta_2 X_t^2) \kappa_{mt-(m-2)}^{-1} \\ \vdots \\ r_{mt+1}^2 \kappa_{mt+1}^{-1} - (\beta_0 + \beta_1 X_t + \beta_2 X_t^2) \kappa_{mt+1}^{-1} \end{pmatrix}, \quad (3.10)$$

where the argument X_t is suppressed in $\kappa(X_t)$. This readily gives

$$\hat{\beta}_{QL} = \left[\sum_{t=1}^n \sum_{i=(2-m)}^1 \kappa_{mt+i}^{-1} \begin{pmatrix} 1 & X_t & X_t^2 \\ X_t & X_t^2 & X_t^3 \\ X_t^2 & X_t^3 & X_t^4 \end{pmatrix} \right]^{-1} \sum_{t=1}^n \sum_{i=(2-m)}^1 \begin{pmatrix} r_{mt+i}^2 \kappa_{mt+i}^{-1} \\ X_t r_{mt+i}^2 \kappa_{mt+i}^{-1} \\ X_t^2 r_{mt+i}^2 \kappa_{mt+i}^{-1} \end{pmatrix}. \quad (3.11)$$

Limit distribution of $\hat{\beta}_{QL}$ and (its one step version $\tilde{\beta}_{QL}$) is identified below.

Theorem 1. Assume that 3×3 non-random matrix Q exists and is invertible where

$$Q = E \left[\begin{pmatrix} 1_m^T \\ X_t 1_m^T \\ X_t^2 1_m^T \end{pmatrix} V_t^{-1}(\beta) \begin{pmatrix} 1_m, X_t 1_m, X_t^2 1_m \end{pmatrix} \right]. \quad (3.12)$$

We then have as n goes to infinity

$$\sqrt{n}(\hat{\beta}_{QL} - \beta) \xrightarrow{d} N(0, Q^{-1}) \quad (3.13)$$

and

$$\sqrt{n}(\tilde{\beta}_{QL} - \beta) \xrightarrow{d} N(0, Q^{-1}). \quad (3.14)$$

Proof: Notice that the quaslikelihood estimating function $Q_n(\beta)$ forms a vector of martingale differences and hence one can verify that

$$n^{-\frac{1}{2}} Q_n(\beta) \xrightarrow{d} N(0, Q). \quad (3.15)$$

A law of large number for martingales provides us with

$$n^{-1} \left(\frac{\partial Q_n(\beta)}{\partial \beta} \right) \xrightarrow{p} Q. \quad (3.16)$$

Here, \xrightarrow{d} and \xrightarrow{p} denote respectively “convergence in distribution” and “convergence in probability”. Since $Q_n(\hat{\beta}_{QL}) = 0$, one may asymptotically expand $Q_n(\beta)$ in (3.6) as

$$n^{-\frac{1}{2}} Q_n(\beta) = \left[n^{-1} \left(\frac{\partial Q_n(\beta)}{\partial \beta} \right) \right] \sqrt{n} (\hat{\beta}_{QL} - \beta),$$

which readily gives the desired result (3.13) using (3.15) and (3.16). Equation (3.14) regarding one-step solution is immediate because $\sqrt{n}(\hat{\beta}_{QL} - \beta)$ is asymptotically negligible due to the property of the Newton-Raphson iteration algorithm in (3.8). \square

Recall Λ being a collection of all multi-casting AR models $\{X_t\}$ satisfying the quadratic BH. Results discussed in Section 3 continues to be valid for all models in Λ and thus implementation of $\hat{\beta}_{QL}$ may be a preliminary action in the sense that $\hat{\beta}_{QL}$ is useful at an early stage of the analysis to identify a quadratic BH in effect for all models in Λ . We have not much discussed on the conditional mean function $\mu_t(X_t)$ defined in (3.1). To evaluate the residual process $\{r_t\}$, one can either estimate $\mu_t(X_t)$ non-parametrically or parametrically specify the conditional mean function as $\mu_t(X_t, \theta)$ involving parameters θ to be estimated. Let $\hat{\theta}$ denote a “good” estimator of θ , and the resulting estimated residual is then given by $r_{mt+i} = X_{mt+i} - \mu_t(X_t, \hat{\theta})$. One may choose $\hat{\theta}$ by minimizing (with respect to θ) $\sum_{t=1}^n \sum_{i=(2-m)}^1 (X_{mt+i} - \mu_t(X_t, \theta))^2$. It is usual for the conditional variance to depend on the parameter, θ say, appearing in the mean function. Consequently, we have $R_t(\beta, \theta)$, $V_t(\beta, \theta)$ and $Q_n(\beta, \theta)$ in place of $R_t(\beta)$, $V_t(\beta)$ and $Q_n(\beta)$ in (3.6). Specifically, the quasilielihood estimating equation is given by

$$Q_n(\beta, \theta) = - \sum_{t=1}^n \begin{pmatrix} 1_m^T \\ X_t 1_m^T \\ X_t^2 1_m^T \end{pmatrix} V_t^{-1}(\beta, \theta) R_t(\beta, \theta). \quad (3.17)$$

A modified quasilielihood estimator $\hat{\beta}_{MQL}$ is obtained by solving $Q_n(\beta, \hat{\theta}) = 0$. Under some regularity conditions, it can be shown that $\hat{\beta}_{MQL}$ has the same limiting distribution as for $\hat{\beta}_{QL}$ addressed in (3.13). Details are omitted. Refer to, for instance, Basawa and Zhou (2004) for a modified quasilielihood estimation.

References

- Baek, J. S., Choi, M. S. and Hwang, S. Y. (2011). A broad class of partially specified autoregressions on multicasting data, *Communications in Statistics-Theory and Methods*, to appear.
- Basawa, I. V. and Zhou, J. (2004). Non-Gaussian bifurcating models and quasi-likelihood estimation, *Journal of Applied Probability*, **Spec. 41A**, 55–64.
- Cowan, R. and Staudte, R. G. (1986). The bifurcating autoregression model in cell lineage studies, *Biometrics*, **42**, 769–783.
- 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 & New Zealand Journal of Statistics*, **42**, 479–495.
- 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. (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 Choi, M. S. (2009). Modeling and large sample estimation for multi-casting autoregression, *Statistics & Probability Letters*, **79**, 1943–1950.

Received July 2011; Accepted September 2011