Dynamic Bayesian smooth transition autoregressive models

Alexandre J. Santos
The Open University, UK

Alvaro E. Faria
The Open University, UK

March 28, 2013

Abstract

In this paper we propose the Gaussian Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models for nonlinear autoregressive time series processes as alternative to both the classical Smooth Transition Autoregressive (STAR) models of Chan and Tong (1986) and the computational Bayesian STAR (CBSTAR) models of Lopes and Salazar (2005). The DBSTAR models are autoregressive formulations of dynamic linear models by West and Harrison (1997) based on polynomial approximations of transition functions of STAR models. Unlike the classical STAR and CBSTAR models, their estimate parameters vary in time, being suitable for modelling non-stationary processes. For being Bayesian, the DBSTAR models do not require extensive historical data for parameter estimation and allow expert intervention via prior distribution assessment of model parameters. For being sequential and analytical, the DBSTAR models avoid potential computational problems associated with the CBSTAR models and allow fast estimation of the dynamic parameters sequentially in time, being thus suitable for real time applications. The application of DBSTAR models to the Canadian Lynx data showed improved fitting performances when compared with both the classical and the CBSTAR models.

Keywords: STAR models, Dynamic Linear Models, Nonlinear autoregressive models, Nonlinear autoregressive time series, Forecasting, Bayesian autoregressive models

1 Introduction

Recently, many statistical models designed to analyse and forecast nonlinear autoregressive time series processes have been proposed in the literature. Moran (1953) was the first to detect limitations of linear models. He drew attention to residuals that for the sample points greater than the mean were significantly smaller than those for the sample points smaller than the mean. Motivated by some economic studies, Quandt (1958) suggested a linear regression method suitable for processes characterised by having two distinct regimes that attributes a different linear regression model to each of the regimes. The determination of two regression lines for a single set of data would also shed additional light on potential nonlinearities of the system. The

Draft: please do not quote

Address for correspondence: Alexandre J. Santos, Department of Mathematics and Statistics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, United Kingdom. E-mail: alexandre.santos@open.ac.uk

possibility of modelling nonlinear Autoregressive (AR) processes with two regimes by considering a model that can detect change and switch between regimes, the so-called regime switching models, was well-investigated by Tong (1978). He proposed the *Threshold Autoregressive* (TAR) models that use an indicator function as the transition function to abruptly switch from one AR model to another. Tong (1990) extended the TAR models by treating lagged endogenous variable as transition variable and thus defining the *Self-Exciting Threshold Autoregressive* (SETAR) models. The TAR and SETAR are piecewise linear and can therefore have their parameters estimated by ordinary least square. One problem with those two models is that their likelihood functions present discontinuities that affect statistical inference due to the abrupt switching between regimes (Terasvirta, 1994).

A generalisation of the TAR and SETAR models incorporating a smooth transition between the two regimes via a pre-defined transition function was proposed by Chan and Tong (1986) and called *Smooth Transition Autoregressive* (STAR) models. One of the main advantages in using models with smooth transition is the possibility of specifying a gradual transition from one regime to the other in order to avoid an abrupt switching between them. However, STAR models require large and stationary data sets for parameter estimation. In addition, those models require that parameters and observational variance be constant.

In this paper we refer to Chan and Tong's models as classical STAR models. In the following subsections we start introducing the notation we will use throughout this paper by reviewing in more detail both the classical STAR models and the computational Bayesian STAR (CBSTAR) models of Lopes and Salazar (2005). These are the models we compare against our DBSTAR models in the application sections of this paper.

1.1 The classical logistic STAR models

Let y_t be a univariate time series, for $t = 1, 2, \ldots, T$. The mathematical expression representing a STAR model of autoregressive order p with two regimes is given by

$$y_t = \underline{z}_t \underline{\phi}_1' + \pi(s_t; \gamma, c)\underline{z}_t \underline{\phi}_2' + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

where for $i = 1, 2$, $\underline{\phi}_i = (\phi_{0i}, \phi_{1i}, \ldots, \phi_{pi})$ are $(p+1)$-dimensional vectors with element ϕ_{ji} $(j = 0, 1, \ldots, p)$ representing an autoregressive coefficient associated with each component j of the AR regime i; $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$ is a $(p+1)$-dimensional vector; $\pi(.)$ is a non-linear smooth transition function in the range [0, 1]. Note that we use underlining to represent vectors and prime for transposition throughout the paper. We will also be using boldface capital characters to represent matrices.

In the literature, we can find STAR models with different transition functions such as the first- and second-order logistic function, the exponential function (Terasvirta, 1994), and others such as the cumulative distribution function (Chan and Tong, 1986) and the hyperbolic tangent function (Bacon and Watts, 1971). The most commonly used transition function, and the one we will be adopting in the application section is the first-order logistic function. It is defined as

$$\pi(s_t; \gamma, c) = [1 + \exp\{-\gamma(s_t - c)\}]^{-1} \tag{2}$$

with parameters $\gamma \in \Re^+$ and $c \in \Re$ defined as smoothness and location, respectively, and $s_t \in \Re$ defined as a transition variable (usually in practice being either an external variable or lagged dependent variable y_{t-d}, where d is a delay parameter). The parameter γ is responsible for the degree of smoothness of $\pi(s_t; \gamma, c)$ and the parameter c represents the threshold between the two regimes. For the same value of γ, the distance between the value of s_t and c determines

the degree of pertinence between the two regimes. When γ tends to zero the logistic function becomes constant ($\pi(s_t; 0, c) = 0.5$) and the *logistic* STAR, or simply LSTAR, model is reduced to an average between two linear AR(p) models. As the values of γ increase, the logistic function tends to a step function and the transition from one regime to the other becomes more abrupt, resembling a TAR model.

An alternative transition function is the exponential function $\pi(s_t; \gamma, c) = 1 - \exp\{-\gamma(s_t - c)^2\}$ where the parameters have similar interpretation to the logistic function above. See e.g. Terasvirta (1994) for more details on alternative transition functions.

A strategy for statistical parametric inference in STAR models was proposed by van Dijk et al. (2002) who used ordinary least squares for estimating the autoregressive coefficients and nonlinear least squares for the parameters within the logistic function. Their nonlinear least squares algorithm is an iterative procedure that requires the choice of initial values for parameters. A problem with this approach is that despite some choices guaranteeing local convergence, global convergence can be very slow, as shown in Terasvirta (2005). An important point to notice is that the nonlinear least squares used by the classical STAR models adopt linear approximations of the nonlinear function at each iteration step. In our proposed models, this approximation is carried out at the modelling rather than at the estimation stage as we shall see.

1.2 The computational Bayesian STAR models

In general, Bayesian parametric inference for non-linear models rely on computational simulation methods due to the loss of analytical tractability in calculating posterior distributions (see, e.g. Bauwens et al. (1999)). They were the first authors to propose computational Bayesian approaches for STAR models where sampling importance resampling algorithms (Gelfand and Smith, 1990) were used for posterior inference of parameters of a classical STAR model with known AR order p. Lopes and Salazar (2005) further developed the CBSTAR models where (i) a Gibbs sampler approach was used for inferences on the unknown parameters $\underline{\phi}_1, \underline{\phi}_2, \gamma, c, d$ and σ^2 of the logistic STAR in (1) and (2) when the AR order p is known, and (ii) a reversible jump Markov Chain Monte Carlo (RJMCMC) algorithm for posterior assessment of unknown p was included. Their approach was based on Lubrano's prior distributions as in Bauwens et al. (1999) where $p\left(\underline{\phi}_1, \gamma, \sigma^2, c\right) \propto \frac{1}{(1+\gamma^2)\sigma^2}$ and $\left(\underline{\phi}_2 \mid \sigma^2, \gamma\right) \sim N\left(\underline{0}, \sigma^2 e^{\gamma} I_{p+1}\right)$ for $\underline{\phi}_1, \underline{\phi}_2 \in \Re^{p+1}, \gamma, \sigma^2 > 0$ and $c \in [c_a, c_b]$, with $c_a = \hat{F}^{-1}(0.15)$ and $c_b = \hat{F}^{-1}(0.85)$, with $\hat{F}^{-1}(q)$ is the empirical cumulative distribution function for the quantile q. The conditional posteriors obtained using the Gibbs sampling algorithm of Gelfand and Smith (1990) for the AR coefficients $\underline{\phi}_1$ and $\underline{\phi}_2$ follow a Normal distribution with the variance σ^2 having an Inverse Gamma distribution. A Metropolis-Hasting algorithm (Gilks et al., 1996) was adopted to obtain the conditional posterior distribution for the nonlinear parameters γ and c. They considered $s_t = y_{t-d}$ in (2) and obtained a conditional posterior distribution for the delay parameter d as $p(d|y, \underline{\Theta}) \propto p(y|d, \underline{\Theta})p(d)$ from a discrete prior set for $d \in \{d_1, d_2, \ldots, d_{max}\}$, where $d_1 \leq \ldots \leq d_{max}$, for a large upper bound d_{max} with $p(d_i) = Pr(d = d_i)$.

It is worth noting that computational Bayesian inference approaches such as the CBSTAR models above are computer intensive numerical simulation models that rely on the availability of extensive data sets and on the possible convergence of chains to obtain approximate posterior distributions of underlying parameters. They are, thus, not generally appropriate for real-time applications that require sequential prior-to-posterior parametric updating and forecasting.

This paper is organised as follows: Section 2 defines the DBSTAR models. Section 3 shows a comparative analysis of the application of classical STAR, CBSTAR and DBSTAR models to the Canadian lynx data. Discussions on both models and results are presented in Section 4.

2 The Gaussian Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models

A DBSTAR model in its simplest form can be seen as polynomial approximation of the classical STAR model as defined by (1), where the smooth transition function, such as the logistic in (2), is represented by its Taylor series expansion. This polynomial approximation allows an alternative approach to both the classical STAR and the CBSTAR models that sequentially updates its dynamic parameters in time in a Bayesian analytical fashion via Kalman filtering based on the Dynamic Linear model (DLM) formulation of West and Harrison (1997). Computationally cheap, sequential analytical approximations for posterior parametric and forecasting distributions are obtained at each time step that includes estimation of the original parameters (by solving a system of equations) to preserve model interpretability.

The following sub-sections define the DBSTAR models and describe the prior-to-posterior algorithm for sequential parametric distribution estimation. A formulation for modelling processes with cycles is also described. The order r of the Taylor series expansion of the transition function is assumed *fixed* as is the autoregressive order p. In cases where data are initially available, both r and p can be determined, for example, by a grid search approach for optimal values that minimize a forecasting error statistic. Alternatively, we adopted a method proposed in West (1986) of monitoring the predictive performance via likelihood function which we describe in section 2.3.

Unlike the classical STAR and CBSTAR model formulations, the observational variance Σ_t of the underlying nonlinear AR process, Y_t, is treated as an *unknown* dynamic parameter which distribution sequentially adapts to data in the usual Normal-Inverse-Gamma conjugate analysis. We adopt a dynamic smoothing transition function $\pi_t(s_t; \gamma_t, c_t)$ where the smoothing and the threshold parameters γ_t and c_t (treated as *unknowns*) are allowed to vary in time and to adapt to data sequentially. The transition variable s_t can be either an exogenous variable or a past value of the process, depending on the application. In the latter case, $s_t = y_{t-d}$ with delay parameter d determined a priori. Again, in cases where initial data is available, a model predictive performance approach is straightforward for determining the unknown delay parameter d.

2.1 Model definition

For a generic dynamic transition function $\pi_t(s_t; \gamma_t, c_t)$ with real values in the interval $[0, 1]$, where s_t is a transition variable, $\gamma_t \in \Re^+$ is a smoothing parameter and $c_t \in \Re$ is a threshold value, a DBSTAR(r, p) model of orders r and p is defined by the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$ as follows.

The observational process Y_t has conditional Normal distribution

$$(Y_t \mid \underline{\theta}_t) \sim N\left(\underline{F}_t'\underline{\theta}_t, \Sigma_t\right) \tag{3}$$

where $\underline{F}_t' = [s_t^r \underline{z}_t, s_t^{r-1}\underline{z}_t, \ldots, s_t\underline{z}_t, \underline{z}_t]$ is a *known* $(r+1)(p+1)$-dimensional vector of polynomial regression variables $s_t^i\underline{z}_t$ $(i = 0, 1, \ldots, r)$ with $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$; $\underline{\theta}_t$ is the state vector containing *unknown* parameters associated with the components of $\underline{F}_t$, i.e, $\underline{\theta}_t' = (\underline{\theta}_{r,t}, \underline{\theta}_{r-1,t}, \ldots, \underline{\theta}_{1t}, \underline{\theta}_{0t})$

with elements $\underline{\theta}_{it} = \alpha_{it}(\gamma_t, c_t)\underline{\phi}_t$, where for $i = 1, ..., r$, $\alpha_{it}(\gamma_t, c_t)$ are polynomial functions of γ_t and c_t obtained from the i-th coefficient of the multivariate Taylor series expansion of $\pi_t(s_t; \gamma_t, c_t)$

$$\beta_{it} = \left. \frac{\partial^i \pi_t(s_t; \gamma_t, c_t)}{\partial s_t^i \partial \gamma_t^i \partial c_t^i} \right|_{s_t = s_0, \gamma_t = \gamma_0, c_t = c_0}$$

where $\beta_{0t} = \pi_t(s_0; \gamma_0, c_0)$, with s_0, γ_0 and c_0 being constant values around which the Taylor series expansion is to be carried out. $\underline{\phi}_t = (\underline{\phi}_{1t}, \underline{\phi}_{2t})$ is a dynamic $2(p+1)$-vector of *unknown* AR coefficients where, for $i = 1, 2$, $\underline{\phi}_{it} = (\phi_{0it}, \phi_{1it}, \ldots, \phi_{pit})$ with element ϕ_{ji} representing coefficient j $(j = 0, 1, \ldots, p)$ of the AR regime i. Note that the Taylor series expansion approximates $\pi_t(s_t; \gamma_t, c_t)$ better in the vicinities of s_0, γ_0 and c_0, so that, at each time t, those values can be conveniently chosen to improve the approximation (e.g. they can be chosen to be closer to the turning points of the transition function when it is known a regime change is about to happen).

The observational variance Σ_t is assumed unknown and defined as $\Sigma_t = k_t V_t$, where $k_t = k(\mu_t)$ is an appropriately chosen variance law (a scaling function of the mean $\mu_t = \underline{F}_t' \underline{a}_t$ of Y_t, where $\underline{a}_t$ is the mean of the prior distribution of $\underline{\theta}_t$ as we shall see) for the (unknown) underlying variance V_t. While a suitable chosen variance law can model systematic changes in the observational variability in time, we assume that Σ_t may change but only slowly and steadily in time (with the use of a variance discounting technique) to avoid potential unpredictable behaviour that can lead to loss of analytical tractability (Broemeling, 1985).

The state vector $\underline{\theta}_t$ has a conditional multivariate Student-t distribution with n_{t-1} degrees of freedom

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}} \left(\mathbf{G}_t \underline{\theta}_{t-1}, \mathbf{W}_t\right) \tag{4}$$

where $\mathbf{G}_t$ is a *fixed* $(r+1)(p+1) \times (r+1)(p+1)$ state evolution matrix with elements g_{ijt} (for row i and column j) chosen according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$, and $\mathbf{W}_t$ is a $(r+1)(p+1) \times (r+1)(p+1)$ *known* evolution covariance matrix, for which a discount factor δ_W, satisfying the condition $0 < \delta_W \leq 1$, is considered, as follows,

$$\mathbf{W}_t = \left(\frac{1 - \delta_W}{\delta_W}\right) \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t'.$$

where $\mathbf{C}_{t-1}$ is a prior covariance matrix for $\underline{\theta}_t$.

The dynamic underlying observational precision Σ_t^{-1} is assumed to relate to Σ_{t-1}^{-1} according to

$$\Sigma_t^{-1} = \frac{\lambda_t}{\delta_V} \Sigma_{t-1}^{-1}$$

with $\lambda_t \sim Beta\left(\frac{\delta_V n_{t-1}}{2}, \frac{(1-\delta_V) n_{t-1}}{2}\right)$ and δ_V is a discount factor for the observational variance taking values in [0,1]. Usually, δ_V assume values between 0.9 and 0.99 and δ_W assume values between 0.8 and 0.99 (West and Harrison, 1997).

Let $D_t = (y_t, D_{t-1})$ represent the information available at time t after observing Y_t. The initial information at time *t-1*, including history, is used to form initial relevant views about the future for all model parameters. In case of no prior knowledge about the parameters, non-informative prior distributions can be used. At time $t = 0$, the state vector is given by the specification of a multivariate Student-t distribution with n_0 degrees of freedom

$$\left(\underline{\theta}_0 \mid D_0\right) \sim T_{n_0} \left(\underline{m}_0, \mathbf{C}_0\right)$$

where, $\underline{m}_0$ and $\mathbf{C}_0$ are the prior mean vector and covariance matrix. Similarly, the initial distribution for the observational variance Σ_0 is a Inverse Gamma with hyperparameters n_0 and S_0

$$(\Sigma_0 \mid D_0) \sim IG\left(\frac{n_0}{2}, \frac{n_0 S_0}{2}\right)$$

Now that the DBSTAR(r,p) model is defined, we proceed to describe how their parameters are sequentially updated by the Kalman filter as well as how estimates of the original parameters $\underline{\phi}_t$, γ_t and c_t can be obtained.

2.1.1 The parametric prior-to-posterior updating

At time $t-1$, the posterior probability distribution for $\underline{\theta}_{t-1}$ is $(\underline{\theta}_{t-1} \mid D_{t-1}) \sim T_{n_{t-1}}\left(\underline{m}_{t-1}, \mathbf{C}_{t-1}\right)$, where n_{t-1} is degree of freedom of a multivariate Student-t distribution, $\underline{m}_{t-1}$ and $\mathbf{C}_{t-1}$ are the posterior mean vector and covariance matrix, respectively, for $\underline{\theta}_{t-1}$. The posterior probability distribution for the observational variance is $(\Sigma_{t-1} \mid D_{t-1}) \sim IG\left(\frac{n_{t-1}}{2}, \frac{n_{t-1}S_{t-1}}{2}\right)$, where S_{t-1} is the point estimate of the observational variance.

At time t, before observing Y_t, the prior probability distribution for the state vector $\underline{\theta}_t$ is $(\underline{\theta}_t \mid D_{t-1}) \sim T_{n_{t-1}}(\underline{a}_t, \mathbf{R}_t)$, where $\underline{a}_t = \mathbf{G}_t \underline{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^{'} + \mathbf{W}_t$. The prior probability distribution for the observational variance Σ_t is $(\Sigma_t \mid D_{t-1}) \sim IG\left(\frac{\delta_V n_{t-1}}{2}, \frac{\delta_V n_{t-1} S_{t-1}}{2}\right)$. The 1-step ahead forecast probability distribution for y_t is $(Y_t \mid D_{t-1}) \sim T_{\delta_V n_{t-1}}(f_t, Q_t)$, with $f_t = \underline{F}_t^{'} \underline{a}_t$ and $Q_t = \underline{F}_t^{'} \mathbf{R}_t \underline{F}_t + k_t S_{t-1}$.

After observing $Y_t = y_t$, the updating recurrence analysis is given by the posterior probability distributions: $(\underline{\theta}_t \mid D_t) \sim T_{n_t}(\underline{m}_t, \mathbf{C}_t)$, where $n_t = \delta_V n_{t-1} + 1$, $\underline{m}_t = \underline{a}_t + (\mathbf{R}_t \underline{F}_t / Q_t) e_t$ and $\mathbf{C}_t = \frac{S_t}{S_{t-1}}(\mathbf{R}_t - \mathbf{R}_t \underline{F}_t / Q_t \underline{F}_t^{'} \mathbf{R}_t^{'})$, and $(\Sigma_t \mid D_t) \sim IG\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right)$, with $S_t = S_{t-1} + \frac{S_{t-1}}{n_t}(\frac{e_t^2}{Q_t} - 1)$ and $e_t = Y_t - f_t$, the one-step-ahead forecasting error.

Conditional forecast distributions for the states and observations at a future time $t+k$, given the data up to time t, are recursively determined. Since these forecast distributions are Student-t, it is enough to compute their means and variances. With $\underline{a}_t(0) = \underline{m}_t$ and $\mathbf{R}_t(0) = \mathbf{C}_t$, i.e., the distributions for the future take information from the filtering distributions at time t, the $k \geq 1$ -step ahead forecast probability distribution for the state vector is $(\underline{\theta}_{t+k} \mid D_t) \sim T_{n_t}(\underline{a}_t(k), \mathbf{R}_t(k))$, where $\underline{a}_t(k) = \mathbf{G}_{t+k} \underline{a}_t(k-1)$ and $\mathbf{R}_t(k) = \mathbf{G}_{t+k} \mathbf{R}_t(k-1) \mathbf{G}_{t+k}^{'} + \mathbf{W}_{t+k}$. And for the observation $(Y_{t+k} \mid D_t) \sim T_{\delta_V n_t}(f_t(k), Q_t(k))$, with $f_t(k) = \underline{F}_{t+k}^{'} \underline{a}_t(k)$ and $Q_t(k) = \underline{F}_{t+k}^{'} \mathbf{R}_t(k) \underline{F}_{t+k} + k_{t+k} S_{t+k}$.

2.1.2 Obtaining estimates of the interpretable parameters

As we have seen, the state parameters $\underline{\theta}_{it} = \alpha_{it} \underline{\phi}_t$ $(i = 0, 1, \ldots, r)$ of the DBSTAR(r,p) are polynomial functions of the interpretable parameters $\underline{\phi}_t, \gamma_t$ and c_t and do not themselves possess contextual interpretations. However, at each time t, after observing $Y_t = y_t$, the first two moments (but not the probability distribution) of the parameters $\underline{\phi}_t, \gamma_t$ and c_t are obtained from the first two moments of the posterior Student-t distribution of $\underline{\theta}_t$ by solving one system of $(r+1)$ polynomial equations on $(2p+4)$ variables for $E(\underline{\theta}_t | D_t) = \underline{m}_t$ and one for $\mathrm{Var}(\underline{\theta}_t | D_t) = \mathrm{diag}(\mathbf{C}_t)$, respectively.

For instance, the system $m_{it} = f_{it}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t), i = 0, \ldots, r$, where f_{it} are polynomial functions of (the Taylor coefficients β_{it}) order $(r-1)$ of their arguments $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t$ and c_t only, may have complex conjugate solutions which the real parts are used as point estimates

of their first moment. That is, $\hat{\underline{\phi}}_{1t} = \text{Re}\{g_{1t}(\underline{m}_t)\}$, $\hat{\underline{\phi}}_{2t} = \text{Re}\{g_{2t}(\underline{m}_t)\}$, $\hat{\gamma}_t = \text{Re}\{g_{3t}(\underline{m}_t)\}$ and $\hat{c}_t = \text{Re}\{g_{4t}(\underline{m}_t)\}$ (where $\hat{\psi} = E(\psi)$), i.e, each original parameter has the real part of a function $g_{jt}(\underline{m}_t), j = 1, \ldots, (r+1)$, a location point estimate. Note that the vector $\underline{\theta}_t$ has dimension $(r+1)(p+1)$ whilst $\underline{\phi}_{1t}$ and $\underline{\phi}_{2t}$ are $(p+1)$-vectors and γ_t and c_t are scalars.

Similarly, second moments can be obtained by solving a second system of equations $v_{it} = f^*_{it}(\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t, c_t, \mathbf{C}_t)$, where $v_{it} = Var(\underline{\theta}_{it}|D_t)$ for polynomials f^*_{it} of order $(r-1)$ of their arguments.

2.2 Harmonic DBSTAR models for processes with cycles

In this sub-section we present the DBSTAR models for modelling observed cyclical behaviour in terms of cyclical components added to the model structure of Sub-section 2.1. It is important for the natural cycle to be included as a component in forecasting models. An useful and functional representation of the cycles is given in terms of trigonometric terms (West and Harrison, 1997). Fourier form representations of a cycle are considered as sine/cosine waves which provide economic characterisation on parameters and easy interpretation of them. Basically, the DBSTAR models for cycles, we call Harmonic DBSTAR (HDBSTAR) models, extend the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$, with $\underline{F}_t = (\underline{F}_{1t}, \underline{F}_{2t}), \underline{G}_t = (\underline{G}_{1t}, \underline{G}_{2t})$ and $\underline{W}_t = (\underline{W}_{1t}, \underline{W}_{2t})$, where $\underline{F}_{1t}, \underline{G}_{1t}$ and $\underline{W}_{1t}$ are associated to the nonlinear autoregressive components as in (3) and (4), and $\underline{F}_{2t}, \underline{G}_{2t}$ and $\underline{W}_{2t}$ are associated to the cyclical components. A HDBSTAR(r, p, h) model for cycles is defined as a DBSTAR(r, p) with an explicit component for cycle with h harmonics as follows

$$\left(Y_t \mid \underline{\theta}_t, \underline{\psi}_t\right) \sim N\left(\underline{F}'_{1t}\underline{\theta}_t + \underline{F}'_{2t}\underline{\psi}_t, \Sigma_t\right) \tag{5}$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{6}$$

$$\left(\underline{\psi}_t \mid \underline{\psi}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\psi}_{t-1}, \mathbf{W}_{2t}\right) \tag{7}$$

where, $\underline{\psi}_t$ is a $2h$-dimensional vector with the cyclical function, $\underline{\psi}_t = [\sum_{j=1}^{h} S_j(t), \sum_{j=1}^{h} S^*_j(t)]$, which is a linear combination of trigonometric terms, i.e., a summation of the h harmonics, $S_j(t) = a_j cos(t\omega_j) + b_j sin(t\omega_j)$ and theirs conjugate at every other position $S^*_j(t) = -a_j sin(t\omega_j) + b_j cos(t\omega_j)$. The $2h$-dimensional vector $\underline{F}_{2t}$ is a canonical vector associated to the harmonics in $\underline{\psi}_t$, with 1 regarding the harmonic positions and 0 otherwise, which gives $\underline{F}_{2t} = [1, 0]$ for 1 harmonic, $\underline{F}_{2t} = [1, 0, 1, 0]$ for 2 harmonics, and so forth. The frequency of each harmonic is defined as $\omega_j = \frac{2\pi j}{s}$, with s the period of the cycle. The quantities a_j and b_j are called the Fourier coefficients. The evolution matrix of the cyclical components $\mathbf{G}_{2t}$ has $|\mathbf{G}_{2t}| = cos^2(t\omega_j) + sin^2(t\omega_j) = 1$. The discrete-time evolution of the jth harmonic from time t to time $t+1$ is given by

$$\begin{pmatrix} S_j(t+1) \\ S^*_j(t+1) \end{pmatrix} = \mathbf{G}_{2t} \begin{pmatrix} S_j(t) \\ S^*_j(t) \end{pmatrix}$$

The $(2h \times 2h)$-matrix $\mathbf{W}_{2,t}$ contains the covariances of the cyclical components.

The first harmonic, the fundamental harmonic, is expected to dominate the cyclical pattern, having a strong sinusoidal signal. The higher frequency harmonics oscillate faster than the fundamental one and more appropriate for modelling higher frequency repetitive behaviour. Obviously that the larger the h the more accurate the modelling of periodic variations in the data. However, adopting the parsimony principle we look for the smallest h that can still provide a good representation of the cycle in the underlying process.

A hypothesis test is recommended to determine the retention of harmonics. Denote the marginal posterior distributions of the coefficients of the h-th harmonic by $(\underline{\psi}_{ht} \mid D_t) \sim T_{n_t}(\underline{m}_{ht}, \mathbf{C}_{ht})$. We have then $(\underline{\psi}_{ht} - \underline{m}_{ht})' \mathbf{C}_{ht}^{-1} (\underline{\psi}_{ht} - \underline{m}_{ht})/2 \sim F_{2,n_t}$. West and Harrison (1997) recommend a formal statistical test for assessing the importance of harmonic components at each time t based on the calculation of HPD (high posterior density) region for each harmonic as $P(F_{2,n_t} \leq \underline{m}_{ht}^{'} \mathbf{C}_{ht}^{-1} \underline{m}_{ht}/2)$.

Alternatively, for cases where a large enough initial dataset is available to enable the investigation of the cyclic behaviour, the parsimonious value of h can be determined for example by applying various models in a stepwise approach. For cases where an initial dataset is available, a Bayesian approach similar to the one described in the following subsection can be adopted.

2.3 Bayesian model selection

In this sub-section, we present a Bayesian approach to obtain value of unknown parameters which cannot be accommodated in the state vector $\underline{\theta}_t$. This method is based on predictive performance of different models which involves a sequential processing of observations Y_t, easily calculated from the one-step ahead predictive distribution $p(Y_t \mid D_{t-1})$, mainly its residuals e_t. For a philosophical viewpoint see Good (1985); for a general discussion see West (1986) and West and Harrison (1997). Note that it is important to keep the same model structure to use this method, differing only in value of parameters.

2.3.1 The unknown discrete parameters

Some unknown parameters in the model do not belong to the state vector for few important reasons: (i) they do not need to be updated sequentially in time, (ii) analytical computations would be intractable, therefore we would lose the conjugacy properties in our prior-to-posterior updating analysis, (iii) they do not belong to the model definition originally. The method is recommended for the following parameters, in case the modeller cannot specify values for those:

- The Taylor series order r, where $r = 1, 2, ..., r_{max}$,
- The autoregressive order p, where $p = 1, 2, ..., p_{max}$,
- The harmonics h, where $h = 1, 2, ..., h_{max}$,
- The delay parameter d of the transition variable $s_t = y_{t-d}$, where $d = 1, 2, ..., d_{max}$,
- The discount factor δ_W of the state variance $\mathbf{W}_t$, where $0 < \delta_W \leq 1$,
- The discount factor δ_V of the observational variance Σ_t, where $0 < \delta_V \leq 1$.

If all the above mentioned parameters can be individually identified, this method should be skipped and the parametric prior-to-posterior updating can be carried out using the given values. In case the modeller needs to decide about two possible values for one parameter, say for example, $p_{max} = 2$, so two candidate models will be under analysis, one for $p = 1$ and another for $p = 2$. Therefore, the Bayes' factor described in section 2.3.2 should be used to identify p. In case there is, however, a need of identifying more than two possible values for one or more parameters, then more than two candidate models will be under analysis, thus the joint log-likelihood predictive should be instead used to select them.

2.3.2 Bayes' factor for determination of unknown discrete parameters

Let τ represent an unknown discrete parameter (e.g., r, p, d or h) that can take integer values for which the most likely value τ^* is to be determined. From a Bayesian parametric prior-to-posterior updating approach, we have for each observation Y_t the one-step ahead predictive distribution $p(Y_t \mid D_{t-1}, \tau)$. The joint predictive distribution

$$p(Y_t, Y_{t-1} \ldots Y_1 \mid D_0, \tau) = \prod_{t=1}^{T} p(Y_t \mid D_{t-1}, \tau) \tag{8}$$

is the predictive likelihood of all observations $Y_t, Y_{t-1}, \ldots, Y_1$ conditional on the unknown parameter τ. We use the Bayes' factor (Jeffreys, 1961), or weights of evidence (Good, 1985), defined as the ratio between the predictive likelihood of two models, say model A and model B, differing only in values of the parameter τ. Thus, from (8), the logarithm of Bayes' factor is

$$log(BF) = \sum_{t=1}^{T} log\left[p_A(Y_t \mid D_{t-1}, \tau = \tau_1)\right] - \sum_{t=1}^{T} log\left[p_B(Y_t \mid D_{t-1}, \tau = \tau_2)\right] \tag{9}$$

This measure gives us evidence in favour of ($log(BF) > 0$) or against ($log(BF) < 0$) model A relative to model B, according to their predictive performances when τ assumes either τ_1 or τ_2. The farther from zero $log(BF)$ is, the stronger the evidence and for $log(BF) = 0$ there is no evidence either way.

Alternatively, we can generalise the idea of the Bayes' factor for choosing values for τ when comparing more than two models. The evidence in favour of a model can also be obtained by checking which log-likelihood predictive, defined in (8), is the largest. So, that one reflects the current choice of model and determines the values for the parameters represented by τ. Operationally, we choose τ^* by calculating the log-likelihood predictive for different models differing only values of *tau* and selecting the model which gives the largest log-likelihood predictive. Note that, contrary to a fully Bayesian approach where all unknown parameters are included in the model parameter set and in principle requires no data initially, our approach here requires an initial dataset.

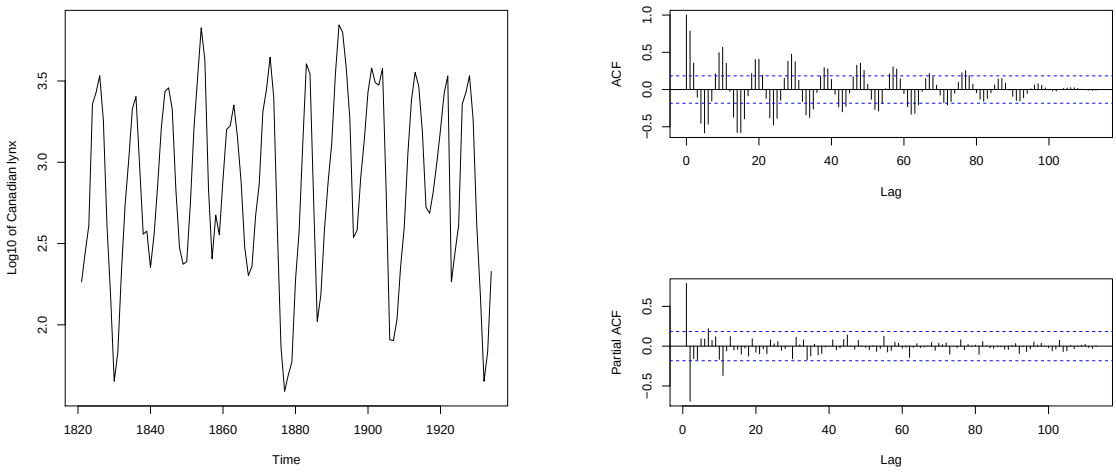
3 Application: the Canadian lynx data

In this section, we apply two formulations of DBSTAR models to the well-known Canadian lynx data. One formulation, a DBSTAR(r, p) model, accounts for the data cyclic behaviour through high AR components order, and the other, a HDBSTAR(r, p, h) model, models the cycle with the inclusion of an explicit component.

The aim here is twofold. In one hand, we aim to validate our proposed DBSTAR models by comparing their performances against the performances of both the classical STAR and the CBSTAR models. In another hand, we aim to illustrate the extra features we can achieve by adopting sequential models with dynamic parameters.

3.1 The dataset and some initial analyses

The Canadian Lynx dataset consists of the annual series of the number of Canadian lynx trapped in the Mackenzie River district of North-west Canada from 1821 to 1934, giving therefore a total



(a) (b)

Figure 1: (a) Log_{10}-transformed series (b) ACF and PACF of $\log_{10}$-transformed series

of 114 observations. It was originally recorded to support understanding the population dynamics of the ecological system in that area. Analysis of this time series can be found, for example, in Tong (1990), Terasvirta (1994) and Lopes and Salazar (2005). The most prominent features of the Canadian lynx time series are (i) the lack of trend, (ii) the presence of irregular changes in the amplitude in time, and (iii) the presence of persistent non-regular cyclic oscillations with periods of 10 or 11 years. These features have been familiar to biologists for a long time and are prominent in historical records of trappings of lynx in Canada (see, for example, Elton and Nicholson (1942) and the references therein).

The original series was $\log_{10}$-transformed to (i) remove the marked right-skewness of its frequency plot, and (ii) allow the comparative analysis in this section as the classical STAR and the CBSTAR models were all applied to the $\log_{10}$-transformed series. Figure 1 shows (a) the $\log_{10}$-transformed series together with the corresponding (b) ACF and PACF plots. Note that there seems to be a slow sinusoidal decay along time in the ACF together with a cut after lag 2 (assuming the significant PACFs on lags 4, 7 and 10 as spurious) in the PACF which suggest an auto-correlation of second order. There is also a significant peak on lag 11 in the PACF that seems to show a cyclical period of 11 years. Note also that the use of ACF and PACF for AR order identification is not appropriate when the relationships between y_t and y_{t-k} $(k = 1, \ldots, 114)$ are not linear. In fact, a graphical analysis of the original Canadian lynx data showed lack of linearity for most lags but those were largely improved after the $\log_{10}$-transformation.

3.2 Modelling

In this application, both formulations the DBSTAR(r, p) and the HDBSTAR(r, p, h) consider the logistic function in (2), represented by its Taylor series expansion, as the smooth transition function between the two regimes. For this specification, Taylor series order has to be odd, as for even orders the derivatives of logistic function evaluated at the point we are expanding around are zero. We observed that for $r = 1$ (a line) does not approximate well the transition function and for high Taylor series order $(r = 5, 7, 9, \ldots)$ the approximations showed only marginal improvements. Therefore, for parsimony reasons we use $r = 3$ in both models. Lagged values of

the $\log_{10}$-transformed series are used as the transition variable $s_t = y_{t-d}$, as in Terasvirta (1994) and Lopes and Salazar (2005).

For the DBSTAR(r, p) model as described in section (2.1), the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$ is considered as follows, (i) $\underline{F}'_t = [y^3_{t-d}\underline{z}_t, y^2_{t-d}\underline{z}_t, y_{t-d}\underline{z}_t, \underline{z}_t]$ is a vector of polynomial regression variables, with $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$; (ii) $\mathbf{G}_t = \mathbf{I}$, the identity matrix, that stands for no specified evolution changes for the state parameters from time $t-1$ to t were detected; (iii) $\Sigma_t = \Sigma$, which means the unknown observational variance of the underlying nonlinear AR process was not detected to be stochastically changing, so variance law was not used in this analyses. This implies we can make use of simpler algorithm as Σ_t is set as fixed and does not have to be updated in time sequentially and the discount factor associated to it does not to have be identified - in this case we set $\delta_V = 1$; and (iv) for the state variance $\mathbf{W}_t$, the use of its discount factor δ_W was important to allow (slow) dynamic parameter adaptation.

The state vector $\underline{\theta}_t = [\underline{\theta}_{3t}, \underline{\theta}_{2t}, \underline{\theta}_{1t}, \underline{\theta}_{0t}]$ accommodates each polynomial coefficients associated to the polynomial regression variables in $\underline{F}'_t$. This vector is sequentially updated in time as shown in section 2.1.1. To start off, non-informative prior distributions were used to form initial relevant views about the future for all model parameters, as we did not have any prior knowledge about them. Thus, we set $\underline{m}_0 = \underline{0}$ and large variances $\mathbf{C}_0 = 100\mathbf{I}$ as the prior mean vector and covariance matrix. Moreover, we used for the observational variance Σ_0, hyperparameters $n_0 = 1$ and $S_0 = 1$.

For the HDBSTAR(r, p, h) model described in section (2.2), $\underline{F}'_{1t}$, $\mathbf{G}_{1t}$, $\Sigma_t = \Sigma$ and $\mathbf{W}_{1t}$, are specified as above, because they are similarly associated to the nonlinear AR components. For the cyclical components, we have (i) $\underline{F}_{2t} = [1, 0, 1, 0]$ for 2 harmonics; (ii) $\mathbf{G}_{2t}$, a pre-specified trigonometric evolution changes for the state parameters from time $t-1$ to t; and (iii) for the covariances of the cyclical components $\mathbf{W}_{2t}$, the use of its discount factor δ_W was important to allow (slow) dynamic parameter adaptation as well.

Note that we carried out a hypothesis test in order to determine the retention of harmonics and $h = 2$ was the most appropriate for this application, with the first harmonic, the fundamental harmonic dominating the cyclical pattern, having a strong sinusoidal signal and one more harmonic oscillating faster than the fundamental one.

As we could not obtain value of some parameters which are not accommodated in the state vector $\underline{\theta}_t$, we used the Bayesian approach based on predictive performance of different models, described in section 2.3.2, as initial data is available. Those parameters were then considered unknown and we had to specify (i) autoregressive order p, where $p = 1, 2, ..., 12$; (ii) delay parameter d of the transition variable $s_t = y_{t-d}$, where $d = 1, 2, ..., 12$; and (iii) discount factor δ_W of the state variance $\mathbf{W}_t$, where $0 < \delta_W \leq 1$. We ran different models differing only in values of these three unknown parameters and we selected those by checking which likelihood predictive on log-scale, defined in (8), was the largest. This step is evaluated jointly with the parametric prior-to-posterior updating described in section 2.1.1 using the above given values. The algorithm gave us evidence for selecting a DBSTAR model with $p = 12$, as well as a HDBSTAR model with $p = 5$. Both the delay parameter and the discount factor assumed same value for both models, $d = 3$ and $\delta_W = 1$, respectively. Note that, our analyses suggest that to model the Canadian lynx data we have to make use of static nonlinear AR models. Nonetheless, we also ran all this models forcing the algorithm to select $\delta_W < 1$ in order to have dynamic models and the identified value was $\delta_W = 0.99$. It then selected $p = 4$ for the HDBSTAR model and we use both the DBSTAR$(3, 12)$ and the HDBSTAR$(3, 4, 2)$ models based on $\delta_W = 0.99$ for further analyses.

For parsimony reasons, again, we analysed the value of autoregressive coefficients in both

regimes and could detected although order $p = 12$ was given to the DBSTAR model, in fact, that order was significant only in one of the regimes. Therefore, the selected DBSTAR model considers AR order 12 in the first regime and 3 in the second, which allows us to have an important reduction in terms of number of parameters to be sequentially in time updated in the second regime. Similarly, for the HDBSTAR model we could detect such behaviour, AR order 5 in the first regime and 2 in the second, although for the latter the economy is less impressive as the AR orders are not high.

A point to note is that the DBSTAR$(3, 12)$ model tries to capture the cycle using high autoregressive order ($p = 12$). On the other hand, the HDBSTAR$(3, 4, 2)$ model uses low autoregressive order ($p = 4$) and model the cycle explicitly using Fourier analyses. The second model gives us an accurate specification of the cyclical behaviour of the Canadian lynx data set which is a key feature for modelling such series.

3.3 Modelling the cycle explicitly

Initial data exploration using Bayesian periodogram analysis (Bretthorst, 1988) gives us an idea about the cycle in this data. Bayesian periodogram uses the marginal log-likelihood of a single sinusoid regression model $y_t = acos(2\pi t/\lambda) + bsin(2\pi t/\lambda) + \epsilon_t$, where $\epsilon_t \sim N(0, \sigma^2)$ and λ is the periodicity or wavelength of the process Y_t. The spike at the wavelength of 11 years is evident in this data, suggesting a sustained and persistent cyclical feature of that period. This analysis indicates that a harmonic cycle with a wavelength of 11 years should be included in the model.

Furthermore, as in Elton and Nicholson (1942), the length of the cycle is considered of about 10 years rather than 11 years. We fitted both models and harmonics 1 and 2 seem to be significant in either model. Both analyses indicate a clear and useful inspection of the contribution of the two harmonics to the models. Once we compared the amplitude of the retrospective estimates of the two harmonics in both models we observed higher magnitude for models with a 10-year period. In addition, we concluded that improved performance was obtained by taking the 10-year period into account and we take this model for further analysis.

Comparing both the DBSTAR$(3, 12)$ and the HDBSTAR$(3, 4, 2)$ models we expected the latter to help improve fitting, even though a good fitting was observed and the residuals are white noises. The expected relative improvement did not actually happen for this data as well shall see. Moreover, we suggest taking the cycle present in the data into account and modelling it explicitly in order to obtain an accurate specification of the cyclical behaviour for a more appropriate exploration of the dynamic model.

3.4 Comparative analyses

Note that in order to perform a comparative analysis with existing results for static (non-dynamic) models that account for the whole Canadian lynx data set, a retrospective analysis using Kalman smoothing was carried out on our dynamic models. A retrospective analysis make static and dynamic models comparable in sense it uses the whole dataset after the filtering is applied. First off, we need the posterior probability distribution for $\underline{\theta}_t$, at each time t, provided by the Kalman filter. Then, the Kalman smoother provides the conditional probability distributions of $\underline{\theta}_t$ given the data D_T, for any time $t < T$ (see e.g. West and Harrison (1997)). As the Kalman filter, the Kalman smoother can be straightforwardly implemented as a backward-recursive algorithm, which depends only on the data used for filtering and one-step-ahead forecast moments. Thus, we go (forwardly) filtering the posterior probability distribution for $(\underline{\theta}_t|D_t)$ and then go

Table 1: Model comparison: Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE)

Model	MAE	RMSE
DBSTAR$(3,12)$	0.107	0.139
HDBSTAR$(3,4,2)$	0.127	0.161
CBSTAR$(11,3)$	0.118	0.153
Classical STAR$(11,2)$	0.142	0.179

(backwardly) smoothing this distribution to provide the smoothing probability distribution for $(\underline{\theta}_t|D_T)$, for $t = T-1, T-2, \ldots, 1$.

The backward-recursive algorithm starts with $(\underline{\theta}_T \mid D_T) \sim T_{nt}(\underline{a}_t(0), \mathbf{R}_t(0))$, where $\underline{a}_t(0) = \underline{m}_T$, the posterior mean vector and $\mathbf{R}_t(0) = \mathbf{C}_T$, the posterior covariance matrix, both computed by the last updating of the Kalman filter. The smoothing probability distributions for the state are $(\underline{\theta}_{t-k}|D_T) \sim T_{nt}\left[\underline{a}_t(-k), \frac{S_t}{S_{t-k}}\mathbf{R}_t(-k)\right]$, where $\underline{a}_t(-k) = \underline{m}_{t-k} + \mathbf{C}_{t-k}\mathbf{G}'_{t-k+1}\mathbf{R}^{-1}_{t-k+1}(\underline{a}_t(-k+1) - \underline{a}_{t-k+1})$ and $\mathbf{R}_t(-k) = \mathbf{C}_{t-k} - \mathbf{C}_{t-k}\mathbf{G}'_{t-k+1}\mathbf{R}^{-1}_{t-k+1}(\mathbf{R}_{t-k+1} - \mathbf{R}_t(-k+1))\mathbf{R}^{-1}_{t-k+1}\mathbf{G}_{t-k+1}\mathbf{C}'_{t-k}$. And the corresponding smoothed probability distribution for the level of the series are $(\underline{\mu}_{t-k}|D_T) \sim T_{nt}\left[\underline{f}_t(-k), \frac{S_t}{S_{t-k}}\underline{F}'_{t-k}\mathbf{R}_t(-k)\underline{F}_{t-k}\right]$, where $\underline{f}_t(-k) = \underline{F}'_{t-k}\underline{a}_t(-k)$.

The DBSTAR$(3,12)$ and HDBSTAR$(3,4,2)$ models were then compared to the classical STAR$(11,2)$ and CBSTAR$(11,3)$ models as in Terasvirta (1994) and Lopes and Salazar (2005), respectively. Table 1 presents the Mean Absolute Errors (MAE) and the Root Mean Squared Errors (RMSE) for the competing models. Comparing both DBSTAR$(3,12)$ and HDBSTAR$(3,4,2)$ models, we have noticed an improved fitting in favour of the DBSTAR$(3,12)$ model which has no explicit cycle component by 15%. Yet the HDBSTAR$(3,4,2)$ model did affect the prediction capability of the model but it still adapts to the slight changes in pattern from year to year. Both the DBSTAR$(3,12)$ and the HDBSTAR$(3,4,2)$ models have improved fitting with respect to the existing classical STAR$(11,2)$ model in the literature, by 25% and 10%, respectively. Furthermore, the DBSTAR$(3,12)$ model which uses high AR order as the CBSTAR$(11,3)$ model, our formulation also improved fitting performance by 9%.

Figure 2 illustrates a comparative analysis of both the DBSTAR$(3,12)$ and HDBSTAR$(3,4,2)$ models with the competing models individually. We conclude that both DBSTAR models fit well the data, although overall any of those models show no much distant points from the line.

3.5 Analysis of dynamic parameters

As we described in section 2.1.2, at each time t, after observing $Y_t = y_t$, we can obtain estimates of the interpretable parameters $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t$ and c_t, which are parameters in the STAR models without approximating by Taylor series and analyse their evolution in time. Note that this analysis cannot be done using neither the classical STAR nor the CBSTAR models as they are static, thus none of their constant parameters has evolution in time.

For this application, the system $\underline{m}_{0t} = \frac{1}{48}(24\underline{\phi}_{2t} - 12\gamma_t c_t \underline{\phi}_{2t} + \gamma_t^3 c_t^3 \underline{\phi}_{2t}) + \underline{\phi}_{1t}$, $\underline{m}_{1t} = \frac{1}{48}(12\gamma_t\underline{\phi}_{2t} - 3\gamma_t^3 c_t^2 \underline{\phi}_{2t})$, $\underline{m}_{2t} = \frac{3}{48}\gamma_t^3 c_t \underline{\phi}_{2t}$ and $\underline{m}_{3t} = -\frac{\gamma_t^3}{48}\underline{\phi}_{2t}$ has complex conjugate solutions $\hat{c}_t = -\frac{1}{3}\underline{m}_{2t}\underline{m}_{3t}^{-1}$, $\hat{\gamma}_t = \sqrt{\frac{12}{3\hat{c}_t^2 - \underline{m}_{1t}\underline{m}_{3t}^{-1}}}$, $\hat{\underline{\phi}}_{2t} = -\frac{48}{\hat{\gamma}^3 t}\underline{m}_{3t}$ and $\hat{\underline{\phi}}_{1t} = \underline{m}_{0t} - \frac{1}{48}(24 - 12\hat{\gamma}_t\hat{c}_t + \hat{\gamma}_t^3\hat{c}_t^3)\hat{\underline{\phi}}_{2t}$, with the real parts used as point estimates.

Figure 3 shows the evolution in time of $\hat{\underline{\phi}}_{1t}$ in both DBSTAR$(3,12)$ and HDBSTAR$(3,4,2)$

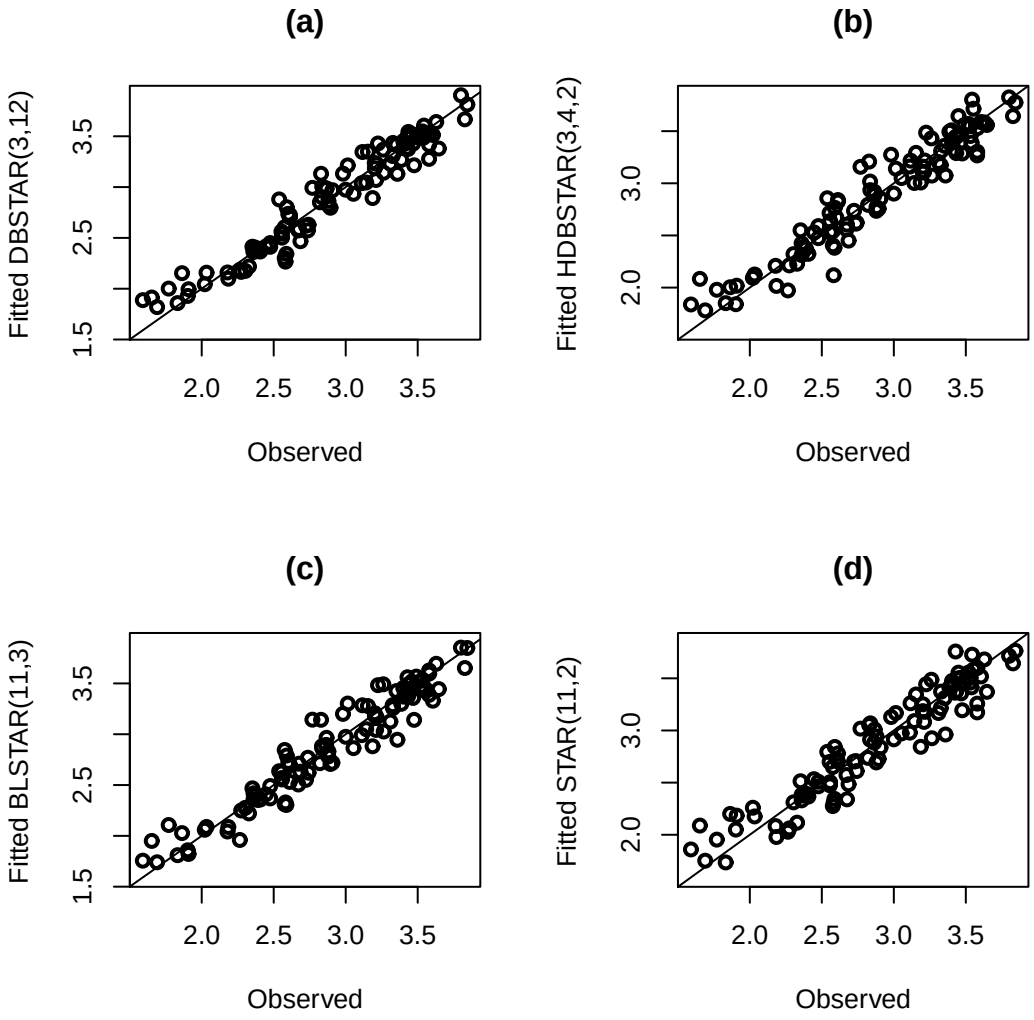
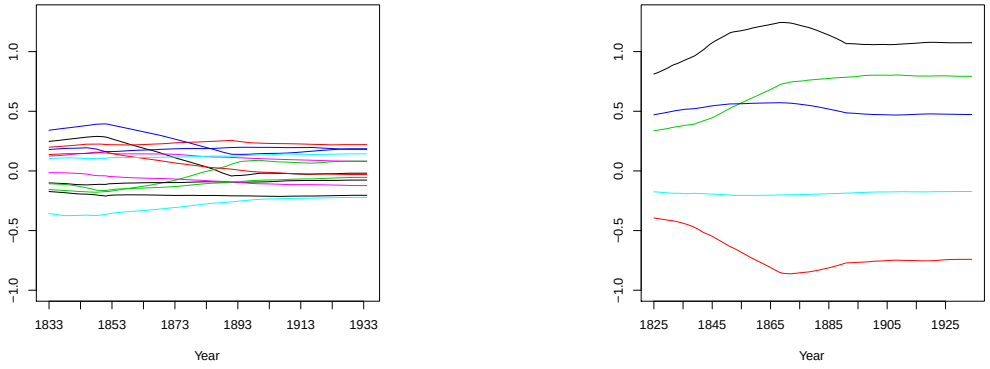
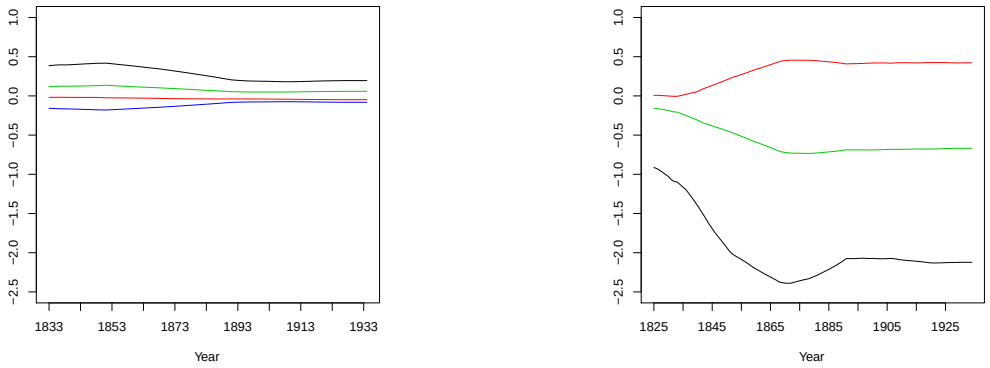


Figure 2: Observed series vs fitted models (a) DBSTAR$(3, 12, 3)$, (b) HDBSTAR$(3, 4, 2)$, (c) CBSTAR(11,3) and (d) LSTAR(11,2)



(a) (b)

Figure 3: Evolution in time of $\hat{\underline{\phi}}_{1t}$ in (a) DBSTAR$(3, 12)$ model and (b) HDBSTAR$(3, 4, 2)$ model



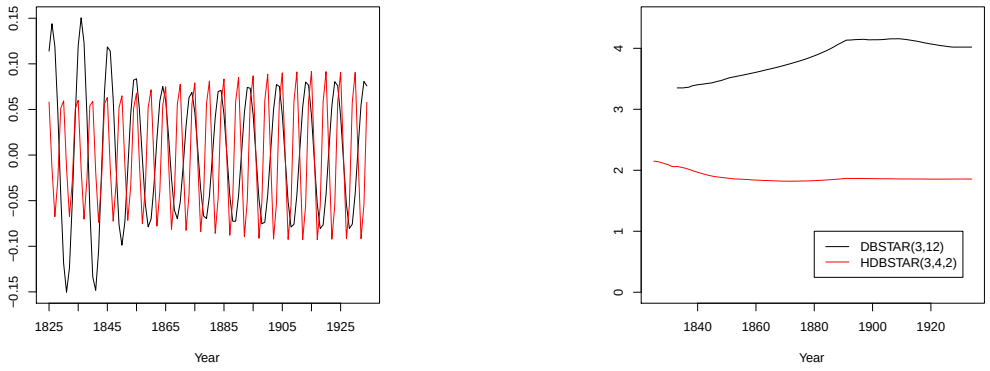
(a) (b)

Figure 4: Evolution in time of $\hat{\underline{\phi}}_{2t}$ in (a) DBSTAR$(3, 12)$ model and (b) HDBSTAR$(3, 4, 2)$ model

with $\delta_W = 0.99$. As discount factor is close to 1, the point estimates show small variation along time. Around decade 1880s, there seem to be an occurred phenomenon which may have caused changes in the dynamic of the parameters. Overall, the amplitude of the estimated autoregressive coefficients is larger for the HDBSTAR than DBSTAR due to higher autoregressive order in the latter model. More weights are given to the past values of the dependent variable in the HDBSTAR models. The coefficients in DBSTAR model are mainly concentrated between -0.5 and 0.5 while for the HDBSTAR the coefficients assume estimate values between -1 and 1.

Similarly, figure 4 shows the evolution in time of $\hat{\underline{\phi}}_{2t}$. The weights in HDBSTAR model are even more pronounced, along time. The coefficients in DBSTAR model are mainly concentrated between -0.5 and 0.5 while for the HDBSTAR the coefficients assume estimate values between -2.5 and 0.5. We see it as an advantage of the HDBSTAR models over the DBSTAR, classical STAR and CBSTAR, for modelling time series presenting cyclical behaviour. The guaranteed parsimony of HDBSTAR is balanced by larger amplitudes in the autoregressive coefficients.

The posterior mean of components associated to cycle are illustrated in figure 5. Note that,



(a) (b)

Figure 5: Evolution in time of (a) harmonics in HDBSTAR$(3, 4, 2)$ model and (b) smoothing parameter γ_t in both DBSTAR$(3, 12)$ and HDBSTAR$(3, 4, 2)$models

along time, the contribution of both harmonics oscillates between -0.2 and 0.2. These values are added up together with the linear combination of autoregressive coefficients and past values of the time series, at each regime. That is the reason we need small autoregressive order (less autoregressive coefficients) in the model.

Figure 5 also illustrates the evolution in time of the smoothing parameter γ_t. We observe that its point estimate is larger in DBSTAR$(3, 12)$ than in HDBSTAR$(3, 4, 2)$, overall, which means that the switching between the regimes occurs more abruptly for DBSTAR model. The reason the point estimate of γ_t is smaller and smoother in time for the HDBSTAR is that this model needs low autoregressive order and extra components for the cycle, so the two autoregressive parts of the model are similar and the switch between regimes may occur smoothly.

4 Discussion

The main goal of this research is to propose the Gaussian Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models for nonlinear autoregressive time series processes. The DBSTAR models consider analytical approximations for STAR models based on Dynamic Linear Models (DLM) of West and Harrison (1997) as alternative to both the classical STAR models of Chan and Tong (1986) and the CBSTAR models of Lopes and Salazar (2005). We used the rth-order Taylor series approximations as a solution for the transition function $\pi_t(s_t; \gamma, c)$, such as the logistic or exponential functions.

The classical STAR models use ordinary least squares for estimating the autoregressive coefficients and nonlinear least squares for the parameters within the transition function. A problem with this approach is that it requires choice of initial values to start off the algorithm which convergence can be very slow and not guaranteed.

Lopes and Salazar (2005) developed the CBSTAR models which use MCMC algorithms for posterior assessment of unknown parameters. As they are computer intensive numerical simulation models that rely on the possible convergence of chains, both the classical STAR and the CBSTAR models, thus, are not generally appropriate for real-time applications.

The DBSTAR models sequentially update their dynamic parameters in time via Kalman filtering. This solution estimates the parameters in analytical closed form sequentially in time.

In this way, we can avoid some computational problems associated with the classical STAR and CBSTAR models, such as, different convergence for different starting values, lack of guaranteed convergence, and others.

At each time t, the first two moments of the original parameters can be obtained from the first two moments of the posterior Student-t distribution of the state vector by solving one system of polynomial equations. It is worth mentioning that this analysis cannot be done using neither the classical STAR nor the CBSTAR models as they are static models and none of their constant parameters has evolution in time.

Due to Bayesian formulation, our proposal compared to the classical STAR models has also the advantage of allowing formal inputs and interventions from experts, where appropriate. In addition, the DBSTAR models can be applied to time series data sets when stationary and/or seasonality are present. It is not necessary to differentiate the time series to achieve stationary and remove possible seasonality. The DBSTAR models can be also applied to data in the presence of heteroskedasticity based on a conjugate analysis. The assumption of constant observational variance might be unrealistic depending on the application. Unlike the classical STAR and CBSTAR model formulations, the DBSTAR models incorporate a variance law, in which the mean is related to the unknown variance. We assume that the observational variance may change but only slowly and steadily in time with the use of a variance discounting technique.

Note that the classical STAR models are a particular case of the DBSTAR models for a constant observational variance $\Sigma_t = \sigma^2$, fixed autoregressive coefficients $\underline{\theta}_t = \underline{\theta}$, an identity matrix set for the evolution of the states $\mathbf{G}_t = \underline{\mathbf{I}}$, a null state covariance matrix $\mathbf{W}_t = \underline{\mathbf{0}}$ (equivalently, for $\delta_W = 1$) and fixed smoothing parameter, $\gamma_t = \gamma$ and, consequently, $\pi_t(s_t; \gamma_t, c) = \pi(s_t; \gamma, c)$. Recall that the classical STAR models estimate the parameters using the whole dataset, so to be equivalent, the modeller should run both the Kalman filter and the Kalman smoother algorithms.

The DBSTAR models can be extended for modelling observed cyclical behaviour in terms of cyclical components added to the model structure. A Harmonic DBSTAR model is defined as a DBSTAR with an explicit component for cycle with h harmonics. Fourier form representations of a cycle are considered as sine/cosine waves which provide economic characterisation on parameters and easy interpretation of them. In general, we use lower autoregressive order (less autoregressive coefficients) in the model when more weights are given to the past values of the dependent variable. The guaranteed parsimony of HDBSTAR is balanced by larger amplitudes in the autoregressive coefficients. This is an advantage of the HDBSTAR models over the DBSTAR, classical STAR and CBSTAR, for modelling time series in the presence of cyclical behaviour.

We applied two formulations of DBSTAR models to the well-known Canadian lynx data. One formulation, a DBSTAR(r, p) model, accounts for the data cyclic behaviour through high AR components order, and the other, a HDBSTAR(r, p, h) model, models the cycle with the inclusion of an explicit component. We concluded that both the DBSTAR and HDBSTAR models fitted well the data. Moreover, the application of DBSTAR models to the Canadian Lynx data showed improved fitting performances when compared with both the classical and the CBSTAR models.

There are some improvements of our current DBSTAR models that can be addressed in order to make our formulations more suitable for nonlinear time series processes. Nonlinear functions may well be approximated by either local linearisation or global spline approximations. The Taylor series expansion could be replaced by other approximation methods, such as splines functions. Also, a fully Bayesian approach that can be used in real-time application would use

a different method, such as, Particle Filtering, for updating all the unknown parameters in the model.

References

Bacon, D. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, 58:525–534.

Bauwens, L., Lubrano, M., and Richard, J. (1999). *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press.

Bretthorst, G. L. (1988). *Bayesian Spectrum Analysis and Parameter Estimation*. Springer-Verlag.

Broemeling, L. D. (1985). *Bayesian Analysis of linear models*. Decker.

Chan, K. and Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis*, 7:179–190.

Elton, C. and Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in canada. *Journal of Animal Ecology*, 11:215–244.

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

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall.

Good, I. J. (1985). Weight of evidence: A brief survey. In Bernardo, J., DeGroot, M., Lindley, D., and Smith, A., editors, *Bayesian Statistics 2*, pages 249–270. Elsevier Science Publisher B V, North–Holland, Amsterdam, and Valencia University Press.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press.

Lopes, H. F. and Salazar, E. (2005). Bayesian model uncertainty in smooth transition autorregressions. *Journal of Time Series Analysis*, 27:99–117.

Moran, P. A. P. (1953). The statistical analysis of the canadian lynx cycle, i: Structure and prediction. *Australian Journal of Zoology*, 1:163–173.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, 53:873–80.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89:208–218.

Terasvirta, T. (2005). Forecasting economic variables with nonlinear models. *Working Paper Series in Economics and Finance*, 598:1–55.

Tong, H. (1978). *On a threshold model*. In Patern Recognition and Signal Processing.

Tong, H. (1990). *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

van Dijk, D., Terasvirta, T., and Franses, P. H. (2002). Smooth transition autoregressive models– a survey of recent developments. *Econometric Reviews*, 21:1–47.

West, M. (1986). Bayesian model monitoring. *Journal of Royal Statistical Society B*, 48:70–78.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.