UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

TRANSFORMED GENERALIZED ARMA MODELS

Gauss M. Cordeiro
Marinho G. de Andrade

Nº 81

NOTAS

São Carlos - SP

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2577

TRANSFORMED GENERALIZED ARMA MODELS

Gauss M. Cordeiro
Marinho G. de Andrade

Nº 81

NOTAS

Série Estatística

São Carlos – SP
Mar./2007

Transformed Generalized ARMA Models

Gauss M. Cordeiro
Departamento de Estatística e Informática,
Universidade Federal Rural of Pernambuco,
52171-900, Recife, PE, Brasil
e-mail: *gauss@deinfo.ufrpe.br*

Marinho G. de Andrade
Departamento de Matematica Aplicada e Estatística,
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo,
C.P. 668, 13560-970, São Carlos, SP, Brasil
e-mail: *marinho@icmc.usp.br*

Abstract

Generalized autoregressive moving average (GARMA) models discussed by Benjamin et al. [1] is a flexible observation-driven model for non-Gaussian time series data. GARMA models are based on exponential family models and enable the fitting of models to a wide range of time series data types. We propose a class of transformed generalized ARMA (TGARMA) models that extend the GARMA models and discuss maximum partial likelihood estimation and inference. We obtain a simple formula to estimate the parameter that index the transformation of the response variable. The TGARMA model is demonstrated by simulation. We give an application to a real time series data set.

Keywords: Dispersion parameter, Family of transformations, Generalized ARMA model, Generalized linear model, Profile likelihood.

1 Introduction

This article considers the problem of extending Box-Cox models to a non-Gaussian framework to deal with non-Gaussian time series models providing an extension of the generalized autoregressive moving average (GARMA) models (Benjamin et al., [1]).

Suppose we observe a pair of jointly distributed time series $\{(x_t, Y_t), t = 1, \ldots, n\}$, where Y_t is the response time series of interest and x_t is a time-dependent random covariate. We work with a rather general parametric family of transformations from the response variable Y_t to

$$Y_t^{(\lambda)} = \Lambda(Y_t; \lambda), \tag{1}$$

where $\Lambda(Y_t; \lambda)$ is a known strictly monotonic real function of Y_t depending on an unknown scalar parameter λ defining a particular transformation. Frequently, we use the Box-Cox [2] power transformation for positive continuous time series, $Y_t^{(\lambda)} = (Y_t^\lambda - 1)/\lambda$ when $\lambda \neq 0$ or $Y_t^{(\lambda)} = \log(Y_t)$ when $\lambda = 0$, and assume that there exists a λ value for the response variable such that $Y_t^{(\lambda)}$ follows a Gaussian autoregressive moving average (ARMA) time series model. Manly [9] proposed the exponential transformation to deal with negative Y_t's: $Y_t^{(\lambda)} = (e^{\lambda Y_t} - 1)/\lambda$ when $\lambda \neq 0$ or $Y_t^{(\lambda)} = \log(Y_t)$ when $\lambda = 0$. Other transformations can be defined for count, binary and categorical time series. The Box-Cox type of power transformations have generated a great deal of interests, both in theoretical work and in practical applications. Inference procedures for the regression coefficients and transformation parameter under this model setting have been studied extensively. The power transformation is only adequate for continuous positive data. Clearly not all data could be power-transformed to normal. Draper and Cox [5] studied this problem and conclude in one example that if the raw data follow an exponential distribution, values of λ close to its estimate will yield Weibull distributions for the transformed data.

Generalized linear models (GLMs), first introduced by Nelder and Wedderburn [10], are based on distributions that are exponential family models. GLMs extend the normal theory linear model, include a general algorithm for computing the maximum likelihood estimates (MLEs) and enable the fitting of different types of models to a wide range of data types. The main ideas of GLMs (exponential family, link function) can be extended quite readily to time series and several authors presented special GLMs for time series analysis. See, for example, the references given in Benjamin et al. [1], Li [8] and Fokianos and Kedem [6], where Poisson, binomial logistic, negative binomial and gamma GARMA models are discussed and applied to real time series data.

In this paper, we work with a rather general family of monotonic transformations (1) (see, for example, Sakia [11]) and incorporate the idea of transforming the response variable to follow the framework of the GARMA model. The transformed generalized ARMA (TGARMA) model assumes that there exists some λ value in (1) such that the transformed time series $Y_t^{(\lambda)}, t = 1, ..., n$, satisfy the full assumptions of the GARMA models.

In Section 2 we define the TGARMA model and give a summary of key results for these models. The maximum partial likelihood estimation is discussed in Section 3 for continuous TGARMA models. Section 4 considers some special continuous TGARMA models. Sections 5 deals with partial likelihood inference for these models. In Section

6 we present simulated gamma TGARMA models and provide an application to a real time series dataset.

2 Model Definition

Let $\{y_t, t = 1, \ldots, n\}$ be the observed time series. The transformation (1) applied to this time series yields the transformed time series $\{y_t^{(\lambda)}, t = 1, \ldots, n\}$. We assume that the conditional distribution of the transformed response $\{Y_t^{(\lambda)}, t = 1, \ldots, n\}$ given the past history of the process belongs to the exponential family distribution. The conditional density function of $Y_t^{(\lambda)}$ is defined given the set $H_t = \{x_t, \ldots, x_1, y_{t-1}, \ldots, y_1, \mu_{t-1}, \ldots, \mu_1\}$ that represents past values of the transformed series and their means and past and possibly present values (when known) of the covariates, meaning all that is known to the observer at time t, where x_t is a specified $1 \times m$ $(m < n)$ vector of explanatory variable. The set H_t then represents covariate and outcome history as a function of $x_t, \ldots, x_1, y_{t-1}, \ldots, y_1$ and $\mu_{t-1}, \ldots, \mu_1$. The conditional distribution of the transformed response given H_t is

$$\pi(y_t^{(\lambda)}|H_t) = \exp\left[\frac{1}{\phi}\left\{y_t^{(\lambda)}\theta_t - b(\theta_t)\right\} + c(y_t^{(\lambda)}, \phi)\right], \tag{2}$$

where θ_t and ϕ are the canonical and dispersion parameters, respectively, and $b(\cdot)$ and $c(\cdot, \cdot)$ are specific functions that define the particular distribution in the exponential family (2). The notation adopted is the same as for GLMs with independent observations, but here conditional rather than marginal distributions are modeled. The conditional mean and variance of $Y_t^{(\lambda)}$ are $E\{Y_t^{(\lambda)}|H_t\} = \mu_t$, where $\mu_t = db(\theta_t)/d\theta_t$ and $Var\{Y_t^{(\lambda)}|H_t\} = \phi V_t$, where $V_t = d\mu_t/d\theta_t$ is the variance function.

As with the standard GLM, the mean μ_t is related to the predictor η_t by a monotone link function $g(\cdot)$ assumed to be known. Unlike the standard GLM linear predictor, in the systematic component of GARMA models there is an additional component τ_t that allows autoregressive moving average terms to be included additively in the linear predictor and then the mean μ_t is given by $g(\mu_t) = \eta_t = x_t\beta + \tau_t$, where $\beta = (\beta_1, \ldots, \beta_m)^T$ is a set of unknown linear parameters to be estimated and τ_t is the ARMA component given by Benjamin [1]. We propose a general model for the mean μ_t of the transformed time series defined by the following expression

$$g(\mu_t) = \eta_t = x_t\beta + \sum_{j=1}^{p}\varphi_j\left\{g(y_{t-j}^{(\lambda)}) - x_{t-j}\beta\right\} + \sum_{j=1}^{q}\psi_j\left\{g(y_{t-j}^{(\lambda)}) - \eta_{t-j}\right\}, \tag{3}$$

which includes many well-known special models. We define the TGARMA(p, q, λ) model by the equations (1), (2) and (3). GARMA model proposed by Benjamin [1] is a special case of the TGARMA model when (1) is independent of λ and given by $\Lambda(Y_t) = Y_t$.

The class of TGARMA models has potentially wide applications since generalizes the common ARMA models by considering a more general family of distributions for the response time series and by including functions of past response and/or past conditional mean response values and GARMA models by incorporating an extra parameter λ. The aim of the transformation (1) is to ensure that the usual assumptions for GARMA models hold for the transformed series $Y_t^{(\lambda)}$.

The function $c(x, \phi)$ in (2) plays a fundamental role in the process of fitting the TGARMA models. When (2) is a two-parameter full exponential family distribution with canonical parameters $1/\phi$ and θ/ϕ, we have the following decomposition

$$c(x, \phi) = \frac{1}{\phi}a(x) + d(\phi) + d_1(x). \tag{4}$$

Equation (4) holds for transformed Gaussian, gamma and inverse Gaussian GARMA models which are discussed in Section 4.

3 TGARMA Model Fitting

We assume that the transformed response $Y_t^{(\lambda)}$ for some unknown transformation parameter λ in (1) satisfies the usual assumptions (2) and (3) of the GARMA models. The $m+p+q+2$ parameters of the TGARMA model to be estimated are then the vector $\gamma = (\beta^T, \varphi^T, \psi^T)^T$, where $\beta = (\beta_1, \ldots, \beta_m)^T$, $\varphi = (\varphi_1, \ldots, \varphi_p)^T$, $\psi = (\psi_1, \ldots, \psi_q)^T$ and the scalars ϕ and λ. The main objective in the analysis of TGARMA models is to make partial likelihood inference on the model parameters. The model fitting procedure described herein is valid only for continuous time series and exclude count time, binary and categorical time series.

The partial likelihood introduced by Cox ([4]) is based entirely on the conditional distribution of the current response, given past responses, and past covariate information and functions thereof and can be used for inference. It is used since is conceptually easy and the profile log-partial likelihood for λ could be computed easily in the most important cases. The log-likelihood for the parameter vector γ and scalars ϕ and λ expressed in terms of the transformed series $y^{(\lambda)} = (y_{r+1}^{(\lambda)}, \ldots, y_n^{(\lambda)})^T$ conditioned on the first r transformed observations, where $r = \max\{p, q\}$, leads to the following partial likelihood

$$PL(\gamma, \phi, \lambda) = \prod_{t=r+1}^{n} \pi(y_t^{(\lambda)}|H_t)J(\lambda, y_t), \tag{5}$$

where $J(\lambda, y)$ is the Jacobian of the transformation from y_t to $y_t^{(\lambda)}$. The log-partial likelihood for the model parameters is simply

$$l(\gamma,\phi,\lambda)=\frac{1}{\phi}\sum_{t=r+1}^{n}\left\{y_t^{(\lambda)}\theta_t-b(\theta_t)\right\}+\sum_{t=r+1}^{n}\left[c(y_t^{(\lambda)},\phi)+\log\left\{J(\lambda,y_t)\right\}\right], \tag{6}$$

and

$$J(\lambda,y_t)=\left|\frac{d\Lambda(y_t;\lambda)}{dy_t}\right|.$$

The Box-Cox power GARMA model, termed here PGARMA model, yields $\log\{J(\lambda,y_t)\}=(\lambda-1)\log|y_t|$.

We define the local matrices A and B of orders $(n-r)\times p$ and $(n-r)\times q$ which are functions of the model parameters by

$$A=\begin{bmatrix} g(y_r^{(\lambda)})-x_r\beta & \cdots & g(y_{r+1-p}^{(\lambda)})-x_{r+1-p}\beta \\ g(y_{r+1}^{(\lambda)})-x_{r+1}\beta & \cdots & g(y_{r+2-p}^{(\lambda)})-x_{r+2-p}\beta \\ \vdots & \ddots & \vdots \\ g(y_{n-1}^{(\lambda)})-x_{n-1}\beta & \cdots & g(y_{n-p}^{(\lambda)})-x_{n-p}\beta \end{bmatrix}_{(n-r)\times p}$$

and

$$B=\begin{bmatrix} g(y_r^{(\lambda)})-\eta_r & \cdots & g(y_{r+1-q}^{(\lambda)})-\eta_{r+1-q} \\ g(y_{r+1}^{(\lambda)})-\eta_{r+1} & \cdots & g(y_{r+2-q}^{(\lambda)})-\eta_{r+2-q} \\ \vdots & \ddots & \vdots \\ g(y_{n-1}^{(\lambda)})-\eta_{n-1} & \cdots & g(y_{n-q}^{(\lambda)})-\eta_{n-q} \end{bmatrix}_{(n-r)\times q}.$$

We can write the systematic component (3) of the TGARMA model corresponding to the observations $y_{r+1}^{(\lambda)},\ldots,y_n^{(\lambda)}$ in matrix notation by

$$\eta=\begin{bmatrix} X & A & B \end{bmatrix}\begin{bmatrix}\beta\\ \varphi\\ \psi\end{bmatrix},$$

where X is the local matrix formed by the rows x_t for $t=r+1,\ldots,n$. The systematic component reduces to $\eta=M\gamma$, the local model matrix being $M=[X\ A\ B]$ of order $(n-r)\times(m+p+q)$.

For maximizing the log-partial likelihood (6), we assume first that the transformation parameter λ is fixed and then obtain the partial likelihood equations for estimating γ and ϕ. Let $\widehat{\gamma}^{(\lambda)}$, $\widehat{\eta}^{(\lambda)}=\widehat{M}^{(\lambda)}\gamma^{(\lambda)}$ and $\widehat{\phi}^{(\lambda)}$ be the maximum partial likelihood estimates (MPLEs) of γ, η and ϕ, respectively, for given λ. The estimate $\widehat{\gamma}^{(\lambda)}$ does not depend on the dispersion parameter ϕ. For fixed λ, the MPLE $\widehat{\gamma}^{(\lambda)}=(\widehat{\beta}^{(\lambda)^T},\widehat{\varphi}^{(\lambda)^T},\widehat{\psi}^{(\lambda)^T})^T$ can be obtained from the fitting of the model (2)-(3) to $y^{(\lambda)}$ by iteratively reweighted least squares

$$\widehat{\gamma}^{(\lambda)}=\left(\widehat{M}^{(\lambda)T}\widehat{W}^{(\lambda)}\widehat{M}^{(\lambda)}\right)^{-1}\widehat{M}^{(\lambda)T}\widehat{W}^{(\lambda)}\widehat{z}^{(\lambda)}. \tag{7}$$

We can construct easily the iterative algorithm (7) from the weight matrix $W = diag\{w_{r+1},\ldots,w_n\}$, $w_t = V_t^{-1}\left(\frac{dg(\mu_t)}{d\mu_t}\right)^{-2}$ and the working variate $z^{(\lambda)} = (z_{r+1}^{(\lambda)},\ldots,z_n^{(\lambda)})^T$ with typical component given by

$$z_t^{(\lambda)} = \eta_t + \left(y_t^{(\lambda)} - \mu_t\right)\frac{dg(\mu_t)}{d\mu_t},$$

where the estimates $\widehat{\eta}_t^{(\lambda)}$ for $t = r+1,\ldots,n$ are obtained from

$$\widehat{\eta}_t^{(\lambda)} = x_t\widehat{\beta}^{(\lambda)} + \sum_{j=1}^{p}\widehat{\varphi}_j^{(\lambda)}\left\{g(y_{t-j}^{(\lambda)}) - x_{t-j}\widehat{\beta}^{(\lambda)}\right\} + \sum_{j=1}^{q}\widehat{\psi}_j^{(\lambda)}\left\{g(y_{t-j}^{(\lambda)}) - \widehat{\eta}_{t-j}^{(\lambda)}\right\}$$

and the adjusted means of the transformed series come from $\widehat{\mu}_t^{(\lambda)} = g^{-1}(\widehat{\eta}_t^{(\lambda)})$ for $t = r+1,\ldots,n$. An initial approximation $\widehat{\gamma}_1^{(\lambda)}$ for the iterative algorithm is used to evaluate $\widehat{M}^{(\lambda)}$, $\widehat{W}^{(\lambda)}$ and $\widehat{z}^{(\gamma)}$ from which equation (7) can be used to obtain the next estimate $\widehat{\gamma}_2^{(\lambda)}$. This new value can update $\widehat{M}^{(\lambda)}$, $\widehat{W}^{(\lambda)}$ and $\widehat{z}^{(\lambda)}$ and so the iterations continue until convergence is observed.

We now move to estimate ϕ. The MPLE $\widehat{\phi}^{(\lambda)}$ of ϕ for fixed λ can be obtained by differentiating (6) with respect to ϕ. We have

$$\widehat{\phi}^{(\lambda)2}\sum_{t=r+1}^{n}\left.\frac{dc(y_t^{(\lambda)},\phi)}{d\phi}\right|_{\phi=\widehat{\phi}} = \sum_{t=r+1}^{n}\left\{y_t^{(\lambda)}\widehat{\theta}_t^{(\lambda)} - b(\widehat{\theta}_t^{(\lambda)})\right\}, \tag{8}$$

where $\widehat{\theta}_t^{(\lambda)} = q(g^{-1}(x_t\widehat{\beta}^{(\lambda)}))$. Given the variance function $V(x)$ we can easily obtain $q(x) = \int V(x)^{-1}dx$ and $b(x) = \int q^{-1}(x)dx$, and then the deviance $D^{(\lambda)}$ of the TGARMA model comes easily as

$$D^{(\lambda)} = 2\sum_{t=r+1}^{n}\left\{y_t^{(\lambda)}q(y_t^{(\lambda)}) - b(q(y_t^{(\lambda)}))\right\} - 2\sum_{t=r+1}^{n}\left\{y_t^{(\lambda)}q(\widehat{\mu}_t^{(\lambda)}) - b(q(\widehat{\mu}_t^{(\lambda)}))\right\}. \tag{9}$$

The MPLE $\widehat{\phi}^{(\lambda)}$ is a function of the deviance (9) and using (8) we obtain

$$\widehat{\phi}^{(\lambda)2}\sum_{t=r+1}^{n}\left.\frac{dc(y_t^{(\lambda)},\phi)}{d\phi}\right|_{\phi=\widehat{\phi}^{(\lambda)}} = \sum_{t=r+1}^{n}e(y_t^{(\lambda)}) - \frac{D^{(\lambda)}}{2}, \tag{10}$$

where $e(x) = xq(x) - b(q(x))$. Equation (10) is in general nonlinear except for normal and inverse Gaussian models.

Substituting the MPLEs $\widehat{\gamma}^{(\lambda)}$ and $\widehat{\phi}^{(\lambda)}$ in (6) yields the profile log-partial likelihood for λ

$$l_P(\lambda) = \frac{1}{\widehat{\phi}^{(\lambda)}}\sum_{t=r+1}^{n}\left\{y_t^{(\lambda)}\widehat{\theta}_t^{(\lambda)} - b(\widehat{\theta}_t^{(\lambda)})\right\} + \sum_{t=r+1}^{n}\left[c(y_t^{(\lambda)},\widehat{\phi}^{(\lambda)}) + \log\{J(\lambda,y_t)\}\right], \tag{11}$$

which in terms of the deviance of the TGARMA model reduces to

$$l_P(\lambda) = \frac{1}{\widehat{\phi}^{(\lambda)}} \sum_{t=r+1}^{n} e(y_t^{(\lambda)}) - \frac{D^{(\lambda)}}{2\widehat{\phi}^{(\lambda)}} + \sum_{t=r+1}^{n} \left[c(y_t^{(\lambda)}, \widehat{\phi}^{(\lambda)}) + \log\{J(\lambda, y_t)\} \right]. \quad (12)$$

For any TGARMA model we can construct equation (12) from the functions $e(x)$, $c(x, \phi)$ obtained from (2), the deviance $D^{(\lambda)}$ and the Jacobian. Then, the MPLE $\widehat{\phi}^{(\lambda)}$ and the profile log-partial likelihood $l_P(\lambda)$ can be computed numerically from equations (10) and (12). The plot of the profile log-partial likelihood $l_P(\lambda)$ in (12) against λ for a trial series of values determines numerically the value of the MPLE $\hat{\lambda}$. Once the $\hat{\lambda}$ is obtained from the plot, it can be substituted into equations (7) and (10) to produce the unrestricted estimates $\hat{\gamma}$ and $\hat{\phi}$. The process of estimating γ, ϕ and λ can be carried out by standard statistical software such as MATLAB, S-PLUS, SAS and R.

For two-parameter full exponential family distribution, the decomposition of the function $c(x, \phi)$ given in (4) yields the equation for $\widehat{\phi}^{(\lambda)}$

$$(n-r)\widehat{\phi}^{(\lambda)2} d'(\widehat{\phi}^{(\lambda)}) = \sum_{t=r+1}^{n} t(y_t^{(\lambda)}) - \frac{D^{(\lambda)}}{2}, \quad (13)$$

where $t(x) = xq(x) - b(q(x)) + a(x)$, and using (8), we can write

$$l_P(\lambda) = (n-r)v(\widehat{\phi}^{(\lambda)}) + \sum_{t=r+1}^{n} \left\{ d_1(y_t^{(\lambda)}) + \log J(\lambda, y_t) \right\}, \quad (14)$$

where $v(x) = xd'(x) + d(x)$.

In Table 1 we give the functions $d(x)$, $t(x)$, $v(x)$ and $d_1(x)$ which enable us to compute $\widehat{\phi}^{(\lambda)}$ in (13) and the profile log-partial likelihood (14) for some TGARMA models, where $\Gamma(\cdot)$ and $\Psi(\cdot)$ are the gamma and digamma functions, respectively. For the power Gaussian ARMA models, (14) is identical to equation (8) given by Box and Cox [2] and can be viewed as a generalization of this equation for some other continuous transformed non-Gaussian ARMA models.

Table 1: Special Transformed ARMA Models

Model	$d(x)$	$t(x)$	$v(x)$	$d_1(x)$
Normal	$-\frac{1}{2}\log x$	0	$-\frac{1}{2}(1+\log x)$	$-\frac{1}{2}\log(2\pi)$
Gamma	$-\frac{\log x}{2} - \log\Gamma(\frac{1}{x})$	-1	$\frac{1}{x}\Psi(\frac{1}{x}) - \frac{1}{x} - \log\Gamma(\frac{1}{x})$	$-\log x$
Inverse Gaussian	$-\frac{1}{2}\log x$	0	$-\frac{1}{2}(1+\log x)$	$-\frac{1}{2}\log(2\pi x^3)$

We can estimate the mean of the untransformed series Y_t using Taylor series expansion of $Y_t = F(Y_t^{(\lambda)}; \lambda)$, where $F(.)$ is the inverse transformation $\Lambda^{-1}(.)$ of (1). Conditioning on the set H_t we obtain

$$\widehat{E}(Y_t) \approx F(\widehat{\mu}_t; \widehat{\lambda}) + \frac{\widehat{\phi}\widehat{V}_t}{2} F''(\widehat{\mu}_t; \widehat{\lambda}).$$

Additional terms can be easily included in this equation since the central moments of $Y_t^{(\lambda)}$ are just given in terms of the derivatives of the variance function. For the Box-Cox power transformation $F''(\mu_t; \lambda) = (1-\lambda)(1+\lambda\mu_t)^{(1-2\lambda)/\lambda}$.

4 Special Continuous TGARMA Models

For transformed Gaussian and inverse Gaussian ARMA models, (13) yields

$$\widehat{\phi}^{(\lambda)} = \frac{D^{(\lambda)}}{n-r}. \tag{15}$$

For transformed gamma ARMA models, (13) reduces to Cordeiro and McCullagh's equation [3] given by

$$\log\left(\widehat{\phi}^{(\lambda)^{-1}}\right) - \Psi\left(\widehat{\phi}^{(\lambda)^{-1}}\right) = \frac{D^{(\lambda)}}{2(n-r)}. \tag{16}$$

As approximate solution for $\widehat{\phi}^{(\lambda)}$ is obtained from

$$\widehat{\phi}^{(\lambda)} = \frac{2D^{(\lambda)}}{(n-r)\left\{1 + \left(1 + \frac{2D^{(\lambda)}}{3(n-r)}\right)^{1/2}\right\}}. \tag{17}$$

For transformed Gaussian and inverse-Gaussian ARMA models, the profile log-partial likelihood reduces to

$$l_p(\lambda) = -\frac{(n-r)}{2}\log\widehat{\phi} - \frac{(n-r)}{2}[1+\log(2\pi)] + \sum_{i=r+1}^{n}\log\left\{\frac{J(\lambda, y_i)}{\sqrt{V(y_i^{(\lambda)})}}\right\}. \tag{18}$$

To maximize the profile log-partial likelihood (18), we only need to find a λ value that minimizes the ratio below

$$\hat{\lambda} = \arg\min_{\lambda} \frac{\left\{D^{(\lambda)}\widetilde{V}(y^{(\lambda)})\right\}^{1/2}}{\widetilde{J}(\lambda, y)}, \tag{19}$$

where $\widetilde{V}(y^{(\lambda)})$ and $\widetilde{J}(\lambda, y)$ are the geometric means of $V(y_t^{(\lambda)})$ and $J(\lambda, y_t)$ for $t = r+1, \ldots, n$, respectively. For PGARMA models, $\widetilde{J}(\lambda, y) = \widetilde{y}^{\lambda-1}$, where $\widetilde{y}$ is the geometric mean of the raw observations $y_{r+1}, \ldots, y_n$ and, in particular, for transformed Gaussian ARMA models ($V = 1$), equation (19) yields a known result given by Yang and Abeysinghe [12]. For transformed gamma ARMA models, the sum of the first two terms in (18) is substituted by $(n-r)h(\widehat{\phi})$, where

$$h(\phi) = \frac{1}{\phi}\left[\Psi\left(\phi^{-1}\right) - 1 - \phi \log \Gamma\left(\phi^{-1}\right)\right].$$

For small ϕ, we can obtain to order $O(\phi^3)$

$$h(\phi) = -\frac{1}{2} - \frac{\phi}{6} - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log\phi,$$

which gives

$$l_p(\lambda) = -\frac{(n-r)}{2}\left\{\log\widehat{\phi} - \frac{\widehat{\phi}}{3}\right\} - \frac{(n-r)}{2}\{1 + \log(2\pi)\} + \sum_{t=r+1}^{n} \log\left\{\frac{J(\lambda, y_t)}{\sqrt{V(y_t^{(\lambda)})}}\right\}. \quad (20)$$

Clearly, (20) converges to the form (18) when $\phi \to 0$ and, therefore, the profile log-partial likelihood for λ for all TGARMA models have the same form for very small dispersion parameter values.

5 Inference

We can make inference about ϕ and the parameters in γ conditioning on the transformed parameter $\lambda = \hat{\lambda}$. Then, the estimated value $\hat{\lambda}$ is viewed as known, and estimates of γ, η_t, μ_t and ϕ, confidence intervals for these parameters, analysis of the deviance, likelihood ratio tests, residuals and diagnostics can be carried over routinely to the TGARMA models in the usual context of GARMA models conditioning on $\lambda = \hat{\lambda}$.

Suppose we wish to test whether the estimate of the parameter of the transformation family (1) conforms to a hypothesized value $\lambda^{(0)}$. We can easily obtain from (12) a likelihood ratio (LR) statistic $w = 2\{l_P(\hat{\lambda}) - l_P(\lambda^{(0)})\}$ for testing $\lambda = \lambda^{(0)}$ and construct a large sample confidence interval for λ by inverting the LR test. The LR statistic w has under the null hypothesis an asymptotic χ_1^2 distribution. Then, an approximate $100(1-\alpha)\%$ confidence interval for λ is given by

$$\{\lambda \mid l_P(\lambda) > l_P(\hat{\lambda}) - \frac{1}{2}\chi_1^2(\alpha)\}.$$

We can also work with the square root of the LR statistic $w^{1/2}$, where the sign of the statistic is that of $\hat{\lambda} - \lambda^{(0)}$, which asymptotically is standard normal, to make inference about λ.

The scaled deviance in TGARMA models is defined conditioning on $\hat{\lambda}$ by $S^{(\hat{\lambda})} = D^{(\hat{\lambda})}/\hat{\phi}^{(\hat{\lambda})}$ and can be used for testing the adequacy of a TGARMA model fitted to a time series data. We can take $S^{(\hat{\lambda})}$ as distributed as $\chi^2_{n-r-m-p-q}$ approximately, but in general the chi-square approximation may not be effective because the dimension of the saturated model $n-r$ depends on n and the usual asymptotic argument does not apply. However, if we are testing two nested TGARMA models, the χ^2 distribution could be a good approximation for the difference of scaled deviances conditioned on the same $\hat{\lambda}$. Indeed, consider two nested TGARMA models A and B ($A \subset B$) conditioning on the same $\lambda = \hat{\lambda}$ which could be estimated under a well-fitted TGARMA model. For given $\hat{\lambda}$, the systematic component of the model B contains the same parameters of the model A plus additional parameters, the models being otherwise identical. For $\hat{\lambda}$ fixed, let $S_A^{(\hat{\lambda})}$ and $S_B^{(\hat{\lambda})}$ be the scaled deviances of the models A and B with ν_A and ν_B linear parameters being fitted. To test model A against model B, the LR statistic is just equal to the difference between the scaled deviances $w^{(\hat{\lambda})} = S_A^{(\hat{\lambda})} - S_B^{(\hat{\lambda})}$ and has an asymptotic $\chi^2_{\nu_B-\nu_A}$ distribution with an error of order n^{-1}.

Finally, consider a set of arbitrary TGARMA models $J = A, \cdots, I$ with log-partial likelihoods $\hat{l}_J$ obtained by maximizing (6) over all $\nu_J + 2$ parameters, ν_J parameters in the systematic component of the model J, and over the scalar parameters ϕ and λ. Evaluation and selection among TGARMA models A, ..., I may be based on Akaike information criterion (AIC) defined for the TGARMA model J by $AIC_J = 2(\nu_J + 2 - \hat{l}_J)$.

6 Application to simulated and real data

6.1 Simulated data

We assume $\psi_j = 0$ for $j = 1, \ldots, q$ and $\beta = 0$ and consider the transformed gamma autoregressive (denoted here by TGAR) model with reciprocal link function

$$\frac{1}{\mu_t} = \eta_t = \sum_{j=1}^{p} \left(\frac{\varphi_j}{y_{t-j}^{(\lambda)}} \right).$$

In this case $r = p$ and the matrix A reduces to

$$A = \begin{bmatrix} \frac{1}{y_p^{(\lambda)}} & \cdots & \frac{1}{y_1^{(\lambda)}} \\ \frac{1}{y_{p+1}^{(\lambda)}} & \cdots & \frac{1}{y_2^{(\lambda)}} \\ \vdots & \ddots & \vdots \\ \frac{1}{y_{n-1}^{(\lambda)}} & \cdots & \frac{1}{y_{n-p}^{(\lambda)}} \end{bmatrix}_{(n-p)\times p}.$$

The estimates $\widehat{\varphi}^{(\lambda)}$ can be obtained from the fitting of the model matrix A to $y^{(\lambda)}$ by iteratively reweighted least squares

$$\widehat{\varphi}^{(\lambda)}_{k+1} = \left(\widehat{A}^{(\lambda)T}_k \widehat{W}^{(\lambda)}_k \widehat{A}^{(\lambda)}_k\right)^{-1} \widehat{A}^{(\lambda)T}_k \widehat{W}^{(\lambda)}_k \widehat{z}^{(\lambda)}_k,$$

where $W = diag\{\eta^{-2}_{p+1}, \ldots, \eta^{-2}_n\}$ and the working variate $z^{(\lambda)}$ has components given by

$$z^{(\lambda)}_t = \eta_t - \frac{(y^{(\lambda)}_t - \mu_t)}{\mu^2_t}.$$

The estimate $\widehat{\lambda}$ comes as the solution of (19).

We simulate this TGAR model assuming a simple power transformation $y^{(\lambda)}_t = y^{\lambda}_t$ and $p = 2$ for specific values of the autoregressive parameters and the dispersion parameter ϕ. We aim to illustrate the use of the profile log-partial likelihood to estimate the transformation parameter λ for different values of the dispersion parameter. First, we consider a time series of small size $n = 50$ and take $\lambda = 2$, $\varphi_1 = 0.01$ and $\varphi_2 = 0.8$. The simulated data was obtained with a statistic tool box of the Matlab and all the estimates were calculated with an structural programming made in Matlab environment. The time series y^{λ}_t, $y^{\widehat{\lambda}}_t$ and the predicted values $\widehat{\mu}_t$ obtained from the fitted TGAR model are shown in Figures 1 and 2 when the dispersion parameter ϕ is equal to 1/10 to 1/20, respectively. The MPLEs of the model parameters are given in Table 1 and the profile log-partial likelihood for λ under theses two models are shown in Figures 3 and 4.

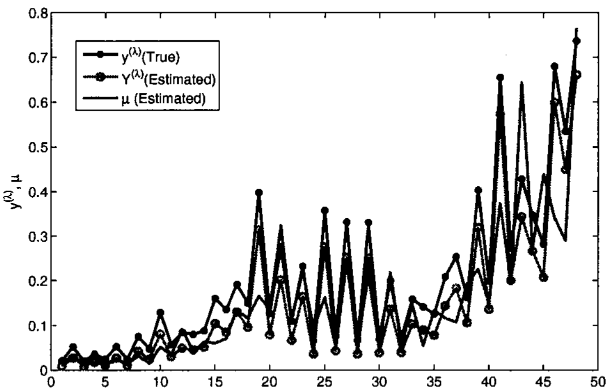

Figure 1: Simulated TGAR model, $n = 50$, $\phi = 1/10$, $\lambda = 2$, $\varphi = (0.01, 0.8)$

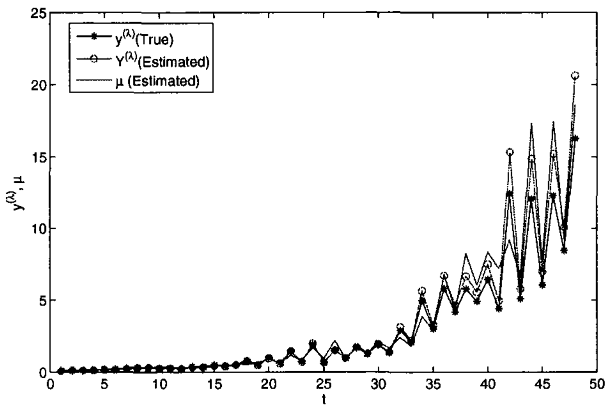

Figure 2: Simulated TGAR model, $n = 50$, $\phi = 1/20$, $\lambda = 2$, $\varphi = (0.01, 0.8)$

Table 1: True and parameter estimates for a small TGAR time series model

sample size	True parameters			Parameter Estimates		
n	ϕ	λ	φ	$\widehat{\phi}$	$\widehat{\lambda}$	$\widehat{\varphi}$
50	0.10	2	(0.01, 0.8)	0.165	2.415	(0.070, 0.690)
50	0.05	2	(0.01, 0.8)	0.041	2.188	(0.058, 0.728)

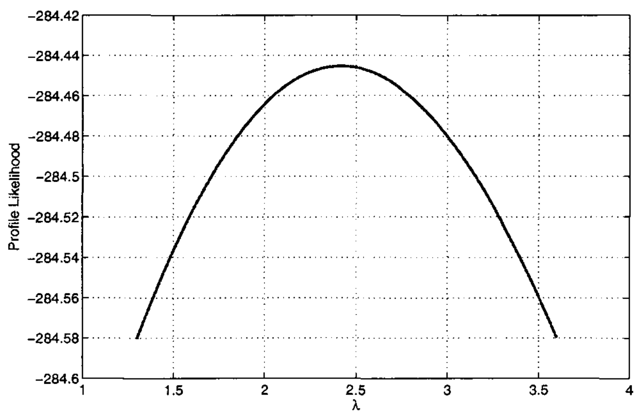

Figure 3: Profile likelihood for the TGAR model, $n = 50$, $\phi = 1/10$, $\lambda = 2$, $\varphi = (0.01, 0.8)$

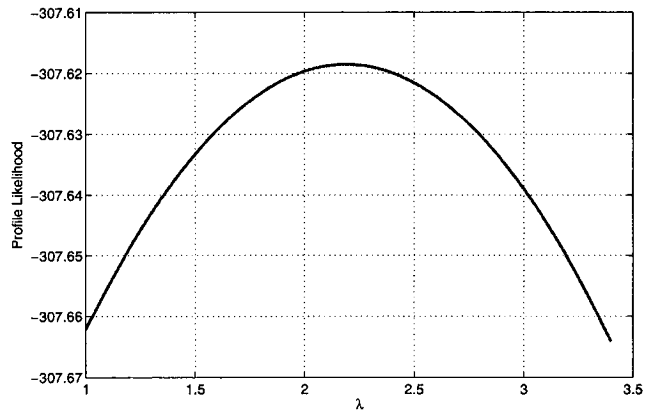

Figure 4: Profile likelihood for the TGAR model,
$n = 50$, $\phi = 1/20$, $\lambda = 2$, $\varphi = (0.01, 0.8)$

Second, we consider a time series with large sample size $n = 300$ and take $\lambda = 1/2$. Figures 5 and 6 present the simulated and predicted values from the fitted of the TGAR model with $p = 2$, different values of the autoregressive parameters and the dispersion parameter $\phi = 1/10$ and $1/20$, respectively. The estimates of the parameters of these two models are given in Table 2 and the profile log-partial likelihoods for λ under these models are shown in Figures 7 and 8, respectively.

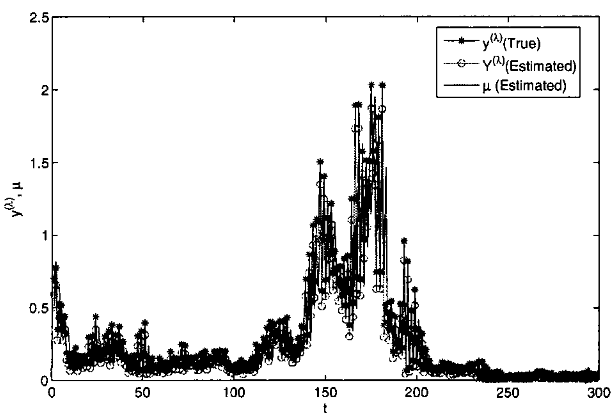

Figure 5: Simulated TGAR model, $n = 300$, $\phi = 1/10$, $\lambda = 0.5$, $\varphi = (0.08, 0.85)$

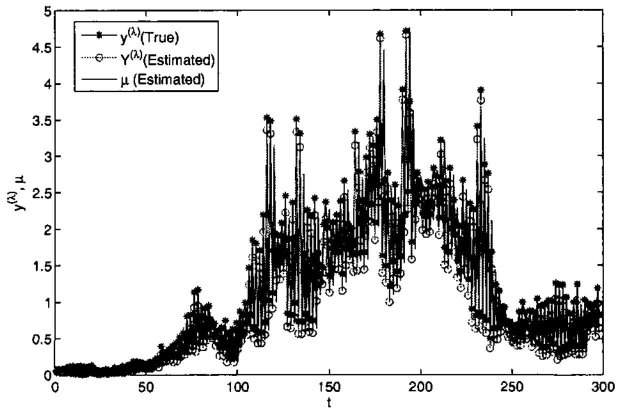

Figure 6: Simulated TGAR model, $n = 300$, $\phi = 1/20$, $\lambda = 0.5$, $\varphi = (0.05, 0.9)$

Table 2: True and estimates parameter for a large sample TGAR model

sample size	True parameters			Parameter Estimates		
n	ϕ	λ	φ	$\widehat{\phi}$	$\widehat{\lambda}$	$\widehat{\varphi}$
300	0.10	0.5	(0.08, 0.85)	0.119	0.54	(0.065, 0.872)
300	0.05	0.5	(0.05, 0.90)	0.0672	0.57	(0.045, 0.888)

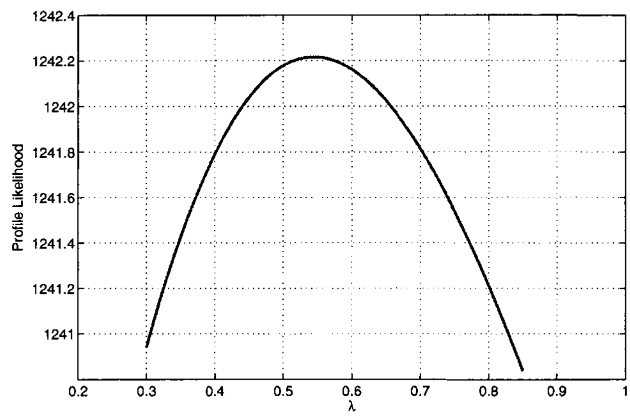

Figure 7: Profile likelihood for the TGAR model, $n = 50$, $\phi = 1/10$, $\lambda = 0.5$, $\varphi = (0.08, 0.85)$

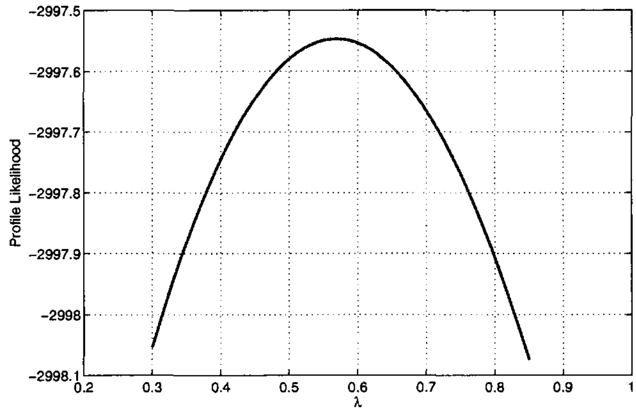

Figure 8: Profile likelihood for the TGAR model, $n = 300$, $\phi = 1/20$, $\lambda = 0.5$, $\varphi = (0.05, 0.90)$

We analyse other TGAR models fitted to time series data generated for different sizes $n = 50, 100, 200$, and 300, by taking the ARMA parameters in the intervals $0.01 \leq \varphi_1 \leq 0.08$ and $0.80 \leq \varphi_2 \leq 0.95$ and for $\phi = 1/10$, $1/20$ and $1/30$. The values of the parameter λ were taken as 0.5, 1.0, 1.5 and 2.0. These values were chosen to provide simulated series with good graphical visibility. The simulations were analyzed and the results were similar to those presented here. The MPLE of the transformation parameter λ is usually more accurate for long series. The dispersion parameter ϕ also affects the estimates of the parameter λ but to a less extent than the sample size of the series. The MPLEs of the parameters are usually more accurate when the dispersion parameter ϕ decreases.

6.2 Real data

The TGAR model proposed in this paper was applied to analyze a real time series representing the population of the common named American Black Bears, (species name: *Ursus americanus*) living in Manitoba, Canada, in the period of 1919 to 1981. The real time series is shown in Figure 9 and an complete description of this data set can be find in the Global Population Dynamics Database (GPDD, [7]).

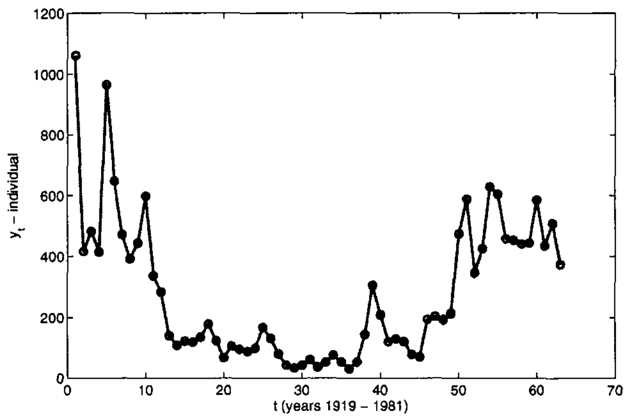

Figure 9: Population of American Black Bears (*Ursus americanus*) in Manitoba, Canada, of the 1919 to 1981

After fitting TGAR models of orders $p = 1, 2$ and 3, we choose the TGAR model of order $p = 1$ which yields a good adjustment. The MPLEs of the model parameters are presented in Table 3. The profile log-partial likelihood for λ is shown in Figure 10 and the transformed data $y_t^{\widehat{\lambda}}$ jointly with the adjusted means $\widehat{\mu}_t$ are given in Figure 11.

Table 3: TGAR model adjusted to the Population data of the *Ursus americanus*

p	$\widehat{\lambda}$	$\widehat{\varphi}$	$\widehat{\phi}$
1	0.573	0.945	0.061

A future line of research will aim to obtain the variance of the estimate of the parameter λ. Making an analogy with the classical ARMA models, it is interesting to establish a criterion for selecting the orders p and q of the model and conditions for choosing the distribution in (2) for a given observed time series and also to know how to predict future observations.

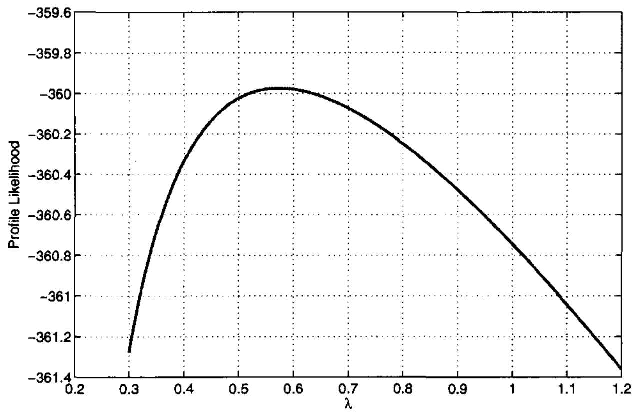

Figure 10: Profile log-partial likelihood for the parameterλ of the *Ursus americanus* population data set

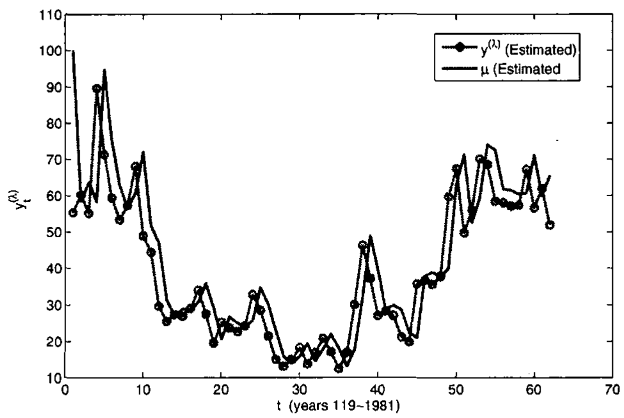

Figure 11: Transformed time series and adjusted means of the TGAR model to the *Ursus americanus* population data series

References

[1] Benjamin, M.A., Rigby, R.A., Stasinopoulos, D.M. (2003). Generalized autoregressive moving average models, J. Amer. Statist. Assoc., 98, 214-223.in the periodo of 1919 to 1981,

[2] Box, G.E.P., Cox, D.R. (1964). An analysis of transformation, J. R. Statist. Soc. B, 26, 211-252.

[3] Cordeiro G.M., McCullagh, P. (1991). Bias correction in generalized linear models, J. R. Statist. Soc. B, 53, 629-643.

[4] Cox, D. R. (1975). Partial likelihood, Biometrika, 62, 69-76.

[5] Draper, N. R., Cox, D.R. (1969). On distributions and their transformation to normality, J. R. Statist. Soc. B, 31, 472-476.

[6] Fokianos, K., Kedem, B. (2004). J. Time Series, 25, 173-197.

[7] The Global Population Dynamics Database, (1999). NERC Centre for Population Biology, Imperial College. http://www.sw.ic.ac.uk/cpb/cpb/gpdd.html.

[8] Li, W. K. (1994). Time series models based on generalized linear models: some further results, Biometrics, 50, 506-511.

[9] Manly, B. F. (1976). Exponential data transformation, The Statistician, 25, 37-42.

[10] Nelder, J.A., Wedderburn, R.W.M. (1972). Generalized linear models, J. R. Statist. Soc. A, 135, 370-384.

[11] Sakia, R.M. (1992). The Box-Cox transformation technique: a review, The Statistician, 41, 168-178.

[12] Yang, Z., Abeysinghe, T. (2002). An explict variance formula for the Box-Cox functional form estimator. Economics Letters, 76, 259-265.

NOTAS DO ICMC

SÉRIE ESTATÍSTICA

080/2007 ANDRADE, M. G.; LOIBEL, S.; ACHCAR, J.A. – Interfailure data with constant hazard function in the presence of change-points.

079/2006 ANDRADE, M.G.; OLIVEIRA, S.C. – Bayesian and maximum likelihood approaches for arch models considering brazilian financial series.

078/2005 LABRA, F.U.; AOKI, R.; ROJAS, F. – An application of the local influence diagnostics to ridge regression under elliptical model.

077/2005 ANDRADE, M.G. – Estimação bayesiana para processos PAR (*pm*) com a transformação de Box-Cox.

076/2004 COSTA, E.F.; VAL, J.B.R. – Generalized LQ problem for infinite jump linear systems and the minimal solution of algebraic Riccati equations.

075/2004 LOIBEL, S.; VAL, J.B.R.; ANDRADE, M.G. – Inference for the Richards Growth model using Box and Cox transformation and Bootstrap techniques.

074/2003 AOKI, R.; BOLFARINE, H.; ACHCAR, J.A.; LEÃO P. JR., D. – Bayesian analysis of a multivariate null intercept errors-in-variables regression model.

073/2002 LEÃO P. JR., D.; AOKI, R.; SILVA, G.F.- Statistical analysis of proficiency testing results.

072/2002 ANDRADE, MARINHO G.; FERREIRA, VALÉRIA A. MARTINS; OLIVEIRA, SANDRA C. – Abordagem bayesiana com MCMC versus máxima verossimilhança para modelos ARCH(p)

071/2002 LOUZADA-NETO, F.; TOMAZELLA, V.L.D.; ANDRADE, M.G. – Bayesian analysis for reccurent lifetime data with a non homogeneous Poisson process with a frailty term.