Bayesian Analysis (2012) **7**, Number 2, pp. 385–410

Bayesian Model Selection for Beta Autoregressive Processes

Roberto Casarin*, Luciana Dalla Valle† and Fabrizio Leisen‡

Abstract. We deal with Bayesian model selection for beta autoregressive processes. We discuss the choice of parameter and model priors with possible parameter restrictions and suggest a Reversible Jump Markov-Chain Monte Carlo (RJMCMC) procedure based on a Metropolis-Hastings within Gibbs algorithm.

Keywords: Bayesian Inference, Beta Autoregressive Processes, Reversible Jump MCMC

1 Introduction

The analysis of time series data defined on a bounded interval (such as rates or proportions) has been a challenging issue for many years and still represents an open issue. For modelling data defined on a bounded interval there are at least two alternative approaches. Historically the main approach applies a transform to the data in order to map the interval to the real line and then uses standard time series models. Typical examples of transformations are the additive log-ratio transformation and the Box-Cox transformation (see Aitchinson (1986)). One of the earlier and relevant contributions to this framework is Wallis (1987).

In this paper, we follow the second approach, which is based on a direct modelling on the original sample space. Among the first contributions along this line we refer to Grunwald et al. (1993), who suggest a multivariate state space model for time series data defined on the standard simplex. Another seminal contribution is McKenzie (1985), who introduces a new beta autoregressive process for time series defined on the standard unit interval $(0, 1)$. In recent years, Ferrari and Cribari-Neto (2004) have introduced a beta regression model, showing that it is more convenient to consider the data in the original sample space instead of using a transformation. Rocha and Cribari-Neto (2009) extend the beta regression model and propose a beta autoregressive moving average process that possibly includes exogenous variables in the dynamics. Beta autoregressive (BAR) processes have also recently been used in Amisano and Casarin (2007) for stochastic correlation and in Taddy (2010) for spatial Poisson processes. Both papers describe sequential Monte Carlo methods for first-order BAR processes. Billio and Casarin (2010, 2011) employ beta processes for modelling the transition probabilities in Markov-switching processes and propose a Bayesian approach for inference on first-order

*University Ca' Foscari and School for Advanced Studies, Venice, Italy, r.casarin@unive.it
†School of Computing and Mathematics, Plymouth University, Plymouth, UK, luciana.dallavalle@plymouth.ac.uk
‡Departamento de Estadstica, Universidad Carlos III de Madrid, Getafe, Spain, fabrizio.leisen@uc3m.es

 DOI:10.1214/12-BA713

BAR.

The main contribution of this paper is to propose a Bayesian method for the estimation of the number of lags in beta autoregressive models of general order. In the proposed framework, model uncertainty can be handled by indexing all the models, including the model order in the parameter vector, and calculating posterior model probabilities. In the literature, it has been recognised that there are many difficulties with the computation of the marginals (or the predictives) involved in posterior model probabilities. When the number of competing models is large, the computation of the marginals involves numerical integration over parameter spaces with different dimensions. In this context, the use of standard Markov Chain Monte Carlo (MCMC) algorithms on the different spaces is computationally expensive, since it is not possible to use the MCMC output of a given dimension for computing the integrals in different dimensions. This calls for the use of tailor-made numerical approximations. For a brief discussion on the alternative approaches based on Monte Carlo and MCMC approximations, we refer the reader to Robert (2007). In this paper we consider a natural extension of the MCMC algorithms to the variable dimension case called Reversible Jump MCMC (RJMCMC) introduced in Green (1995) (see also Fan and Sisson (2011) for an updated review). One of the drawbacks that is often encountered by RJMCMC is that the acceptance probability of moves between different parameter spaces can be very low. We circumvent this problem by following Ehlers and Brooks (2008) and propose a very efficient RJMCMC for the simultaneous estimation of both the parameters and the model order of the autoregressive process.

This paper extends the RJMCMC strategy for Gaussian models given in Brooks et al. (2003) and Ehlers and Brooks (2008) to the non-Gaussian case. The Gaussian assumption allows Ehlers and Brooks (2008) to have a closed form solution of the efficient proposal calibration problem, whereas in the case of the beta processes this result is no longer valid. In order to overcome this difficulty, we employ a posterior mode approximation to calibrate the proposal in the RJMCMC algorithm. To the best of our knowledge, only a few studies exist on the application of RJMCMC algorithms to non-Gaussian autoregressive models. Among them, we refer the interested reader to Enciso-Mora et al. (2009), who proposed a model selection algorithm for integer valued ARMA.

The outline of the paper is as follows. In Section 2 the beta autoregressive process of order k (BAR(k)) is introduced with a suitable parametrisation. In Section 3 the Bayesian inference is developed for the BAR(k) model under the choice of some priors and the RJMCMC algorithm is illustrated. In Section 4 some simulation results are shown, and in Section 5 the model selection procedure for beta processes is applied to unemployment and capacity utilisation data. Concluding remarks are given in Section 6.

2 The beta AR(k) model

Let us define a beta autoregressive process $\{x_t\}_{t\geq 0}$ of the order k as follows

$$x_t|\mathcal{F}_{t-1}^X \sim \mathcal{B}e(\eta_{1t}(k), \eta_{2t}(k)) \tag{1}$$

where the $\mathcal{F}_t^X = \sigma(\{x_s\}_{s\leq t})$ is the σ-algebra generated by the process, $\mathcal{B}e(\eta_{1t}(k), \eta_{2t}(k))$ denotes the type I beta distribution, and $\eta_{1t}(k) > 0$, $\eta_{2t}(k) > 0$ are the two parameters of the distribution, usually referred to as shape parameters. The two parameters are $\mathcal{F}_{t-1}$-measurable functions of the k last values of the process. The beta process has the following transition density

$$f(x_t|x_{t-1}, \ldots, x_{t-k}) = \frac{1}{B(\eta_{1t}(k), \eta_{2t}(k))} x_t^{\eta_{1t}(k)-1}(1-x_t)^{\eta_{2t}(k)-1}\mathbb{I}_{(0,1)}(x_t) \tag{2}$$

where $\mathbb{I}_A(x)$ is the indicator function, $B(a, b)$ is the beta function with $a, b > 0$, k is the order of the process and $\eta_{1t}(k)$ and $\eta_{2t}(k)$ will be defined later. In the following we will denote with BAR(k) the k-order beta autoregressive process and assume $k \leq k_{\max}$, with $k_{\max} < \infty$ the maximum order of the process.

We consider the beta distribution of the first type (see Kotz and van Dorp (2004) for a review on beta distributions) and the parametrisation suggested in Robert and Rousseau (2002), Ferrari and Cribari-Neto (2004), and Rocha and Cribari-Neto (2009). In such parametrisation, the conditional distribution of the process at time t is

$$x_t|\mathcal{F}_{t-1} \sim \mathcal{B}e(\eta_t\phi, (1-\eta_t)\phi) \tag{3}$$

where ϕ and η_t represent the precision and location parameters respectively (see Rocha and Cribari-Neto (2009)).

We define $\eta_t = \varphi(\boldsymbol{\alpha}'\mathbf{z}_t)$, where $\varphi(v)$, with $v \in \mathbb{R}$, is a twice differentiable strictly monotonic link function $\varphi : (0, 1) \to \mathbb{R}$, $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_k)'$, $\mathbf{x}_{s:t} = (x_s, \ldots, x_t)'$, with $s < t$ and $\mathbf{z}_t = (1, \varphi^{-1}(x_{t-1}), \ldots, \varphi^{-1}(x_{t-k}))'$. It should be noted that for a linear mean process $\varphi(v) = v$, and it is necessary to assume that $\boldsymbol{\alpha} \in \Delta_{k+1}$ where $\Delta_{k+1} = \{\boldsymbol{\alpha} \in (0, 1)^{k+1} | \sum_{i=0}^k \alpha_i \in (0, 1)\}$ for the process conditional mean to belong to the unit interval. We will refer to this as a convexity constraint.

In Figure 1 there are some sample paths of a beta process of the third order, with $\varphi(v) = v$. The paths are given for different values of the precision parameter ϕ and of the constant term α_0. For higher values of ϕ the process exhibits less volatility (top-left chart). The larger the value of the constant term α_0, the greater the conditional mean of the process (top-right chart). Finally we observe that the conditional mean of the process $\eta_t \in [\alpha_0, \sum_{i=1}^k \alpha_i]$. When $\alpha_0 < 1/\phi$ and $\sum_{i=1}^k \alpha_i > (\phi - 1)/\phi$, then $\phi\eta_t < 1$ and $\phi(1 - \eta_t) < 1$, and the transition density of the process is anti-unimodal. In this case the process exhibits a switching-type behaviour (see the left and right charts at the bottom of Figure 1).

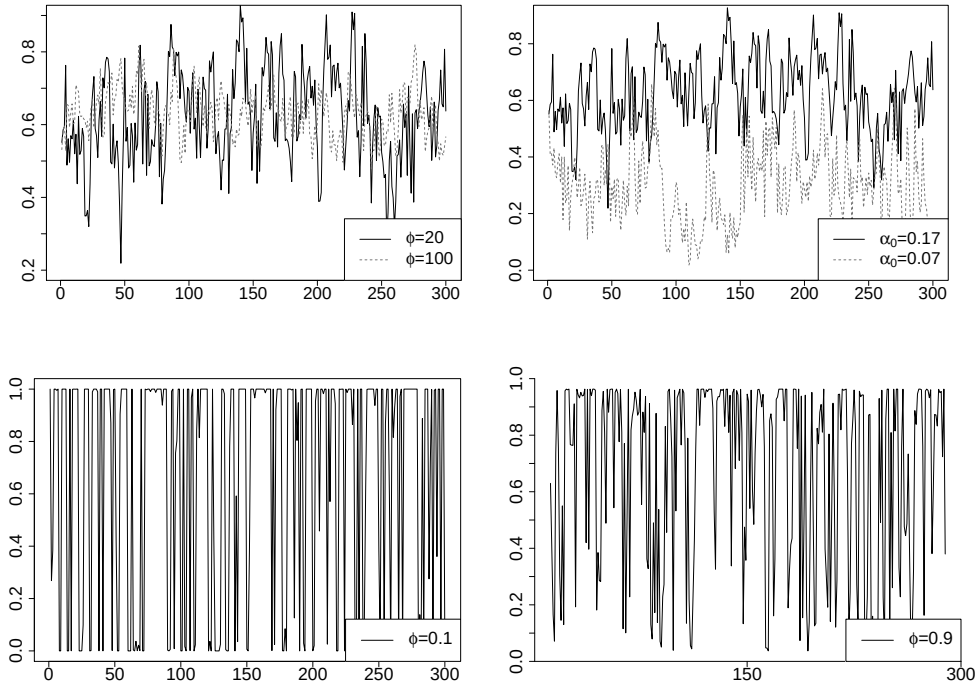

Figure 1: Simulated trajectories of a BAR(3) process for different parameter settings. *Top-left*: the effect of the precision parameter $\phi \in \{20, 100\}$ for $\boldsymbol{\alpha} = (0.17, 0.03, 0.1, 0.60)$. *Top-right*: effect of the constant term α_0 for $\phi = 100$ and $(\alpha_1, \alpha_2, \alpha_3) = (0.03, 0.1, 0.60)$. *Bottom-left* and *bottom-right*: anti-unimodal transition distribution and switching-type trajectories of the BAR(3) process for different values of ϕ ($\phi \in \{0.1, 0.9\}$) and with $\boldsymbol{\alpha} = (0.46, 0.03, 0.01, 0.30)$. Both of the cases correspond to $\alpha_0 = 0.46 < 1/\phi$ and $\sum_{i=1}^{k} \alpha_i = 0.8 > (\phi - 1)/\phi$.

3 Bayesian inference

The likelihood function of the model is

$$\mathcal{L}(\boldsymbol{\alpha}, \phi, k | \mathbf{x}_{t_0:T}) = \prod_{t=t_0}^{T} B(\eta_t \phi, (1-\eta_t)\phi)^{-1} x_t^{\eta_t \phi - 1} (1 - x_t)^{(1-\eta_t)\phi - 1} \tag{4}$$

where $\mathbf{x}_{t_0:T} = (x_{t_0}, \ldots, x_T)'$ and $t_0 = k_{\max} + 1$. Note that we consider an approximated likelihood because, for a beta process of the order $k \leq k_{\max}$, we assume the observations start in $t = k_{\max} + 1$ and thus forget the first $(k_{\max} - k)$ observations on x_t. Moreover, in the following we will assume that the first $k_{\max}$ initial values of the process are known. It is possible to include the initial values in the inference process following, for example, the approach given in Vermaak et al. (2004) for the Gaussian autoregressive processes.

3.1 The priors

Assuming that the constant term and the coefficients of a BAR(k) belong to the set $\mathbb{R}^{k+1}$, then we consider different prior specifications.

For general $\varphi(v)$ we use a multivariate normal with mean $\boldsymbol{\nu}$ and variance Υ, truncated to the set $\mathbb{R}^{k+1}$. In the following we will denote

$$f(\boldsymbol{\alpha}|k) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\nu})'\Upsilon^{-1}(\boldsymbol{\alpha}-\boldsymbol{\nu})\right\}\mathbb{I}_{\mathbb{R}^{k+1}}(\boldsymbol{\alpha}) \tag{5}$$

the density function of the prior on $\boldsymbol{\alpha}$.

For the linear mean process (i.e., $\varphi(v) = v$), a slightly different prior should be considered. In this setting it is not easy to have a diffuse prior and at the same time to guarantee the convexity constraints, i.e., $\eta_t\phi > 0$ and $(1-\eta_t)\phi > 0\ \forall t$. The truncation on the simplex of the normal prior distribution can generate numerical problems in the evaluation of the posterior distribution. In order to prevent the posterior from taking values near the boundaries of the parameter space, Robert and Rousseau (2002) introduce a repulsive factor around the boundaries of the standard simplex defined in the previous section. We observe that $\alpha_0 \leq \eta_t \leq \sum_{i=1}^{k}\alpha_i$ and propose the following prior distribution conditional on ϕ

$$f(\boldsymbol{\alpha}|\phi,k) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}-\boldsymbol{\nu})'\Upsilon^{-1}(\boldsymbol{\alpha}-\boldsymbol{\nu})\right\}\exp\left\{-\frac{\kappa}{\phi^2\alpha_0(1-\sum_{i=1}^{k}\alpha_i)}\right\}\mathbb{I}_{\Delta_{k+1}}(\boldsymbol{\alpha}) \tag{6}$$

where κ is a hyperparameter. In the left and right charts of Figure 2 we show, for the bivariate case, the shape of this prior distribution conditional on $\phi = 10$ for two values of the hyperparameters ($\kappa = 5$ and $\kappa = 10$). The multiplicative factor creates low density regions near the boundaries of the simplex. In both the simulation experiments and the real data applications considered in Sections 4 and 5, this kind of prior contributes considerably to avoiding the numerical problems in the evaluation of the posterior density and its gradient and Hessian, which are needed for the simulation based inference procedures.

The precision parameter ϕ is positive, thus we assume a gamma prior

$$\phi \sim \mathcal{G}a(c,d) \tag{7}$$

with hyperparameters c and d, and we denote with $f(\phi|k)$ the associated density function.

Finally for the model index k, which will be included in the inference process, we assume a discrete uniform prior on $\{1,\ldots,k_{\max}\}$ and denote with $f(k)$ the associated density function.

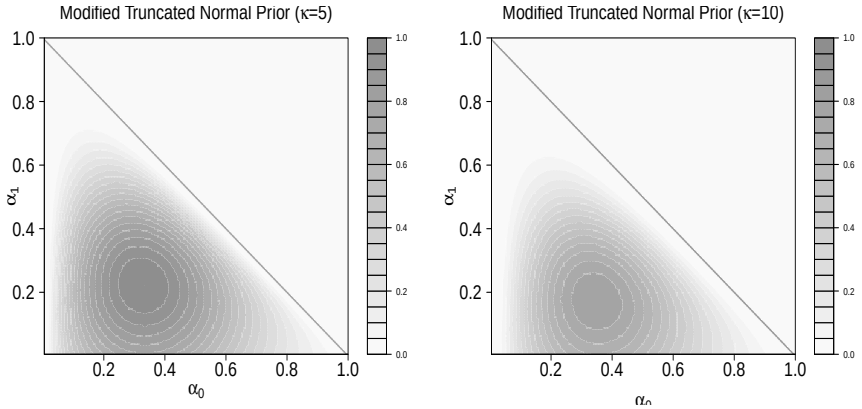

Figure 2: Prior distributions for $\boldsymbol{\alpha}$ in the simplex Δ_{k+1} for $k = 1$. Each graph shows the level set (*gray areas*) and the sum-to-one constraint (*solid line*). Modified normal prior conditional on $\phi = 10$, with parameters $\boldsymbol{\nu} = (1/(k+2), 1/(k+2))'$ and $\Upsilon = 0.1I_2$ for $\kappa = 5$ (*left*) and $\kappa = 10$ (*right*).

3.2 The Reversible Jump algorithm

According to the RJMCMC approach, a set of models $\mathcal{M} = \{\mathcal{M}_k\}_{k\in\mathcal{K}}$ is assumed, where the model $\mathcal{M}_k$ is a BAR(k) with a vector of parameters $\boldsymbol{\theta}$ of dimension n_k, where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \phi)$ and $n_k = k + 1$. Then the joint distribution of $(k, \boldsymbol{\theta})$, given the observed data $\mathbf{x}_{t_0:T}$ is

$$\pi(k, \boldsymbol{\theta} \mid \mathbf{x}_{t_0:T}) \propto \mathcal{L}(\boldsymbol{\theta}, k|\mathbf{x}_{t_0:T}) f(k, \boldsymbol{\theta}) \tag{8}$$

where $\mathcal{L}(\boldsymbol{\theta}, k|\mathbf{x}_{t_0:T})$ is the likelihood given in Equation (4) and $f(k, \boldsymbol{\theta}) = f(\boldsymbol{\alpha}|\phi, k) f(\phi|k) f(k)$ is the joint prior for the model order and the parameter from the previous section.

The joint posterior distribution in Equation (8) is the target distribution of the RJMCMC sampler over the state space $\Theta = \cup_{k\in\mathcal{K}}(k, \mathbb{R}^{n_k})$. Within each iteration, the RJMCMC algorithm updates the parameters, given the model order, and then the model order given the parameters.

The full conditional distributions of the parameters given the model cannot be sampled exactly and are simulated by the following Metropolis-Hastings (M.-H.) within Gibbs steps.

The full conditional distribution of $\boldsymbol{\alpha}$ is

$$\pi(\boldsymbol{\alpha}|\phi, k, \mathbf{x}_{t_0:T}) \propto \exp\left(-\sum_{t=t_0}^{T} \log B(\eta_t\phi, (1-\eta_t)\phi) + \sum_{t=t_0}^{T} A_t\eta_t\phi \right) f(\boldsymbol{\alpha}|k) \tag{9}$$

where $A_t = \log(x_t/(1-x_t))$. To simulate from this distribution we employ a M.-H. algorithm with a proposal distribution that makes use of the information on the local structure of the posterior surface (see Albert and Chib (1993) and Lenk and

DeSarbo (2000)). Consider the second-order Taylor expansion of the log-posterior, $g_k(\boldsymbol{\alpha}) = \log \pi(\boldsymbol{\alpha}|\phi, k, \mathbf{x}_{t_0:T})$, centred around $\tilde{\boldsymbol{\alpha}}^{(j)}$,

$$g_k(\boldsymbol{\alpha}) \approx g_k(\tilde{\boldsymbol{\alpha}}^{(j)}) + (\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}^{(j)})'\nabla^{(1)} g_k(\tilde{\boldsymbol{\alpha}}^{(j)}) + \frac{1}{2}(\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}^{(j)})'\nabla^{(2)} g_k(\tilde{\boldsymbol{\alpha}}^{(j)})(\boldsymbol{\alpha} - \tilde{\boldsymbol{\alpha}}^{(j)}) \quad (10)$$

where $\tilde{\boldsymbol{\alpha}}^{(j)}$ represents the approximated mode of the posterior. Then at the j-th iteration of the M.-H. step we generate a candidate as follows

$$\boldsymbol{\alpha}^{(*)} \sim \mathcal{N}(\tilde{\boldsymbol{\alpha}}^{(j)}, \Sigma^{(j-1)}) \quad (11)$$

where $\Sigma^{(j-1)} = -\left(\nabla^{(2)} g_k(\tilde{\boldsymbol{\alpha}}^{(j-1)})\right)^{-1}$. We remark that in the linear case the candidate distribution has to be truncated to the set Δ_{k+1}, in order to satisfy the convexity constraints.

The full conditional distribution of ϕ is

$$\pi(\phi|\boldsymbol{\alpha}, k, \mathbf{x}_{t_0:T}) \propto \exp\left(-\sum_{t=t_0}^{T}\Big(\log B(\eta_t\phi, (1-\eta_t)\phi) + \phi(A_t\eta_t + \log(1-x_t))\Big)\right) f(\phi|k). \quad (12)$$

We simulate from the full conditional with a M.-H. step. We consider a gamma random walk proposal and, at the j-th step of the algorithm, given the previous value $\phi^{(j-1)}$ of the chain, we simulate

$$\phi^{(*)} \sim \mathcal{G}a(\zeta(\phi^{(j-1)})^2, \zeta\phi^{(j-1)}) \quad (13)$$

where ζ represents the scale of the random walk. After tuning, $\zeta = 0.01$ was finally adopted, achieving an acceptance rate approximately equal to 80% with an autocorrelation lower than 0.2 for lag ≥ 5. The proposal distribution for ϕ is centered on the previous value with a fixed variance of $1/\zeta$. An alternative proposal distribution for positive support quantities would be $\mathcal{G}a(\zeta,\zeta/\phi^{(j-1)})$, which has mean $\phi^{(j-1)}$ and variation coefficient $1/\sqrt{\zeta}$.

The proposal is accepted with probability

$$\min\left\{1, \frac{\pi(\phi^{(*)}|\boldsymbol{\alpha}, k, \mathbf{x}_{t_0:T})}{\pi(\phi^{(j-1)}|\boldsymbol{\alpha}, k, \mathbf{x}_{t_0:T})} \frac{\Gamma(\zeta(\phi^{(j-1)})^2)(\phi^{(j-1)})^{\zeta(\phi^{(*)})^2-1}(\zeta\phi^{(*)})^{\zeta(\phi^{(*)})^2}}{\Gamma(\zeta(\phi^{(*)})^2)(\phi^{(*)})^{\zeta(\phi^{(j-1)})^2-1}(\zeta\phi^{(j-1)})^{\zeta(\phi^{(j-1)})^2}}\right\}. \quad (14)$$

We now describe the model update step. We start with a BAR(k) with parameters $\boldsymbol{\alpha}$ and ϕ. Suppose the RJMCMC proposal is to move to the model BAR(k'), with parameters $\boldsymbol{\alpha'}$ and ϕ. Let $p_{k\,k'}$ denote the probability of proposing to move to BAR(k'), given that the chain is currently in BAR(k), with an analogous notation for the reverse move. The RJMCMC proposes a jump to the space $\mathbb{R}^{k'+1}$ through three steps. First, generate k' from a discrete uniform distribution $p_{k\,k'}$, secondly draw a k'-dimensional parameter vector $\mathbf{u} \in \mathbb{R}^{k'+1}$ from the distribution $q(\mathbf{u})$ and, finally, let $\boldsymbol{\alpha'}$ be a deterministic function of $\boldsymbol{\alpha}$ and $\mathbf{u}$. Conversely, in the reverse move, we generate a k-dimensional parameter vector $\mathbf{u}' \in \mathbb{R}^{k+1}$, setting $\boldsymbol{\alpha}$ to be a function of $\boldsymbol{\alpha'}$ and $\mathbf{u}'$. For the dimension

matching, we consider the linear map $(\boldsymbol{\alpha}', \mathbf{u}') = (\mathbf{u}, \boldsymbol{\alpha})$, which has the Jacobian equal to one. The acceptance probability is $\min(1, A_{k,k'})$, where

$$A_{k,k'} = \frac{\mathcal{L}(\mathbf{u}, \phi, k'|\mathbf{x}_{t_0:T}) f(\mathbf{u}|k') f(\phi|k') p_{k'} p_{k',k}}{\mathcal{L}(\boldsymbol{\alpha}, \phi, k|\mathbf{x}_{t_0:T}) f(\boldsymbol{\alpha}|k) f(\phi|k) p_k p_{k\,k'}} \frac{q(\boldsymbol{\alpha})}{q(\mathbf{u})},$$

and $p_{k'}$ and p_k are the prior probabilities of, respectively, model BAR(k') and BAR(k).

As a proposal for ϕ and $\mathbf{u}$, we consider a parametric family of distributions and choose the parameters in such a way that the first and higher-order derivatives of the log-acceptance ratio are approximately equal to zero. This implies an approximately constant acceptance ratio. We consider a second-order method and focus on the gradient and the Hessian of the log-acceptance ratio with respect to $(\mathbf{u}, \boldsymbol{\alpha})$. The gradient naturally splits into two subvectors, which are the gradients with respect to $\mathbf{u}$ and $\boldsymbol{\alpha}$. The cross derivatives in the Hessian are null. In the following, we consider the case for $\mathbf{u}$; the case for $\boldsymbol{\alpha}$ being similar.

Let $\varphi^{(1)}$ and $\varphi^{(2)}$ be, respectively, the first derivative and second derivative of the function φ. We choose the parameters of the proposal such that

$$\begin{aligned}
\nabla^{(1)} \log A_{k,k'} =& \phi \sum_{t=t_0}^{T} \left(A_t - \Psi^{(0)}(\eta_t \phi) + \Psi^{(0)}((1-\eta_t)\phi) \right) \varphi^{(1)}(\alpha' \mathbf{z}_t) \mathbf{z}_t \\
&+ \nabla^{(1)} \log f(\mathbf{u}|k') - \nabla^{(1)} \log q(\mathbf{u}) = \mathbf{0} \\
\nabla^{(2)} \log A_{k,k'} =& -\phi^2 \sum_{t=t_0}^{T} \left(\Psi^{(1)}(\eta_t \phi) + \Psi^{(1)}((1-\eta_t)\phi) \right) (\varphi^{(1)}(\alpha' \mathbf{z}_t))^2 \mathbf{z}_t \mathbf{z}_t' \\
&- \phi \sum_{t=t_0}^{T} \left[\Psi^{(0)}(\eta_t \phi) - \Psi^{(0)}((1-\eta_t)\phi) \right] \varphi^{(2)}(\alpha' \mathbf{z}_t) \mathbf{z}_t \mathbf{z}_t' \\
&+ \nabla^{(2)} \log f(\mathbf{u}|k') - \nabla^{(2)} \log q(\mathbf{u}) = 0
\end{aligned}$$

where $\Psi^{(0)}$ and $\Psi^{(1)}$ are the digamma and trigamma functions respectively.

In contrast to the Gaussian autoregressive model by Ehlers and Brooks (2008), here the gradient and the Hessian depend on $\mathbf{u}$. We propose to evaluate the derivatives at the approximated posterior mode $\tilde{\mathbf{u}}_{k'}$, defined on the k'-dimensional space.

As an example, let us consider a Gaussian prior distribution truncated on the simplex, and a Gaussian proposal distribution with mean $\boldsymbol{\mu}_{k'}$ and variance $\Sigma_{k'}$. We solve

the above system of equations with respect to $\boldsymbol{\mu}_{k'}$ and $\Sigma_{k'}$ and find that

$$\begin{aligned}\Sigma_{k'}^{-1} =&\Upsilon_{k'}^{-1} + \sum_{t=t_0}^{T} \left(\Psi^{(1)}(\eta_t\phi) + \Psi^{(1)}((1-\eta_t)\phi)\right) \phi^2(\varphi^{(1)}(\alpha'\mathbf{z}_t))^2\mathbf{z}_t\mathbf{z}_t' \\ &- \phi \sum_{t=t_0}^{T} \left[\Psi^{(0)}(\eta_t\phi) - \Psi^{(0)}((1-\eta_t)\phi)\right] \varphi^{(2)}(\alpha'\mathbf{z}_t)\mathbf{z}_t\mathbf{z}_t'\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}_{k'} =&\tilde{\mathbf{u}}_{k'} + \Sigma_{k'}\Big[\Upsilon_{k'}^{-1}(\tilde{\mathbf{u}}_{k'} - \boldsymbol{\nu}_{k'}) \\ &- \phi \sum_{t=t_0}^{T} \left(A_t - \Psi^{(0)}(\eta_t\phi) + \Psi^{(0)}((1-\eta_t)\phi)\right) \varphi^{(1)}(\alpha'\mathbf{z}_t)\mathbf{z}_t\Big].\end{aligned}$$

By applying the same procedure for the derivatives with respect to $\boldsymbol{\alpha}$ we obtain a procedure that promotes jumps between the modes of the posterior distributions of $\mathbf{u}$ and $\boldsymbol{\alpha}$, which are defined in the k- and k'-dimensional spaces respectively.

We do not know the value of the posterior mode and in order to find an approximation we suggest to use a Newton-Raphson procedure. Each time the RJMCMC chain visits the posterior in a given dimension, say k, then the approximated mode $\tilde{u}_k^{(j)}$, which is dimension-specific, is updated following a Newton-Raphson procedure. The approximated modes of all the other dimensions are kept unchanged. At the $(j-1)$-th iteration the updating recursion for the approximated modes is

$$\tilde{\mathbf{u}}_k^{(j)} = \begin{cases} \tilde{\mathbf{u}}_k^{(j-1)} - \Sigma^{(j-1)}\nabla^{(1)}g_k(\tilde{\mathbf{u}}_k^{(j-1)}) & \text{if } k^{(j-1)} = k \\ \tilde{\mathbf{u}}_k^{(j)} & \text{if } k^{(j-1)} \neq k \end{cases} \tag{15}$$

where $k^{(j-1)}$ is the value of the model index at iteration $j-1$ and g_k is the log-posterior distribution of $\mathbf{u}$ as defined in Eq. (10).

After an initial learning period, the posterior mode in the various spaces of different dimensions is reached with a certain tolerance value, and the log-acceptance rate of the jump move is close to zero as an effect of the choice of the proposal parameters.

Finally it should be noted that the approximation $\tilde{\mathbf{u}}_{k'}^{(j)}$ of the posterior mode in the jump step is also used in the M.-H. steps within the Gibbs sampler for simulating $\mathbf{u}$ and ϕ; thus no further computational burden is required for this approximation.

4 Simulation Results

In this section we study, through some simulation experiments, the efficiency of the proposals. First, we focus on the parameter inference for given BAR model orders, checking the mixing of the chains through the acceptance rate of the M.-H. steps, the effective sample size, and the convergence diagnostic statistics. Then, we describe parameter inference results for BAR models of unknown order through posterior model probabilities.

4.1 Inference with fixed AR order

In a first set of experiments, we verify the efficiency of the Metropolis-Hastings proposals for different prior distributions and different parameter settings.

In each experiment we proceed as follows. For each model order k and parameter setting $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \phi)$, we generate 50 independent random samples with size $T = 300$ from a BAR(k). On each dataset we iterate the RJMCMC algorithm defined in the previous sections $N = 10,000$ times.

As regards the values of the parameters, we consider two scenarios. In the first scenario we set the precision parameter to $\phi = 20$, which corresponds to the case of high variability in the data (see Figure 1); we will refer to this as low precision data. We expect that parameter estimation will be more challenging in this context.

In the second scenario, we set the precision parameter to $\phi = 100$. In this case the data exhibit less variability (see Figure 1); we will refer to this as high precision data.

For each scenario we consider different values of the parameters and different autoregressive orders. The parameter settings used in the MCMC experiments are summarized in Table 1. We also evaluate the efficiency of the MCMC algorithm for different choices of the prior distributions. We consider the truncated normal distribution with parameters $\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$ ($\boldsymbol{\iota}$ being the unit vector of dimension k) and $\Upsilon = 100 I_k$ (I_k being the identity matrix of dimension k) and the modified truncated normal distribution with parameters $\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$ and $\kappa = 10$.

k	$\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \phi)$	$\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \phi)$
1	$(0.32, 0.5, 20)$	$(0.32, 0.5, 100)$
2	$(0.32, 0.5, 0.1, 20)$	$(0.32, 0.5, 0.1, 100)$
3	$(0.32, 0.5, 0.1, 0.03, 20)$	$(0.32, 0.5, 0.1, 0.03, 100)$
4	$(0.32, 0.4, 0.1, 0.03, 0.1, 20)$	$(0.32, 0.4, 0.1, 0.03, 0.1, 100)$

Table 1: The parameter settings employed in the MCMC experiments. *First column*: the autoregressive order. *Second and third columns*: conditional mean parameters for low precision data (i.e., $\phi = 20$) and high precision data (i.e., $\phi = 100$).

Figure 3 shows a typical output and progressive averages of $N = 10,000$ iterations of the MCMC chain for $\boldsymbol{\alpha}$ (left) and ϕ (right). In these figures, we include the initial value of the Gibbs sampler and the burn-in sample in order to show the convergence of the MCMC progressive averages to the true values of the parameters. In this example, we considered a sample of $T = 300$ observations simulated from a BAR(3) model with parameters $\boldsymbol{\alpha} = (0.32, 0.5, 0.1, 0.03)'$ and $\phi = 20$. As it is clear from Figure 3, the autoregressive parameters $\boldsymbol{\alpha}$ seem to reach convergence in a slightly lower number of iterations than the precision parameter ϕ. However, a burn-in period of 1,000 iterations is enough to guarantee convergence for all of the chains. All of the sample paths denote a good chain mixing and the chains show a moderate variability, with the exception of the α_2 parameter, with true value 0.1, that exhibits a higher variability. Furthermore, the posterior estimates are slightly higher than the true values except for the constant

term α_0, with true value 0.32.

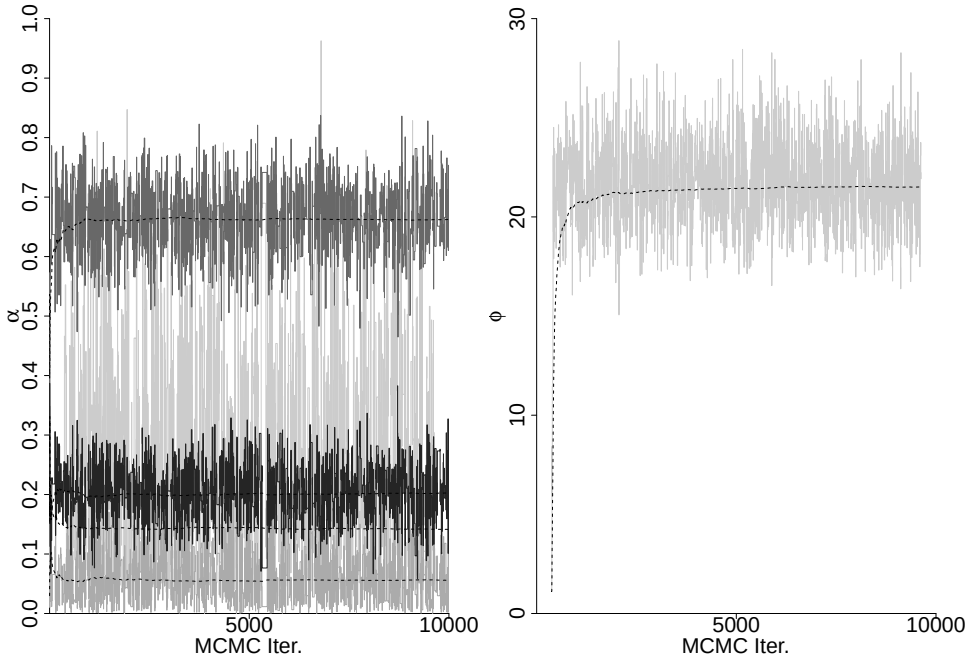

Figure 3: MCMC raw output (*gray solid lines*) and progressive averages (*dashed lines*) for $\boldsymbol{\alpha}$ (*left chart*) and ϕ (*right chart*) estimated on a dataset of $T = 300$ observations simulated from a BAR(3) process with parameters $\boldsymbol{\alpha} = (0.32, 0.5, 0.1, 0.03)'$ and $\phi = 20$. The initial value of the Gibbs sampler and the burn-in sample are included in the MCMC sample in order to show the convergence of the MCMC progressive averages to the true values of the parameters.

For each set of 50 independent MCMC experiments of length N we estimate the root mean square error (RMSE), the average acceptance probability (ACC) of the M.-H. within Gibbs steps and the effective sample size (ESS).

We estimate the RMSE for each component θ of the parameter vector $\boldsymbol{\theta}$.

The ACC index is the acceptance probability, averaged for all the parameters

$$ACC = \frac{1}{50}\frac{1}{N}\left\{\sum_{i=1}^{50}\sum_{j=1}^{N} A_{ij}\right\}$$

where A_{ij} is the acceptance probability for each sample i, with $i = 1, \ldots, 50$ and for each iteration j, with $j = 1, \ldots, N$.

The ESS for a parameter θ is

$$ESS = N \left(1 + \sum_{t=1}^{\infty} \text{corr}(\theta^{(0)}, \theta^{(t)}) \right)^{-1}.$$

In order to evaluate the RMSE and the other statistics we consider $N = 10,000$ MCMC iterations and discard the output of the first 1,000 iterations.

In the various experiments we also evaluate the number of MCMC iterations necessary for our MCMC algorithm to reach convergence. For that purpose, we here combine a graphical inspection of the progressive averages of the parameter posterior distribution with the evaluation of a convergence criterion based on the Kolmogorov-Smirnov (KS) test (see Robert and Casella (2004), Ch. 12). In the simulation experiments, for each component θ of the parameter vector $\boldsymbol{\theta}$, we split the associated MCMC sample $\theta^{(j)}$, $j = 1, \ldots, N$ into two subsamples $\theta_1^{(g)}$ and $\theta_2^{(g)}$ with $g = 1, \ldots, M$ and evaluate

$$KS = \frac{1}{M} \sup_{\eta} \left| \sum_{g=1}^{M} \mathbb{I}_{(0,\eta)}(\theta_1^{(gG)}) - \sum_{g=1}^{M} \mathbb{I}_{(0,\eta)}(\theta_2^{(gG)}) \right|$$

where G is the batch size. The use of batches is necessary in order to obtain quasi-independent samples. The independence of the samples is one of the assumptions to have a known limit distribution for the KS statistics. For each experiment we show the average p-value of the KS statistics over the vector of parameters and over the last 100 iterations of the MCMC chain.

We summarise the output of the MCMC experiments in Table 2 for the truncated-Gaussian and modified truncated-Gaussian priors. From Table 2 we note that, generally, the higher the model order, the worse the performance of the algorithm. The RMSE of the autoregressive coefficients and of the ϕ parameter tends to grow with the model order, denoting a lower precision of the estimates. Furthermore, the ESS decreases as the order of the process increases, denoting a worse mixing of the Markov chain.

Moreover, the results (see Table 2) show that high precision data have a better performance than low precision data. This is evident from the autoregressive coefficients, whose RMSEs are generally lower for high precision data. We get the same conclusion for the ϕ parameter. Furthermore, data with higher precision show a noticeable improvement in terms of ACC compared to data with lower precision. From Table 2 we can conclude that the improvement is mainly due to the higher ACC value of the ϕ parameter, since the ACC of the autoregressive coefficients is very similar for both types of data. Moreover, as we see in the fourth column of Table 2, ESS values are higher in high precision data, showing a better mixing than in low precision data.

The average p-values of the KS statistic take values close to 0.5 in all cases, suggesting the acceptance of the null hypothesis that the subsamples associated with the Markov chain have the same distribution, guaranteeing convergence. Besides, since Table 2 displays only the average p-values, we stress that over the last 100 iterations of the Markov chains the KS p-values improved, getting closer to 1.

Finally, the modified truncated-Gaussian prior performs better than the truncated-Gaussian prior for both the low and high precision data. In particular, we note that the ACC and the ESS values of the modified prior are slightly higher, and the RMSE of the ϕ parameter is noticeably lower than the corresponding values of the truncated-Gaussian.

Low precision data									
k	Estimated RMSE						*ACC*	*ESS*	*KS*
	α_0	α_1	α_2	α_3	α_4	ϕ			
	Truncated-Gaussian Prior ($\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$)								
1	0.032	0.058				0.376	0.176	704	0.534
2	0.033	0.043	0.023			0.996	0.172	630	0.552
3	0.087	0.094	0.026	0.051		2.092	0.163	584	0.523
4	0.041	0.011	0.019	0.075	0.032	3.727	0.155	538	0.541
	Modified Truncated-Gaussian Prior ($\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$, $\kappa = 10$)								
1	0.033	0.051				0.392	0.181	1013	0.556
2	0.032	0.055	0.021			0.916	0.183	853	0.563
3	0.015	0.077	0.018	0.059		1.701	0.192	783	0.574
4	0.030	0.023	0.013	0.059	0.034	2.564	0.189	740	0.539
High precision data									
k	Estimated RMSE						*ACC*	*ESS*	*KS*
	α_0	α_1	α_2	α_3	α_4	ϕ			
	Truncated-Gaussian Prior ($\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$)								
1	0.011	0.018				0.964	0.392	923	0.513
2	0.029	0.047	0.031			1.815	0.402	827	0.546
3	0.038	0.071	0.032	0.002		3.122	0.422	798	0.593
4	0.021	0.038	0.037	0.007	0.029	6.430	0.538	778	0.511
	Modified Truncated-Gaussian Prior ($\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$, $\kappa = 10$)								
1	0.017	0.021				0.392	0.403	1198	0.511
2	0.020	0.028	0.001			1.101	0.420	1012	0.534
3	0.031	0.063	0.003	0.002		1.539	0.428	941	0.529
4	0.029	0.033	0.033	0.001	0.018	3.955	0.509	830	0.542

Table 2: Estimation results for different model orders, parameter settings, and prior distributions. The results are averages over a set of 50 independent MCMC experiments on 50 independent datasets of $T = 300$ observations. On each dataset we ran the proposed MCMC algorithm for $N = 10,000$ iterations and then discarded the first 1,000 iterations. In each panel: model order (*first column*); estimated root mean square error (RMSE) (*second column*) for each parameter; average acceptance rate (ACC) and effective sample size (ESS) (*third and fourth columns*) averaged over the two M.-H. chains in the Gibbs sampler; KS convergence diagnostic statistics (*last column*) with batch size $G = 50$ (average over the last 100 iterations).

Figure 4 gives a more detailed description of the behaviour of the RMSE and of the ESS in the MCMC experiments. The large circles denote the truncated-Gaussian prior

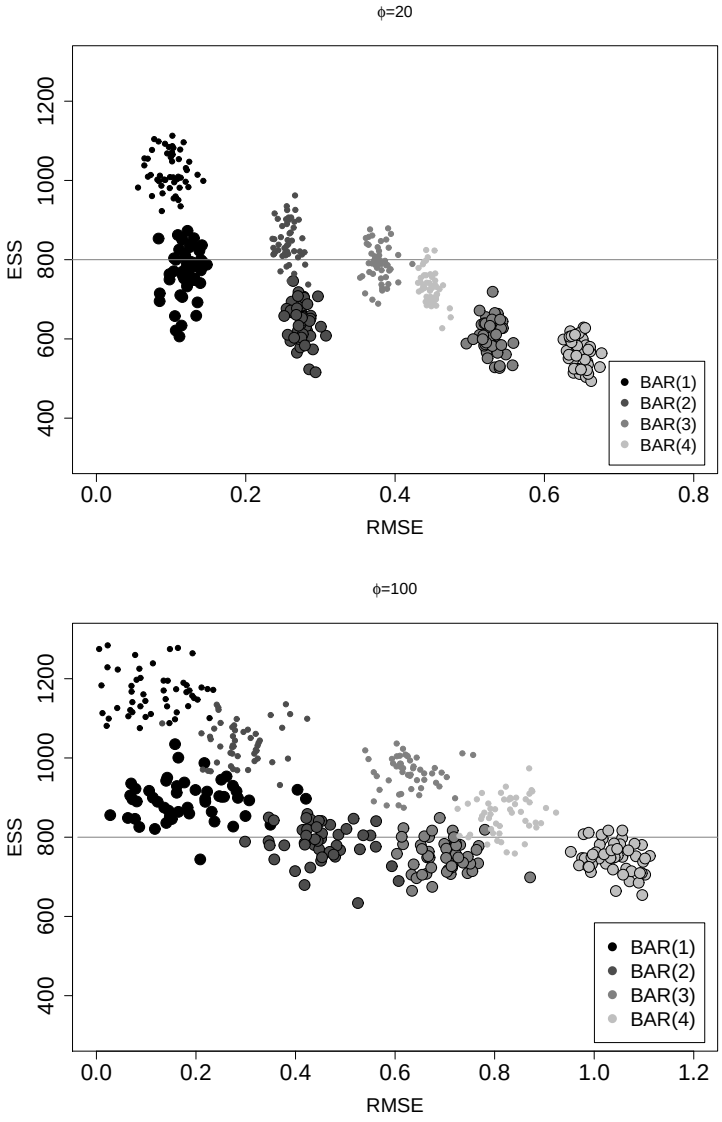

Figure 4: ESS and RMSE of the 50 MCMC simulation experiments (averages over the parameter vector) for different autoregressive orders of the BAR(k) (different gray levels of the circles) and different choices of the prior (bigger circles for the truncated Gaussian and smaller circles for the modified Gaussian prior). The rows show the ESS and the RMSE statistics for two different values of ϕ.

results, while the small circles denote the modified truncated-Gaussian prior results.

The different colors of the circles indicate the different order of the BAR(k). The top plot displays low precision results ($\phi = 20$) and the bottom plot displays high precision results ($\phi = 100$).

For both high and low precision data, the RMSE grows with the order of the model, the ESS decreases. The modified prior gives more efficient estimates in terms of RMSE and ESS.

The RMSEs calculated with the low precision data ($\phi = 20$) are lower than the RMSEs calculated with the high precision data ($\phi = 100$), but this is because the circles depicted in Figure 4 are averages over all the parameters, including the parameter ϕ.

4.2 Inference with unknown AR order

The aim of the second set of experiments is to study the interplay between the sample size and model posterior probability. We consider a dataset of 500 observations simulated from a BAR(k) with $k = 3$ and parameter values $(\boldsymbol{\alpha}', \phi) = (0.37, 0.4, 0.1, 0.03, 100)$. We assume a modified truncated-Gaussian prior with parameters $\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$ and apply the RJMCMC algorithm presented in the previous section for estimating k to subsamples of different sizes (from 100 to 500 observations), with $k_{\max} = 15$. Our aim is to analyse how the RJMCMC approximates the autoregressive order estimate when we have samples of increasing dimension.

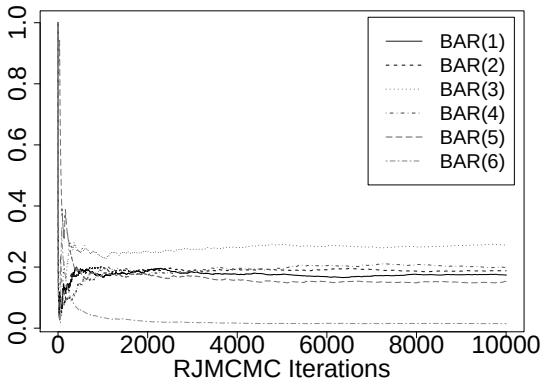

Figure 5: RJMCMC progressive estimates of the model posterior probabilities, when the true model is a BAR(3) with $\boldsymbol{\alpha} = (0.37, 0.4, 0.1, 0.03)'$ and $\phi = 100$.

A typical RJMCMC estimation of model posterior probabilities for a dataset of 300 observations is given in Figure 5, when the true model is BAR(3). The chain moves frequently between the different subspaces, but it visits the BAR(3) model more often. The least visited model is instead BAR(6).

Estimated Model Probabilities					
k \ T	100	200	300	400	500
1	0.06667	0.09143	0.10291	0.09806	0.00129
2	0.19208	0.38611	0.29345	0.19384	0.17828
3	0.18939	0.43252	0.35730	0.65473	0.75391
4	0.13245	0.03387	0.24522	0.05255	0.04449
5	0.10674	0.02291	0.00002	0.00011	0.00018
6	0.06682	0.01910	0.00001	0.00033	0.00013
7	0.05682	0.00691	0.00001	0.00001	0.00663
8	0.04682	0.00149	0.00001	0.00001	0.00012
9	0.04283	0.00139	0.00001	0.00001	0.00018
10	0.03891	0.00109	0.00042	0.00001	0.00001
11	0.01488	0.00080	0.00014	0.00001	0.00001
12	0.01455	0.00055	0.00022	0.00001	0.00001
13	0.01109	0.00062	0.00012	0.00001	0.00001
14	0.01078	0.00027	0.00011	0.00001	0.00001
15	0.00917	0.00092	0.00005	0.00030	0.00023
Mode	2	3	3	3	3
Mean	4.80121	2.65058	2.75522	2.66801	2.85226
s.d.	3.1253	1.2256	0.9854	0.76153	0.70101

Table 3: Relation between sample size T (*first row*) and model order posterior (*columns from one to six*) for $k_{\max} = 15$ when data are simulated from a BAR(3) with $(\boldsymbol{\alpha}', \phi) = (0.37, 0.4, 0.1, 0.03, 100)$. We assume a modified truncated-Gaussian prior with parameters $\boldsymbol{\nu} = (k+2)^{-1}\boldsymbol{\iota}$, $\Upsilon = 100 I_k$. Approximation of the model order posterior and of its mode, mean and standard deviation (*last three rows*) is based on 100,000 RJMCMC iterations.

The results of the model selection procedure for the different sample sizes are given in Table 3. All estimates are based on 100,000 iterations of the RJMCMC chain. In the specific parameter setting considered, the last three rows in the table show two interesting results. First, the estimated model order is not correct ($\hat{k} = 2$ with $k = 3$) for a sample of 100 observations. Furthermore, the standard deviation is 3.1253 for $T = 100$ and decreases for increasing T, which means that the model posterior distribution is less concentrated around the mode for the samples of smaller size.

Small-sample bias in the model order estimates has also been observed in RJMCMC-based model selection procedures for other non-Gaussian autoregressive processes, such as the integer valued ARMA processes (see Enciso-Mora et al. (2009)); however, the high dispersion of the model order estimates in small samples is a characteristic of the beta autoregressive processes.

Both of these results may contribute to the explanation of the shape of the model posterior distributions presented in Section 5.

5 Production factor data

The utilisation rate of the production factors is a key variable for policy makers. Better modelling and forecasting of factor utilisation can improve the ability of policy makers to make effective decisions. In this empirical analysis, we focus on two well studied series for factor utilisation: the unemployment rate and the capacity utilisation rate.

The dynamics of the unemployment rate is certainly one of the most studied in time series econometrics. See for example Deschamps (2008) for some recent modeling advances. The unemployment rate is usually characterised by a pattern of relatively short periods of rapid economic contraction and by relatively long periods of slow expansion. We do not aim to model the asymmetric behaviour of time series, but focus instead on another fundamental feature of this variable, that is, that the unemployment rate is naturally defined on the $(0, 1)$ interval. Data transformation is usually applied to map data on the real interval (see Wallis (1987)). We follow Rocha and Cribari-Neto (2009) and apply beta autoregressive processes for modelling the unemployment rate.

While unemployment rate deals with utilisation of labour as a production factor, capacity utilisation deals with all of the production factors (i.e., labour force and stock of capital) of an economic system or sector. A detailed definition of capacity utilisation and a discussion on the different ways to obtain a statistical measure of this quantity can be found in Klein and Su (1979). A decreasing capacity utilisation is usually interpreted as a slowdown of the aggregated demand and consequently a reduction in the inflation level. An increase in capacity utilisation reveals an expansion of the level of economic activity (see Baghestani (2008)).

5.1 Unemployment Data

Here we consider two interesting datasets (see Figure 6). The first one is the US unemployment rate (source: Datastream (2010)) sampled at a monthly frequency from February 1971 to December 2009. We are mainly interested in modelling the economic cycle and thus consider deseasonalised data. This dataset is quite large (467 observations) when compared to other macroeconomic datasets and this variable is one of the most studied in econometrics (see for example Nickel (1997)). The other dataset is the deseasonalised unemployment rate of the Euro Area sampled at a monthly frequency (source: Datastream (2010)), from January 1995 to December 2009. We consider the aggregated unemployment rate for the 13 countries in the area. This is another well studied variable (see Bean (1994)) and inference on this dataset could be challenging due to the limited amount of observations (180 observations). Moreover, modelling and forecasting of this variable represents one of the most important issues for the European Central Bank and for the European Institute of Official Statistics (Eurostat).

We assume a modified Gaussian prior for the autoregressive parameters with hyperparameters $\kappa = 5$ and $\Upsilon = 100 I_k$, a Gamma prior for the precision with hyperparameters $c = 1$ and $d = 1$, and a uniform prior for the autoregressive order with maximum order $k_{max} = 15$. For the RJMCMC algorithm we consider $N = 100,000$ iterations

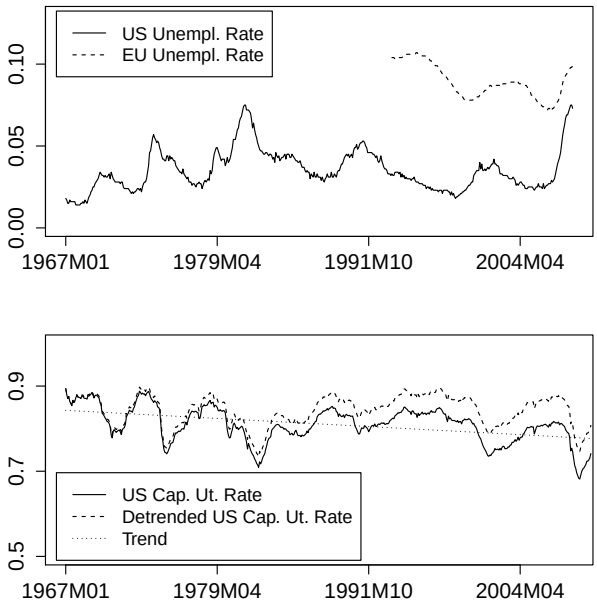

Figure 6: *Top*: US (*solid line*) and EU (*dashed line*) unemployment rates at a monthly frequency. *Bottom*: US capacity utilisation rate (*solid line*), its estimated linear trend (*dotted line*) with intercept $\hat{\gamma}_0 = 0.843$ and slope $\hat{\gamma}_1 = -0.066$ and the detrended capacity rate (*dashed line*).

and discard the first 10,000 samples. The top two plots of Figure 7 illustrate the model posterior probabilities of BAR(k), with $k = 1, \ldots, 15$. Table 4 gives the estimated parameters, the 95% credible interval, and the average acceptance probability (ACC) of the transdimensional jump move in the RJMCMC chain, for the estimated model order.

We compare our results with the results available in the literature. For comparison purposes we apply an inverse logistic transformation to the data (see Deschamps (2008)) and then estimate a Gaussian autoregressive model on the transformed data. In other words, the dependent variable is $y_t = \ln[0.01x_t/(1 - 0.01x_t)]$, where x_t is the monthly unemployment rate. We selected the order of the model, for $k = 1, \ldots, 15$, with the Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC), listing the results in Table 5.

The RJMCMC algorithm strongly favours the BAR(2) model amongst the others. Table 4 shows the parameter estimates for the estimated model order $\hat{k} = 2$. The estimated value of the constant term α_0 is close to 0, and its posterior distribution has a very low variability. The other two autoregressive parameters are instead characterised by higher estimate values and higher variability. Finally, the ACC value lies in the range of 10-60%; that is a good sign of efficiency for most algorithms, as suggested, for

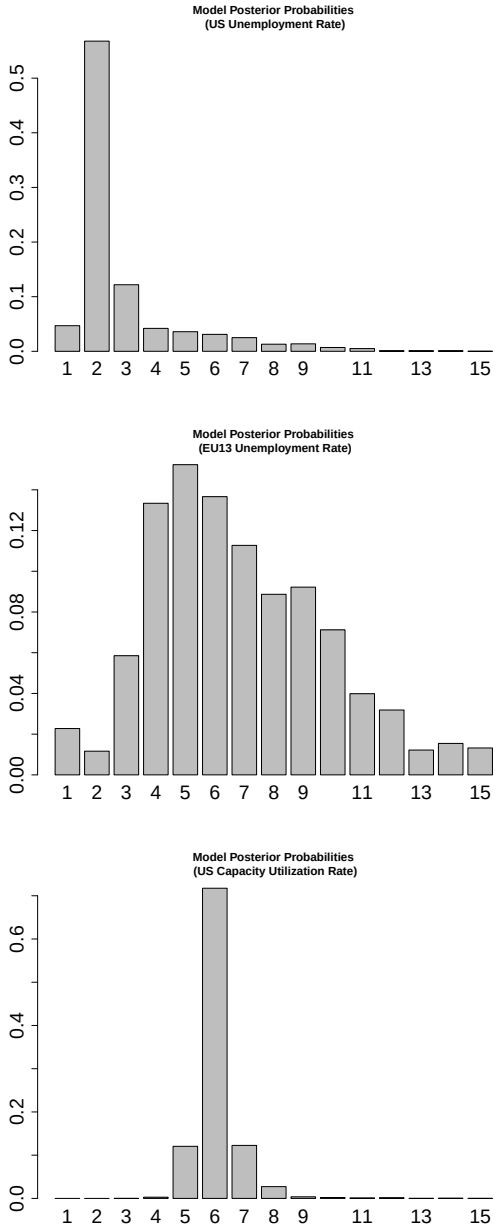

Figure 7: Model probabilities for BAR(k) on the US (*top chart*) and EU13 (*middle chart*) unemployment rates and on the US capacity utilisation rate (*bottom chart*).

US Unemployment Rate ($\hat{k}=2$)								
$\boldsymbol{\theta}_k$	α_0	α_1	α_2				ϕ	ACC
$\hat{\boldsymbol{\theta}}_k$	0.011	0.547	0.272				130	0.392
$Q_{0.025}$	0.006	0.118	0.007				126	
$Q_{0.975}$	0.017	0.820	0.708				131	
EU Unemployment Rate ($\hat{k}=5$)								
$\boldsymbol{\theta}_k$	α_0	α_1	α_2	α_3	α_4	α_5	ϕ	ACC
$\hat{\boldsymbol{\theta}}_k$	0.029	0.544	0.076	0.010	0.024	0.011	128	0.278
$Q_{0.025}$	0.024	0.140	0.001	0.004	0.010	0.004	123	
$Q_{0.975}$	0.034	0.572	0.096	0.016	0.029	0.016	132	
US Capacity Utilisation Rate ($\hat{k}=6$)								
$\boldsymbol{\theta}_k$	α_0	α_1	α_2	α_3	α_4	α_5	α_6	ϕ	ACC
$\hat{\boldsymbol{\theta}}_k$	0.397	0.196	0.075	0.065	0.045	0.053	0.049	130	0.398
$Q_{0.025}$	0.390	0.189	0.058	0.058	0.040	0.049	0.030	126	
$Q_{0.975}$	0.408	0.207	0.079	0.074	0.067	0.060	0.051	130	

Table 4: Estimated parameters $\boldsymbol{\theta}$ and the model order k for the US (*first panel*) and EU13 (*second panel*) unemployment rates and for the US capacity utilisation rate (*last panel*). In the last column, the acceptance rate (ACC) of the trans-dimensional jump move of the RJMCMC chain.

example, by Rosenthal (2011).

Considering instead the results obtained by applying the AR model to the transformed data y_t, the AIC and BIC select the model orders $\hat{k}=7$ and $\hat{k}=4$ respectively (second and fifth columns of Table 5). This result for the BIC is in line with that in Deschamps (2008), whereby a Bayes factor criterion and a log-transformed US unemployment rate, from February 1965 to December 2004, are considered to choose among Gaussian and linear AR models with lags from 1 to 6. The author found that the AR(5) is preferred. Our empirical findings also exhibit another similarity with the findings in Deschamps (2008). In his comparison between linear and different nonlinear Gaussian models, such as Markov-switching AR and smooth-transition AR, Deschamps (2008) found that numbers of lags $\hat{k}=3$ and $\hat{k}=2$ are preferred for nonlinear models. Our results show that in a nonlinear and non-Gaussian model, such as a BAR(k) model, a lower number of lags than in the Gaussian and linear case is needed for fitting the US unemployment rate.

We carry out the same kind of comparison for the EU unemployment rates. The RJMCMC results are shown in the middle chart of Figure 7 and in the second panel of Table 4. The estimated model order is $\hat{k}=5$ (in Figure 7). However, the choice of the model in this case is not as straightforward as it was for the US data, since the posterior probabilities of models such as BAR(4) and BAR(6), are very close to that of BAR(5). As we explained in Section 4, the model posterior distribution is less concentrated around the mode for smaller samples, as in the EU unemployment time

series data, with only 180 observations. The model posterior distribution is instead more concentrated around the mode for samples of larger size, consistent with the results from the US unemployment time series data. Considering the estimated parameter results, in Table 4, we note that all parameter posteriors exhibit low variability, and their estimates are all close to 0, except for α_1 which has a value of about 0.5. Moreover, the data show high precision, since the ϕ parameter estimate is equal to 128. Finally, the average acceptance probability is slightly lower than the one calculated for US data, but it is still fully satisfactory, since it belongs to the range 10-60%. Calculating the AIC and BIC of Gaussian linear AR for the transformed data y_t, we find that $\hat{k} = 7$ and $\hat{k} = 5$ respectively are preferred (third and sixth columns of Table 5). Our RJMCMC approach for the BAR model selects $\hat{k} = 5$. This result may also be due to the sample size bias described in Section 4.2. However, we should note that the BAR model still requires a lower number of lags than the AR model, as already noticed for the US unemployment rate.

5.2 Capacity Utilisation

We consider the capacity utilisation rate series for the US sampled at a monthly frequency from January 1967 to May 2010. The dataset is quite large (521 observations); the series refers to all the industry sectors and is seasonally adjusted (source: Datastream).

From a graphical inspection (bottom of Figure 6) we note that the series exhibits a negative trend. A deterministic trend could be naturally included in the beta regression model with linear conditional mean by imposing some constraints on the slope and the intercept of the linear trend. These constraints can be imposed by a suitable specification of the prior distribution. However, we focus on the autoregressive components; thus we follow a two-step procedure.

First we define a normalised linear trend t/T, where T is the sample size, and introduce the constrained linear regression model

$$x_t = \gamma_0 + \gamma_1 \frac{t}{T} + \varepsilon_t, \quad \text{with } t = 1, \dots, T \tag{16}$$

with $\gamma_0 \in (0,1)$ and $(\gamma_1 + \gamma_0) \in (0,1)$. These parameter constraints ensure that the residuals of the regression are in the interval $(0,1)$. In the first step, we calculate the detrended capacity utilisation rate $\tilde{x}_t = x_t - \hat{\gamma}_1 \frac{t}{T}$. The results of the trend extraction are given in Figure 6, bottom chart. In the second step, we estimate a beta process on the variable $\tilde{x}_t$.

We applied the RJMCMC algorithm with $N = 100,000$ iterations and a burn-in of 10,000 in order to obtain an estimate of the parameters and of the model posterior. We assume a modified Gaussian prior for the autoregressive parameters with hyperparameters $\kappa = 5$ and $\Upsilon = 100 I_k$, a Gamma prior for the precision with hyperparameters $c = 1$ and $d = 1$, and a uniform prior for the autoregressive order with maximum order $k_{max} = 15$.

	AIC			BIC		
k	US Un	EU Un	US Cap	US Un	EU Un	US Cap
1	-2947.934	-1621.598	-3227.065	-2939.493	-1615.212	-3218.554
2	-2946.183	-1670.723	-3271.669	-2933.522	-1661.144	-3258.902
3	-2972.176	-1706.379	-3287.777	-2955.294	-1693.607	-3270.754
4	-2999.256	-1712.201	-3299.520	**-2978.153**	-1696.237	**-3278.241**
5	-3002.027	-1718.446	-3302.093	-2976.703	**-1699.288**	-3276.558
6	-3006.093	-1717.617	-3300.535	-2976.548	-1695.266	-3270.745
7	**-3008.355**	**-1719.267**	-3299.398	-2974.590	-1693.723	-3265.352
8	-3006.355	-1717.612	-3297.507	-2968.370	-1688.876	-3259.205
9	-3006.492	-1716.549	-3300.804	-2964.286	-1684.620	-3258.246
10	-3004.574	-1714.804	**-3306.622**	-2958.147	-1679.682	-3259.809
11	-3003.247	-1716.888	-3304.701	-2952.600	-1678.573	-3253.632
12	-3002.365	-1715.103	-3302.718	-2947.497	-1673.594	-3247.393
13	-3004.718	-1715.291	-3300.796	-2945.630	-1670.590	-3241.216
14	-3004.694	-1713.473	-3299.106	-2941.386	-1665.578	-3235.270
15	-3002.695	-1712.969	-3297.328	-2935.165	-1661.881	-3229.236

Table 5: AIC (columns from 2 to 4) and BIC (columns from 5 to 7) values for Gaussian AR(k) models on transformed data, for US Unemployment Rate (second and fifth columns), EU Unemployment Rate (third and sixth columns), and US Capacity Utilisation Rate (fourth and seventh columns). The values for the selection criteria for the chosen models are in boldface. The chosen models are, respectively, of order 7, 7, and 10 for AIC, and of order 4, 5, and 4 for BIC.

The results of the model selection procedure for the beta process are in the bottom chart of Figure 7 and in the third panel of Table 4. The preferred model order is $\hat{k} = 6$, with the highest posterior model probability, as illustrated in Figure 7. Here, as for the US unemployment data, the sample size of more than 500 observations guarantees that the model posterior distribution is concentrated around the mode.

Table 4 shows that the posterior estimates of all the parameters have a small variability. All of the autoregressive parameter estimates are close to 0, with the exception of α_1 and the constant term α_0, which had the highest value. Moreover, as in the previous examples, the estimate of the parameter ϕ denotes a dataset with high precision. Furthermore, the value of ACC is very similar to the one calculated for the US unemployment data, showing a good value of acceptance rate.

Finally, we applied a logistic transform to the data and estimated the AR model to the transformed time series. Table 5 shows the AIC and BIC results for different model orders. The model orders chosen by the AIC and BIC are $\hat{k} = 10$ and $\hat{k} = 4$ respectively (fourth and seventh columns of Table 5). This result shows that the US capacity rate seems to exhibit a longer memory than what was obtained with a Gaussian linear AR model when using BIC. As a side remark, we note that, even for this data, the AIC for AR is less conservative than the BIC.

6 Conclusion

We proposed an efficient RJMCMC algorithm for conducting Bayesian inference and model selection on BAR processes of general order. The methodology is given through a very general algorithm that depends on the choice of the beta as the conditional distribution of the process. The degrees of freedom in the choice of such a distribution suggest the application of this algorithm to a wide class of autoregressive models, making the methodology very appealing. Moreover, the RJMCMC methodology can be used to analyse other BAR models that include explanatory variables and change-points in the parameters.

In the simulation studies, the algorithm has been shown to be successful in estimating true parameters and number of lags. We found that, in the sub-class of conditionally linear BAR processes, the choice of the prior may have effects on the mixing property of the RJMCMC chain by reducing the probability of the RJMCMC chain moves near the boundaries of the parameter space. We also apply our inference procedure to real-life data such as unemployment and capacity utilisation rates. Our results suggest that, for some series, such as the EU unemployment data, there is a lot of uncertainty in the selection of the number of lags. This uncertainty should be taken into account when using these models in economics decisions. Model and parameter uncertainty can be naturally included in the decision process by using the posterior probabilities of our RJMCMC algorithm for Bayesian Model Averaging.

References

Aitchinson, J. (1986). *The Statistical Analysis of Compositional Data*. London: Chapman & Hall. 385

Albert, J. H. and Chib, S. (1993). "Bayesian analysis of binary and polychotomous response data." *Journal of American Statistical Association*, 88: 669–679. 390

Amisano, G. and Casarin, R. (2007). "Particle filters for Markov-Switching Stochastic Correlation Models." In *Proceedings of the Italian Statistical Society Conference, 2007 Intermediate Conference, Risk and Prediction*, 305–316. CLEUP Padova. 385

Baghestani, H. (2008). "Predicting capacity utilization: Federal Reserve vs time-series models." *Journal of Economics and Finance*, 32: 47–57. 401

Bean, C. (1994). "European unemployment: A survey." *Journal of Economic Literature*, 32: 573–619. 401

Billio, M. and Casarin, R. (2010). "Identifying Business Cycle Turning Points with Sequential Monte Carlo: An Online and Real-Time Application to the Euro Area." *Journal of Forecasting*, 29: 145–167. 385

— (2011). "Beta Autoregressive Transition Markov-switching Models for Business Cycle Analyis." *Studies in Nonlinear Dynamics and Econometrics*, 15(4): 1–32. 385

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003). "Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions." *Journal of the Royal Statistical Society, B*, 65: 3–55. 386

Datastream (2010). "Datastream International." Available: www.datastream.com. 401

Deschamps, P. J. (2008). "Comparing smooth transition and Markov switching autoregressive models of US unemployment." *Journal of Applied Econometrics*, 23: 435–462. 401, 402, 404

Ehlers, R. S. and Brooks, S. P. (2008). "Adpative Proposal Construction for Reversible Jump MCMC." *Scandinavian Journal of Statistics*, 35: 677–690. 386, 392

Enciso-Mora, V., Neal, P., and Subba Rao, T. (2009). "Efficient order selection algorithms for integer-valued ARMA processes." *Journal of Time Series Analysis*, 30: 1–18. 386, 400

Fan, Y. and Sisson, S. (2011). "Reversible Jump Markov chain Monte Carlo." In S.P. Brooks, G. J., A. Gelman and Meng, X.-L. (eds.), *Handbook of Markov Chain Monte Carlo*. Chapman & Hall. 386

Ferrari, S. L. P. and Cribari-Neto, F. (2004). "Beta regression for modelling rates and proportions." *Journal of Applied Statistics*, 31: 799–815. 385, 387

Green, P. J. (1995). "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination." *Biometrika*, 82: 711–732. 386

Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). "Time series of continuous proportions." *Journal of the Royal Statistical Society, Series B*, 55: 103–116. 385

Klein, L. R. and Su, V. (1979). "Direct estimates of unemployment rate and capacity utilization in macroeconometric models." *International Economic Review*, 20: 725–740. 401

Kotz, S. and van Dorp, J. R. (2004). *Beyond Beta: Other Continuous Families of Distributions with Bounded Support and Applications*. New Jersey: World Scientific. 387

Lenk, P. J. and DeSarbo, W. S. (2000). "Bayesian Inference for Finite Mixtures of Generalized Linear Models with Random Effects." *Psycometrika*, 65: 93–119. 390

McKenzie, E. (1985). "An autoregressive process for beta random variables." *Management Science*, 31: 988–997. 385

Nickel, S. A. (1997). "Unemployment and labour market rigidieties: Europe versus North America." *Journal of Economic Perspective*, 11: 55–74. 401

Robert, C. P. (2007). *The Bayesian Choice, 2nd Edition*. New York: Springer Verlag. 386

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods, 2nd Edition*. New York: Springer Verlag. 396

Robert, C. P. and Rousseau, J. (2002). "A Mixture Approach To Bayesian Goodness of Fit." Technical Report 02009, CEREMADE, University Paris Dauphine. 387, 389

Rocha, V. A. and Cribari-Neto, F. (2009). "Beta autoregressive moving averages models." *Test*, 18: 529–545. 385, 387, 401

Rosenthal, J. S. (2011). "Optimal Proposal Distributions and Adaptive MCMC." In S.P. Brooks, G. J., A. Gelman and Meng, X.-L. (eds.), *Handbook of Markov Chain Monte Carlo: Methods and Applications*, 93–112. Chapman & Hall. 404

Taddy, M. (2010). "Autoregressive mixture models for dynamic spatial Poisson processes." *Journal of the American Statistical Association*, 105: 1403–1427. 385

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. J. (2004). "Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes." *Journal of Time Series Analysis*, 25: 785–809. 388

Wallis, F. K. (1987). "Time series analysis of bounded economic variables." *Journal of Time Series Analysis*, 8: 115–123. 385, 401

Acknowledgments

The authors wish to thank Eleni Bakra, Alessandra Luati, Francesco Ravazzolo, and the anonymous referees, for the useful comments on earlier versions of this article and the participants at: the 5th International conference on Computational and Financial Econometrics, 2011, London; Yeditepe International Research Conference on Bayesian Learning, Istanbul, 2010 and the Norges Bank seminar series, 2010, Oslo.