Beta autoregressive moving average models

Article *in* Test · November 2009

DOI: 10.1007/s11749-008-0112-z

Francisco Cribari-Neto
Federal University of Pernambuco

Test (2009) 18: 529–545
DOI 10.1007/s11749-008-0112-z

ORIGINAL PAPER

Beta autoregressive moving average models

Andréa V. Rocha · Francisco Cribari-Neto

Received: 1 November 2007 / Accepted: 26 May 2008 / Published online: 13 June 2008

Abstract We build upon the class of beta regressions introduced by Ferrari and Cribari-Neto (J. Appl. Stat. 31:799–815, 2004) to propose a dynamic model for continuous random variates that assume values in the standard unit interval $(0, 1)$. The proposed βARMA model includes both autoregressive and moving average dynamics, and also includes a set of regressors. We discuss parameter estimation, hypothesis testing, goodness-of-fit assessment and forecasting. In particular, we give closed-form expressions for the score function and for Fisher's information matrix. An application that uses real data is presented and discussed.

Keywords ARMA · Beta distribution · Beta ARMA · Forecasts

Mathematics Subject Classification (2000) 62M10 · 91B84

1 Introduction

The beta distribution is commonly used for modeling experiments in which the variable of interest is continuously distributed in the interval (a, b), where a and b are known scalars, and $a < b$, since its density can assume quite different shapes depending on the values of the two parameters that index the distribution. A particularly useful situation occurs when $a = 0$ and $b = 1$ so that the random variable assumes values in the standard unit interval, (0, 1); this is the case, e.g., of rates or proportions.

The beta probability density function is given by

$$\pi(y; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1}(1-y)^{q-1}, \quad 0 < y < 1, \tag{1}$$

A.V. Rocha (✉) · F. Cribari-Neto
Departamento de Estatística, Universidade Federal de Pernambuco. Cidade Universitária, Recife, PE, 50740-540, Brazil
e-mail: andrea@cox.de.ufpe.br

where $p > 0, q > 0$, and $\Gamma(\cdot)$ is the gamma function. The mean and variance of y are, respectively,

$$\mathbb{E}(y) = \frac{p}{p+q} \quad \text{and} \quad \mathrm{Var}(y) = \frac{pq}{(p+q)^2(p+q+1)}.$$

The mode of the distribution exists when both p and q are greater than one, in which case, $\mathrm{mode}(y) = (p-1)/(p+q-2)$. The uniform distribution is a special case of (1) when $p = q = 1$. Estimation of p and q can be carried out by maximum likelihood. Small sample bias adjustments to the maximum likelihood estimators of p and q were obtained by Cribari-Neto and Vasconcellos (2002).

Ferrari and Cribari-Neto (2004) proposed a regression model in which the dependent variable is beta distributed. Their parameterization is as follows.[1] Let $\mu = p/(p+q)$ and $\phi = p+q$, i.e., $p = \mu\phi$ and $q = (1-\mu)\phi$; here, $0 < \mu < 1$ and $\phi > 0$. It then follows that the mean and the variance of y are, respectively,

$$\mathbb{E}(y) = \mu \quad \text{and} \quad \mathrm{Var}(y) = \frac{V(\mu)}{1+\phi},$$

where $V(\mu) = \mu(1-\mu)$. Note that ϕ can be interpreted as a precision parameter in the sense that, for a given value of μ, the larger the value of ϕ, the smaller the variance of y.

Using this parameterization, Ferrari and Cribari-Neto (2004) defined a regression model which in many aspects resembles the class of generalized linear models (see, for example, Nelder and Wedderburn 1972, and McCullagh and Nelder 1989). Their model, however, is not a generalized linear model (GLM).

Our chief goal in this paper is to propose a time series model for random variables that assume values in the standard unit interval. The approach is based on the class of beta regression models of Ferrari and Cribari-Neto (2004). Our approach is also similar to those of Benjamin et al. (2003) and Shephard (1995) (see also Li 1994, and Fokianos and Kedem 2004), who have developed dynamic models for random variables in the exponential family. We note that Zeger and Qaqish (1988) proposed the so-called Markov regression models (which extends the class of GLMs) and that Li (1991) developed goodness-of-fit tests for such models. In this paper, we propose the beta autoregressive moving average model (βARMA). It can be used to model and forecast variates that assume values in the standard unit interval, such as rates and proportions. The use of the βARMA model avoids the need to transform the data prior to modeling. Moreover, the distributions of rates and proportions are typically asymmetric and, hence, Gaussian-based inference is not appropriate. The βARMA model naturally accommodates asymmetries and also non-constant dispersion.

The paper unfolds as follows. Section 2 introduces the proposed model, Sect. 3 focuses on parameter estimation, Sect. 4 considers further inference strategies and prediction, and Sect. 5 illustrates the methodology by applying the model to real data. Finally, concluding remarks are given in Sect. 6.

[1] For an alternative formulation of the class of beta regressions, see Vasconcellos and Cribari-Neto (2005).

2 The model

Our goal is to define a dynamic model for beta distributed random variables observed over time. For both regression and time series analysis it is typically more convenient to work with the mean response and also with a precision (or dispersion) parameter. Therefore, we shall employ the beta parameterization given in Ferrari and Cribari-Neto (2004).

We shall assume that the response is continuous and takes values in the standard unit interval (0, 1). We note, however, that the proposed model is also useful in situations where the response is restricted to the interval (a, b), where a and b are known scalars ($a < b$). In this case, one can model $(y - a)/(b - a)$ instead of modeling y directly. We shall also assume that the covariates x_t, $t = 1, \ldots, n$, where $x_t = (x_{t1}, \ldots, x_{tk})'$, are non-random. Here, n denotes the sample size and $k < n$.

Let y_t, $t = 1, \ldots, n$, be random variables and assume that the conditional distribution of each y_t, given the previous information set $\mathcal{F}_{t-1}$ (i.e., the smallest σ-algebra such that the variables $y_1, \ldots, y_{t-1}$ are measurable), follows the beta distribution. That is, the conditional density of y_t given $\mathcal{F}_{t-1}$ is

$$f(y_t \mid \mathcal{F}_{t-1}) = \frac{\Gamma(\phi)}{\Gamma(\mu_t\phi)\Gamma((1-\mu_t)\phi)} y_t^{\mu_t\phi-1}(1-y_t)^{(1-\mu_t)\phi-1}, \quad 0 < y_t < 1, \quad (2)$$

where $\mathbb{E}(y_t \mid \mathcal{F}_{t-1}) = \mu_t$ and $\mathrm{Var}(y_t \mid \mathcal{F}_{t-1}) = V(\mu_t)/(1+\phi)$ are, respectively, the conditional mean and the conditional variance of y_t; here, $V(\mu_t) = \mu_t(1-\mu_t)$.

In the class of beta regression models (see Ferrari and Cribari-Neto 2004), μ_t is related to a linear predictor, η_t, through a twice differentiable strictly monotonic link function $g : (0, 1) \to \mathbb{R}$. The most commonly used link functions are the logit, probit, and complementary log–log links. Unlike the linear predictor of the beta regression model, in the systematic component of the βARMA specification there is an additional component, τ_t, which allows autoregressive and moving average terms to be included additively. Thus, a general model for μ_t is given by

$$g(\mu_t) = \eta_t = x_t'\beta + \tau_t,$$

where $\beta = (\beta_1, \ldots, \beta_k)'$ is a set of unknown linear parameters and τ_t is an ARMA component which shall be described below and is similar to what is given in Benjamin et al. (2003).

We shall now motivate the definition of the ARMA component τ_t. Consider an ARMA(p, q) model initially as function of a term ξ_t, such that $\xi_t = g(y_t) - x_t'\beta$. Then,

$$\xi_t = \alpha + \sum_{i=1}^{p} \varphi_i \xi_{t-i} + \sum_{j=1}^{q} \theta_j r_{t-j} + r_t, \quad (3)$$

where r_t denotes a random error and $\alpha \in \mathbb{R}$ is a constant. Although we have not defined r_t, it is assumed that $\mathbb{E}(r_t \mid \mathcal{F}_{t-1}) = 0$. Taking conditional expectations with respect to the σ-algebra $\mathcal{F}_{t-1}$ in (3) we obtain the approximate model

$$\tau_t = \alpha + \sum_{i=1}^{p} \varphi_i \xi_{t-i} + \sum_{j=1}^{q} \theta_j r_{t-j}.$$

Note that ξ_{t-i} with $i > 0$ is $\mathcal{F}_{t-1}$-measurable, and $\mathbb{E}(\xi_t \mid \mathcal{F}_{t-1}) \approx \tau_t$. Therefore, we obtain the following expression for τ_t:

$$\tau_t = \alpha + \sum_{i=1}^{p} \varphi_i \left\{ g(y_{t-i}) - x'_{t-i}\beta \right\} + \sum_{j=1}^{q} \theta_j r_{t-j},$$

where $x_t \in \mathbb{R}^k$, $\beta = (\beta_1, \ldots, \beta_k)'$, $k < n$, and $p, q \in \mathbb{N}$ are, respectively, the autoregressive and moving average orders. The φ's and the θ's are the autoregressive and moving average parameters, respectively, and r_t is an error. Finally, since $\tau_t = g(\mu_t) - x'_t\beta$, we propose the following general model for the mean μ_t:

$$g(\mu_t) = \alpha + x'_t\beta + \sum_{i=1}^{p} \varphi_i \left\{ g(y_{t-i}) - x'_{t-i}\beta \right\} + \sum_{j=1}^{q} \theta_j r_{t-j}. \tag{4}$$

The βARMA(p, q) model is defined by (2) and (4). It is noteworthy that both the fitted values and the out-of-sample forecasts obtained using the βARMA model will belong to the standard unit interval. There are several choices for the moving average error terms; for example, errors measured on the original scale (i.e., $y_t - \mu_t$), on the predictor scale (i.e., $g(y_t) - \eta_t$), etc. What is required is that the error r_t be measurable with respect to $\mathcal{F}_t$.

Let us obtain the mean and variance of two errors. For $y_t - \mu_t$, we have

$$\mathbb{E}(y_t - \mu_t \mid \mathcal{F}_{t-1}) = 0 \quad \text{and} \quad \operatorname{Var}(y_t - \mu_t \mid \mathcal{F}_{t-1}) = \frac{V(\mu_t)}{1+\phi},$$

where $V(\mu_t) = \mu_t(1 - \mu_t)$. In particular, $\mathbb{E}(y_t - \mu_t) = 0$ and $\operatorname{Var}(y_t - \mu_t) = V(\mu_t)/(1 + \phi)$. Note that the errors are orthogonal, since for $i < j$

$$\mathbb{E}\big((y_i - \mu_i)(y_j - \mu_j)\big) = \mathbb{E}\big((y_i - \mu_i)\mathbb{E}(y_j - \mu_j \mid \mathcal{F}_{j-1})\big) = 0. \tag{5}$$

Since $g(\cdot)$ is continuously differentiable, we can Taylor-expand it as

$$g(y_t) \approx g(\mu_t) + g'(\mu_t)(y_t - \mu_t) \quad \Rightarrow \quad g(y_t) - g(\mu_t) \approx g'(\mu_t)(y_t - \mu_t).$$

Moreover, $\eta_t = g(\mu_t)$, then, for the error $g(y_t) - \eta_t$,

$$\mathbb{E}\big(g(y_t) - \eta_t \mid \mathcal{F}_{t-1}\big) \approx \mathbb{E}\big(g'(\mu_t)(y_t - \mu_t) \mid \mathcal{F}_{t-1}\big) = 0.$$

Given that $g(\cdot)$ is twice differentiable, it follows from the delta method that

$$\operatorname{Var}\big(g(y_t) - \eta_t \mid \mathcal{F}_{t-1}\big) \approx \big(g'(\mu_t)\big)^2 \frac{V(\mu_t)}{1+\phi}.$$

In particular, $\mathbb{E}(g(y_t) - \eta_t) \approx 0$ and $\operatorname{Var}(g(y_t) - \eta_t) \approx (g'(\mu_t))^2 V(\mu_t)/(1+\phi)$. With an analogous argument to the one used with (5), we conclude that these errors are also approximately orthogonal.

3 Parameter estimation

The estimation of the parameters that index the βARMA model can be carried out by maximum likelihood. Let us denote the vector of parameters as $\gamma = (\alpha, \beta', \phi, \varphi', \theta')'$, where $\varphi = (\varphi_1, \ldots, \varphi_p)'$ and $\theta = (\theta_1, \ldots, \theta_q)'$. As noted earlier, we assume that the covariates x_t are non-stochastic.

The log-likelihood function for the parameter vector γ conditional on the first m observations, where $m = \max\{p, q\}$, is $\ell = \sum_{t=m+1}^{n} \log f(y_t \mid \mathcal{F}_{t-1})$, with $f(y_t \mid \mathcal{F}_{t-1})$ given in (2). Expectations are also taken in conditional fashion. Note that, conditioned to $\mathcal{F}_m$, the first m errors are zero (or approximately zero). Thus, in the construction of the conditional log-likelihood function, the first q errors are assumed to equal zero.

3.1 Score vector

Let $\log f(y_t \mid \mathcal{F}_{t-1}) = \ell_t(\mu_t, \phi)$. Then

$$\ell_t(\mu_t, \phi) = \log\Gamma(\phi) - \log\Gamma(\mu_t\phi) - \log\Gamma\big((1-\mu_t)\phi\big) + (\mu_t\phi - 1)\log y_t \\ + \big\{(1-\mu_t)\phi - 1\big\}\log(1-y_t).$$

Therefore, the conditional log-likelihood function is

$$\ell = \sum_{t=m+1}^{n} \ell_t(\mu_t, \phi).$$

Thus,

$$\frac{\partial \ell}{\partial \alpha} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \alpha}.$$

Note that $d\mu_t/d\eta_t = 1/g'(\mu_t)$. We also have that

$$\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} = \phi\left[\log\frac{y_t}{1-y_t} - \big\{\psi(\mu_t\phi) - \psi\big((1-\mu_t)\phi\big)\big\}\right], \tag{6}$$

where $\psi(\cdot)$ is the digamma function, i.e., $\psi(z) = d\log\Gamma(z)/dz$ for $z > 0$. Let $y_t^\star = \log\{y_t/(1-y_t)\}$ and $\mu_t^\star = \psi(\mu_t\phi) - \psi((1-\mu_t)\phi)$.Then,

$$\frac{\partial \ell}{\partial \alpha} = \phi \sum_{t=m+1}^{n} (y_t^\star - \mu_t^\star)\frac{1}{g'(\mu_t)}.$$

Additionally, for $l = 1, \ldots, k$,

$$\frac{\partial \ell}{\partial \beta_l} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \beta_l}.$$

Then,

$$\frac{\partial \ell}{\partial \beta_l} = \phi \sum_{t=m+1}^{n} (y_t^\star - \mu_t^\star) \frac{1}{g'(\mu_t)} \left(x_{tl} - \sum_{i=1}^{p} \varphi_i x_{(t-i)l} \right).$$

Furthermore,

$$\frac{\partial \ell}{\partial \phi} = \sum_{t=m+1}^{n} \left\{ \mu_t (y_t^\star - \mu_t^\star) + \log(1 - y_t) - \psi\big((1-\mu_t)\phi\big) + \psi(\phi) \right\}.$$

Note also that, for $i = 1, \ldots, p$,

$$\frac{\partial \ell}{\partial \varphi_i} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \varphi_i},$$

which yields

$$\frac{\partial \ell}{\partial \varphi_i} = \phi \sum_{t=m+1}^{n} (y_t^\star - \mu_t^\star) \frac{1}{g'(\mu_t)} \big(g(y_{t-i}) - x'_{t-i}\beta \big).$$

Finally, for $j = 1, \ldots, q$,

$$\frac{\partial \ell}{\partial \theta_j} = \sum_{t=m+1}^{n} \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \theta_j}.$$

Therefore,

$$\frac{\partial \ell}{\partial \theta_j} = \phi \sum_{t=m+1}^{n} (y_t^\star - \mu_t^\star) \frac{1}{g'(\mu_t)} r_{t-j}.$$

It is now possible to obtain the score vector $U(\gamma)$. Let $y^\star = (y_{m+1}^\star, \ldots, y_n^\star)'$, $\mu^\star = (\mu_{m+1}^\star, \ldots, \mu_n^\star)'$ and $T = \operatorname{diag}\{1/g'(\mu_{m+1}), \ldots, 1/g'(\mu_n)\}$. Let also $\mathbf{1}$ be an $n \times 1$ vector of ones, M be the $(n-m) \times k$ matrix with (i, j)th element given by $x_{(i+m)j} - \sum_{l=1}^{p} \varphi_l x_{(i+m-l)j}$, P be the $(n-m) \times p$ matrix whose (i, j)th element equals $g(y_{i+m-j}) - x'_{i+m-j}\beta$ and R be the $(n-m) \times q$ matrix with (i, j)th element given by r_{i+m-j}. Hence,

$$\begin{aligned}
U_\alpha(\gamma) &= \phi \mathbf{1}' T (y^\star - \mu^\star), \\
U_\beta(\gamma) &= \phi M' T (y^\star - \mu^\star), \\
U_\phi(\gamma) &= \sum_{t=m+1}^{n} \left\{ \mu_t (y_t^\star - \mu_t^\star) + \log(1 - y_t) - \psi\big((1-\mu_t)\phi\big) + \psi(\phi) \right\}, \\
U_\varphi(\gamma) &= \phi P' T (y^\star - \mu^\star),
\end{aligned}$$

and

$$U_\theta(\gamma) = \phi R' T (y^\star - \mu^\star).$$

Therefore, the score vector is

$$U(\gamma) = \begin{pmatrix} U_\alpha(\gamma) \\ U_\beta(\gamma) \\ U_\phi(\gamma) \\ U_\varphi(\gamma) \\ U_\theta(\gamma) \end{pmatrix},$$

which is of dimension $(k + p + q + 2) \times 1$. The conditional maximum likelihood estimator (CMLE) of γ is obtained as the solution of the system of equations given by $U(\gamma) = 0$. Note that it does not have closed-form. Hence, it has to be numerically obtained by maximizing the conditional log-likelihood function using a nonlinear optimization algorithm, such as a Newton or quasi-Newton algorithm (see Nocedal and Wright 1999).

3.2 Conditional Fisher's information matrix

In what follows λ_i and δ_i will be used as surrogates for β_i, φ_i or θ_i. We have

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \lambda_i \partial \delta_j} &= \sum_{t=m+1}^{n} \frac{\partial}{\partial \mu_t} \left(\frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \delta_j} \right) \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \lambda_i} \\ &= \sum_{t=m+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \delta_j} + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \delta_j} \right) \right] \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \lambda_i}. \end{aligned}$$

Since we are working with the conditional likelihood, we know, from the regularity conditions, that $\mathbb{E}(\partial \ell_t(\mu_t, \phi)/\partial \mu_t \mid \mathcal{F}_{t-1}) = 0$; in particular, we have that $\mathbb{E}(\partial \ell_t(\mu_t, \phi)/\partial \mu_t) = 0$.

We also note that $\partial \eta_t/\partial \beta_l = x_{tl} - \sum_{i=1}^{p} \varphi_i x_{(t-i)l}$, $\partial \eta_t/\partial \varphi_i = g(y_{t-i}) - x'_{t-i}\beta$ and $\partial \eta_t/\partial \theta_j = r_{t-j}$ are $\mathcal{F}_{t-1}$-measurable (since $\mathcal{F}_t$ is a filtration). Thus, it follows from the regularity conditions that

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \delta_j} \middle| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^{n} \mathbb{E}\left(\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \middle| \mathcal{F}_{t-1} \right) \left(\frac{d\mu_t}{d\eta_t} \right)^2 \frac{\partial \eta_t}{\partial \delta_j} \frac{\partial \eta_t}{\partial \lambda_i}.$$

From (6) we obtain

$$\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} = -\phi \big\{ \psi'(\mu_t \phi) + \psi'\big((1 - \mu_t)\phi\big) \big\}.$$

Furthermore,

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \delta_j} \middle| \mathcal{F}_{t-1} \right) = -\phi \sum_{t=m+1}^{n} \frac{\{\psi'(\mu_t \phi) + \psi'((1 - \mu_t)\phi)\}}{g'(\mu)^2} \frac{\partial \eta_t}{\partial \delta_j} \frac{\partial \eta_t}{\partial \lambda_i}.$$

Note that

$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \alpha} = \sum_{t=m+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \lambda_i} + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \lambda_i} \right) \right] \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \alpha}.$$

Hence,

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \alpha} \middle| \mathcal{F}_{t-1} \right) = -\phi \sum_{t=m+1}^{n} \frac{\{\psi'(\mu_t \phi) + \psi'((1-\mu_t)\phi)\}}{g'(\mu)^2} \frac{\partial \eta_t}{\partial \lambda_i}.$$

Moreover,

$$\frac{\partial^2 \ell}{\partial \alpha^2} = \sum_{t=m+1}^{n} \left[\frac{\partial^2 \ell_t(\mu_t, \phi)}{\partial \mu_t^2} \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \alpha} + \frac{\partial \ell_t(\mu_t, \phi)}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left(\frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \alpha} \right) \right] \frac{d\mu_t}{d\eta_t} \frac{\partial \eta_t}{\partial \alpha}.$$

Thus,

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \alpha^2} \middle| \mathcal{F}_{t-1} \right) = -\phi \sum_{t=m+1}^{n} \frac{\{\psi'(\mu_t \phi) + \psi'((1-\mu_t)\phi)\}}{g'(\mu)^2}.$$

We have that

$$\frac{\partial \ell}{\partial \lambda_j} = \phi \sum_{t=m+1}^{n} (y_t^\star - \mu_t^\star) \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \lambda_j}.$$

Therefore,

$$\frac{\partial^2 \ell}{\partial \lambda_i \partial \phi} = \sum_{t=m+1}^{n} \left[(y_t^\star - \mu_t^\star) - \phi \frac{\partial \mu_t^\star}{\partial \phi} \right] \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \lambda_i}.$$

It also follows from the regularity conditions that $\mathbb{E}(y_t^\star \mid \mathcal{F}_{t-1}) = \mu_t^\star$. Given that $\partial \mu_t^\star / \partial \phi = \psi'(\mu_t \phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)$, we have

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \lambda_i \partial \phi} \middle| \mathcal{F}_{t-1} \right) = -\phi \sum_{t=m+1}^{n} \frac{\psi'(\mu_t \phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)}{g'(\mu_t)} \frac{d\eta_t}{d\lambda_i}.$$

We also have that

$$\frac{\partial^2 \ell}{\partial \alpha \partial \phi} = \sum_{t=m+1}^{n} \left[(y_t^\star - \mu_t^\star) - \phi \frac{\partial \mu_t^\star}{\partial \phi} \right] \frac{1}{g'(\mu_t)} \frac{\partial \eta_t}{\partial \alpha},$$

which yields

$$\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \alpha \partial \phi} \middle| \mathcal{F}_{t-1} \right) = -\phi \sum_{t=m+1}^{n} \frac{\psi'(\mu_t \phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)}{g'(\mu_t)}.$$

Finally, $\partial^2\ell/\partial\phi^2$ follows from the differentiation of $U_\phi(\gamma)$ with respect to ϕ. We obtain

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\phi^2}\middle|\mathcal{F}_{t-1}\right) = -\sum_{t=m+1}^{n}\left(\psi'(\mu_t\phi)\mu_t^2 + \psi'\big((1-\mu_t)\phi\big)(1-\mu_t)^2 - \psi'(\phi)\right).$$

Using

$$\frac{d\eta_t}{d\beta_l} = x_{tl} - \sum_{i=1}^{p}\varphi_i x_{(t-i)l}, \qquad \frac{d\eta_t}{d\varphi_i} = g(y_{t-i}) - x_{t-i}'\beta, \quad \text{and} \quad \frac{d\eta_t}{d\theta_j} = r_{t-j},$$

we can obtain Fisher's information matrix for γ. Let $W = \operatorname{diag}\{w_{m+1}, \dots, w_n\}$, with

$$w_t = \phi\frac{\{\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)\}}{g'(\mu_t)^2},$$

$c = (c_{m+1}, \dots, c_n)'$, with $c_t = \phi\{\psi'(\mu_t\phi)\mu_t - \psi'((1-\mu_t)\phi)(1-\mu_t)\}$, and $D = \operatorname{diag}\{d_{m+1}, \dots, d_n\}$, with $d_t = \psi'(\mu_t\phi)\mu_t^2 + \psi'((1-\mu_t)\phi)(1-\mu_t)^2 - \psi'(\phi)$. Thus,

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\alpha^2}\middle|\mathcal{F}_{t-1}\right) = -\phi\operatorname{tr}(W), \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\beta\partial\alpha}\middle|\mathcal{F}_{t-1}\right) = -\phi M'W\mathbf{1},$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\alpha\partial\phi}\middle|\mathcal{F}_{t-1}\right) = -\mathbf{1}'Tc, \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\varphi\partial\alpha}\middle|\mathcal{F}_{t-1}\right) = -\phi P'W\mathbf{1},$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\beta\partial\beta'}\middle|\mathcal{F}_{t-1}\right) = -\phi M'WM, \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\beta\partial\phi}\middle|\mathcal{F}_{t-1}\right) = -M'Tc,$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\phi^2}\middle|\mathcal{F}_{t-1}\right) = -\operatorname{tr}(D), \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\varphi\partial\varphi'}\middle|\mathcal{F}_{t-1}\right) = -\phi P'WP,$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\varphi\partial\phi}\middle|\mathcal{F}_{t-1}\right) = -P'Tc, \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\theta\partial\theta'}\middle|\mathcal{F}_{t-1}\right) = -\phi M'WM,$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\theta\partial\phi}\middle|\mathcal{F}_{t-1}\right) = -R'Tc, \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\beta\partial\varphi'}\middle|\mathcal{F}_{t-1}\right) = -\phi M'WP,$$

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\beta\partial\theta'}\middle|\mathcal{F}_{t-1}\right) = -\phi M'WR, \qquad \mathbb{E}\left(\frac{\partial^2\ell}{\partial\varphi\partial\theta'}\middle|\mathcal{F}_{t-1}\right) = -\phi P'WR,$$

and

$$\mathbb{E}\left(\frac{\partial^2\ell}{\partial\theta\partial\alpha}\middle|\mathcal{F}_{t-1}\right) = -\phi R'W\mathbf{1}.$$

Therefore, Fisher's information matrix can be expressed as

$$K = K(\gamma) = \begin{bmatrix} K_{\alpha\alpha} & K_{\alpha\beta} & K_{\alpha\phi} & K_{\alpha\varphi} & K_{\alpha\theta} \\ K_{\beta\alpha} & K_{\beta\beta} & K_{\beta\phi} & K_{\beta\varphi} & K_{\beta\theta} \\ K_{\phi\alpha} & K_{\phi\beta} & K_{\phi\phi} & K_{\phi\varphi} & K_{\phi\theta} \\ K_{\varphi\alpha} & K_{\varphi\beta} & K_{\varphi\phi} & K_{\varphi\varphi} & K_{\varphi\theta} \\ K_{\theta\alpha} & K_{\theta\beta} & K_{\theta\phi} & K_{\theta\varphi} & K_{\theta\theta} \end{bmatrix},$$

where $K_{\alpha\alpha} = \phi\mathrm{tr}(W)$, $K_{\beta\alpha} = K'_{\alpha\beta} = \phi M'W\mathbf{1}$, $K_{\alpha\phi} = K_{\phi\alpha} = \mathbf{1}'Tc$, $K_{\varphi\alpha} = K'_{\alpha\varphi} = \phi P'W\mathbf{1}$, $K_{\theta\alpha} = K'_{\alpha\theta} = \phi R'W\mathbf{1}$, $K_{\beta\beta} = \phi M'WM$, $K_{\beta\phi} = K'_{\phi\beta} = M'Tc$, $K_{\phi\phi} = \mathrm{tr}(D)$, $K_{\beta\varphi} = K'_{\varphi\beta} = \phi M'WP$, $K_{\beta\theta} = K'_{\theta\beta} = \phi R'WR$, $K_{\varphi\varphi} = \phi P'WP$, $K_{\varphi\phi} = K'_{\phi\varphi} = P'Tc$, $K_{\theta\theta} = \phi R'WR$, $K_{\theta\phi} = K'_{\phi\theta} = R'Tc$, and $K_{\varphi\theta} = K'_{\theta\varphi} = \phi R'MP$.

Note that Fisher's information matrix is not block-diagonal, which implies that our model is not a dynamic GLM. Under the usual regularity conditions for maximum likelihood estimation and when the sample size is large,

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \\ \widehat{\phi} \\ \widehat{\varphi} \\ \widehat{\theta} \end{pmatrix} \sim \mathcal{N}_{(k+p+q+2)} \left(\begin{pmatrix} \alpha \\ \beta \\ \phi \\ \varphi \\ \theta \end{pmatrix}, K^{-1} \right)$$

approximately, where $\mathcal{N}_r$ denotes the r-dimensional normal distribution, and $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\phi}$, $\widehat{\varphi}$, and $\widehat{\theta}$ are the CMLEs of α, β, ϕ, φ, and θ, respectively.

4 Hypothesis testing and prediction

Consider the following null and alternative hypotheses:

$$\mathcal{H}_0 : t\gamma = 0 \quad \text{and} \quad \mathcal{H}_1 : t\gamma \neq 0, \tag{7}$$

where t is an $r \times (k+p+q+2)$ matrix ($r < k+p+q+2$) of rank r. For instance, consider the following partition of the $(k+p+q+2) \times 1$ parameter vector: $\gamma = (\gamma_1', \gamma_2')'$, where γ_2 is $r \times 1$ ($r < k+p+q+2$). Note that by letting $t\gamma = \gamma_2$ in (7) one can test whether γ_2 equals zero.

Let $\widetilde{\gamma}$ be the CMLE of γ under the null hypothesis in (7) and let $\widehat{\gamma}$ be the unrestricted CMLE of γ. The test statistic commonly used to test $\mathcal{H}_0 : t\gamma = 0$ is the conditional log-likelihood ratio statistic (CLR):

$$\lambda_n = 2\{\ell(\widehat{\gamma}) - \ell(\widetilde{\gamma})\},$$

where $\ell(\cdot)$ is the conditional log-likelihood function. Under mild regularity conditions and under $\mathcal{H}_0$, $\lambda_n \xrightarrow{\mathcal{D}} \chi^2_r$, where $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, so that the test can be performed using approximate critical values from the limiting χ^2_r null distribution. One can also base the testing inference on the square root of the CLR statistic, where the sign of the statistic is that of $(\widehat{\gamma} - \widetilde{\gamma})$, which is asymptotically standard normal under the null hypothesis. It is also noteworthy that by using the asymptotic normality of the CMLE of γ, $\widehat{\gamma}$, one can easily construct approximate confidence intervals for the elements of γ.

We shall now consider tests of model adequacy. Li (1991) proposed Portmanteau and score statistics for Markov regression models. We shall now follow his approach to provide Portmanteau and score statistics for the βARMA model. At the outset, consider the standardized score errors defined as

$$a_t = \frac{y_t^\star - \mu_t^\star}{\sqrt{\psi'(\mu_t\phi) + \psi'((1-\mu_t)\phi)}},$$

and note that it follows from the regularity conditions that $\mathbb{E}(a_t|\mathcal{F}_{t-1}) = 0$ and $\mathrm{Var}(a_t|\mathcal{F}_{t-1}) = 1/\phi$; also, $\mathbb{E}(a_i a_j) = 0$ whenever $i \neq j$. Then, the lag k innovation autocorrelation of a_t is

$$C_k = \frac{1}{n}\sum_{t=k+1}^{n} \phi a_t a_{t-k}.$$

The corresponding kth residual autocorrelation, $\hat{C}_k$, can be written as

$$\hat{C}_k = \frac{1}{n}\sum_{t=k+1}^{n} \hat{\phi}\hat{a}_t\hat{a}_{t-k}.$$

We shall work with the subsets of the score vector and Fisher's information matrix relative to β, ϕ and θ. The following quantity will be useful:

$$V = \lim_{n\to\infty}\frac{1}{n}K,$$

where K denotes Fisher's information matrix. Consider the following partition of V^{-1}:

$$V^{-1} = \lim_{n\to\infty}\frac{1}{n}\begin{bmatrix} K^{\alpha\alpha} & K^{\alpha\beta} & K^{\alpha\phi} & K^{\alpha\varphi} & K^{\alpha\theta} \\ K^{\beta\alpha} & K^{\beta\beta} & K^{\beta\phi} & K^{\beta\varphi} & K^{\beta\theta} \\ K^{\phi\alpha} & K^{\phi\beta} & K^{\phi\phi} & K^{\phi\varphi} & K^{\phi\theta} \\ K^{\varphi\alpha} & K^{\varphi\beta} & K^{\varphi\phi} & K^{\varphi\varphi} & K^{\varphi\theta} \\ K^{\theta\alpha} & K^{\theta\beta} & K^{\theta\phi} & K^{\theta\varphi} & K^{\theta\theta} \end{bmatrix},$$

and let $V^{\alpha\beta\varphi\theta}$ be the block of V^{-1} which corresponds to α, β, φ, and θ:

$$V^{\alpha\beta\varphi\theta} = \lim_{n\to\infty}\frac{1}{n}\begin{bmatrix} K^{\alpha\alpha} & K^{\alpha\beta} & K^{\alpha\varphi} & K^{\alpha\theta} \\ K^{\beta\alpha} & K^{\beta\beta} & K^{\beta\varphi} & K^{\beta\theta} \\ K^{\varphi\alpha} & K^{\varphi\beta} & K^{\varphi\varphi} & K^{\varphi\theta} \\ K^{\theta\alpha} & K^{\theta\beta} & K^{\theta\varphi} & K^{\theta\theta} \end{bmatrix}.$$

Let $\hat{C} = (\hat{C}_1, \ldots, \hat{C}_m)'$ for some $m > 0$. Then, following Sect. 2 and the Appendix of Li (1991), it can be shown that, under correct model specification, $\sqrt{n}\hat{C}$ is asymptotically normally distributed with mean zero and variance $I_m - \phi X' V^{\alpha\beta\varphi\theta} X$, where I_m is the $m \times m$ identity matrix and

$$X = \lim_{n\to\infty} \frac{1}{n} \begin{bmatrix} \sum h_t a_{t-1} & \cdots & \sum h_t a_{t-m} \\ \sum (x_{t1} - \sum_{i=1}^{p} \varphi_i x_{(t-i)1}) h_t a_{t-1} & \cdots & \sum (x_{t1} - \sum_{i=1}^{p} \varphi_i x_{(t-i)1}) h_t a_{t-m} \\ \vdots & \ddots & \vdots \\ \sum (x_{tk} - \sum_{i=1}^{p} \varphi_i x_{(t-i)k}) h_t a_{t-1} & \cdots & \sum (x_{tk} - \sum_{i=1}^{p} \varphi_i x_{(t-i)k}) h_t a_{t-m} \\ \sum (g(y_{t-1}) - x_{t-1}'\beta) h_t a_{t-1} & \cdots & \sum (g(y_{t-1}) - x_{t-1}'\beta) h_t a_{t-m} \\ \vdots & \ddots & \vdots \\ \sum (g(y_{t-p}) - x_{t-p}'\beta) h_t a_{t-1} & \cdots & \sum (g(y_{t-p}) - x_{t-p}'\beta) h_t a_{t-m} \\ \sum r_{t-1} h_t a_{t-1} & \cdots & \sum r_{t-1} h_t a_{t-m} \\ \vdots & \ddots & \vdots \\ \sum r_{t-q} h_t a_{t-1} & \cdots & \sum r_{t-q} h_t a_{t-m} \end{bmatrix},$$

with

$$h_t = \frac{(\psi'(\mu_i\phi) + \psi'((1-\mu_i)\phi))^{1/2}}{g'(\mu_t)}.$$

Hence, a test for the joint significance of the first m autocorrelations can be based on $n\hat{C}'(I_m - \hat{\phi}\hat{X}'\hat{V}^{\alpha\beta\varphi\theta}\hat{X})^{-1}\hat{C}$, which is asymptotically χ^2_m under the null hypothesis of no serial correlation.

Score tests on the parameter vector can be performed using the approach proposed by Li (1991). Let

$$\beta = (\beta_1', \beta_2')', \qquad \varphi = (\varphi_1', \varphi_2')', \qquad \theta = (\theta_1', \theta_2')', \quad \text{and} \quad \tau = (\beta_2', \varphi_2', \theta_2')'.$$

The null hypothesis is $\tau = 0$ which is to be tested against the alternative that the number of parameters is $1 + k_1 + k_2 + p_1 + p_2 + q_1 + q_2$, where, k_i, p_i, and q_i are the number of parameters in β_i, φ_i, and θ_i, respectively, for $i = 1, 2$.[2] The corresponding score function is

$$\tilde{U}(\gamma) = \begin{pmatrix} U_\alpha(\gamma) \\ U_\beta(\gamma) \\ U_\varphi(\gamma) \\ U_\theta(\gamma) \end{pmatrix},$$

where $\gamma = (\alpha, \beta, \phi, \varphi, \theta)$. It is possible to show that $\tilde{U}/\sqrt{n}$ is asymptotically normally distributed with mean zero and variance $V_{\alpha\beta\varphi\theta}$ when the null hypothesis

[2]Note that we do not include α in τ; one can, however, consider the case where $\tau = (\alpha, \beta_2', \varphi_2', \theta_2')'$ when the null hypothesis also imposes $\alpha = 0$.

is true, where $V_{\alpha\beta\varphi\theta}$ is the part of V that corresponds to α, β, φ, and θ. Then, following Li (1991), and noting that under the null hypothesis $\tilde{U}(\gamma_1) = 0$, where $\gamma_1 = (\alpha, \beta_1', \phi, \varphi_1', \theta_1')'$, a score test statistic is

$$Q = n^{-1}\tilde{U}(\hat{\gamma})'\hat{V}^{\alpha\beta\varphi\theta}\tilde{U}(\hat{\gamma}),$$

where the estimates are obtained under the null hypothesis. Let $\hat{a} = (\hat{a}_1, \ldots, \hat{a}_n)$, $S_t = h_t Z_t$, $S = (S_1, \ldots, S_n)'$ and

$$Z_t = \begin{pmatrix} 1 \\ (x_{t1} - \sum_{i=1}^{p} \varphi_i x_{(t-i)1}) \\ \vdots \\ (x_{tk} - \sum_{i=1}^{p} \varphi_i x_{(t-i)k}) \\ (g(y_{t-1}) - x_{t-1}'\beta) \\ \vdots \\ (g(y_{t-p}) - x_{t-p}'\beta) \\ r_{t-1} \\ \vdots \\ r_{t-q} \end{pmatrix}.$$

We can rewrite Q as

$$Q = \frac{\hat{\phi}^2\hat{a}'\hat{S}\hat{V}^{\alpha\beta\varphi\theta}\hat{S}'\hat{a}}{n}.$$

An asymptotically equivalent statistic is

$$\tilde{Q} = \hat{\phi}^2\hat{a}'\hat{S}K^{\alpha\beta\varphi\theta}\hat{S}'\hat{a},$$

where $K^{\alpha\beta\varphi\theta}$ is the block of Fisher's information matrix inverse which corresponds to α, β, φ, and θ:

$$K^{\alpha\beta\varphi\theta} = \begin{bmatrix} K^{\alpha\alpha} & K^{\alpha\beta} & K^{\alpha\varphi} & K^{\alpha\theta} \\ K^{\beta\alpha} & K^{\beta\beta} & K^{\beta\varphi} & K^{\beta\theta} \\ K^{\varphi\alpha} & K^{\varphi\beta} & K^{\varphi\varphi} & K^{\varphi\theta} \\ K^{\theta\alpha} & K^{\theta\beta} & K^{\theta\varphi} & K^{\theta\theta} \end{bmatrix}.$$

Under the null hypothesis, Q is asymptotically $\chi^2_{k_2+p_2+q_2}$. We note that the conditional score test that we just developed only requires the estimation of the null model.

In order to produce forecasts, the CMLE of γ, $\widehat{\gamma}$, must be used to obtain estimates for μ_t, $t = m+1, \ldots, n$, say $\widehat{\mu}_t$. By using $\widehat{\mu}_t$ one can obtain the estimates of r_t, $\widehat{r}_t$, for $t = m+1, \ldots, n$ (based on the functional structure of the error). For $N > n$, the forecast of the error r_N equals zero. Thus, to predict the mean value of the process at $T > n$, one should use the CMLE of γ, $\widehat{\gamma}$; the estimates of μ_t, $t = m+1, \ldots, n$; the estimates of r_t, $t = m+1, \ldots, n$; replace r_t by zero if $t > n$ (these suffice to obtain

$\widehat{\mu}_{n+1}$, and one can then proceed analogously to obtain $\widehat{\mu}_{n+2}$, and so on); and replace y_t by $\widehat{\mu}_t$ if $n < t < T$. For instance, the mean response estimate at $n+1$ is

$$\widehat{\mu}_{n+1} = g^{-1}\left(\widehat{\alpha} + x'_{n+1}\widehat{\beta} + \sum_{i=1}^{p}\widehat{\varphi}_i\left\{g(y_{n+1-i}) - x'_{n+1-i}\widehat{\beta}\right\} + \sum_{j=1}^{q}\widehat{\theta}_j\widehat{r}_{n+1-j}\right).$$

At time $n+2$, we obtain

$$\widehat{\mu}_{n+2} = g^{-1}\left(\widehat{\alpha} + x'_{n+2}\widehat{\beta} + \sum_{i=2}^{p}\widehat{\varphi}_i\left\{g(y_{n+2-i}) - x'_{n+2-i}\widehat{\beta}\right\}\right.$$
$$\left. + \widehat{\varphi}_1\left\{g(\widehat{\mu}_{n+1}) - x'_{n+1}\widehat{\beta}\right\} + \sum_{j=2}^{q}\widehat{\theta}_j\widehat{r}_{n+2-j}\right),$$

and so on.

Finally, we note that model selection can be performed using the Akaike information criterion (AIC) introduced by Akaike (1973, 1974) or, alternatively, the Bayesian information criterion (BIC) of Schwarz (1978). For a detailed discussion of information criteria and their properties, see Choi (1992).

5 An application

This section contains an application of the βARMA model proposed in Sect. 2. The estimations and computations were carried out using the free statistical software R; see http://www.r-project.org. We used the quasi-Newton algorithm known as BFGS to maximize the conditional log-likelihood function. The data refers to the rate of hidden unemployment due to substandard work conditions in São Paulo, Brazil (TDOP-RMSP). Hidden unemployment due to substandard work conditions relates to people who work illegally, who perform unpaid work for relatives, and also who have been seeking employment for the past 12 months. The data were obtained from the database of the Applied Economic Research Institute (IPEA) from the Brazilian Federal Government[3] and covers a period of 179 months (January 1991 through November 2005). The maximum and minimum values are 0.057 and 0.024, respectively, and the average unemployment rate equals 0.044. A time series plot of the data is given in Fig. 1.

We shall consider four βAR models ($p = 1, \ldots, 4$); see Table 1. The link function is logit and model selection is carried out using the AIC (Akaike information criterion) and the BIC (Bayesian information criterion):

$$\mathrm{AIC} = -2\widehat{\ell} + 2p \quad \text{and} \quad \mathrm{BIC} = -2\widehat{\ell} + p\log(n),$$

where $\widehat{\ell}$ denotes the log-likelihood function evaluated at the maximum likelihood estimates, p is the number of autoregressive parameters, and n is the sample size.

[3]See http://www.ipeadata.gov.br or obtain directly from http://beta.arma.googlepages.com/beta-arma-data.txt.

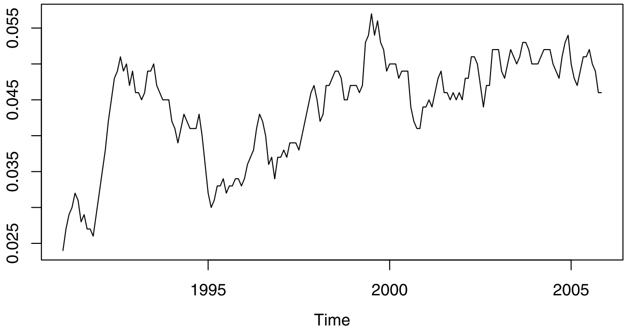

Fig. 1 Rate of hidden unemployment in São Paulo, Brazil

Table 1 βAR models

Model 1	$\eta_t = \alpha + \varphi_1 g(y_{t-1})$
Model 2	$\eta_t = \alpha + \varphi_1 g(y_{t-1}) + \varphi_2 g(y_{t-2})$
Model 3	$\eta_t = \alpha + \varphi_1 g(y_{t-1}) + \varphi_2 g(y_{t-2}) + \varphi_3 g(y_{t-3})$
Model 4	$\eta_t = \alpha + \varphi_1 g(y_{t-1}) + \varphi_2 g(y_{t-2}) + \varphi_3 g(y_{t-3}) + \varphi_4 g(y_{t-4})$

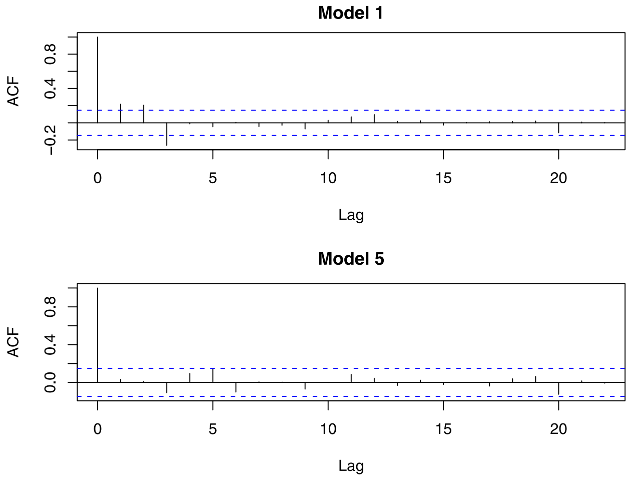

Fig. 2 Sample autocorrelation functions of the standardized residuals obtained from Models 1 and 5

The AIC selects Model 4 whereas the BIC picks Model 1. We note that the coefficient of $g(y_{t-2})$ (Model 4) is not statistically significant at the usual significance levels, since the corresponding p-value equals 0.898. We thus consider a new model, namely, the model with AR terms $g(y_{t-1})$, $g(y_{t-3})$, and $g(y_{t-4})$ (Model 5). Figure 2 shows the residual correlograms corresponding to Models 1 and 5. It is clear that the residuals from Model 1 are serially correlated, unlike the residuals obtained using Model 5. We thus select Model 5 as the best model.

The estimated model is

$$\widehat{\mu}_t = \frac{\exp\{\widehat{\alpha} + \widehat{\varphi}_1 g(y_{t-1}) + \widehat{\varphi}_3 g(y_{t-3}) + \widehat{\varphi}_4 g(y_{t-4})\}}{1 + \exp\{\widehat{\alpha} + \widehat{\varphi}_1 g(y_{t-1}) + \widehat{\varphi}_3 g(y_{t-3}) + \widehat{\varphi}_4 g(y_{t-4})\}},$$

where $(\widehat{\alpha}, \widehat{\varphi}_1, \widehat{\varphi}_3, \widehat{\varphi}_4) = (-0.16726, 1.18317, -0.57566, 0.33718)$, and the respective asymptotic standard errors obtained from the inverse of Fisher's information matrix are (0.0611, 0.0479, 0.0918, 0.0692).

As a final step in the analysis, we turn to forecasting. We remove the final six observations from the series, fit the model (Model 5) and produce six out-of-sample forecasts. The observed values are 0.051, 0.052, 0.050, 0.049, 0.046, and 0.046, and the corresponding forecasts are 0.052, 0.052, 0.051, 0.050, 0.050, and 0.049. The βAR forecasts are, overall, quite accurate.

6 Concluding remarks

In this paper we proposed a dynamic beta regression model: the βARMA model. It can be used to model random variates that are continuous, assume values in the standard unit interval (0, 1) and are observed over time. The proposed model is particularly useful for the time series modeling of rates and proportions. The model is built upon the assumption that the conditional distribution of the variable of interest given its past behavior is beta. As is well known, the beta distribution is very flexible for modeling data that are restricted to the standard unit interval, since the beta density can display quite different shapes depending on the values of the parameters that index the distribution. Parameter estimation is performed by maximum likelihood, and we derived closed-form expressions for the score function and Fisher's information matrix. Hypothesis testing inference can be carried out using standard asymptotic tests. The proposed βARMA yields fitted values and out-of-sample forecasts which belong to the standard unit interval, unlike the standard ARMA model fitted to rates and proportion time series data.

Acknowledgements We thank two referees for their comments and suggestions. We also gratefully acknowledge partial financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Petroc BN, Kaski F (eds) Second international symposium in information theory. Akademiai Kiado, Budapest, pp 267–281

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Automat Control AC-19:716–723

Benjamin MA, Rigby RA, Stasinopoulos M (2003) Generalized autoregressive moving average models. J Am Stat Assoc 98:214–223

Choi B (1992) ARMA model identification. Springer, New York

Cribari-Neto F, Vasconcellos KLP (2002) Nearly unbiased maximum likelihood estimation for the beta distribution. J Stat Comput Simul 72:107–118

Ferrari SLP, Cribari-Neto F (2004) Beta regression for modelling rates and proportions. J Appl Stat 31:799–815

Fokianos K, Kedem B (2004) Partial likelihood for time series following generalized linear models. J Time Ser Anal 25:173–197

Li WK (1991) Testing model adequacy for some Markov regression models for time series. Biometrika 78:83–89

Li WK (1994) Time series models based on generalized linear models: some further results. Biometrics 50:506–511

McCullagh P, Nelder JA (1989) Generalized linear models, 2nd edn. Chapman and Hall, London

Nelder JA, Wedderburn RWM (1972) Generalized linear models. J R Stat Soc A 135:370–384

Nocedal J, Wright SJ (1999) Numerical optimization. Springer, New York

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464

Shephard N (1995). Generalized linear autoregressions. Technical report, Nuffield College, Oxford University. Manuscript available at http://www.nu.ox.ac.uk/economics/papers/1996/w8/glar.ps

Vasconcellos KLP, Cribari-Neto F (2005) Improved maximum likelihood estimation in a new class of beta regression models. Braz J Probab Stat 19:13–31

Zeger SL, Qaqish B (1988) Markov regression models for time series: a quasi-likelihood approach. Biometrics 44:1019–1031