BAYESIAN MODEL UNCERTAINTY IN SMOOTH TRANSITION AUTOREGRESSIONS

By Hedibert F. Lopes and Esther Salazar

University of Chicago and Federal University of Rio de Janeiro

First Version received November 2004

Abstract. In this paper, we propose a fully Bayesian approach to the special class of nonlinear time-series models called the logistic smooth transition autoregressive (LSTAR) model. Initially, a Gibbs sampler is proposed for the LSTAR where the lag length, k, is kept fixed. Then, uncertainty about k is taken into account and a novel reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm is proposed. We compared our RJMCMC algorithm with well-known information criteria, such as the Akaikes̀ information criteria, the Bayesian information criteria (BIC) and the deviance information criteria. Our methodology is extensively studied against simulated and real-time series.

Keywords. Markov Chain Monte Carlo; nonlinear time-series model; model selection; reversible jump MCMC; deviance information criterion.

1. introduction

Even though regime switching models can be traced back to Bacon and Watts' (1971) seminal work on estimating the transition between two intersecting straight lines, a rejuvenated and steady interest on nonlinear time-series models have emerged approximately over the last 15 years. Special attention has been directed towards measuring, testing and forecasting regime-switching models, such as Chan and Tong's (1986) threshold autoregressive models (TAR), Hamilton's (1989) Markov switching-regime models and Teräsvirta's (1994) smooth transition autoregressive models (STAR).

In this paper, we concentrate our attention on STAR models. STAR models generalize TAR models [Tong (1978), Tong and Lim (1980), Tsay (1989) and Tong (1990)] and self-exciting autoregressions (SETAR, Petruccelli and Woolford, 1984; Chen and Tsay, 1991; Wong and Li, 1998), by replacing the step transition functions by a smooth transition function. More explicitly, let y_t be an univariate time-series and $x_t = (y_{t-1}, \ldots, y_{t-k})'$ intercept plus the k most recent values of the series, then standard versions of TAR(k) and SETAR(k, d) models are, respectively,

$$\text{TAR} : y_t = x_t'\theta_1 + I[s_t > c]x_t'\theta_2 + \varepsilon_t,$$
$$\text{SETAR} : y_t = x_t'\theta_1 + I[y_{t-d} > c]x_t'\theta_2 + \varepsilon_t,$$

0143-9782/06/01 99–117 JOURNAL OF TIME SERIES ANALYSIS Vol. 27, No. 1

where c is a location or 'threshold' parameter and d is the 'delay' parameter. Therefore, a STAR model is obtained by replacing the step function, $I[s_t > c]$, by a smooth transition function, $\pi(\gamma, c, s_t)$, with smoothness parameter γ (explicit functional form is presented in Section 2). A STAR(k) model can be viewed as a continuous mixture of AR(k) models, with the smooth transition function mimicking the degree of nonlinearity. STAR models were initially proposed, in its univariate form, by Chan and Tong (1986) and further developed by Luukkonen *et al.* (1988) and Teräsvirta (1994). van Dijk *et al.* (2002) make an extensive review of recent developments related to the STAR model and its variants.

In this article, we focus on the full Bayesian analysis of an important subclass of the STAR model, the logistic STAR model of order k [hereafter, LSTAR(k) model], where the smooth transition function has the form of a logistic function. We focus our attention to the logistic transition function for two main reasons. First, it can be observed as a natural extension of TAR models where the smoothness of the transition between regimes is controlled by one parameter. Second, several of the current research on STAR models focus on the logistic transition function (e.g. see Teräsvirta, 1994; van Dijk *et al.*, 2002). Nonetheless, the algorithms we develop here apply to other transition functions such as the exponential, the second-order logistic functions, or any other continuous transition function taking values in [0, 1].

Several Bayesian approaches in modelling TAR, STAR and their variants have appeared recently. Geweke and Terui (1993) and Chen and Lee (1995) are among the first to propose Bayesian estimation of SETAR models. Chen (1998) proposes the generalized TAR model by incorporating exogenous variables, which Koop and Potter (1999, 2003) compared, Markov-switching and Endogenous Delay TAR models with SETAR models through Bayes factors. So and Chen (2003) developed an Markov Chain Monte Carlo (MCMC) algorithm for SETAR models based on stochastic search ideas, and Campbell (2004) develops a reversible jump MCMC (RJMCMC, Green, 1995) algorithm for SETAR models with different orders in each one of multiple regimes. As for STAR models, Lubrano (2000) uses a sampling importance resampling algorithm (Gelfand and Smith, 1990).

Our main contribution is to propose an MCMC algorithm that fully accounts for model order uncertainty in STAR(k) models, i.e. we assume that k is one of the parameters of the model. To this end, a customized RJMCMC algorithm is proposed. As an important by-product, we devise an MCMC algorithm for fixed-k LSTAR(k) models. The rest of the article is discussed as follows. In Section 2 the general LSTAR(k) model is defined, while Section 3 discusses prior specification and lays down a MCMC algorithm for posterior inference when the model order, k, is kept fixed. Uncertainty about k is explained in Section 4, where our novel RJMCMC algorithm is introduced. Finally, Section 5 includes several simulated and real-time series to substantiate our methods, with final thoughts and perspectives listed in Section 6.

2. THE LOGISTIC STAR MODEL

Recalling the notation from Section 1, let y_t be the observed value of a time-series at time t and $x_t = (1, y_{t-1}, \ldots, y_{t-k})'$ the vector of regressors corresponding to the intercept plus k lagged values, for $t = 1, \ldots, n$. The logistic smooth transition autoregressive model of order k, or simply LSTAR(k), is defined as follows,

$$y_t = x_t'\theta_1 + \pi(\gamma, c, s_t)x_t'\theta_2 + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2) \tag{1}$$

with π playing the role of a smooth transition continuous function bounded between 0 and 1. In this paper, we focus on the logistic transition, i.e.

$$\pi(\gamma, c, s_t) = \{1 + \exp(-\gamma(s_t - c))\}^{-1}.$$

The parameter $\gamma > 0$ is responsible for the smoothness of π, while c is a location or threshold parameter and d is the delay parameter. When $\gamma \to \infty$, the LSTAR model reduces to the TAR model (or, SETAR model when $s_t = y_{t-d}$) (Tong, 1990) and when $\gamma = 0$ the standard AR(k) model arises. Finally, s_t is called the transition variable, with $s_t = y_{t-d}$ commonly used (Teräsvirta, 1994). Even though we use y_{t-d} as the transition variable throughout this paper, it is worth emphasizing that any other linear/nonlinear functions of exogenous/endogenous variables can be easily included with minor changes in the algorithms presented and studied in this paper. We will also assume throughout this paper, without loss of generality, that $d \leq k$ and that $y_{-k+1}, \ldots, y_0$ are known and fixed quantities. This is common in the time-series literature and would only marginally increase the complexity of our computation and could easily be performed by introducing a prior distribution for those missing values. Also, several other functions could be easily accommodated, such as the exponential or the second-order logistic function (van Dijk *et al.*, 2002). Finally, we will denote by LSTAR(k, d) a LSTAR(k) model that uses y_{t-d} as smoothing variable.

The LSTAR(k, d) model can be observed as a model that allows smooth and continuous shifts between two extreme regimes. More specifically, if $\delta = \theta_1 + \theta_2$, then $x_t'\theta_1$ and $x_t'\delta$ represent the conditional means under the two extreme regimes,

$$y_t = (1 - \pi(\gamma, c, y_{t-d}))x_t'\theta_1 + \pi(\gamma, c, y_{t-d})x_t'\delta + \varepsilon_t \tag{2}$$

or,

$$y_t = \begin{cases} x_t'\theta_1 + \varepsilon_t & \text{if } \pi = 0 \\ (1 - \pi)x_t'\theta_1 + \pi x_t'\delta + \varepsilon_t & \text{if } 0 < \pi < 1. \\ x_t'\delta + \varepsilon_t & \text{if } \pi = 1 \end{cases}$$

Therefore, the model (2) can be expressed as (1) with $\theta_2 = \delta - \theta_1$ being the contribution to the regression of considering a second regime. We consider the following parameterization that will be very useful in the computation of the posterior distributions:

$$y_t = z_t'\theta + \varepsilon_t, \tag{3}$$

where $\theta' = (\theta_1', \theta_2')$ represent the linear parameters, (γ, c) are the nonlinear parameters, $\Theta = (\theta, \gamma, c)$ and $z_t' = (x_t', \pi(\gamma, c, y_{t-d})x_t')$. The dependence of z_t on γ, c and y_{t-d} will be made explicit whenever necessary.

3. INFERENCE

We derive an MCMC algorithm for posterior assessment based on two distinct prior specifications: (i) standard conditionally conjugate prior distribution, and (ii) Lubrano's prior. We develop inferential procedures for both linear and nonlinear parameters and for the observable variance, σ^2, so that the number of parameters to be estimated is $2k + 6$, since we also consider the estimation of the delay parameter d. Fully Bayesian inference, including k, is delayed until the Section 4.

3.1. *Prior distribution*

We investigate posterior sensitivity to two functional forms for the prior distribution. In the first case, we adopt the following conditionally conjugate priors: $\theta \sim N(m_\theta, \sigma_\theta^2 I_{2k+2})$, $\gamma \sim G(a, b)$, $c \sim N(m_c, \sigma_c^2)$ and $\sigma^2 \sim IG(v/2, vs^2/2)$, with hyperparameters m_θ, σ_θ^2, a, b, m_c, σ_c^2, v and s^2 chosen to represent relatively little prior information. In the second case, we adopt Lubrano's (2000) prior distributions where $(\theta_2|\sigma^2, \gamma) \sim N(0, \sigma^2 e^\gamma I_{k+1})$ and $\pi(\theta_1, \gamma, \sigma^2, c) \propto (1 + \gamma^2)^{-1}\sigma^{-2}$ for $\theta_1 \in \Re^{k+1}$, $\gamma, \sigma^2 > 0$ and $c \in [c_a, c_b]$, such that the conditional prior for θ_2 becomes more informative as γ approaches zero, relatively noninformative prior about θ_1 and γ, noninformative prior about σ^2 and relatively noninformative prior about c, with $c_a = \hat{F}^{-1}(0.15)$, $c_b = \hat{F}^{-1}(0.85)$, and $\hat{F}$ the data's empirical cumulative distribution function. In both cases, a discrete prior is set for $d \in \{d_1, d_2, \ldots, d_{\max}\}$, where $d_1 \leq \cdots \leq d_{\max}$, for $d_{\max}$ a large upper bound, so $p(d_i) = \Pr(d = d_i)$.

3.2. *Posterior assessment*

We assume that k is known. So, the posterior simulation for θ, γ, c and σ^2 was carried out using the MCMC algorithm. More specifically, the full conditional posterior of θ and σ^2 (not depending on the form of the prior: subjective or objective) are normal and inverse gamma, respectively, in this case it is possible to use the Gibbs sampling algorithm (Gelfand and Smith, 1990). On the contrary, the full conditional of γ and c is unknown, in this case we use the Metropolis–Hastings algorithm (Metropolis *et al.*, 1953; Hastings, 1970). For more details about Gibbs sampling and Metropolis–Hastings algorithms see Gilks *et al.* (1996) or the books by Gamerman (1997) and Robert and Casella (1999). Below, $[\xi]$ denotes the full conditional distribution for the parameter ξ conditional on all the other parameters, and $\sum$ is short for $\sum_{t=1}^{n}$.

3.2.1. *Conditionally conjugate prior specifications*

The full conditional distributions when the conditionally conjugate prior distribution is used are:

[θ] Combining $y_t \sim N(z_t'\theta, \sigma^2)$ with $\theta \sim N(m_\theta, \sigma_\theta^2 I_{2k+2})$, it is easy to see that $[\theta] \sim N(\tilde{m}_\theta, C_\theta)$, where

$$\tilde{m}_\theta = C_\theta\left\{\sigma^{-2}\sum z_t y_t + \sigma_\theta^{-2} m_\theta\right\} \tag{4}$$

and

$$C_\theta^{-1} = \sigma_\theta^{-2} I_{2k+2} + \sigma^{-2}\sum z_t z_t' \tag{5}$$

[σ^2] By noticing that $\varepsilon_t = y_t - z_t'\theta \sim N(0, \sigma^2)$, it can be seen that $[\sigma^2] \sim IG[(n + v)/2, (vs^2 + \sum \varepsilon_t^2)/2]$;

[γ, c] The joint full conditional distribution of γ and c has no known form, so we update them jointly by using a random-walk Metropolis step, with $\gamma^* \sim G[(\gamma^{(i)})^2/\Delta_\gamma, \gamma^{(i)}/\Delta_\gamma]$ and $c^* \sim N(c^{(i)}, \Delta_c)$ the proposal densities, where $\gamma^{(i)}$ and $c^{(i)}$ are the current values of γ and c. The pair (γ^*, c^*) is accepted with probability $\alpha = \min\{1, A\}$, where

$$A = \underbrace{\frac{\prod_{t=1}^n p_N(y_t|z_{t*}'\theta, \sigma^2)}{\prod_{t=1}^n p_N(y_t|z_t'\theta, \sigma^2)}}_{\text{likelihood ratio}} \underbrace{\frac{p_g(\gamma^*|a,b)p_N(c^*|m_c, \sigma_c^2)}{p_g(\gamma^{(i)}|a,b)p_N(c^{(i)}|m_c, \sigma_c^2)}}_{\text{prior ratio}}$$
$$\times \underbrace{\frac{p_g(\gamma^{(i)}|(\gamma^*)^2/\Delta_\gamma, \gamma^*/\Delta_\gamma)}{p_g(\gamma^*|(\gamma^{(i)})^2/\Delta_\gamma, \gamma^{(i)}/\Delta_\gamma)}}_{\text{proposal ratio}}$$

where $z_{t*}' = z_t'(\gamma^*, c^*, d)$, p_g and p_N are the probability density functions of the gamma and the normal respectively. Δ_γ and Δ_c are chosen such that the acceptance probability is between 10% and 50%, for instance.

[d] Inference about the delay parameter d is rather trivial. Since, $d \in \{d_1, d_2, \ldots, d_{\max}\}$, the posterior conditional posterior distribution for d is $p(d|y, \Theta) \propto p(y|d, \Theta)p(d)$.

3.2.2. *Lubrano's prior specifications*

The full conditional distributions when Lubrano's prior distribution is used are:

[θ] Combining $y_t \sim N(z_t'\theta, \sigma^2)$ with $(\theta_2|\sigma^2, \gamma) \sim N(0, \sigma^2 e^\gamma I_{k+1})$, it is easy to view that $[\theta] \sim N(\tilde{m}_\theta^*, C_\theta^*)$, where

$$\tilde{m}_\theta^* = C_\theta^*\left(\sum z_t y_t \sigma^{-2}\right) \tag{6}$$

and

$$C_\theta^* = \left(\sum z_t z_t' \sigma^{-2} + \Sigma^{-1}\right) \tag{7}$$

for $\Sigma^{-1} = \text{diag}(0, \sigma^{-2}e^{-\gamma} I_{k+1})$.

[σ^2] From model 3, $\varepsilon_t = y_t - z_t'\theta \sim N(0, \sigma^2)$, which combined with the noninformative prior $\pi(\sigma^2) \propto \sigma^{-2}$, leads to

$$[\sigma^2] \sim IG\left(\frac{(n+k+1)}{2}, \frac{(e^{-\gamma}\theta_2'\theta_2 + \sum \varepsilon_t^2)}{2}\right).$$

[γ, c] Again, $\gamma^* \sim G[(\gamma^{(i)})^2/\Delta_\gamma, \gamma^{(i)}/\Delta_\gamma]$ and $c^* \sim N(c^{(i)}, \Delta_c)$, truncated at the interval $[c_a, c_b]$, and $\gamma^{(i)}$ and $c^{(i)}$ current values of γ and c. The pair (γ^*, c^*) is accepted with probability $\alpha = \min\{1, A\}$, where

$$A = \frac{\prod_{t=1}^{T} p_N(\varepsilon_t^*|0, \sigma^2)}{\prod_{t=1}^{T} p_N(\varepsilon_t^{(i)}|0, \sigma^2)} \frac{p_N(\theta_2|0, \sigma^2 e^{\gamma^*} I_{k+1})}{p_N(\theta_2|0, \sigma^2 e^{\gamma^{(i)}} I_{k+1})} \frac{\pi(\gamma^*)\pi(c^*)}{\pi(\gamma^{(i)})\pi(c^{(i)})}$$
$$\times \frac{\left[\Phi\left(\frac{c_b - c^{(i)}}{\sqrt{\Delta_c}}\right) - \Phi\left(\frac{c_a - c^{(i)}}{\sqrt{\Delta_c}}\right)\right] f_G(\gamma^{(i)}|(\gamma^*)^2/\Delta_\gamma, \gamma^*/\Delta_\gamma)}{\left[\Phi\left(\frac{c_b - c^*}{\sqrt{\Delta_c}}\right) - \Phi\left(\frac{c_a - c^*}{\sqrt{\Delta_c}}\right)\right] f_G(\gamma^*|(\gamma^{(i)})^2/\Delta_\gamma, \gamma^{(i)}/\Delta_\gamma)},$$

where $\varepsilon_t^* = y_t - z_t'(\gamma^*, c^*, y_{t-d})\theta$, $\varepsilon_t^{(i)} = y_t - z_t'(\gamma^{(i)}, c^{(i)}, y_{t-d})\theta$ and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

[*d*] Similar to Section 3.2.1.

4. CHOOSING THE MODEL ORDER *K*

We start by introducing the traditional information criteria, Akaike information criteria (AIC) (Akaike, 1974) and Bayesian information criteria (BIC) (Schwarz, 1978), widely used in model selection/comparison. To this toolbox we add the deviance information criterion, (DIC) i.e. a criterion recently developed and widely discussed in Spiegelhalter *et al.* (2002). In the second part of this section we adapted the RJMCMC, proposed by Green (1995), to LSTAR(k) models, where k is unknown. Standard MCMC algorithms cannot be used since the dimension of the parameter vector is also a parameter.

4.1. *Information Criteria*

Traditionally, model order or model specification are compared through the computation of information criteria, which are well known for penalizing likelihood functions of overparametrized models. AIC (Akaike, 1974) and BIC (Schwarz, 1978) are the most used ones. For data y and parameters θ, these criteria are defined as follows: $\text{AIC} = -2\ln(p(y|\hat{\theta})) + 2d$ and $\text{BIC} = -2\ln(p(y|\hat{\theta})) + d\ln n$, d is the dimension of θ, with $d = 2k + 5$ for LSTAR(k), sample size n and maximum-likelihood estimator, $\hat{\theta}$. One of the major problems with AIC/BIC is that defining k is not trivial, mainly in Bayesian hierarchical models, where the priors act like reducers of the effective

number of parameters through its interdependencies. To overcome this limitation, Spiegelhalter *et al.* (2002) developed an information criterion that properly defines the effective number parameter by $p_D = \bar{D} - D(\tilde{\theta})$, where $D(\theta) = -2 \ln p(y|\theta)$ is the deviance, $\tilde{\theta} = E(\theta|y)$ and $\bar{D} = E(D(\theta)|y)$. As a by-product, they proposed the DIC as $\text{DIC} = D(\tilde{\theta}) + 2p_D = \bar{D} + p_D$. One could argue that the most attractive and appealing feature of the DIC is that it combines model fit (measured by $\bar{D}$) with model complexity (measured by p_D). Besides, DICs are more attractive than Bayes factors since the former can be easily incorporated into MCMC routines. For successful implementations of DIC we refer to Zhu and Carlin (2000) (spatio-temporal hierarchical models) and Berg *et al.* (2004) (stochastic volatility models), to name a few.

4.2. *RJMCMC in LSTAR(k) models*

Proposals to incorporate k in a fully Bayesian modelling already exist in the literature. For example, Huerta and Lopes (2000) analysed the Brazilian Industrial Production through AR(k) model where $k = 2c + s$, i.e. c pairs of complex roots and s real roots, consider c and s, and therefore k unknown. Troughton and Godsill (1997) suggest an RJMCMC algorithm for AR(k). Our main contribution is to generalize the work of Troughton and Godsill (1997) by proposing a RJMCMC (Green, 1995) that permits the incorporation of k in the parameter vector for LSTAR(k, d) models.

In our LSTAR(k, d) model from eqn (1), θ is a $2(k + 1)$-dimensional parameter vector, whose dimension depends on k. Therefore, sampling k involves changing the dimensionality of θ. The idea behind RJMCMC is to propose a move from a k-dimensional parameter space to a k'-dimensional one, with the additional aspect that the vector $\phi = (\gamma, c, \sigma^2, d)$ has the same interpretation under both models, k and k'. If the current state of the Markov chain is (ϕ, k), then one cycle of the RJMCMC algorithm proceeds as follows. Specifically, a candidate model, k', is sampled from a proposal density, $q(\cdot|k)$, and conditional upon k', candidate parameter vector, $\theta^{(k')}$, is sampled from $p(\cdot|y, k', \phi)$, and accepted with probability $\alpha \equiv \alpha\{(k, \theta^{(k)}) \rightarrow (k', \theta^{(k')})\}$

$$\begin{aligned}\alpha &= \min\left\{1, \frac{p(k', \theta^{(k')}|y, \phi)}{p(k, \theta^{(k)}|y, \phi)} \frac{q(k|k')}{q(k'|k)} \frac{p(\theta^{(k)}|y, k, \phi)}{p(\theta^{(k')}|y, k', \phi)}\right\} \\ &= \min\left\{1, \frac{p(y|k, \phi)p(k|\phi)}{p(y|k', \phi)p(k'|\phi)} \frac{q(k|k')}{q(k'|k)}\right\},\end{aligned}$$

with the second equality following from the candidate's identity (Besag, 1989):

$$\frac{p(k, \theta^{(k)}|y, \phi)}{p(\theta^{(k)}|y, k, \phi)} = \frac{p(y|k, \phi)p(k|\phi)}{p(y|\phi)}.$$

In the particular case when $p(k'|\phi) = p(k|\phi)$, such as in uniformly discrete priors, the acceptance probability becomes

$$\alpha = \min\left\{1, \frac{p(y|k,\phi)}{p(y|k',\phi)}\frac{q(k|k')}{q(k'|k)}\right\}.$$

Conditional on ϕ, this step of the RJMCMC can be thought of as a Metropolis-step algorithm that moves according to the marginal-likelihood ratio times a correction factor, and will more often visit those models with higher marginal-likelihood, which is an appealing and desirable property. The remaining marginal likelihoods, $p(y|k,\ \phi)$, are easily derived once one realizes that, conditional on k and ϕ, the parameters of θ follows standard posterior normal distributions. More precisely, under the standard conditionally conjugate prior distribution (Section 3.2.1),

$$\begin{aligned} p(y|k,\phi) &\propto \int p(y|k,\phi,\theta)p(\theta|k,\phi)\mathrm{d}\theta \\ &\propto \sigma_{\theta^{(k)}}^{-(2k+2)}|C_{\theta^{(k)}}|^{\frac{1}{2}} \\ &\quad \times \exp\left\{-\frac{1}{2}\left(\sigma^{-2}y'y + \sigma_{\theta^{(k)}}^{-2}m'_{\theta^{(k)}}m_{\theta^{(k)}} - \tilde{m}^T_{\theta^{(k)}}C^{-1}_{\theta^{(k)}}\tilde{m}_{\theta^{(k)}}\right)\right\}, \end{aligned} \tag{8}$$

where $\tilde{m}_{\theta^{(k)}}$ and $C_{\theta^{(k)}}$ were defined by eqns (4) and (5). Similarly, under Lubrano's prior (Section 3.2.2),

$$\begin{aligned} p(y|k,\phi) \propto (2\pi)^{\frac{k+1}{2}}\left[\sigma^2\exp(\gamma)\right]^{-\frac{k+1}{2}}|C^*_{\theta^{(k)}}|^{\frac{1}{2}} \\ \times \exp\left\{-\frac{1}{2}(\sigma^{-2}y^Ty - \tilde{m}^{*'}_{\theta^{(k)}}{C^*_{\theta^{(k)}}}^{-1}\tilde{m}^*_{\theta^{(k)}})\right\} \end{aligned} \tag{9}$$

where $\tilde{m}^*_{\theta^{(k)}}$ and $C^*_{\theta^{(k)}}$ were defined by eqns (6) and (7). Finally, conditional on k and y, $\phi = (\gamma, c, \sigma^2, d)$ can be easily sampled from by using the MCMC steps presented in Section 3, completing, therefore, the MCMC cycle. The following chart schematically represents one cycle of our RJMCMC.

RJMCMC algorithm
Current state: $(k, \theta^{(k)}, \phi)$,
Candidate model: Sample model order k' from $q(\cdot|k)$,
Candidate parameters: Sample $\theta^{(k')}$ from $p(\theta^{(k')}|y, k', \phi)$,
Model move: new state $(\tilde{k}, \theta^{(\tilde{k})})$

$$(\tilde{k}, \theta^{(\tilde{k})}) = \begin{cases} (k', \theta^{(k')}) & \text{w.p. } \alpha\left\{(k,\theta^{(k)}) \to (k',\theta^{(k')})\right\} \\ (k, \theta^{(k)}) & \text{w.p. } 1-\alpha\left\{(k,\theta^{(k)}) \to (k',\theta^{(k')})\right\} \end{cases}$$

Other parameters: Sample ϕ from $p(\phi|\tilde{k}, \theta^{(\tilde{k})}, y)$,
Final state: $(\tilde{k}, \theta^{(\tilde{k})}, \phi)$

5. APPLICATIONS

Our methodology is extensively studied in this section against simulated and real-time series. We start with an extensive simulation study, which is followed by the analysis of two well-known dataset: (i) the Canadian Lynx series, which stands for the number of Canadian Lynx trapped in the Mackenzie River district of North-west Canada, and (ii) the USIPI Index, which stands for the US Industrial Production Index.

5.1. *A simulation study*

We performed a study by simulating 50 replicates of time series with 400 and 1000 observations from the following LSTAR(2) structure:

$$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})\pi(y_{t-2}) + \varepsilon_t,$$

where $\pi(y_{t-2}) = [1 + \exp\{-100(y_{t-2}-0.02)\}]^{-1}$ and $\varepsilon_t \sim N(0, 0.02^2)$. We used Lubrano's prior in this simulated exercise. The initial values were $k = 5$, $\phi_1 = \phi_2 = (0, 1, 1, 1, 1, 1)$, $\gamma = 150$, $c = \bar{y}$ and $\sigma^2 = 0.01^2$. We replicated this study 50 times for sample sizes $n = 400$ and $n = 1000$. Table I summarizes the experiment by counting number of times each one of the entertained models have the highest posterior model probability [model $\mathcal{M}_k$ is an LSTAR(k, 2) model].

For each one of the 100 simulations (50 for $n = 400$ and 50 for $n = 1000$), we consider 5000 MCMC runs with the first half used as burn-in. We found similar results when running longer chains. As expected, better results are found for the larger dataset. However, in both cases the right model is selected almost always. We randomly picked one of the simulated series with $n = 1000$ for further detailed illustrations. Figure 1 exhibits the time series with AIC, BIC and DIC in Table II. Posterior mean values and standard deviations for the model with highest posterior model probability are shown in Table III.

5.2. *Revisiting the Canadian Lynx Data*

We analyse a Canadian Lynx series (see Figure 2) that represents logarithm of the number of Canadian Lynx trapped in the Mackenzie River district of North-west Canada over the period from 1821 to 1934. Some previous analyses of this series can be found in Ozaki (1982), Tong (1990), Teräsvirta (1994), Medeiros and Veiga (2000) and Xia and Li (1999).

TABLE I

MONTE CARLO EXPERIMENT FOR 50 SIMULATED SERIES LOGISTIC SMOOTH TRANSITION AUTOREGRESSIVE (LSTAR) (2, 2) OF SIZE $N = 400$ AND $N = 1000$

	Models				
	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$
400	0	45	5	0	0
1000	0	49	1	0	0

$\mathcal{M}_k$ is an LSTAR(k, 2) model.

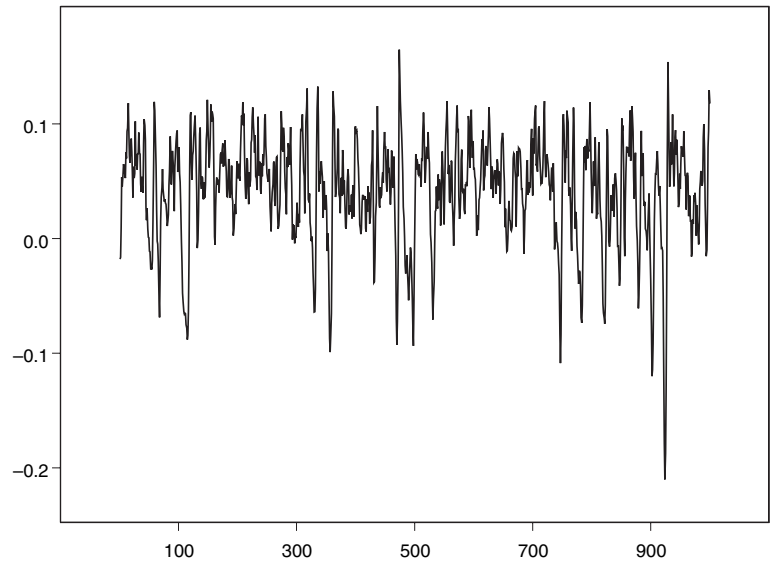

FIGURE 1. Simulated series of 1000 observations from logistic smooth transition autoregressive (2, 2).

TABLE II

MODEL COMPARISON USING INFORMATION CRITERIA (AKAIKE INFORMATION CRITERIA, BAYESIAN INFORMATION CRITERIA AND DEVIANCE INFORMATION CRITERIA) FOR THE SIMULATED LOGISTIC SMOOTH TRANSITION AUTOREGRESSIVE (2, 2) PROCESS

k	d	AIC	BIC	DIC
1	1	−4654.6	−4620.2	−8622.2
	2	−4728.6	−4694.3	−8731.2
	3	−4682.7	−4648.4	−8606.8
2	1	−3912.0	−3867.9	−7434.2
	2	−5038.9	−4994.8	−9023.6
	3	−4761.2	−4717.0	−8643.8
3	2	−5037.0	−4983.0	−9023.3
	3	−4850.4	−4796.5	−8645.9

TABLE III

POSTERIOR MEAN VALUES AND STANDARD DEVIATIONS FOR THE PARAMETERS OF THE LOGISTIC SMOOTH TRANSITION AUTOREGRESSIVE (2, 2) MODEL

Parameter	True value	Mean	SD
θ_{01}	0	−0.0028	0.0021
θ_{11}	1.8	1.7932	0.0525
θ_{21}	−1.06	−1.0809	0.0654
γ	100	100.87	4.9407
σ^2	0.0004	0.00037	0.000016
θ_{02}	0.02	0.023	0.0036
θ_{12}	−0.9	−0.8735	0.0637
θ_{22}	0.795	0.7861	0.0746
c	0.02	0.0169	0.0034

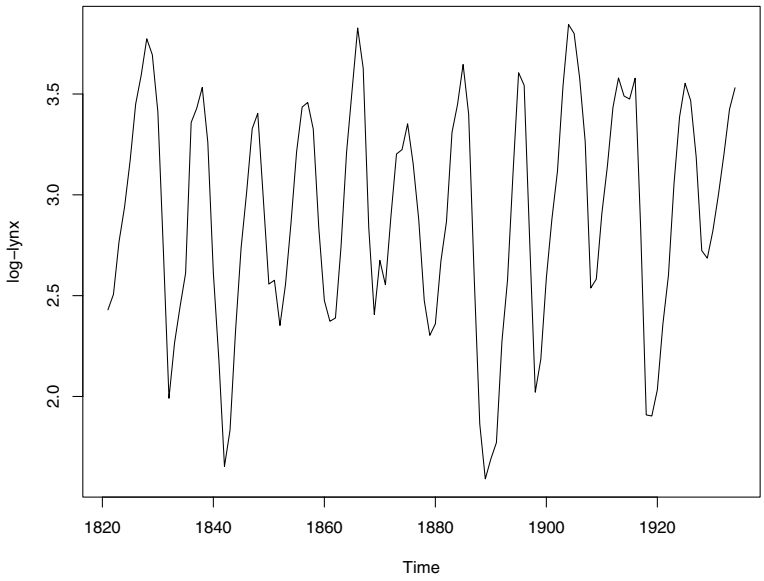

FIGURE 2. Série Canadian Lynx series.

We consider the objective priors and compute the posterior model probabilities using our RJMCMC algorithm. We run our RJMCMC algorithm for 50,000 iterations and discard the first half as burn-in. The LSTAR(11) model has the highest posterior model probability (see the histogram on Figure 3). Note that the density of k is bimodal, which is concordant with Medeiros and Veiga estimates,

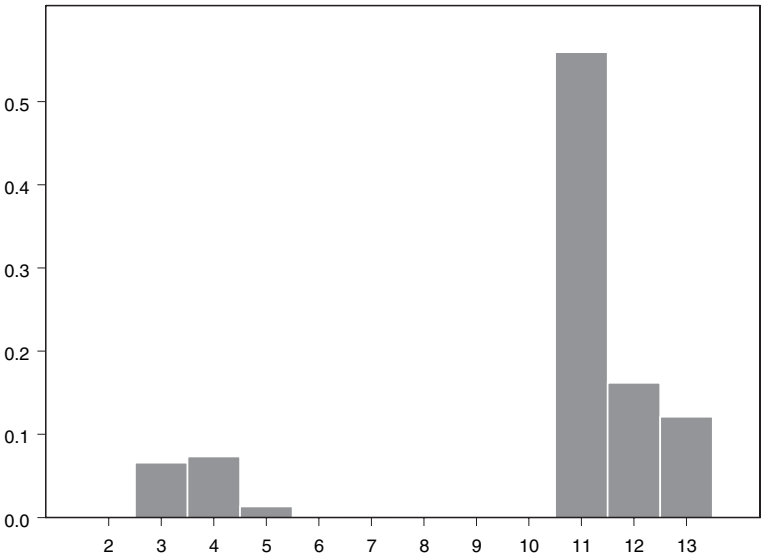

FIGURE 3. Monte Carlo estimate of the marginal posterior density $p(k|y)$ for Canadian Lynx series.

TABLE IV
RESULTS OF THE REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO ALGORITHM FOR CANADIAN LYNX SERIES

k	$d=1$	$d=2$	$d=3$	$d=4$	$p(k\|y)$
1	0	0	0	0	0
2	0	20	42	0	0.00124
3	6	18	3267	7	0.06596
4	0	8	3661	2	0.07342
5	0	0	658	4	0.01342
6	0	0	48	1	0.00098
7	0	0	74	6	0.00160
8	0	0	4	4	0.00016
9	0	0	34	1	0.00070
10	0	0	11	0	0.00022
11	4	0	27745	222	0.55942
12	0	0	8073	27	0.16200
13	41	0	5990	22	0.12106
$p(d\|y)$	0.00102	0.00092	0.99214	0.00592	1.00000

who fit an LSTAR(2,2). Teräsvirta (1994) estimates a LSTAR(11,3). The results of RJMCMC algorithm are shown in Table IV. The modal model is the one where $k = 11$ since $\Pr(k = 11|y) = 56\%$, but the AR(2), AR(3), AR(4), AR(12) and AR(13) also have non-negligible posterior model probabilities (6.6%, 7.3%, 13.4%, 16.2% and 12.1% respectively). We can observe that most two-step estimation schemes, the ones that usually estimates d and assume it is known when estimating k, will most certainly overestimate the uncertainty about k. Therefore, one can argue that our Bayesian estimation scheme fully accounts for all parameter uncertainty, including the uncertainty associated with estimating d and k. The specification of the highest posterior probability model, LSTAR(11, 3), is

$$
\begin{aligned}
\hat{y}_t = {} & \overset{(0.307)}{0.987} + \overset{(0.111)}{0.974} y_{t-1} - \overset{(0.151)}{0.098} y_{t-2} - \overset{(0.142)}{0.051} y_{t-3} - \overset{(0.137)}{0.155} y_{t-4} + \overset{(0.143)}{0.045} y_{t-5} \\
& - \overset{(0.146)}{0.0702} y_{t-6} - \overset{(0.158)}{0.036} y_{t-7} + \overset{(0.167)}{0.179} y_{t-8} + \overset{(0.159)}{0.025} y_{t-9} + \overset{(0.144)}{0.138} y_{t-10} \\
& - \overset{(0.096)}{0.288} y_{t-11} + (\overset{(2.04)}{-3.688} + \overset{(0.431)}{1.36} y_{t-1} - \overset{(0.744)}{3.05} y_{t-2} + \overset{(1.111)}{4.01} y_{t-3} \\
& - \overset{(0.972)}{2.001} y_{t-4} + \overset{(0.753)}{1.481} y_{t-5} + \overset{(0.657)}{0.406} y_{t-6} - \overset{(0.735)}{0.862} y_{t-7} - \overset{(0.684)}{0.666} y_{t-8} \\
& + \overset{(0.539)}{0.263} y_{t-9} + \overset{(0.486)}{0.537} y_{t-10} - \overset{(0.381)}{0.569} y_{t-11}) \\
& \times (1 + \exp\{-\overset{(0.688)}{11.625}(y_{t-3} - \overset{(0.017)}{3.504})\})^{-1}
\end{aligned}
$$

with $E(\sigma^2|y) = 0.025$ and with posterior standard deviations in parentheses. Figure 3 shows that our maximal model, the LSTAR(11, 3) captures the decennial seasonality exhibited by the data. Within each decade the transition cycles according to both extreme regimes as y_t gets away from $c = 3.504$, or the number of trapped lynx gets away from 3191, which happens around 15% of the time.

Finally, we compare our best model, LSTAR(11, 3), Teräsvirta's (1994) LSTAR(11, 2) model, and Tong's (1990, pages 377 and 387) TAR(2, 2) and TAR(7, 2) models:

- Teräsvirtas' (1994) LSTAR(11,2)

$$\begin{aligned}\hat{y}_t = {} & \overset{(0.035)}{1.17}y_{t-1} + (-\overset{(0.20)}{0.92}y_{t-2} + \overset{(0.31)}{1.00}y_{t-3} - \overset{(0.16)}{0.41}y_{t-4} \\ & + \overset{(0.055)}{0.27}y_{t-9} - \overset{(0.079)}{0.21}y_{t-11}) \\ & \times \left(1 + \exp\{-\overset{(0.65)}{1.73} \times 1.8(y_{t-3} - \overset{(0.13)}{2.73})\}\right)^{-1}\end{aligned}$$

- Tong's (1990, p. 377) TAR(2,2)

$$\hat{y}_t = \begin{cases} 0.62 + 1.25y_{t-1} - 0.43y_{t-2}, & y_{t-2} \leq 3.25, \\ 2.25 + 1.52y_{t-1} - 1.24y_{t-2}, & y_{t-2} > 3.25. \end{cases}$$

- Tong's (1990, p. 387) TAR(7,2)

$$\hat{y}_t = \begin{cases} 0.546 + 1.032y_{t-1} - 0.173y_{t-2} + 0.171y_{t-3} & \\ \quad -0.431y_{t-4} + 0.332y_{t-5} - 0.284y_{t-6} + 0.210y_{t-7}, & y_{t-2} \leq 3.116, \\ 2.632 + 1.492y_{t-1} - 1.324y_{t-2}, & y_{t-2} > 3.116. \end{cases}$$

Tables V and VI present, respectively, estimates of the standard deviations, σ^2, and mean absolute error (MAE) and root-mean square error (RMSE) for the four competing models. Besides being the model with the highest posterior model probability, our LSTAR(11, 3) model exhibits the lowest variability (smallest $\hat{\sigma}_\varepsilon$), and smallest MAE and RMSE (See Figure 4).

TABLE V

CANADIAN LYNX DATA: ESTIMATED RESIDUAL STANDARD DEVIATION

Authors	Model	$\hat{\sigma}_\varepsilon$
Lopes and Salazar	LSTAR(11, 3)	0.158
Teräsvirta (1994)	LSTAR(11, 2)	0.187
Tong (1990, p. 377)	TAR(2, 2)	0.220
Tong (1990, p. 387)	TAR(7, 2)	0.210

LSTAR, logistic smooth transition autoregressive; TAR, threshold autoregressive.

TABLE VI

CANADIAN LYNX DATA: MEAN ABSOLUTE ERROR (MAE) AND ROOT MEAN SQUARED ERROR (RMSE)

Authors	Model	MAE	RMSE
Lopes and Salazar	LSTAR(11, 3)	0.118	0.153
Teräsvirta (1994)	LSTAR(11, 2)	0.142	0.179
Tong (1990, p. 377)	TAR(2, 2)	0.160	0.206
Tong (1990, p. 387)	TAR(7, 2)	0.152	0.190

LSTAR, logistic smooth transition autoregressive; TAR, threshold autoregressive.

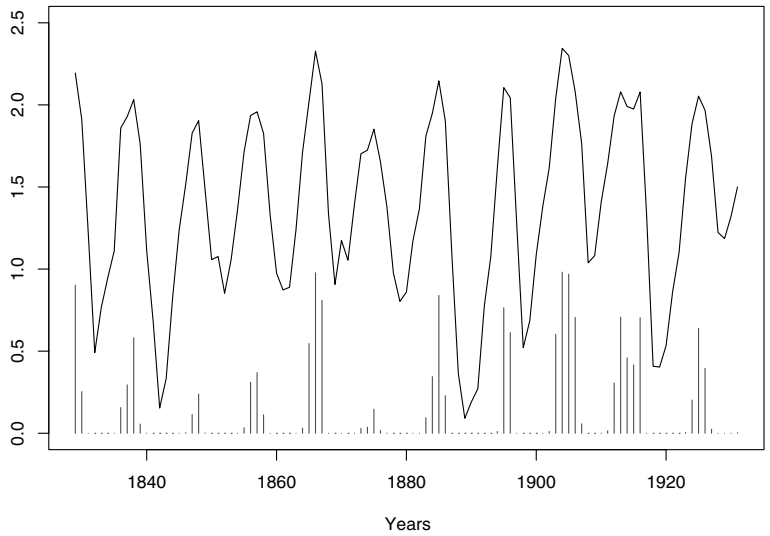

FIGURE 4. Posterior distribution of $\pi(y_{t-d}, \gamma, c)$ (vertical lines) for the Canadian Lynx.

5.3. *Analysing the US IPI*

We analysed the US IPI, which can be downloaded from the internet at http://www.economagic.com/sub-info/. The series comprises monthly observations from 1960 to 2001. Following previous analysis, we converted the original series to quarterly by averaging, after we took the fourth difference of the series so as to achieve stationarity and remove possible seasonality. The resulting series is shown in Figure 5.

We use the RJMCMC algorithm with Lubrano's priors and analyse several LSTAR models. We consider 50,000 runs and then we discarded the values from the first 25,000 iterations as burn-in. The histogram from Figure 6 and Table VII show that the LSTAR(5, 3) model has the highest posterior model probability, $\Pr(k = 5, d = 3|y) \approx 55\%$, and is specified as follows,

$$\begin{aligned} y_t = & \overset{(0.005)}{-0.007} + \overset{(0.182)}{0.909} y_{t-1} - \overset{(0.214)}{0.411} y_{t-2} + \overset{(0.247)}{0.51} y_{t-3} - \overset{(0.225)}{1.265} y_{t-4} \\ & + \overset{(0.185)}{0.541} y_{t-5} + (\overset{(0.005)}{0.008} + \overset{(0.199)}{0.555} y_{t-1} - \overset{(0.261)}{0.153} y_{t-2} \\ & - \overset{(0.297)}{0.412} y_{t-3} + \overset{(0.269)}{1.124} y_{t-4} - \overset{(0.199)}{0.488} y_{t-5}) \\ & \times \left(1 + \exp\{ - \overset{(16.424)}{960.58} (y_{t-3} + \overset{(0.0029)}{0.0076}) \} \right)^{-1} + \varepsilon_t, \quad E(\sigma^2|y) = 0.00019 \end{aligned}$$

with posterior standard deviations in parentheses.

We used the LSTAR(5, 3) to check the forecast performance from the first quarter of 2002 to the second quarter of 2004. However, as can be seen from

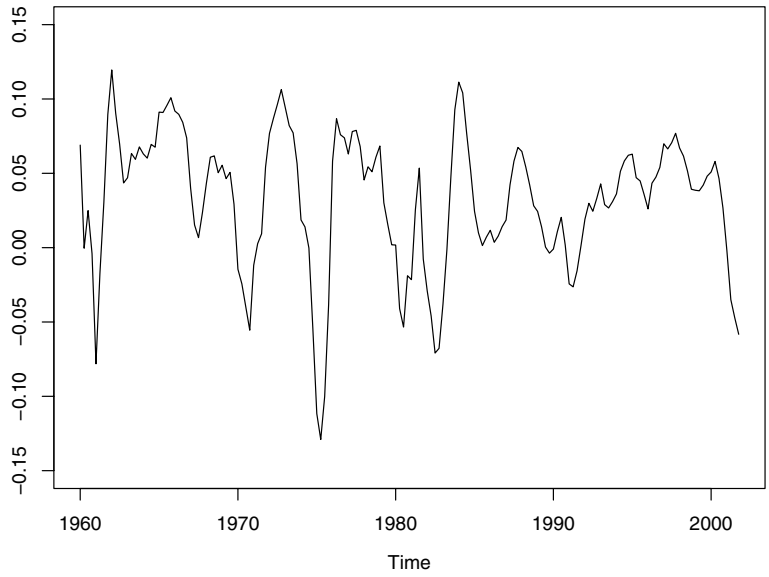

FIGURE 5. US Industrial Production Index.

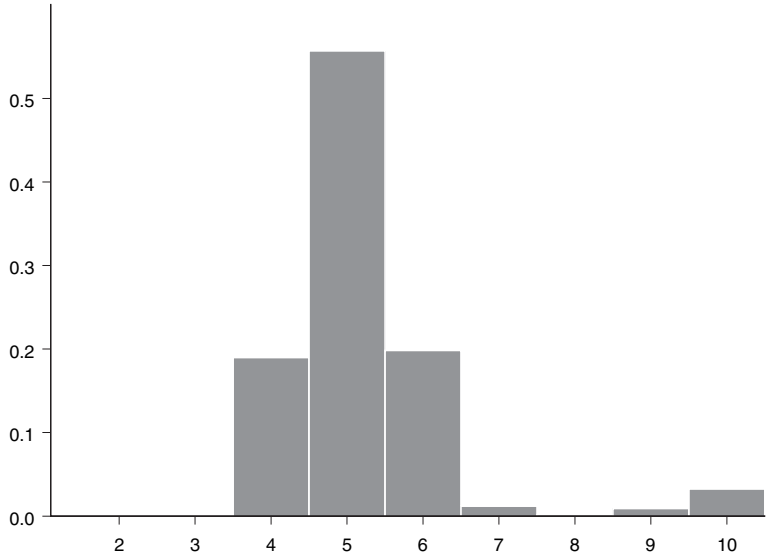

FIGURE 6. Monte Carlo estimate of the marginal posterior density $p(k|y)$ for US Industrial Production Index series.

Table VII, models LSTAR(4, 3) and LSTAR(6, 3) also have relatively high posterior model probability (18.7% and 19.1% respectively), so choosing only one model is not as obvious as it was with the Canadian Lynx data. As far as

TABLE VII

RESULTS OF THE REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO ALGORITHM FOR US INDUSTRIAL PRODUCTION INDEX SERIES

k	$d = 1$	$d = 2$	$d = 3$	$d = 4$	$p(k\|y)$
1	0	0	0	0	0.00000
2	0	0	1	2	0.00006
3	5	0	0	1	0.00012
4	134	1	9328	31	0.18988
5	81	7	27679	77	0.55688
6	246	3	9528	141	0.19836
7	8	1	531	50	0.01180
8	1	0	70	5	0.00152
9	19	0	422	7	0.00896
10	25	0	1591	5	0.03240
$p(d\|y)$	0.01038	0.00024	0.98300	0.00638	1.00000

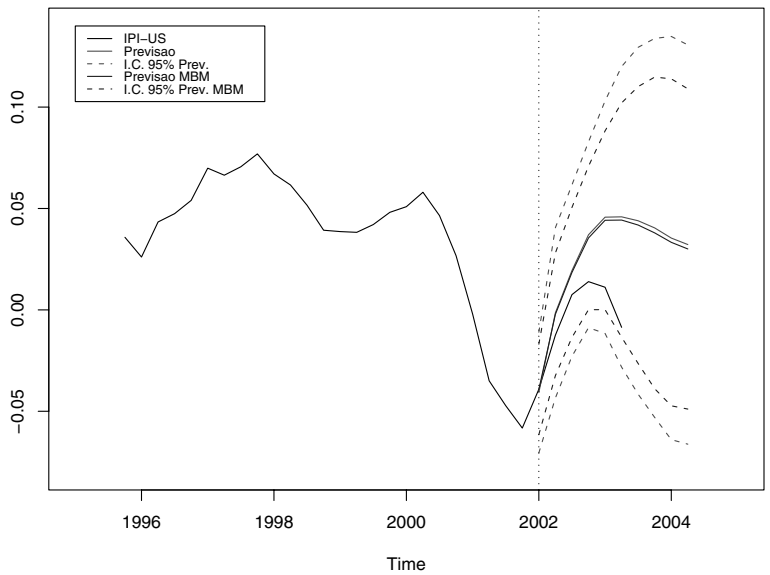

FIGURE 7. Forecast values from 2002–1 to 2004–2 for US-Industrial Production Index series using model (10) and Bayesian model averaging method. The vertical line denotes the start of forecasting.

forecasting is concerned, a naturally Bayesian strategy is to average forecasts from competing models with weights defined by their respective posterior model probabilities, a procedure that has become well known as Bayesian model averaging (BMA) (for more details see Raftery *et al.*, 1997; Clyde, 1999; Hoeting *et al.*, 1999). Figure 7 shows the forecasts for our maximal model LSTAR(5, 3) and from the BMA model. One could argue that the BMA produces more accurate forecasting for short periods, with a possible explanation for this improvement being that individual models are over-parameterized, so suffer the curse of over-fitting and yielding poor predictions with larger credible intervals.

6. CONCLUSIONS

In this paper, we develop MCMC and RJMCMC algorithms for posterior inference and model selection in the broad class of nonlinear time-series models known as LSTAR. Our developments are checked against simulated and real datasets with encouraging results in favour of our RJMCMC scheme.

Even though we concentrated our computations and examples to the logistic transition function, the algorithms we developed can be straightforwardly adapted to other transition functions such as the exponential, the second-order logistic functions, and with (virtually) any continuous function taking values in [0, 1]. Similarly, even though we chose to work with y_{t-d} as the transition variable and accounted for the uncertainty about d, our findings are easily extended to situations where y_{t-d} is replaced by, say, $s(y_{t-1}, \ldots, y_{t-d}, \alpha)$ for α and d unknown.

Our simulation study and applications presented encouraging results. On the one hand, our RJMCMC algorithm selected the correct model with high posterior model probability in the simulated study. On the other hand, by revisiting well-known datasets, such as the Canadian Lynx series and the US IPI, we showed that our estimation scheme outperforms those currently used. We have also showed, through the US IPI series, that mixture of model forecasts is a natural Bayesian tool and a straightforward by-product of our RJMCMC algorithm.

In this paper, we focused on modelling the level of nonlinear time-series. Our current research agenda includes adapting the methods we propose here to model both univariate and multivariate (factor) stochastic volatility problems (Lopes and Migon, 2002; Lopes *et al.*, 2000) and comparison of smooth transition autoregressions with Markov switching models (Carvalho and Lopes, 2004, Abanto-Valle *et al.*, 2004).

All the algorithms were programmed in the student's version of the software Ox, which is publicly available and downloadable from Jurgen Doornik's website at http://www.doornik.com/download_oxcons.html.

ACKNOWLEDGEMENTS

Hedibert Lopes thank the Graduate School of Business, University of Chicago and the Department of Statistical Methods, Federal University of Rio de Janeiro for supporting his research on this paper. Esther Salazar was partially supported by a scholarship from CNPq to pursue a Master Degree in Statistics from the Federal University of Rio de Janeiro.

REFERENCES

ABANTO-VALLE, C., MIGON H., and LOPES, H. F. (2004) Bayesian modelling of financial returns: a relationship between volatility and trading volume. In *Proceedings of the XIV Brazilian Meeting of Econometrics Paraíba Brazil*.

AKAIKE, H. (1974) New look at the statistical model identification. *IEEE Transactions in Automatic Control AC* 19, 716–23.

BACON, D. and WATTS, D. (1971) Estimating the transition between two intersecting straight lines. *Biometrika* 58, 525–34.

BERG, A., MEYER, R. and YU, J. (2004) DIC as a model comparison criterion for stochastic volatility models. *Journal of Business and Economic Statistics* 22, 107–20.

BESAG, J. (1989) A candidate's formula – a curious result in Bayesian prediction. *Biometrika* 76(1), 183.

CAMPBELL, E. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis* 25, 467–82.

CARVALHO, C. and LOPES, H. F. (2004) Simulation-based sequential analysis of Markov switching stochastic volatility models. Working Paper, Graduate School of Business, University of Chicago, Chicago, IL.

CHAN, K. S. and TONG, H. (1986) On estimating thresholds in autoregressive models. *Journal of Time Series Analysis* 7, 179–90.

CHEN C. (1998). A Bayesian analysis of generalized threshold autoregressive models. *Statistics and Probability Letters* 40, 15–22.

CHEN, C. and LEE, J. (1995) Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis* 16, 483–92.

CHEN, R. and TSAY, R. (1991) On the ergodicity of TAR (1) processes. *Annals of Applied Probability* 1, 613–34.

CLYDE M. (1999) Bayesian model averaging and model search strategies (with discussion). In *Bayesian Statistics 6* (eds J. BERNARDO, A. DAWID, J. BERGER and A. SMITH). Oxford: Oxford University Press, pp. 157–85.

VAN DIJK, D., TERÄSVIRTA T. and FRANSES, P. (2002) Smooth transition autoregressive models–a survey of recent developments. *Econometrics Reviews* 21, 1–47.

GAMERMAN D. (1997) *Markov Chain Monte Carlo: Stochastic simulation for Bayesian inference*. London: Chapman and Hall.

GELFAND, A. E. and SMITH, A. M. F. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.

GEWEKE, J., TERUI, N. (1993) Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14, 441–54.

GILKS, W. R., RICHARDSON, S. and SPIEGELHALTER, D. J. (eds.) (1996) *Markov Chain Monte Carlo in Practice* (3rd edn). London: Chapman and Hall.

GREEN, P. J. (1995) Reversible Jump Markov chain Monte Carlo computation Bayesian model determination. *Biometrika* 57(1), 97–109.

HAMILTON J. (1989) A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica* 57, 357–84.

HASTINGS, W. R. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109.

HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999) Bayesian model averaging: a tutorial. *Statistical Science* 14, 401–4.

HUERTA, G. and LOPES, H. (2000) Bayesian forecasting and inference in latent structure for the brazilian industrial production index. *Brazilian Review of Econometrics* 20, 1–26.

KOOP, G. and POTTER S. (1999) Bayes factor and nonlinearity: evidence from economic time series. *Journal of Econometrics* 88, 251–81.

KOOP, G. and POTTER, S. (2003) Bayesian analysis of endogenous delay threshold models. *Journal of Business and Economic Statistics* 21, 93–103.

LOPES, H. F. and MIGON H. (2002) Comovements and contagion in emergent markets: stock indexes volatilities. *Case Studies in Bayesian Statistics* 6, 285–300.

LOPES, H. F., AGUILAR, O. and WEST, M. (2000) Time-varying covariance structures in currency markets. In *Proceedings of the XXII Brazilian Meeting of Econometrics, Campinas, Brazil*.

LUBRANO, M. (2000) Bayesian analysis of nonlinear time-series models with a threshold. In *Nonlinear Econometric Modeling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory* (eds W. A. BARNETT, D. F. HENDRY, S. HYLLEBERG, T. TERASVIRTA, D. TJOSTHEIM and A. W. WURTZ). Cambridge University Press, 2000.

LUUKKONEN, R. SAIKKONEN, P. and TERÄSVIRTA, T. (1988) Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–9.

MEDEIROS, M. C. and VEIGA, A. (2000) A flexible coefficient smooth transition time series model. Working Paper Series in Economics and Finance 360, Stockholm School of Economics.

METROPOLIS, N., ROSENBLUTH, A. W., ROSENBLUTH, M. N., TELLER, A. H. and TELLER E. (1953) Equation of state calculations by fast computing machine. *Journal of Chemical Physics* 21, 1087–91.

OZAKI, T. (1982) The statistical analysis of perturbed limit cycle process using nonlinear time series models. *Journal of time-series Analysis* 3, 29–41.

PETRUCCELLI, J. and WOOLFORD, S. (1984) A threshold AR(1) model. *Journal of Applied Probability* 21, 270–86.

RAFTERY, A. E., MADIGAN, D. and HOETING, J. A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179–191.

ROBERT, C. and CASELLA, G. (1999) *Monte Carlo Statistical Methods*. New York: Springer-Verlag.

SCHWARZ, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–4.

SO, M. and CHEN, C. (2003) Subset threshold autoregression. *Journal of Forecasting* 22, 49–66.

SPIEGELHALTER, D., BEST, N., CARLIN, B. and VAN DER LINDE A. (2002) Bayesian measures of model complexity and fit (with discussion). *Journal of The Royal Statistical Society* 64, 583–639.

TERÄSVIRTA T. (1994) Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89(425), 208–18.

TONG, H. (1978) On a threshold model. In *Pattern Recognition and Signal Processing* (ed. C. CHENG). Amsterdam: Sijthoff and Noordhoff, pp. 101–41.

TONG, H. (1990) *Non-linear Time Series: A Dynamical Systems Approach, Volume 6 of Oxford Statistical Science Series*. Oxford: Oxford University Press.

TONG, H. and LIM K., (1980) Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Society* 42, 245–92.

TROUGHTON, P. and GODSILL, S. (1997) A reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical CUED/F-INFENG/TR.304, Department of Engineering, University of Cambrigde, Cambrigde.

TSAY, R. (1989) Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84, 431–52.

WONG, C. and LI, W. (1998) A note on the corrected Akaike information criterion for threshold autoregressive models. *Journal of Time Series Analysis* 19, 113–24.

XIA, Y. and LI, W. K. (1999) On single-index coefficient regression models. *Journal of the American Statistical Association* 94, 1275–85.

ZHU, L. and CARLIN, B. (2000) Comparing hierarquical models for spatio-temporally misaligned data using the deviance information criterion. *Statistics in Medicine* 19, 2265–78.