Time series mean level and stochastic volatility modeling by smooth transition autoregressions: A Bayesian approach[1]

Hedibert Freitas Lopes and Esther Salazar

Graduate School of Business, University of Chicago and

Instituto de Matemática, Universidade Federal do Rio de Janeiro

Abstract

In this paper we propose a Bayesian approach to model the level and the variance of (financial) time series by the special class of nonlinear time series models known as the logistic smooth transition autoregressive models, or simply the LSTAR models. We first propose a Markov Chain Monte Carlo (MCMC) algorithm for the levels of the time series and then adapt it to model the stochastic volatilities. The LSTAR order is selected by three information criteria: the well known AIC and BIC, and by the deviance information criteria, or DIC. We apply our algorithm to a synthetic data and two real time series, namely the canadian lynx data and the SP500 returns.

JEL-Codes: C11,C15,C22,C63.

Keywords: Markov Chain Monte Carlo, stochastic volatility, nonlinear time series model, Deviance information criterion, mixture models.

1 Introduction

Smooth transition autoregressions (STAR), initially proposed in its univariate form by Chan and Tong (1986) and further developed by Luukkonen, Saikkonen, and Teräsvirta (1988) and

[1] *Corresponding address*: Hedibert Freitas Lopes, Graduate School of Business, University of Chicago, 5807 Woodlawn Avenue, Chicago, IL 60637, USA. *E-mail*: `hlopes@ChicagoGSB.edu`.

Teräsvirta (1994), have been extensively studied over the last twenty years in association to the measurement, testing and forecasting of nonlinear financial time series. The STAR model can be seen as a continuous mixture of two AR(k) models, with the weighting function defining the degree of nonlinearity. In this article, we focus on an important subclass of the STAR model, the logistic STAR model of order k, or simply the LSTAR(k) model, where the weighting function has the form of a logistic function.

Dijk, Teräsvirta, and Franses (2002) make an extensive review of recent developments related to the STAR model and its variants. From the Bayesian point of view, Lubrano (2000) used weighted resampling schemes (Smith & Gelfand, 1992) to perform exact posterior inference of both linear and nonlinear parameters. Our main contribution is to propose an MCMC algorithm that allows easy and practical extensions of LSTAR models to modeling univariate and multivariate stochastic volatilities. Therefore, we devote Section 2 to introducing the general LSTAR(k) to model the levels of a time series. Prior specification, posterior inference and model choice are also discussed in this section. Section 3 adapts the MCMC algorithm presented in the previous section to modeling stochastic volatility. Finally, in Section 4, we apply our estimation procedures to a synthetic data and two real time series, namely the canadian lynx data and the SP500 returns, with final thoughts and perspectives listed in Section 5.

2 LSTAR: mean level

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,

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

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

$$F(\gamma, c, s_t) = \{1 + exp\,(-\gamma(s_t - c))\}^{-1} \tag{2}$$

Several other functions could be easily accommodated, such as the exponential or the second-order logistic function (Dijk et al., 2002). The parameter $\gamma > 0$ is responsible for the smoothness of F, while c is a location or *threshold* parameter. When $\gamma \to \infty$, the LSTAR model reduces to the well known self-exciting TAR (SETAR) model (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), and d a delay parameter. 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 the paper, without loss of generality, that $d \leq k$ and $y_{-k+1}, \ldots, y_0$ are known and fixed quantities. This is common practice in the time series literature and would only marginally increase the complexity of our computation and could easily be done by introducing a prior distribution for those missing values.

The LSTAR(k) model can be seen as 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,

$$y_t = (1 - F(\gamma, c, y_{t-d}))x_t'\theta_1 + F(\gamma, c, y_{t-d})x_t'\delta + \varepsilon_t \tag{3}$$

or,

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

Therefore, the model (3) can be expressed as (1) with $\theta_2 = \delta - \theta_1$ being the contribution to the regression of considering a second regime. Huerta, Jiang, and Tanner (2003) also address issues about nonlinear time series models through logistic functions. We consider the following parameterization that will be very useful in the computation of the posterior distributions:

$$y_t = z_t'\theta + \varepsilon_t \tag{4}$$

where $\theta' = (\theta_1', \theta_2')$ represent the linear parameters, (γ, c) are the nonlinear parameters, $\Theta = (\theta, \gamma, c, \sigma^2)$, and $z_t' = (x_t', F(\gamma, c, y_{t-d})x_t')$. The dependence of z_t on γ, c and y_{t-d} will be made explicit whenever necessary. The likelihood is then written as

$$p(\boldsymbol{y}|\Theta) \propto \sigma^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - z_t'\theta)^2\right\} \tag{5}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$.

2.1 Prior specification and posterior inference

In this section we derive MCMC algorithm for posterior assessment of both linear and nonlinear parameters and for the observations' variance, σ^2, so that the number of parameters to be estimated is $2k+5$. We also consider the estimation of the delay parameter d. We adopt Lubrano's (2000) formulation, ie. $(\theta_2|\sigma^2, \gamma) \sim N(0, \sigma^2 e^{\gamma} I_{k+1})$ 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 $c_a = \hat{\mathcal{F}}^{-1}(0.15)$, $c_b = \hat{\mathcal{F}}^{-1}(0.85)$, and $\hat{\mathcal{F}}$ the data's empirical cumulative distribution function. The prior density is then,

$$\pi(\Theta) \propto \sigma^{-3}(1+\gamma^2)^{-1} \exp\left\{-\frac{1}{2}[\gamma + \sigma^{-2}e^{-\gamma}\theta_2'\theta_2]\right\} \tag{6}$$

which combined with the likelihood (Equation 5) yields the posterior distribution

$$\pi(\Theta) \propto \frac{\sigma^{-(n+6)/2}}{(1+\gamma^2)} \exp\left\{-\frac{1}{2\sigma^2}[\gamma\sigma^2 + e^{-\gamma}\theta_2'\theta_2 + \sum_{t=1}^{n}(y_t - z_t'\theta)^2]\right\} \tag{7}$$

with Θ in $\mathcal{A} = \{\Re^{2k+2} \times \Re^{+} \times \Re \times \Re^{+}\}$.

Needless to mention that the posterior distribution has no known form and that inference is facilitated by Markov Chain Monte Carlo methods (Gilks, Richardson, and Spiegelhalter (1996)). The full conditional posterior of θ and σ^2 are respectively, normal and inverse gamma, in which case the application of Gibbs steps are straigthforward (Gelfand & Smith, 1990). However, the full conditional distributions of γ and c are of unknown form, so we use Metropolis-Hastings steps (Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), Hastings (1970)). For more details about Gibbs and Metropolis-Hastings algorithms see Gilks, Richardson, and Spiegelhalter (1996) and the books by Gamerman (1997) and Robert and Casella (1999).

2.2 Choosing the model order k

In this section we started introducing the traditional information criteria, AIC (Akaike, 1974) and BIC (Schwarz, 1978), widely used in model selection/comparison. To this toolbox we add the DIC (Deviance Information Criterion) that is a criterion recently developed and widely discussed in Spiegelhalter, Best, Carlin, and Linde (2002).

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 parameter θ, these criteria are defined as follows: $AIC = -2\ln(p(\boldsymbol{y}|\hat{\theta})) + 2p$ and $BIC = -2\ln(p(\boldsymbol{y}|\hat{\theta})) + p\ln n$, p is the dimension of θ, with $p = 2k+5$ for LSTAR(k), sample size n and maximum likelihood estimator, $\hat{\theta}$. One of the major problems with AIC/BIC is that define 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, Best, Carlin, and Linde (2002) developed an information criterion that properly defines the effective number parameter by $p_D = \overline{D} - D(\widetilde{\theta})$, where

$D(\theta) = -2 \ln p(\boldsymbol{y}|\theta)$ is the *deviance*, $\widetilde{\theta} = E(\theta|\boldsymbol{y})$ and $\overline{D} = E(D(\theta)|y)$. As a by-product, they proposed the *Deviance Information Criterion*: $DIC = D(\widetilde{\theta}) + 2p_D = \overline{D} + p_D$. One could argue that the most attractive and appealing feature of the DIC is that it combines model fit (measured by $\overline{D}$) 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 implemenations of DIC we refer to Zhu and Carlin (2000) (spatio-temporal hierarchical models) and Berg, Meyer, and Yu (2004) (stochastic volatility models), to name a few.

3 LSTAR: stochastic volatility

In this section we adopt the LSTAR structure introduced in Section 2 to model time-varying variances, or simply the *stochastic volatility*, of (financial) time series. It has become standard to assume that y_t, the observed value of a (financial) time series at time t, for $t = 1, \ldots, n$, is normally distributed conditional on the unobservable volatility h_t, ie. $y_t|h_t \sim N(0, e^{h_t})$, and that the log-volatilities, $\lambda_t = \log h_t$, follow an autoregressive process of order one, ie. $\lambda_t \sim N(\theta_0 + \theta_1 \lambda_{t-1}; \sigma^2)$, with θ_0, θ_1 and σ^2 interpreted as the volatility's level, persistence and volatility, respectively. Several variations and generalizations based on this AR(1) structure were proposed over the last two decades to accommodate asymmetry, heavy tails, strong persistence and other features claimed to be presented in the volatility of financial time series (Kim, Shephard, & Chib, 1998).

As mentioned earlier, our contribution in this section is to allow the stochastic volatility to evolve according to the LSTAR(k) model, Equation 1 from Section 2,

$$\lambda_t = x_t'\theta_1 + F(\gamma, c, \lambda_{t-d})x_t'\theta_2 + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2) \tag{8}$$

with $x_t = (1, \lambda_{t-1}, \ldots, \lambda_{t-k})$, θ_1 and θ_2 $(k+1)$-dimensional vectors and $F(\gamma, c, \lambda_{t-d})$ the logistic function introduced in Equation 2. We assume, without loss of generality, that d

is always less than or equal to k. In the particular case when $k = 1$, the AR(1) model is replaced by a LSTAR(1),

$$\lambda_t \sim N\{\theta_{10} + \theta_{11}\lambda_{t-1} + [\theta_{20} + \theta_{21}\lambda_{t-1}]F(\gamma, c, \lambda_{t-d}); \sigma^2\}$$

Conditional on the vector of log-volatilities, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, we adopt, for the static parameters describing the LSTAR model, the same prior structure presented in Section 2.1. In other words, $(\theta_2|\sigma^2, \gamma) \sim N(0, \sigma^2 e^{\gamma} I_{k+1})$ 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 $c \in [\hat{\mathcal{F}}^{-1}(0.15), \hat{\mathcal{F}}^{-1}(0.85)]$, with $\hat{\mathcal{F}}$ an approximation to the log-volatilities' empirical cdf. Therefore, conditional on $\boldsymbol{\lambda}$, sampling $\theta_1, \theta_2, c, \gamma$ and σ^2 follows directly from the derivations of Section 2.1. Posterior inference to this stochastic volatility LSTAR(k), or simply SV-LSTAR(k), follows closely the developments of Section 2.1 and the Appendix, with the additional burden of sampling the hidden vector of log-stochastic volatilities, $\boldsymbol{\lambda}$. We use the single parameter move introduced by Jacquier, Polson, and Rossi (1994). Specifically, we rely on the fact the full conditional distribution of λ_t given the parameters in Θ plus $\boldsymbol{\lambda}_{-t} = \{\lambda_1, \ldots, \lambda_{t-1}, \lambda_{t+1}, \ldots, \lambda_n\}$ is given by,

$$p(\lambda_t|\boldsymbol{\lambda}_{-t}, \Theta, \boldsymbol{y}) \propto g(\lambda_t|\boldsymbol{\lambda}_{-t}, \Theta, \boldsymbol{y}) \equiv p(y_t|\lambda_t)\prod_{i=0}^{k} p(\lambda_{t+i}|\lambda_{t+i-1}, \ldots, \lambda_{t+i-k}, \Theta) \qquad (9)$$

which has no known closed form. We sample λ_t^* from $N(\lambda_t^{(j)}, \Delta)$, for Δ a problem-specific tuning parameter, and $\lambda_t^{(j)}$ the current value of λ_t in this Random-Walk Markov chain. The draw λ_t^* is accepted with probability $\alpha = \min\left\{1, \frac{g(\lambda_t^*|\boldsymbol{\lambda}_{-t},\Theta,\boldsymbol{y})}{g(\lambda_t^{(j)}|\boldsymbol{\lambda}_{-t},\Theta,\boldsymbol{y})}\right\}$. We do not claim that our proposal is optimal or efficient by any formal means, but we argue that they are useful from a practical viewpoint as our simulated and real data applications reveal.

4 Applications

In this Section our methodology is extensively studied against simulated and real time series. We start with an extensive simulation study, which if 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 USPI Index, which stands for the US Industrial Production Index.

4.1 A simulation study

We performed a study by simulating a time series with 1000 observations from the following the LSTAR(2),

$$y_t = 1.8y_{t-1} - 1.06y_{t-2} + (0.02 - 0.9y_{t-1} + 0.795y_{t-2})F(100, 0.02, y_{t-2}) + \varepsilon_t \qquad (10)$$

where $F(100, 0.02, y_{t-2}) = [1 + exp\{-100(y_{t-2} - 0.02)\}]^{-1}$ and $\varepsilon_t \sim N(0, 0.02^2)$. The initial values were $k = 5, \theta_1 = \theta_2 = (0, 1, 1, 1, 1, 1)$, $\gamma = 150$, $c = \overline{y}$ and $\sigma^2 = 0.01^2$. We consider 5000 MCMC runs with the first half used as burn-in. AIC, BIC and DIC statistics are presented in Table (1).

Table 1 about here

Posterior means and standard deviations for the model with highest posterior model probability are shown in Table (2). The three information criteria point to the correct model, while our MCMC algorithm produces fairly accurate estimates of posterior quantities.

Table 2 about here

4.2 Canadian Lynx

We analyze the well known *Canadian Lynx* series. Figure 1 shows the logarithm of the number of Canadian Lynx trapped in the Mackenzie River district of North-west Canada over the period from 1821 to 1934 (data can be found in Tong (1990), page 470). For further details and previous analysis of this time series, see Ozaki (1982), Tong (1990), Teräsvirta (1994), Medeiros and Veiga (2005), and Xia and Li (1999), among others.

Figure 1 about here

We run our MCMC algorithm for 50000 iterations and discard the first half as burn-in. The LSTAR(11) model with $d = 3$ has the highest posterior model probability when using the BIC as a proxy to Bayes factor,

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

with posterior standard deviations in parenthesis. The LSTAR($k = 11$) captures the decennial seasonality exhibit by the data. The vertical lines on part (a) of Figure 1 are estimates of the logistic transition function, $F(\gamma, c, y_{t-3})$ from Equation 2, for $\gamma = -11.625$ and $c = 3.504$ posterior means of γ and c, respectively. Roughly speaking, 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. Part (b) of Figure 1 exhibit the transition function as a function of $y - 3.5$, which is relatively symmetric around zero corroborating with the fact that the decennial cycles are symmetric as well. Similar results were found in Medeiros and Veiga (2005) who fit an LSTAR(2) with $d = 2$ and Teräsvirta (1994) who fit an LSTAR(11) with $d = 3$ for the same time series.

4.3 S&P500 index

Here, we fit the LSTAR(k)-stochastic volatility model proposed in Section 3 to the North American Standard and Poors 500 index, or simply the SP500 index, which was observed from January 7th, 1986 to December 31st, 1997, a total of 3127 observations. We entertained six models for the stochastic volatility and did model comparison by using the three information criteria presented in Section 2.2, ie. AIC, BIC and DIC. The six models are: $\mathcal{M}_1 : AR(1)$, $\mathcal{M}_2 : AR(2)$, $\mathcal{M}_3 : LSTAR(1)$ with $d = 1$, $\mathcal{M}_4 : LSTAR(1)$ with $d = 2$, $\mathcal{M}_5 : LSTAR(2)$ with $d = 1$, and $\mathcal{M}_6 : LSTAR(2)$ with $d = 2$.

Table 3 present the performance of the six models, in which all three information criteria agree upon the choice of the *best* model for the log-volatiltiies: $LSTAR(1)$ with $d = 1$. One can argue that the linear relationship prescribed by an AR(1) structure is insufficient to capture the dynamic behavior of the log-volatilities. The LSTAR structure brings more flexibility to the modeling. Table 4 present the posterior mean and standard devations of all parameters for each one of the six models listed above.

5 Conclusions

In this paper we develop Markov chain Monte Carlo (MCMC) algorithms for posterior inference in a broad class of nonlinear time series models known as logistic smooth transition autoregressions, LSTAR. Our developments are checked against simulated and real dataset with encouraging results when modeling both the levels and the variance of univariate time series with LSTAR structures. We used standard information criteria such as the AIC and

BIC to compare models and introduced a fairly new Bayesian criterion, the Deviance Information Criteria (DIC), all of which point to the same direction in the examples we explored in this paper.

Even though we concentrated our computations and examples to the logistic transition function, the algorithms we developed can be easily adapted to other functions such as the exponential or the second-order logistic functions, or even combinations of those. Similarly, even though we chose to work with y_{t-d} as the transition variable, our findings are naturally 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 quantities.

In this paper we focused on modeling the level and the variance of univariate time series. Our current research agenda includes (i) the generalization of the methods proposed here to model factor stochastic volatility problems (Lopes and Migon (2002), Lopes, Aguilar, and West (2000)), (ii) fully treatment of k as another parameter and posterior inference based on reversible jump Markov Chain Monte Carlo (RJMCMC) algorithms for both the levels and the stochastic volatility of (financial) time series (Lopes & Salazar, 2005), and (iii) comparing smooth transition regressions with alternative jump models, such as Markov switching models (Carvalho & Lopes, 2002).

All our computer code are available upon request and were fully programmed in the student's version of Ox, a statistical language that can be downloadable free of charge from `http://www.nuff.ox.ac.uk/Users/Doornik/index.html`.

Acknowledgement

Hedibert Lopes would like to 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, *Conselho Nacional de Desenvolvimento Científico e Tecnológico*, to pursue her Masters Degree in Statistics at the Federal University of Rio de Janeiro. We thank the comments of the participants of III Annual Advances in Econometrics Conference (USA), Science of Modeling and Symposium on Recent Developments in Latent Variables Modelling (Japan), II Congreso Bayesiano de America Latina (Mexico), and VII Brazilian Meeting of Bayesian Statistics, and the participants of talks delivered at the Catholic University of Rio de Janeiro and Federal University of Paraná (Brazil), and Northern Illinois University (USA). We also would like to thank the invaluable comments made by the referee.

Appendix: LSTAR full conditional distributions

In this appendix we provide in detail the steps of our MCMC algorithm for the LSTAR models of Section 2. In what follows, $[\xi]$ denotes the full conditional distribution of ξ conditional on the other parameters and the data. Also, $\sum$ is short for $\sum_{t=1}^{n}$. The full conditional distributions are:

$[\theta]$ Combining Equation 4, $y_t \sim N(z_t'\theta, \sigma^2)$, with $(\theta_2|\sigma^2, \gamma) \sim N(0, \sigma^2 e^{\gamma} I_{k+1})$, it is easy to see that $[\theta] \sim N(\widetilde{m}_\theta^*, C_\theta^*)$ where $C_\theta^* = (\sum z_t z_t' \sigma^{-2} + \Sigma^{-1})$, $\widetilde{m}_\theta^* = \Sigma_\theta \left(\sum z_t y_t \sigma^{-2}\right)$ and, $\Sigma^{-1} = \text{diag}(0, \sigma^{-2} e^{-\gamma} I_{k+1})$.

$[\sigma^2]$ From Equation 4, $\varepsilon_t = y_t - z_t'\theta \sim N(0, \sigma^2)$, which combined with the non-informative prior $\pi(\sigma^2) \propto \sigma^{-2}$, leads to $[\sigma^2] \sim IG((T + k + 1)/2, (e^{\gamma}\theta_2'\theta_2 + \sum \varepsilon_t^2)/2)$.

$[\gamma, c]$ We sample γ^* and c^*, respectively from $G[(\gamma^{(i)})^2/\Delta_\gamma, \gamma^{(i)}/\Delta_\gamma]$ and $TN(c^{(i)}, \Delta_c)$, a normal 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

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

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

References

Akaike, H., 1974. New look at the statistical model identification. IEEE Transactions in Automatic Control AC–19, 716–723.

Berg, A., Meyer, R., Yu, J., 2004. DIC as a model comparison criterion for stochastic volatility models. Journal of Business and Economic Statistics 22, 107–120.

Carvalho, C., Lopes, H. F., 2002. Simulation-based sequential analysis of markov switching stochastic volatility models (Technical Report). Department of Statistical Methods, Federal University of Rio de Janeiro.

Chan, K. S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7, 179–190.

Dijk, D. van, Teräsvirta, T., 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. Chapman & Hall, London.

Gelfand, A. E., Smith, A. M. F., 1990. Sampling-based approaches to calculating marginal densities. Journal of the American Statiscal Asociation 85, 398–409.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. Markov Chain Monte Carlo in Practice. Chapman & Hall, London.

Hastings, W. R., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 97–109.

Huerta, G., Jiang, W., Tanner, M., 2003. Time series modeling via hierarchical mixtures. Statistica Sinica 13, 1097–1118.

Jacquier, E., Polson, N., Rossi, P., 1994. Bayesian analysis of stochastic volatility models (with discussion). Journal Bussines of Econometrics Statistics 12, 371–417.

Kim, S. N., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Lopes, H.F., Aguilar, O., West, M., 2000. Time-varying covariance structures in currency markets. In Proceedings of the XXII Brazilian Meeting of Econometrics, Campinas, Brazil.

Lopes, H.F., Migon, H., 2002. Comovements and contagion in emergent markets: stock indexes volatilities, in: Gatsonis, C., Carriquiry, A., Gelman, A., Higdon, D., Kass, R.E., Pauler, D., Verdinelli, I. (Eds.), Case Studies in Bayesian Statistics 6, 285–300. Springer-Verlag, New York.

Lopes, H.F., Salazar, E., 2005. Bayesian inference for smooth transition autoregressive time series models. Journal of Time Series Analysis (to appear).

Lubrano, M., 2000. Bayesian analysis of nonlinear time series models with a threshold, in: Barnett, W.A., Hendry, D.F., Hylleberg, S., Teräsvirta, T., Tjostheim, D., Würtz, A. (Eds.), Proceedings of the Eleventh International Symposium in Economic Theory. Cambridge University Press, Helsinki.

Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.

Medeiros, M. C., Veiga, A., 2005. A flexible coefficient smooth transition time series model. IEEE Transactions on Neural Networks 16, 97–113.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E., 1953. Equation of state calculations by fast computing machine. Journal of Chemical Physics

21, 1087–1091.

Ozaki, T., 1982. The statistical analysis of perturbed limit cycle process using nonlinear time series models. Journal of Time Series Analysis 3, 29–41.

Robert, C., Casella, G., 1999. Monte Carlo statistical methods. Springer-Verlag, New York.

Schwarz, G., 1978. Estimating the dimension of a model. Annals of Statistics 6, 461–464.

Smith, A., Gelfand, A., 1992. Bayesian statistics without tears. American Statistician 46, 84–88.

Spiegelhalter, D., Best, N., Carlin, B., 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, 208–218.

Tong, H., 1990. Non-linear time series: A dynamical systems approach. Oxford University Press, Oxford.

Xia, Y., Li, W. K., 1999. On single-index coefficient regression models. Journal of the American Statistical Association 94, 1275–1285.

Zhu, L., Carlin, B., 2000. Comparing hierarquical models for spatio-temporally misaligned data using the deviance information criterion. Statistics in Medicine 19, 2265–2278.

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 1: Model comparison using information criteria for the simulated LSTAR(2) process.

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

Table 2: Posterior means for the parameters by LSTAR(2) model using 2500 MCMC runs.

Models	AIC	BIC	DIC
$\mathcal{M}_1 : AR(1)$	12795	31697	7223.1
$\mathcal{M}_2 : AR(2)$	12624	31532	7149.2
$\mathcal{M}_3 : LSTAR(1, d = 1)$	**12240**	**31165**	**7101.1**
$\mathcal{M}_4 : LSTAR(1, d = 2)$	12244	31170	7150.3
$\mathcal{M}_5 : LSTAR(2, d = 1)$	12569	31507	7102.4
$\mathcal{M}_6 : LSTAR(2, d = 2)$	12732	31670	7159.4

Table 3: S&P500: Model comparison based on information criteria: AIC (Akaike information criterion) , BIC (Bayesian information criterion) and DIC (Deviance information criterion).

	Models					
Parameter	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$
	Posterior mean (standard deviation)					
θ_{01}	-0.060 (0.184)	-0.066 (0.241)	0.292 (0.579)	-0.354 (0.126)	-4.842 (0.802)	-6.081 (1.282)
θ_{11}	0.904 (0.185)	0.184 (0.242)	0.306 (0.263)	0.572 (0.135)	-0.713 (0.306)	-0.940 (0.699)
θ_{21}	-	0.715 (0.248)	-	-	-1.018 (0.118)	-1.099 (0.336)
θ_{02}	-	-	-0.685 (0.593)	0.133 (0.092)	4.783 (0.801)	6.036 (1.283)
θ_{12}	-	-	0.794 (0.257)	0.237 (0.086)	0.913 (0.314)	1.091 (0.706)
θ_{22}	-	-	-	-	1.748 (0.114)	1.892 (0.356)
γ	-	-	118.18 (16.924)	163.54 (23.912)	132.60 (10.147)	189.51 (0.000)
c	-	-	-1.589 (0.022)	0.022 (0.280)	-2.060 (0.046)	-2.125 (0.000)
σ^2	0.135 (0.020)	0.234 (0.044)	0.316 (0.066)	0.552 (0.218)	0.214 (0.035)	0.166 (0.026)

Table 4: S&P500: Posterior means and posterior standard deviations for the parameters from all six entertained models: $\mathcal{M}_1 : AR(1)$, $\mathcal{M}_2 : AR(2)$, $\mathcal{M}_3 : LSTAR(1)$ with $d = 1$, $\mathcal{M}_4 : LSTAR(1)$ with $d = 2$, $\mathcal{M}_5 : LSTAR(2)$ with $d = 1$, and $\mathcal{M}_6 : LSTAR(2)$ with $d = 2$.

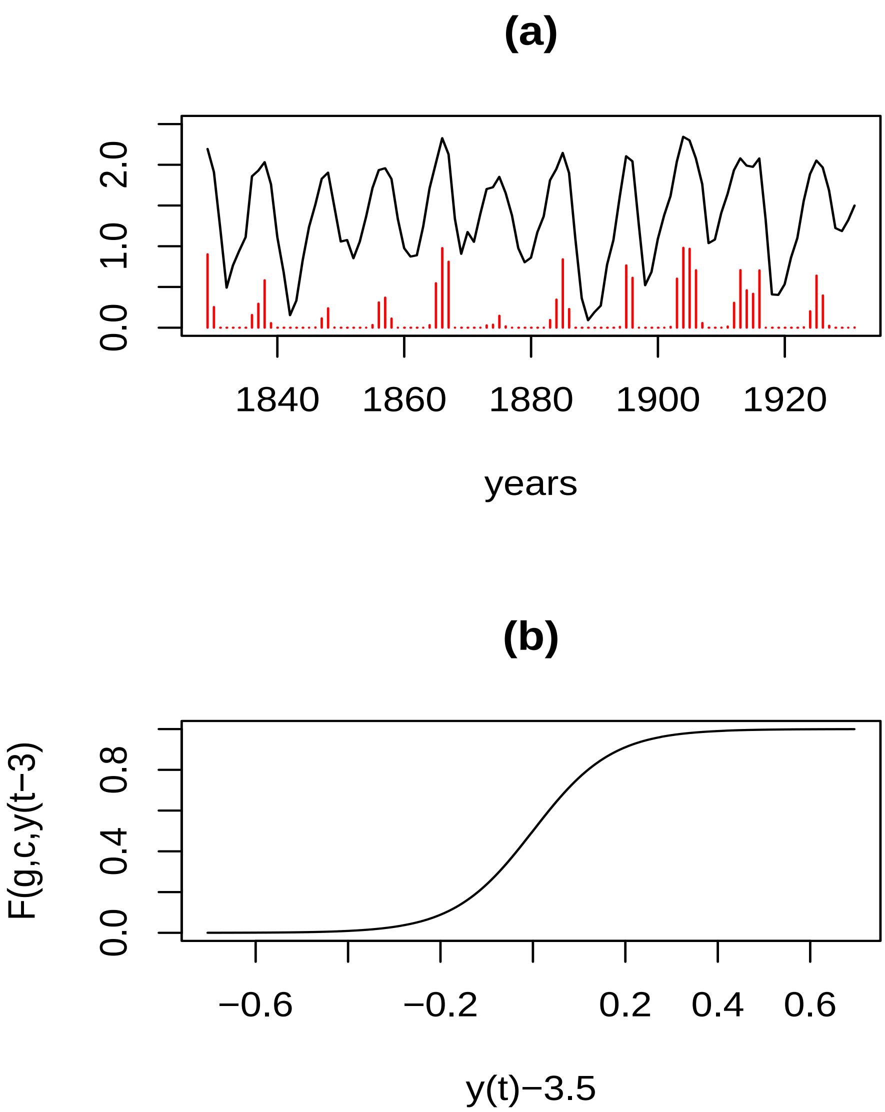

Figure 1: *Canadian Lynx series*: (a) Logarithm of the number of Canadian Lynx trapped in the Mackenzie River district of North-west Canada over the period from 1821 to 1934 (time series is shifted 1.5 units down to accommodate the transition function). The vertical lines are estimates of the logistic transition function, $F(\gamma, c, y_{t-3})$ from Equation 2, for $\gamma = -11.625$ and $c = 3.504$ posterior means of γ and c, respectively. (b) The posterior mean of the transition function as a function of $y - 3.5$.