Dynamic Bayesian smooth transition autoregressive models for nonlinear nonstationary time series

Álvaro E. Faria[1]
and
Alexandre J. Santos

April 5, 2018

Abstract

Dynamic Bayesian Smooth Transition Autoregressive (DBSTAR) models are proposed for nonlinear autoregressive time series processes as alternative to both the classical Smooth Transition Autoregressive (STAR) models of Chan and Tong (1986) and the Bayesian Simulation STAR (BSTAR) models of Lopes and Salazar (2005). Unlike those, DBSTAR models are sequential dynamic analytical models suitable for inherently nonstationary time series with nonlinear characteristics such as asymmetric cycles. As they are analytical, they also avoid potential computational problems associated with BSTAR models and allow fast sequential estimation of parameters.

Two types of DBSTAR models are defined here based on the method adopted to approximate the transition function of their autoregressive components, namely the Taylor and the B-splines DBSTAR models.

In an application to the well-known series of annual Canadian lynx trappings, a harmonic DBSTAR model that accounts for the cyclical component explicitly in a flexible yet parsimonious way, showed improved fitting when compared to both classical STAR and BSTAR models. Another application to a series of hourly electricity loading in southern Brazil illustrates the short-term forecasting accuracy as well as the flexibility of model formulation and fast computing DBSTAR models that account for various characteristics such as periodic behaviour within-the-day and within-the-week as well as temperature.

Keywords: Bayesian dynamic STAR models, polynomial forecasting models, nonlinear autoregressive models, Bayesian autoregressive forecasting models, short-term electricity load forecasting, B-splines approximation.

[1] School of Mathematics and Statistics, Faculty of Science Technology, Engineering and Mathematics, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK.

1 Introduction

The DBSTAR models proposed in this paper consist of Gaussian Bayesian state-space formulations based on polynomial *dynamic linear models* (DLMs) of West and Harrison (1997). They extend *smooth transition autoregressive* (STAR) models of Chan and Tong (1986) to allow parameters such as autoregressive, smoothing and observational variance to change in time. Parametric assessment is analytical with sequential prior-to-posterior distributional updating carried out by Kalman filtering for fast computation.

STAR models were developed from the *threshold autoregressive* (TAR) models proposed by Tong (1978) to address the failures of single linear models to represent certain properties of nonlinear stationary autoregressive time series processes such as asymmetric cycles, amplitude dependent frequencies and sudden changes (see Tong, 2011 for a review of developments of TAR type models over the last 30 years).

Basically, a STAR (and a TAR) model can be seen as a convex linear combination of two (or more) distinct linear *autoregressive* (AR) models of the same order. Each of the models is given a weight $w_i \in [0,1]$ $(i = 1, 2)$ in the combination that will average out the models. As the combination is convex, $\sum_i w_i = 1$, when a weight w_i is either 0 or 1, one of the combining AR models will be selected to operate. For that reason, a TAR or a STAR model is called a *regime-switching* model. Usually, the combining weights are specified through a conveniently chosen function which type defines the type of model. For instance, if a logistic function is adopted, the model is called a logistic STAR model.

For a stationary time series Y_t $(t = 1, 2, \ldots, T)$, a Gaussian STAR model of order p, STAR(p), $p \in \Re^+$, with two *regimes*, can be represented by the combination

$$Y_t = \pi(\cdot)\underline{z}_t\underline{\phi}_1' + [1 - \pi(\cdot)]\underline{z}_t\underline{\phi}_2' + \epsilon_t \quad ; \quad \epsilon_t \sim N(0, \sigma^2) \tag{1}$$

where for $i = 1, 2$, $\underline{\phi}_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})$ are $(p + 1)$-dimensional vectors with element ϕ_{ij} $(j = 0, 1, \ldots, p)$ representing an AR coefficient associated with each component j of the regime i; $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$ is a $(p + 1)$-dimensional vector with element y_{t-j} representing a realisation of the process Y_{t-j} at time $t - j$. The weight $\pi(\cdot)$, called transition function, is a function (of its arguments only) in the range [0, 1]. Note that in this paper, an underlined character is used to represent a vector, a matrix is represented by a boldface capital character and a prime is used to denote transposition.

The original TAR model of Tong (1978) adopted an indicator function of predefined threshold values $r_1, \ldots, r_n$ (for $r_i \in \Re$) as the weight $\pi(r_1, \ldots, r_n) = 1$ when at least one $r_j = 1$ $(j = 1, ..., n)$ and $\pi(0, \ldots, 0) = 0$ for model transition. Tong and Lim (1980) introduced the *self-exciting* TAR (SETAR) models where either a past observational value y_{t-d}, or an exogenous variable s_{t-d}, could be used as a threshold value in the indicator transition function, that is, $\pi(y_{t-d}) = 1$ for $Y_t <= y_{t-d}$ and $\pi(y_{t-d}) = 0$ for $Y_t > y_{t-d}$ for a given integer $d \in \Re^+$. The abrupt transitions between AR models in TAR and SETAR models prevent the use of maximum likelihood estimation methods due to the discontinuities introduced on their likelihood functions at transition points (see e.g. Terasvirta, 1994) and ordinary least squares approaches are usually adopted for inference. To avoid abrupt transitions, Chan and Tong (1986) proposed the STAR models with use of conveniently

chosen smooth transition functions, such as the commonly used logistic function

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

This function has 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, the transition variable is either a chosen external variable or a lagged dependent past value 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 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 at $1/2$ and the *logistic* STAR model is reduced to the average between the combining AR models. As γ increases, the logistic function tends to a step function and the transition from one regime to the other becomes more abrupt. Note that during a transition period, the combining model is nonlinear in form and usually nonlinear least square (NLS) methods are adopted for parametric estimation as described by van Dijk et al. (2002) and Terasvirta (2005). The NLS methods are based on linear approximations of the nonlinear transition function. We refer to those STAR models as *classical* STAR models throughout. Other, less common, transition functions are the second-order logistic function, the exponential function (Terasvirta, 1994), the cumulative distribution function (Chan and Tong, 1986) and the hyperbolic tangent function (Bacon and Watts, 1971). Each of those functions has its own particular set of parameters that characterises how it effects the transition.

Bayesian approaches for TAR models and their variants were initially proposed by Geweke and Terui (1993) and Chen and Lee (1995), and further developed by Chen (1998), Campbell (2004), Lubrano (2000) and Lopes and Salazar (2005). All those approaches are based on Markov Chain Monte Carlo (MCMC) simulation methods such as sampling importance resampling (Gelfand and Smith, 1990) for Bayesian inference due to the loss of analytical tractability in calculating posterior distributions (see, e.g. Bauwens et al., 1999).

Except for Lopes and Salazar (2005), that treat the order p of the combining AR models as unknown and proceed to estimate it's value from data, all the other proposed methods assume p to be fixed a priori. In fact, Lopes and Salazar (2005) adopted a Gibbs sampler approach for inferences about $\underline{\phi}_1, \underline{\phi}_2, \gamma, c, d$ and σ^2 of the logistic STAR when p is known, and a reversible jump MCMC algorithm (Green, 1995) for posterior assessments when p is unknown.

A common characteristic of all those approaches is that they are, without exception, static non-sequential methods for nonlinear but stationary AR processes. In this paper, the models of Lopes and Salazar (2005) are referred to as *Bayesian Simulation STAR* (BSTAR) models. It is also worth noting that computational Bayesian inference approaches such as the BSTAR models above are non-parsimonious computer intensive numerical simulation models relying 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 applications that require fast sequential prior-to-posterior parametric estimation and forecasting.

The DBSTAR models proposed here address some of those limitations. Like the classical STAR models, the AR order p and the delay parameter d are fixed a priori in a

DBSTAR model. When those are unknown, and initial data is available, a model selection approach can be used for determining the optimal values for those. Similarly to BSTAR models, prior distributions must be specified for the state parameters of a DBSTAR model. Those are functions of the AR coefficients and the smoothness parameter as well as the observational variance associated with STAR type models. Unlike both classical STAR and BSTAR models, a DBSTAR model can be formulated to explicitly account for (and parsimoniously represent) components observed in the underlying time series, that is, level, trend, seasonality and cycle. DBSTAR models are based on particular formulations of the polynomial *dynamic linear models* (DLM's) of West and Harrison (1997) that allow hierarchical component modelling by superposition of models appropriate for specific components.

The observational variance in a DBSTAR model is allowed to change in time albeit slowly and steadily to account for possible extra variation in the series as those caused by either additional stochastic factors that have not been modelled (e.g. inaccuracies in the timing of observations, change in data handling procedures, etc), or by improper specification of a variance function for deterministic changes in observational variability. The later cause may include the extra variability not accounted for by the polynomial approximation of the transition function adopted by a DBSTAR model. The slow changes in the observational variance of a DBSTAR model are determined sequentially via data assimilation approach as described in Section 2.

In this paper, two types of DBSTAR models are defined, the Taylor DBSTAR models based on approximating the adopted transition function by a Taylor series expansion, an the B-splines DBSTAR models which use B-spline functions for that purpose as we shall see. B-splines are constructed from piecewise polynomial functions joined at knots created by dividing the underlying interval into parts. Those functions satisfy weak differentiability conditions that guarantee the continuity and smoothness of the resulting function. The reader can refer to de Boor (1978), Dierckx (1993) and Eilers and Marx (1996) amongst others for more details on B-splines.

This paper is organised as follows. Section 2 defines DBSTAR models in general including the harmonic DBSTAR model that accounts for a cyclical component explicitly as well as Taylor and B-splines types. The algorithm for sequential parameter estimation and forecasting is also described. Section 3 shows a comparative analysis of the application of STAR, BSTAR and DBSTAR models to the well known Canadian lynx data. Section 4 shows the analysis and application of harmonic DBSTAR models to the hourly series of electricity loading in southern Brazil to illustrate a case where both classical STAR and the BSTAR models are not appropriate. Discussions on the models and results are presented in Section 5.

2 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 a dynamic smooth transition function, $\pi(s_t; \gamma_t, c_t)$, similar to the logistic in (2) but with both the smoothing parameter γ and the threshold parameter c of $\pi(s_t; \gamma, c)$ allowed to change in time, is represented by a polynomial approximation. This paper considers two distinct approximations, the Taylor series expansion and the B-spline function, that characterise the Taylor and the B-spline DBSTAR models, respectively. Despite based on the logistic transition function, the development below can without loss be adopted for any other transition function that can be approximated by a polynomial function.

So, for a dynamic logistic transition function $\pi(s_t; \gamma_t, c_t)$ 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. We use underlined characters to represent vectors, boldface to represent matrices and prime to denote transposition throughout.

The observational and the system equations of a DBSTAR (r, p) are respectively given by

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

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

where $\underline{F}_t' = [\underline{z}_t, B_1(s_t)\underline{z}_t, \ldots, B_{r-1}(s_t)\underline{z}_t, B_r(s_t)\underline{z}_t]$ is a *known* $(r+1)(p+1)$-dimensional vector of polynomial regression variables $B_i(s_t)\underline{z}_t$ $(i = 0, 1, \ldots, r)$ with $B_i(s_t)$ known functions of s_t which form depends on the approximation used for $\pi(s_t; \gamma_t, c_t)$ and $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-p})$; $\underline{\theta}_t$ is the state vector containing *unknown* parameters associated with the components of $\underline{F}_t$, i.e, $\underline{\theta}_t^{'} = (\underline{\theta}_0, \underline{\theta}_1, \ldots, \underline{\theta}_r)_t$ with elements $\underline{\theta}_{it} = (\theta_{i0}, \theta_{i1}, ..., \theta_{ip})_t$ where the elements of $\underline{\theta}_{0t}$ are $\theta_{0jt} = \phi_{1jt} + \beta_{0t}\phi_{2jt}$ $(j = 0, ..., p)$ and $\underline{\theta}_{it} = \beta_{it}\underline{\phi}_{2t}$ $(i = 1, ..., r)$ such that $\theta_{ijt} = \beta_{it}\phi_{2jt}$; and β_{it} $(i = 0, 1, ..., r)$ are polynomial functions of γ_t and c_t only such that

$$\pi(s_t; \gamma_t, c_t) \simeq \sum_{i=0}^{r} \beta_{it}(\gamma_t, c_t)B_i(s_t) \ . \tag{5}$$

Note that in this form, the parameters γ_t and c_t of $\pi(s_t; \gamma_t, c_t)$ are separated from the observable s_t such that they can be included together with the dynamic AR coefficients $\underline{\phi}_{it} = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})_t$ where ϕ_{ijt} is the coefficient j $(j = 0, 1, \ldots, p)$ of the AR regime i $(i = 1, 2)$, into the *unknown* state vector $\underline{\theta}_t$ in (3) above, while s_t can be included together with past values of y_t into the *known* vector $\underline{F}_t$. For simplicity, the number of regimes is restricted to two AR(p) models here, albeit the methodology can be relatively straightforwardly extended to multiple regimes and differing AR orders.

The observational variance Σ_t in (3) is assumed *unknown* and defined as $\Sigma_t = k_t V$, 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). V is the unknown variance scale parameter that is allowed to change stochastically. While a

suitable chosen variance law can model systematic changes in the observational variability in time, we assume that Σ_t may change stochastically 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 dynamic underlying observational precision Σ_t^{-1} is assumed to relate to Σ_{t-1}^{-1} according to

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

where $\lambda_t \sim Beta\left(\frac{\delta_V n_{t-1}}{2}, \frac{(1-\delta_V)n_{t-1}}{2}\right)$, n_{t-1} is the number of observations up to $t-1$ and δ_V $(0 < \delta_V \leq 1)$ is a discount factor for the observational scale variance. In practice, δ_V typically assumes a value in the range from 0.9 to 0.99 (West and Harrison, 1997). The closer to unity, the less the variance changes due to changes in the data. A discount factor of one corresponds to a constant variance.

The state vector $\underline{\theta}_t$, conditional on $\underline{\theta}_{t-1}$, is assumed to follow a multivariate Student-t distribution with n_{t-1} degrees of freedom with mean vector $\mathbf{G}_t\underline{\theta}_{t-1}$ and covariance matrix $\mathbf{W}_t$. $\mathbf{G}_t$ is a *fixed* $(r+1)(p+1) \times (r+1)(p+1)$ state evolution matrix with elements g_{ijt} chosen according to the desired structural form of association between $\underline{\theta}_t$ and $\underline{\theta}_{t-1}$, and $\mathbf{W}_t$ is a $(r+1)(p+1) \times (r+1)(p+1)$ system evolution covariance matrix, for which a discount factor δ_W $(0 < \delta_W \leq 1)$ can be used to control the degree of change allowed in time, 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 $\underline{\theta}_t$ as we shall see in Section 2.3. In practice, δ_W typically assumes values between 0.8 and 0.99.

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 $t = 0$, D_0 includes the following initial prior information. The state vector is given by the specification of a multivariate Student-t distribution with n_0 degrees of freedom

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

where, $\underline{m}_0$ and $\mathbf{C}_0$ are the prior mean vector and covariance matrix, respectively. Similarly, Σ_0 follows an *inverse Gamma* distribution (IG) with hyperparameters n_0 and S_0

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

We now proceed to define the Taylor and the B-splines DBSTAR models depending on the approximation method used for the dynamic smooth transition function in (5).

2.1 Taylor DBSTAR models

A *Taylor DBSTAR* model is defined as the DBSTAR model for which $B_i(s_t)$ in (5) is a polynomial function of order i of the form

$$B_i(s_t) = s_t^i \tag{6}$$

and $\beta_{it}(\gamma_t, c_t)$ is obtained by expressing the Taylor series expansion of $\pi(s_t; \gamma_t, c_t)$ in the vicinities of $s_t = c_t$. So, a Taylor DBSTAR(r, p) model is characterised by the observational equation (3) above where the Taylor series expansion of the transition function is truncated at order r. Thus, at each time t a Taylor DBSTAR(r, p) model corresponds to a STAR model of order p which transition function is approximated by its Taylor expansion truncated at order r. Notice that the Taylor series expansion approximates $\pi(s_t; \gamma_t, c_t)$ better in the vicinities of $s_t = c_t$, so that, at each time t, the approximation changes for changes in c_t.

2.2 B-splines DBSTAR models

A *B-spline DBSTAR* model on its turn is defined as the DBSTAR model for which the function in (5) is such that $B_i(s_t)$ is a B-spline basis function (a piecewise polynomial function) and $\beta_{it}(\gamma_t, c_t)$ are the associated coefficients. The B-spline of basis functions $B_i(s_t)$ of degree q, given a number n of knots in an interval, say, $[s_{min}, s_{max}]$ around s_t, can be defined as (see e.g. Wold, 1974)

$$B_i(s_t) = \frac{\sum_{k=i-q}^{i+(q-2)} (s_t - \xi_k)_+^q}{\prod_{j=i-q, j\neq k}^{i+(q-2)} (\xi_k - \xi_j)} \ , \tag{7}$$

where $(s_t - \xi_k)_+$ is $(s_t - \xi_k)$ when $s_t > \xi_k$ and zero otherwise, ξ_k is such that

$$\xi_k = \begin{cases} \xi_1 - (1-k)(\xi_1 - s_{min}) & ; k \leq 0 \\ \xi_n + (k-n)(s_{max} - \xi_n) & ; k \geq n+1. \end{cases} \tag{8}$$

Computationally, the B-splines are obtained recursively by the Cox-de Boor algorithm (de Boor, 1978) as

$$B_{i,q}(s_t) = \begin{cases} \frac{s_t - \xi_i}{(\xi_{i+q-1} - s_t)} B_{i,q-1}(s_t) + \frac{\xi_{i+q} - s_t}{\xi_{i+q} - \xi_i} B_{i+1,q-1}(s_t) & ; \xi_i \leq s_t < \xi_{i+1} \\ 0 & ; \text{otherwise,} \end{cases} \tag{9}$$

where $\xi_1, ...\xi_n$ is an ordered non-decreasing set of knots in the interval, and the number of terms in the approximation (5) is $r = q + n - 2$. This recurrence allows the computation of $B_{i,q}$ (or a B-spline B_i of degree q in our notation) by repeatedly forming positive linear combinations of positive quantities, starting with the B-spline of degree 1, $B_{i,1}(s_t) = I_{(\xi_i, \xi_{i+1}]}(s_t)$ where the indicator function $I_{(a,b]}(x) = 1$ for $a \leq x < b$ and has value zero values of x outside the interval $(a, b]$.

Except for $\beta_{0t} = 0$, the remaining coefficients β_{it} in (5) are control points to be estimated. Note that, as pointed out by Eilers and Marx (1996), B-splines are rather

attractive as base functions for univariate regression in which a linear combination of third-degree (or cubic) B-splines gives a curve smooth enough to provide a good fit. In particular this is also the case here for the logistic transition function (2) or indeed for any of the usual alternative transition function such the second-order logistic and the exponential. So, in our case, an appropriate order r in (5) is usually determined by the number n of knots chosen. In the splines regression context, this choice can be a complex task and statistics with penalties for overfitting are used in determining the optimal number. In our context n is determined via model selection approach.

Some general properties of B-splines that guarantee continuity at knots are that (i) each subinterval will be covered by $q+1$ B-splines (polynomial pieces) of degree q such that they join at q inner knots; (ii) at the joining points derivatives up to the order $q-1$ are continuous; (ii) the B-spline is positive on a domain spanned by $q+2$ knots and is zero elsewhere; and (iii) at a given s_t, $q+1$ B-splines are nonzero. See e.g. Eilers and Marx (1996) for more details.

Note that some parameters in a Taylor or B-splines DBSTAR model are not treated as unknown parameters to be estimated but fixed a priori. The main reason for that is to preserve analytical tractability in the Bayesian parametric updating that allows fast sequential computations. Namely, the following parameters must be fixed a priori (i) the autoregressive order p, (ii) the delay parameter d when the transition variable $s_t = y_{t-d}$, (iii) the discount factor δ_W of the state variance $\mathbf{W}_t$, (iv) the discount factor δ_V of the observational variance Σ_t, (v) the order r of the Taylor series expansion in a Taylor DBSTAR model, (vi) the number n of knots and the degree q of the polynomial B-spline functions in a B-splines DBSTAR model, and (vii) the harmonic order h for DBSTAR models with a cyclic component. For the Canadian lynx trappings and electricity loading applications in Sections 3 and 4, optimal values of those were obtained via model selection based on conditional log-smoothing likelihoods similar to Bayes' factors (see e.g. West, 1986) as we shall see.

Also, notice 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.

Now that DBSTAR(r, p) models have been defined, the algorithm for their sequential parametric updating and forecasting is described in the following section.

2.3 Sequential parametric updating and the forecasting function

Similarly to a univariate DLM with variance discounting (West and Harrison, 1997), the sequential prior-to-posterior parametric updating and forecasting can be obtained from the theory of the normal distribution and the normal-inverse-gamma conjugate analysis to give the following algorithm.

At time $t-1$, assume that the posterior distribution of $\underline{\theta}_{t-1}$ conditional on D_{t-1} is a

multivariate Student-t distribution with n_{t-1} degrees of freedom ($T_{n_{t-1}}$), mean $\underline{m}_{t-1}$ and covariance $\mathbf{C}_{t-1}$, that is,

$$(\underline{\theta}_{t-1} \mid D_{t-1}) \sim T_{n_{t-1}} \left(\underline{m}_{t-1}, \mathbf{C}_{t-1}\right).$$

The posterior distribution of the observational variance is an inverse-gamma (IG) distribution, that 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 a point estimate of the observational variance at time $t-1$.

At time t, before observing Y_t, let the prior distribution for the state vector $\underline{\theta}_t$ be

$$(\underline{\theta}_t \mid D_{t-1}) \sim T_{n_{t-1}} \left(\underline{a}_t, \mathbf{R}_t\right),$$

where $\underline{a}_t = \mathbf{G}_t \underline{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t^{'} + \mathbf{W}_t$. Also let the prior distribution for the observational variance be given by

$$(\Sigma_t \mid D_{t-1}) \sim IG\left(\frac{\delta_V n_{t-1}}{2}, \frac{\delta_V n_{t-1} S_{t-1}}{2}\right).$$

The one-step-ahead forecast distribution for Y_t is given by

$$(Y_t \mid D_{t-1}) \sim T_{\delta_V n_{t-1}} \left(f_t, Q_t\right),$$

where $f_t = \underline{F}_t^{'} \underline{a}_t$ and $Q_t = \underline{F}_t^{'} \mathbf{R}_t \underline{F}_t + k_t S_{t-1}$ are the forecasting mean and variance respectively.

After observing $Y_t = y_t$, the posterior distribution of $\underline{\theta}_t$ at time t is

$$(\underline{\theta}_t \mid D_t) \sim T_{n_t} \left(\underline{m}_t, \mathbf{C}_t\right),$$

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

$$\mathbf{C}_t = \frac{S_t}{S_{t-1}} [\mathbf{R}_t - (\mathbf{R}_t \underline{F}_t / Q_t) \underline{F}_t^{'} \mathbf{R}_t^{'}].$$

The posterior distribution for the observational variance is given by

$$(\Sigma_t \mid D_t) \sim IG\left(\frac{n_t}{2}, \frac{n_t S_t}{2}\right),$$

with $S_t = S_{t-1} + \frac{S_{t-1}}{n_t}(\frac{e_t^2}{Q_t} - 1)$ and $e_t = Y_t - f_t$ is the one-step-ahead forecasting error.

At a future time $t + \kappa$, the conditional forecast distributions for the states and observations, given the data up to time t, are recursively determined. Since these forecast distributions are Student-t distributions, it is enough to compute their means and variances.

For $\underline{a}_t(0) = \underline{m}_t$ 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 κ-steps ahead forecasting distribution ($\kappa \geq 1$) for $Y_{t+\kappa}$ is

$$(Y_{t+\kappa} \mid D_t) \sim T_{\delta_V n_t}\left(f_t(\kappa), Q_t(\kappa)\right),$$

with forecasting function

$$f_t(\kappa) = \underline{F}'_{t+\kappa}\underline{a}_t(\kappa)$$

and variance

$$Q_t(\kappa) = \underline{F}'_{t+\kappa}\mathbf{R}_t(\kappa)\underline{F}_{t+\kappa} + k_{t+\kappa}S_{t+\kappa},$$

where $\underline{a}_t(\kappa) = \mathbf{G}_{t+\kappa}\underline{a}_t(k-1)$ and $\mathbf{R}_t(\kappa) = \mathbf{G}_{t+\kappa}\mathbf{R}_t(\kappa-1)\mathbf{G}'_{t+\kappa} + \mathbf{W}_{t+\kappa}$ are the mean and the covariance matrix of the state vector $(\underline{\theta}_{t+\kappa} \mid D_t) \sim T_{n_t}(\underline{a}_t(\kappa), \mathbf{R}_t(\kappa))$.

The formulation of harmonic DBSTAR models used in the applications in Sections 3 and 4 is introduced next.

2.4 Harmonic DBSTAR models for processes with cycles

This sub-section describes DBSTAR models for modelling observed cyclic behaviour in terms of cyclical components explicitly. Similarly to seasonality, the explicit modelling of long term cyclic behaviour allows accounting for changes in that behaviour in a forecasting model. Applications of *dynamic linear models* (DLM's) with cyclical components to contexts in which periods of cyclical components are known are relatively straightforward in the contexts of seasonal modelling and time-varying harmonic analysis (West and Harrison, 1997) as analytical parametric updating is viable. Problems of inference about (uncertain) wavelengths of time-varying cyclical components in those models have been addressed albeit at the expense of loss of analytical tractability that requires the use of stochastic simulation methods (West, 1995).

The formulation of the *Harmonic* DBSTAR (HDBSTAR) model here, like the harmonic component DLMs with Fourier form of West and Harrison (1997), use Fourier form representations of the periodic behaviour that allows for modelling changes in amplitude and phase for fixed number of harmonics preserving thus the analytical tractability of parametric posterior distributions and forecasting functions that allows the use of the algorithm described in Section 2.3.

Basically, HDBSTAR models extend the set of quadruple $\{\underline{F}_t, \mathbf{G}_t, \Sigma_t, \mathbf{W}_t\}$, with $\underline{F}_t = (\underline{F}_{1t}, \underline{F}_{2t})$, $\mathbf{G}_t = (\mathbf{G}_{1t}, \mathbf{G}_{2t})$ and $\mathbf{W}_t = (\mathbf{W}_{1t}, \mathbf{W}_{2t})$, where $\underline{F}_{1t}, \mathbf{G}_{1t}$ and $\mathbf{W}_{1t}$ are associated to the nonlinear autoregressive components as in (3) and (4), and $\underline{F}_{2t}, \mathbf{G}_{2t}$ and $\mathbf{W}_{2t}$ are associated to the cyclical component. 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

$$\left(Y_t \mid \underline{\theta}_t, \underline{\psi}_t\right) \sim N\left(\underline{F}'_{1t}\underline{\theta}_t + \underline{F}'_{2t}\underline{\psi}_t, \Sigma_t\right) \tag{10}$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{11}$$

$$\left(\underline{\psi}_t \mid \underline{\psi}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\psi}_{t-1}, \mathbf{W}_{2t}\right) \tag{12}$$

where $\underline{\psi}'_t = [\underline{\psi}_{1t}, ..., \underline{\psi}_{ht}]$, $\underline{\psi}_{jt} = (a_j, b_j)_t$ with a_{jt} and b_{jt} being the unknown Fourier coefficients of each harmonic $S_j(t) = a_{jt}cos(\omega_j t) + b_{jt}sin(\omega_j t)$ $(j = 1, ..., h)$. The $2h$-dimensional vector $\underline{F}_{2t}$ is a canonical partitioned vector associated to the harmonics in $\underline{\psi}_t$, with 1 in an harmonic position and 0 otherwise. For example, $\underline{F}_{2t} = [1, 0]$ for $h = 1$ harmonic, $\underline{F}_{2t} = [1, 0, 1, 0]$ for $h = 2$ harmonics, and so forth. The frequency of each harmonic is $\omega_j = 2j\pi/\tau_c$, where τ_c is the period of the cycle. The evolution matrix $\mathbf{G}_{2t}$ of the cyclical component is a block diagonal matrix $\mathbf{G}_{2t} = diag(\mathbf{H}_{1t}, ..., \mathbf{H}_{ht})$ where $\mathbf{H}_{jt}$ is the harmonic matrix with trigonometric elements such that $|\mathbf{H}_{jt}| = sin^2(w_j t) + cos^2(w_j t) = 1$. 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 cyclical component of the underlying process. For cases where a large enough initial dataset is available (as are the cases in this paper) to enable the investigation of the cyclic behaviour, an optimal value of h can be determined, for example, by a stepwise model selection approach.

3 The Canadian lynx data application

In this section, three distinct DBSTAR models are applied to the well-known Canadian lynx series. Two formulations, a Taylor and a B-splines DBSTAR model, account for the series cyclic behaviour through high order AR components, and another, a B-splines HDBSTAR model, accounts for the cycle explicitly through low order harmonic components. The purpose of this application is twofold. On one hand, it compares and validates the proposed DBSTAR models against the classical STAR (applied by Terasvirta, 1994) and the Bayesian simulation STAR (applied by Lopes and Salazar, 2005) models, and on the other hand, it illustrates some of the main advantages of adopting Bayesian sequential models with dynamic parameters such as quantifying the changes in the cyclic component along time. Terasvirta (1994) and Lopes and Salazar (2005) make use of the whole Canadian lynx series to identify and estimate their models. Similarly, a retrospective (smoothing) analysis using the whole series was carried out to identify the three DBSTAR models to be used in the sequential fitting (or filtering). Also, like Terasvirta (1994) and Lopes and Salazar (2005), the *mean absolute error* (MAE) and *root mean squared error* (RMSE) statistics were adopted as measures of goodness of fit.

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 of 114 observations. It was originally recorded to support understanding of 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

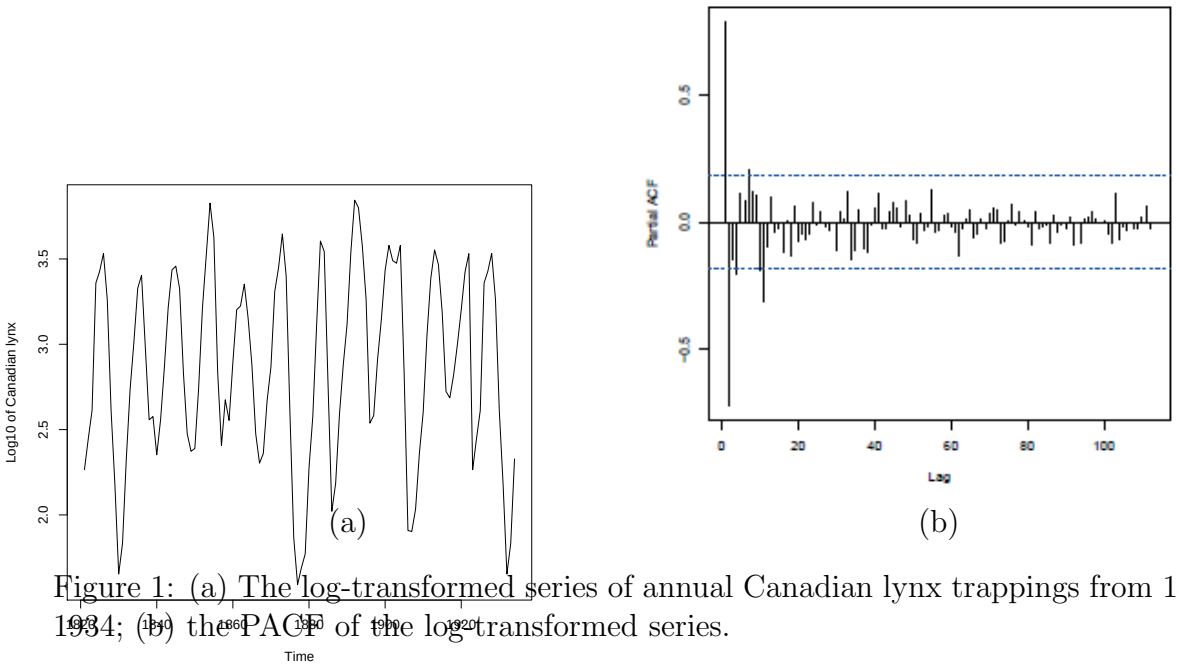


Figure 1: (a) The log-transformed series of annual Canadian lynx trappings from 1821 to 1934; (b) the PACF of the log-transformed series.

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, e.g. Elton and Nicholson, 1942, and references therein).

Similarly to other studies of this series, including Chan and Tong (1986) and Lopes and Salazar (2005), the original series was $\log_{10}$-transformed here to remove the marked right-skewness of the data as well as to allow the comparative analysis with the classical STAR and the BSTAR models. The transformed series, as can be seen in Figure 1(a), shows no evident trend but a clear periodic repetitive behaviour. The corresponding *partial auto-correlation function* (PACF), Figure 1(b), indicates significant autocorrelations at lags 1, 2 and 11. However, note that the PACF assumes that the relationship between y_t and y_{t-k} $(k = 1, \dots)$ is linear, and thus may not be an appropriate tool for model order identification in this case. In fact, a graphical analysis of the data showed lack of linearity for most lags of the original series although some improvement was observed for the transformed data. Consistently with Elton and Nicholson (1942), a periodogram of the whole data showed a spike around the 0.1 frequency indicating a persistent cyclical behaviour with a wavelength of 10-years.

3.1 Model specification and initial values

In this application, logistic DBSTAR models based on dynamic logistic smooth transition functions were considered. A number of candidate models were initially formulated and

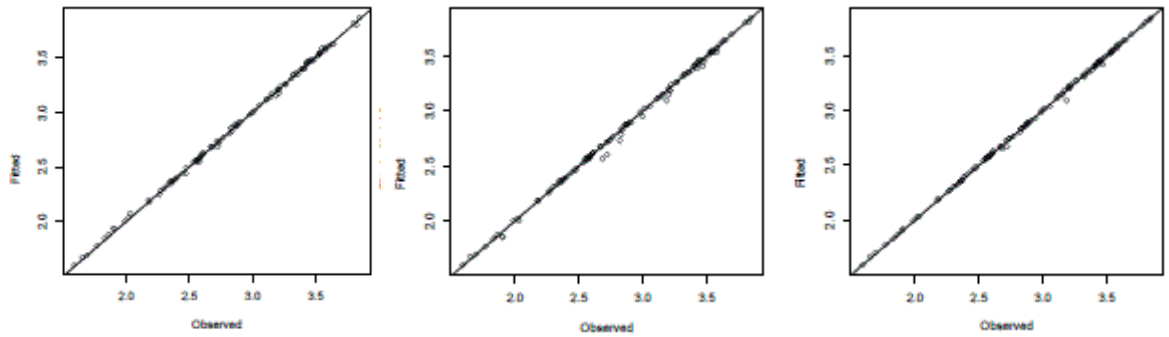
their log-smoothing likelihoods calculated conditionally on different values of d, p, r and h. A Taylor DBSTAR$(3, 12)$, a cubic B-splines DBSTAR$(7, 12)$ with five knots and a cubic B-splines HDBSTAR$(7, 2, 2)$ with five knots were the models selected in the smoothing analysis. The optimum delay parameter of the dynamic logistic transition function was found to be $d = 3$ for all models, that is $s_t = y_{t-3}$. Note that both Terasvirta (1994) and Lopes and Salazar (2005) found the same result, the former via model selection and the later via reversible jump MCMC. However, differently from Lopes and Salazar (2005) who like Terasvirta (1994) found $p = 11$ with the largest posterior probability, our approach found $p = 12$ as the optimum AR order for the Taylor and the B-splines DBSTAR models. The optimal values for the number of harmonics and AR order were $h = 2$ and $p = 2$ respectively for the B-splines HDBSTAR model. For parsimony, the Taylor DBSTAR model with $r = 3$ was also chosen as higher orders only provided marginal improvements. Similarly, a cubic B-splines DBSTAR model with five knots was also chosen for the same reason.

Therefore, for the Taylor DBSTAR model $\underline{F}_t' = [\underline{z}_t, y_{t-3}\underline{z}_t, y_{t-3}^2\underline{z}_t, y_{t-3}^3\underline{z}_t]$ is a vector of polynomial regression variables, where $\underline{z}_t = (1, y_{t-1}, \ldots, y_{t-12})$; the state vector $\underline{\theta}_t = [\underline{\theta}_{0t}, \underline{\theta}_{1t}, \underline{\theta}_{2t}, \underline{\theta}_{3t}]$ has elements that are functions of $\underline{\phi}_{1t}$, $\underline{\phi}_{2t}$, γ_t and c_t determined by the Taylor series expansion (of order 3) associated with a logistic STAR of order 12 as described in Section 2. For the B-splines DBSTAR model, $\underline{F}_t' = [\underline{z}_t, B_1(y_{t-3})\underline{z}_t, ..., B_7(y_{t-3})\underline{z}_t]$ where $B_i(y_{t-3})$ from (7) is a basis function of degree $q = 3$ with $n = 5$ knots.

The state evolution matrix, for both the Taylor and the B-spline DBSTAR models above, was set as $\mathbf{G}_t = \mathbf{I}$, the identity matrix, that is, no deterministic evolution changes in time for the state parameters was assumed. Also, the optimal values of the discount factors associated with the observational variance, Σ_t, and the state covariance $\mathbf{W}_t$ of those models were found to be $\delta_V = 1$ and $\delta_W = 0.85$ respectively. Those discount values were consistent with the fact that the observational variance for this series was not detected to be stochastically changing in time, and the system variance-covariance changes albeit slowly to account for changing variability in the dynamic parameter adaptation to data. For the B-splines HDBSTAR model, $\underline{F}_{1t}'$, $\mathbf{G}_{1t}$, $\Sigma_t = \Sigma$ and $\mathbf{W}_{1t}$, are specified similarly to the B-spline DBSTAR above, while the cyclical component has $\underline{F}_{2t} = [1, 0, 1, 0]$ and $\mathbf{G}_{2t}$ with two diagonal harmonic matrices associated with the 2 harmonics. $\delta_W = 0.8$ was the optimal discount value associated with the variance-covariances $\mathbf{W}_{2t}$ of the cyclical component. The period of the cycle was selected as $\tau_c = 10$ years as in Elton and Nicholson (1942). A period of 11 years as adopted by Chan and Tong (1986) and Lopes and Salazar (2005) was also tested but produced inferior results. Once the amplitudes of the retrospective estimates of the two harmonics in both models were compared, higher magnitudes for models with a 10-year period were observed. Initial values for the hyper-parameters of non-informative prior distributions were set as $\underline{m}_0 = \underline{0}$ and $\mathbf{C}_0 = 100\mathbf{I}$ for $(\underline{\theta}_0|D_0)$, and $n_0 = 1$ and $S_0 = 1$ for Σ_0.

3.2 Comparative fitting analysis

Figure 2 shows the plots of fitted against observed values for each of the tested DBSTAR models. Notice that in all three plots the points lie very close to the diagonal line in-



(a) (b) (c)

Figure 2: Observed series vs fitted models (a) Taylor DBSTAR$(3, 12)$, (b) B-splines DBSTAR $(7, 12)$ and (c) B-splines HDBSTAR$(7, 2, 2)$.

dicating all three models produced fitted values that were quite close to the observed data. However, a visual inspection indicates that the Taylor DBSTAR model apparently produced the closest overall fit, followed by the B-splines HDBSTAR and the B-splines DBSTAR models in that order. In particular, note that there were a few points in the range 2.5-3.5 in Figures 2(b) and (c) that seemed further apart from the diagonal line comparatively to Figure 2(a).

Similar plots, produced for the static versions of each of the formulated DBSTAR models (that is, with discount factors set to unity), showed points that were more scattered along the diagonal line indicating worse fitting to data.

Table 1 displays the MAE and the RMSE of the fitted logistic classical STAR(11) of Terasvirta (1994) and the logistic BSTAR(11) of Lopes and Salazar (2005) as well as those of the Taylor DBSTAR$(3, 12)$, the B-splines DBSTAR(7, 12) and the B-splines HDBSTAR$(7, 2, 2)$ models. Fitness measures of the static versions of each DBSTAR model (those set with $\delta_V = \delta_W = 1$) are also included. It is clear that the B-splines HDBSTAR(7, 2, 2) model with a MAE of 0.006 and a RMSE of 0.013 produced the best fit of all models. Its MAE is 20% lower than the second best fit of the Taylor DBSTAR(3, 12) although its RMSE is only marginally lower. The visual analysis related to the plots in Figure 2 where the Taylor DBSTAR(3, 12) seemed a better fit is thus not supported by those measures. Amongst the static models, the B-splines DBSTAR(7, 12) model outperformed both the Taylor DBSTAR(3, 12) and the B-splines HDBSTAR(7, 2, 2) models. The static Taylor DBSTAR(3, 12), with the largest MAE and RMSE of all DBSTAR models, still outperformed the classical STAR(11) and the BSTAR(11) models, albeit marginally.

3.3 Parametric time evolution

At each time t, after observing $Y_t = y_t$, the parameters $\underline{\phi}_{1t}, \underline{\phi}_{2t}, \gamma_t$ and c_t associated with a DBSTAR model may be estimated from the mean $\underline{m}_t$ of the posterior distribution of $\underline{\theta}_t$ by solving a system of polynomial equations. In this application, we used Wolfram's Mathematica within R to obtain the estimated AR coefficients for $\underline{\phi}_{1t}$ and $\underline{\phi}_{2t}$ of the

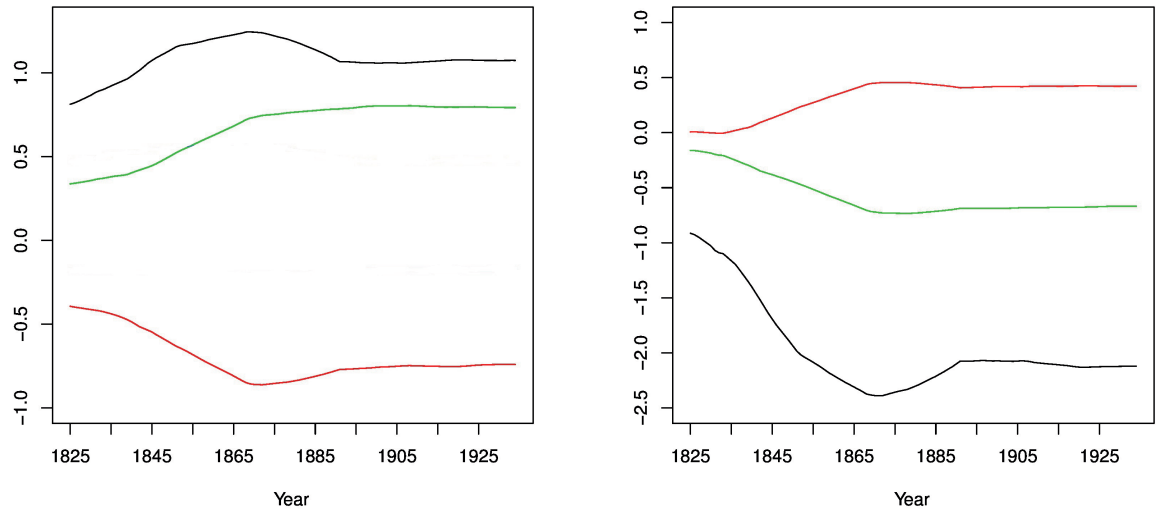
Model	MAE	RMSE
Lopes and Salazar (2005) - BSTAR(11)	0.118	0.153
Terasvirta (1994) - STAR(11)	0.142	0.179
Static Taylor DBSTAR$(3, 12)$	0.109	0.141
Static B-splines DBSTAR$(7, 12)$	0.061	0.082
Static B-splines HDBSTAR$(7, 2, 2)$	0.105	0.139
Taylor DBSTAR$(3, 12)$	0.012	0.015
B-splines DBSTAR$(7, 12)$	0.014	0.027
B-splines HDBSTAR$(7, 2, 2)$	**0.006**	**0.013**

Table 1: Mean Absolute Errors (MAE) and Root Mean Squared Errors (RMSE) of compared models

HDBSTAR$(7, 2, 2)$ model that can be seen in Figures 3 (a) and 3 (b), respectively. The coefficients $\underline{\phi}_{1t}$ fluctuated in the range from approximately -1.0 to 1.3 with no sharp changes along time. Most of the changes occurred from 1824 to the late 1880s after when they seem to have settled in their values with practically no changes. The coefficients $\underline{\phi}_{2t}$, however, had values in the range from approximately -2.5 to 0.5 with larger variations comparatively to $\underline{\phi}_{1t}$. Similarly, the changes in their values occurred before the 1880s after when they remained practically unchanged.

The posterior means of the harmonic coefficients were used to obtain the values of the combined harmonic function used to represent the cyclic component at each time. Figure 4 shows the resulting irregularly changing cyclic component estimated from the data. The amplitude of the cyclic component fluctuates in the range from approximately -0.76 to 0.70 and shows an increase from approximately 0.2 in absolute value in the period 1824-1831 to about 0.7 in 1932-1846 (with peaks of -0.76 and 0.69 in 1839 and 1944 respectively), followed by a decrease back to approximately 0.12 until 1870. From 1871 to 1891 the amplitude decreases even further to approximately 0.06 in absolute value. This is followed by gradual increases to peaks at approximately 0.6 levels (in absolute values) in the period 1904-1916 and fluctuations at slightly lower levels of approximately 0.3 thereafter. Those changes in time reflect the changes of the influence of the cyclic component on the one-step-ahead forecasts of lynx trappings by the HDBSTAR(7, 2, 2) model. The explicit account of that influence by the model seems to have a significant contribution to its much improved fitting over other STAR models as shown the in the section above.

Figure 5(a) illustrates the evolution in time of the smoothing parameter γ_t of the dynamic logistic transition function associated with the B-splines HDBSTAR(7, 2, 2) model. The smoothing parameter start at its largest value of 1.13 in 1821 and decreases steadily to approximately 1.08 in 1830 after when it slightly but steadily increases to 1.085 until 1865. From 1866 it decreases steadily to its lowest value of 1.032 in 1905 after when it starts to increase again to approximately 1.046 in 1934. The lower the value



(a) (b)

Figure 3: Evolution of (a) $\hat{\underline{\phi}}_{1t}$ and (b) $\hat{\underline{\phi}}_{2t}$ of the HDBSTAR$(7, 4, 2)$ model.

the larger the degree of smoothing and vice-versa. Note that the smoothing parameter ranges from 1.032 to 1.13 corresponding to quite smooth transitions and its change of value in time are thus not large enough to reflect a major change in the smoothness of transition function for this model. This is probably due to the differences between the estimated AR coefficients of each each regime not being too far apart as seen in the plots of Figure 3 with no abrupt switching in their combining weights. The smoothing parameters estimated by the classical STAR(11) and the BSTAR(11), 3.114 and 11.625 respectively, were considerably larger than those above.

Figure 5 (b) shows how the estimated threshold parameter, c_t, changes in time. Similarly to γ_t, c_t changes, in an apparently symmetrically opposite evolution direction from that of γ_t, within a limited range of values from approximately 1.04 to 1.22 (a maximum variation of 0.18 from minimum to maximum), corresponding to threshold values in the range 10.96 to 16.60 approximately in the scale of the original series. The threshold values estimated by the classical STAR(11) and the BSTAR(11), 2.73 and 3.504 respectively, corresponding to 537.03 and 3191.54 for the untransformed series, were considerably larger than those above. In a sense, the lower values of c_t in the HDBSTAR model, that are comparatively further apart from typical observed values y_t, partly compensates for its lower γ_t values (or comparatively larger degree of smoothing) by inflating the value of $\gamma_t(y_t - c_t)$ in the logistic transition function.

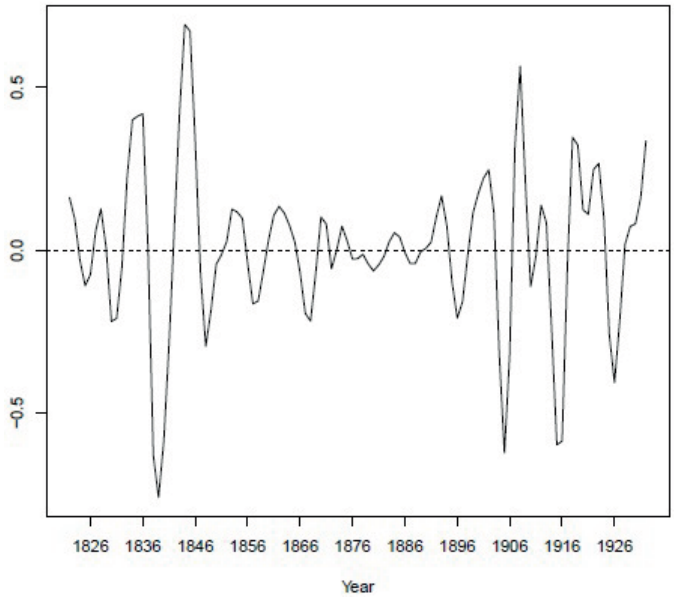


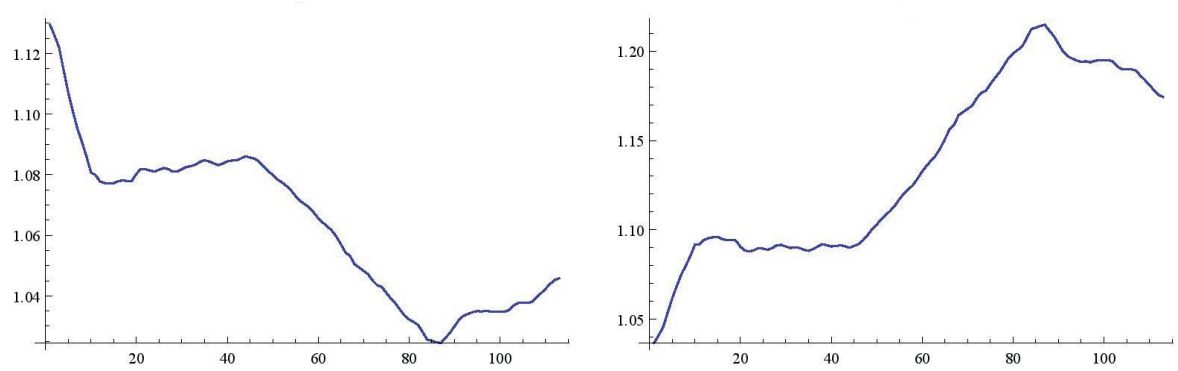
Figure 4: The cyclic component of the HDBSTAR$(7, 2, 2)$ model

4 Hourly electricity load application

The electricity data set analysed in this section consists of the hourly electricity load, measured in MegaWatts (MW), and temperature, in degree Celsius (°C), in the Southeast and Central-West regions of Brazil from the first hour on 1 June 2003 to the last hour on 30 June 2010, with a total of 62,088 hourly observations in the 7 year period. The hourly electricity load data are aggregated while the hourly temperature data are averaged across all the states in the Southeast and Central-West regions of Brazil. Calendar variables indicating normal weekdays, national holidays and bridge-holidays (i.e. days between midweek bank holidays that are near a weekend) were also used to account for their effects on the load. In order to measure the short-term forecasting performances of the adopted DBSTAR models, the last 720 observations (corresponding to one month of hourly data from 1 June 2010 to 30 June 2010) was used as out-of-sample so that 61,368 observations were used as in-sample data (from 1 June 2003 to 31 May 2010).

A preliminary analysis of the data has shown the following main characteristics: (i) a non-linear S-shaped relationship between load and temperature; (ii) different effects of weekdays, weekends and bridge-holidays on the load; (iii) a long-term positive trend of the load with increasing variability; and (iv) within-day and within-week changing periodic behaviour and variability of load.

In fact, the load tends to increase with increases in temperature although on a non-linear fashion resembling an S-shape as can be seen in Figure 6. The rate of increase in weekly average load is lower at lower levels of weekly average temperature (16-20°C) as compared with higher levels (26-30°C) with higher a rate of increase. Note that according to Pardo et al. (2002) and Dordonnat et al. (2008) an U-shape is usually observed in associations between electricity load and temperature in northern hemisphere countries



(a) (b)

Figure 5: Evolution in time of the (a) smoothing parameter γ_t and (b) the threshold value c_t of the dynamic logistic transition function associated with the B-Spline HDBSTAR(7, 2, 2) model.

where usually temperatures are very low in the winter and not very hot in the summer.

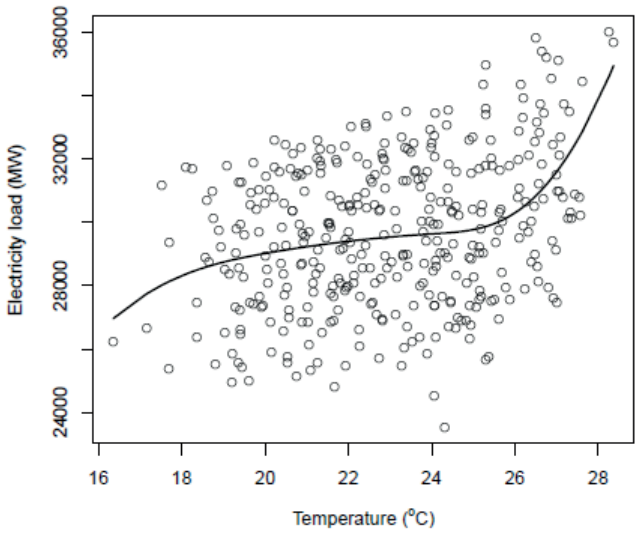
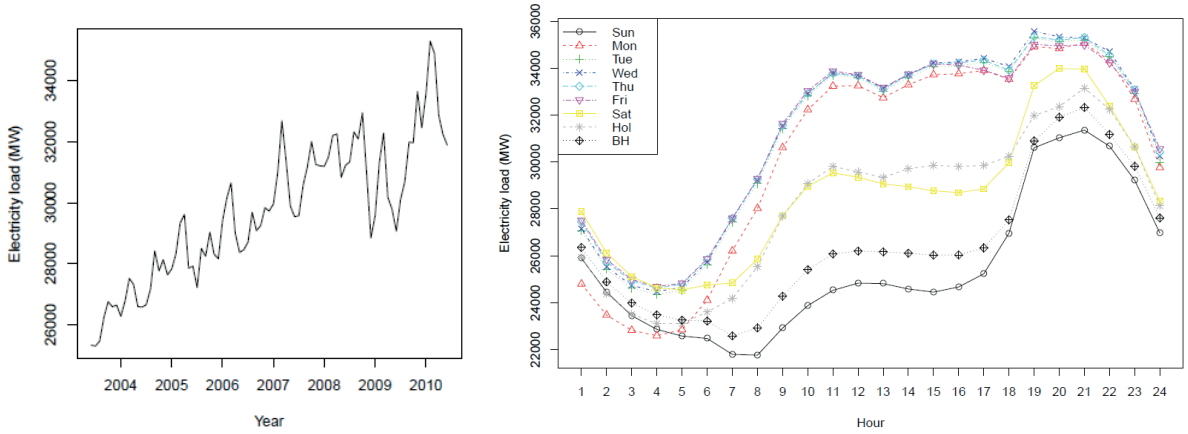


Figure 6: Average weekly electricity load (in MW) versus average weekly temperature (in °C) for the Southeast and Central-western regions in Brazil from 1st June 2003 to 30 June 2010.

The calendar effect shown in Figure 7(b) is such that the general shape of the load is similar for all days, that is the lowering of load from the early hours to a minimum followed by a sharp increase until the middle of the day, a stabilisation at a level until early afternoon, followed by a peak in the evening. Sundays and bridge-holidays show not only a little shift to the right in their location of peaks and troughs but also the lowest levels at almost all times except in the early hours when Mondays and normal holidays show higher levels. Similarly, Saturdays and holidays group together showing a smaller shift to the right and intermediate levels except in the early hours when the load is the largest.

Except for Monday that shows slightly lower loads at all times (except in the early hours when it shows the lowest of all), all the other weekdays are practically undistinguishable between themselves showing the largest levels of load overall with troughs at hour 4 and lower peaks at 11 and 15-16 and largest peaks from 19-21 hours.



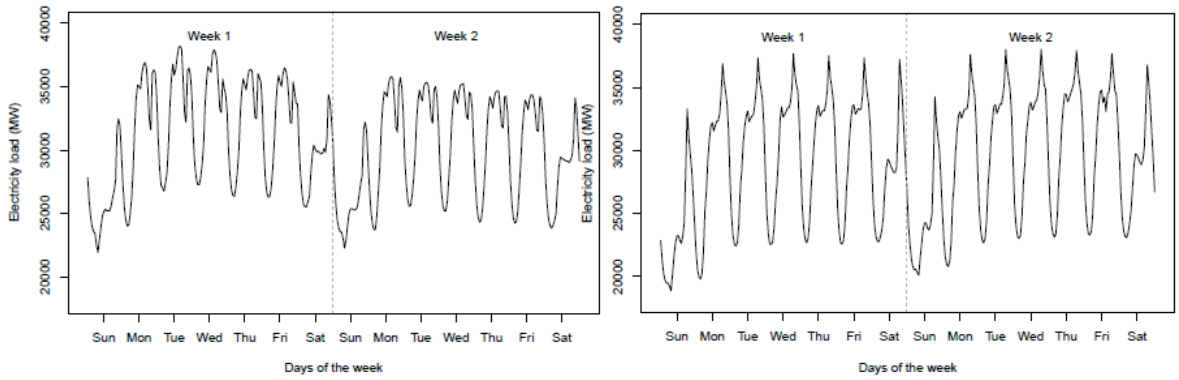
(a) (b)

Figure 7: (a) Time series plot of the monthly average electricity load, and (b) the average within-day electricity loads for weekdays, bank holidays and bridge holidays.

The load series has a long-term trend as can be seen from Figure 7(a) with the average monthly load showing a positive linear trend with increasing variability. The hourly load series on its turn shows within-day periodic variations that are rather similar for weekdays but change slightly for weekends, holidays and bridge-holidays as seen in Figure 7(b). The within-week variations are such that daily peaks change their patterns (or shape) along the year according to the season. As an example, the plots in Figure 8(a) and (b) show the daily loads in a two-week period in the summer and in a two-week period in the winter in 2009 respectively. Note the sharp pointy aspect in the evening peaks at 19 hours that is present during the winter is not present in the summer when the observed smoother peaks are thought to be influenced by the availability of natural lights early in the evenings helped by the change in the Brazilian summer time (when clocks are moved forward by one hour from October to February). There is also a strong within-year seasonal effect on load that is very much in line with the strong seasonal behavior of the temperature.

The Taylor and the B-splines HDBSTAR models formulated in this application account for the above described characteristics of the electricity load series by including a trend and a seasonal component, two cyclical components (one for the within-day and one for the within-week periodic variations), one calendar component and one non-linear AR component with temperature used as the transition variable (associated with a logistic STAR) to account for the non-linear effects of temperature on load. Note that the cyclical components aim to account for the non-linear cyclic behaviour not accounted for by the seasonal component.

A Taylor HDBSTAR(3, 1, 2) and a cubic B-splines HDBSTAR(3, 1, 2) with $n = 1$ knot (located at the median temperature of 23°C) that included the components mentioned above, with delay $d = 1$ for the temperature series as transition variable, were selected



(a) (b)

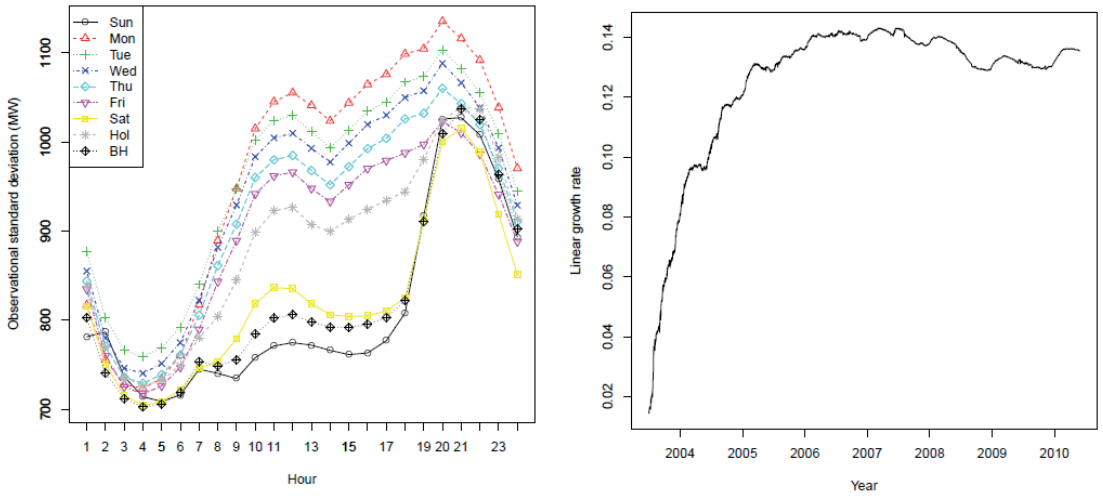
Figure 8: Time series plots of the within-week daily loads during two-week periods (a) in the summer of 2009 (11-24 January 2009), and (b) in the following winter (14-27 June 2009).

as the optimal models. They presented the largest in-sample conditional log-predictive likelihoods in a grid search of models of varying orders. Initial non-informative prior distributions were attributed to the hyperparameters. Discount factors of $\delta_V = 0.90$ and $\delta_W = 0.99$ were also found as optimal values. Two harmonics for each cyclical component were found to represent the non-linear short-term periodic patterns adequately.

The B-splines HDBSTAR(3, 1, 2) model fitted the data slightly better than the Taylor HDBSTAR(3, 1, 2) model with a larger log-posterior likelihood (of -559751 against -550437), lower MAPE (0.00527 against 0.00547) and RMSE (208.19 against 227.68) as can be seen in the one-hour-ahead row of Table 2 corresponding to the in-sample column. As the B-splines model fitted the series slightly better, we proceed to describe the main results from its in-sample application below.

Figure 9 shows the estimated average observational standard deviations ($\sqrt{S_t}$) by day of the week (Sun–Sat), holidays (Hol) and bridge-holidays (BH) of the hourly load. Generally, the standard deviations display similar patterns to the observed loads shown in Figure 7(b) with largest values (in the range 1000–1180 MW approximately) occurring at the evening peak times and the lowest (in the range 720–760 MW approximately) at the early morning through times. Weekdays tend to group together showing larger standard deviations than weekends and bridge-holidays in general.

The monthly average of the posterior means of the growth rates of the linear trend component can be seen in Figure 9(b). Notice that the sharp increase from the rather low rate of 0.01 at the initial time in 2003 to the maximum 0.14 level by 2006 can at least in part be explained by the use of a non-informative prior distribution for that parameter. After 2006, the estimated growth fluctuates around 0.14 until 2007 when it decreases slightly fluctuating around 0.13 until 2010. Those later values are apparently consistent with the overall positive trend of the load with increasing variability in Figure 7(a) which rate of increase seems to lower after 2007.



(a) (b)

Figure 9: (a) In-sample average observational standard deviation of hourly loads (in MW) by day of the week (Sun–Sat), holidays (Hol) and bridge-holidays (BH), and (b) the monthly averages of the in-sample posterior means of the growth rate associated with the linear trend component by the cubic B-splines HDBSTAR(3, 1, 2).

Both the Taylor HDBSTAR(3, 1, 2) and the cubic B-splines HSDBSTAR(3, 1, 2) models were used to produce rolling forecasts sequentially from 1 to 72-hours-ahead horizons during the out-of-sample period (1-30 June 2010). For that, forecasts of temperatures from a multiplicative Winters model (selected by SPSS's expert model amongst a number of exponential smoothing and ARIMA models) were used. The models were implemented in R and their running times, for fitting and forecasting on a desktop PC with i7 processor at 3.30GHz with 32GB of memory and SSD disk drive, varied from 8.93 (for a forecasting horizon $h = 24$ hours-ahead) to 14.50 minutes (for $h = 72$) for the Taylor HDBSTAR model and from 10.23 ($h = 24$) to 17.65 minutes ($h = 72$) for the B-splines HDBSTAR model.

Table 2 shows the MAPE (%) and the RMSE for the Taylor and the B-splines moels for each forecasting horizon of 1, 12, 24, 48 and 72 hours-ahead. The numbers in bold represent the lowest values between the two models for each horizon and error measurement. It can be seen that the B-splines model outperformed the Taylor model for 1 and 12-hours-ahead horizons but for the 24, 48 and 72-hours-ahead horizons the Taylor outperformed the cubic B-splines model on both MAPE and RMSE criteria. Overall, with only a few exceptions, the differences in forecasting performances were relatively small with both models displaying fairly similar performances.

Note that the MAPE and RMSE for a SARIMA(2,1,10)(2,1,1) fitted by SPSS as the best classical forecasting model using the temperature series and the calendar indicators as explanatory variables were 107.80 % and 455.993 respectively. Those results compare with the much smaller MAPE and RMSE of 59.06 % and 221.29, respectively, for the B-splines HSDBSTAR(3,1,2) and 61.20 % and 231.77 for the Taylor HSDBSTAR(3,1,2), indicating improved fitting of nearly 50% by the DBSTAR models.

The following results are shown for the 24-hours-ahead forecasting horizon that is usually of particular interest to practitioners. Figure 10 displays the plots of observed

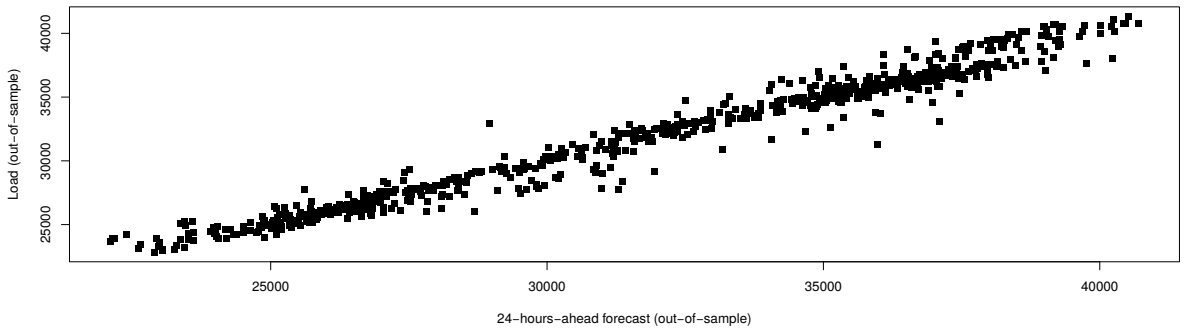
Horizon	Model	In-sample		Out-of-sample	
		MAPE (%)	RMSE	MAPE (%)	RMSE
1	Taylor	61.20	231.77	54.65	227.68
	B-splines	**59.06**	**221.29**	**52.74**	**208.19**
12	Taylor	210.28	915.90	173.62	804.68
	B-splines	**111.90**	**401.20**	**119.06**	**448.55**
24	Taylor	204.27	**890.51**	**184.98**	**850.83**
	B-splines	**204.22**	893.91	199.29	877.58
48	Taylor	**203.29**	**889.03**	**186.95**	**876.09**
	B-splines	204.18	894.67	201.75	897.90
72	Taylor	**197.07**	**880.96**	**193.12**	**897.88**
	B-splines	198.72	887.22	209.41	925.56

Table 2: Mean Absolute Percentage Errors (MAPE) and Root Mean Squared Errors (RMSE) of the Taylor and the B-splines HDBSTAR(3, 1, 2) models for each of the 1, 12, 24, 48 and 72 hours-ahead forecasting horizons

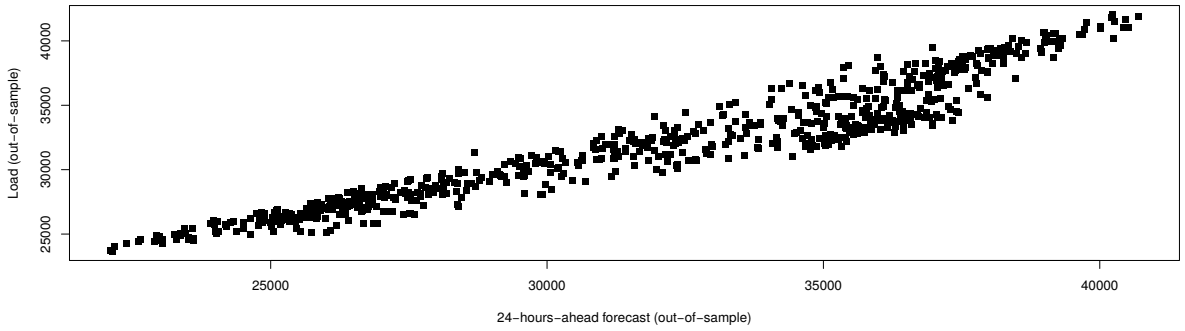
values of electricity load versus the 24-hours-ahead forecasts during the out-of-sample period for (a) the Taylor and (b) the B-splines HDBSTAR(3,1,2) models. Note that both models produced forecasts that were quite close to the observed loads with most points close to a diagonal line indicating their generally good forecasting accuracy. The B-splines model tough showed slightly higher degree of scatter than the Taylor model as expected from their MAPE and RMSE out-of-sample performance measures for the 24 hours horizon in 2.

Figure 11 (a)–(d) shows the observed loads (solid square points), their 24-hours-ahead rolling forecasts (solid line) and 95% predictive intervals (dashed lines) during the out-of-sample period by the Taylor HDBSTAR(3, 1, 2) model. At each hour during the out-of-sample period, the forecast at that hour was made 24-hours earlier using the 24-hour-ahead temperature forecast as the transition variable. Those temperature forecasts were obtained by a Winters' exponential smoothing models with parameters $\alpha = 0.973$ for level and $\delta = 0.07$ for seasonal (the trend parameter $\gamma = 0.001$ was non-significant with a p-value of 0.387).

Note that in general, consistently with the plot in Figure 10 (a), the forecasts are quite close to the observed loads with 95% prediction intervals that are particularly larger at peak times (18-21 hours) than at late night and early morning times in most days. The forecasts were well within the bounds of the 95% prediction intervals at almost all times. The selected out-of-sample period of June 2010 had a number of special days such as the Corpus Christie holidays on Thursday and Friday, 3^{rd} and 4^{th} June, in Figure 11(a), and the South African football world cup when on Tuesday, 15^{th} June, in Figure 11(b), Brazil played North Korea at 15:30 hours as well as at 09:00 on Sunday, 20th June, and at 11:00



(a)

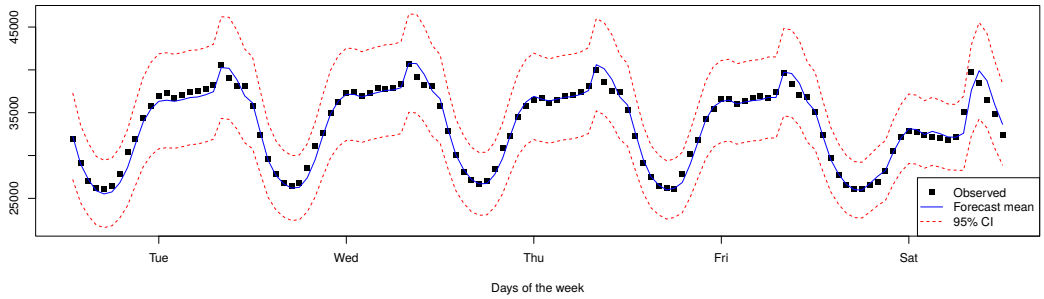


(b)

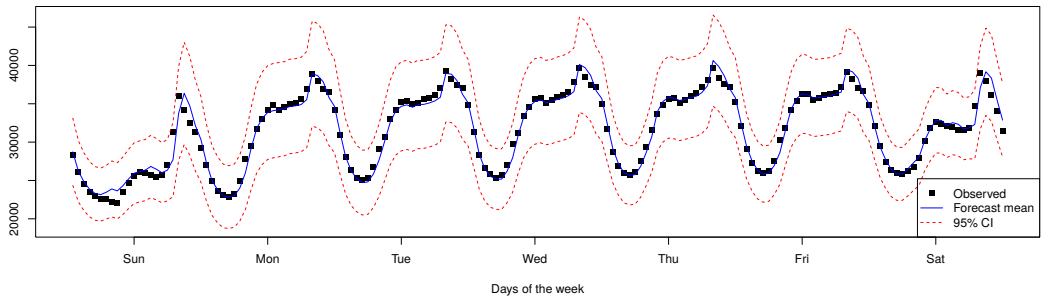
Figure 10: Out-of-sample observed against 24-hours-ahead forecasts for the (a) Taylor and (b) B-splines HDBSTAR(3, 1, 2) models.

on Friday, 25^{th} June, both in Figure 11(c) when Brazil played Ivory Coast and Portugal respectively, followed by Chile at 15:30 on Monday, 28^{th} June, in Figure 11(d). It can noticed that on those events the electricity load decreased comparatively with similar times at similar days. This can be explained by the fact that large amounts of people in urban areas tend to group together with family members on religious holidays and with friends in bars and restaurants or with crowds of people in public places such as squares with large screens to watch the national football team. Notice that the effect of the football match events lasted mainly on the hours of those events with the load levels increasing back to higher levels soon after.

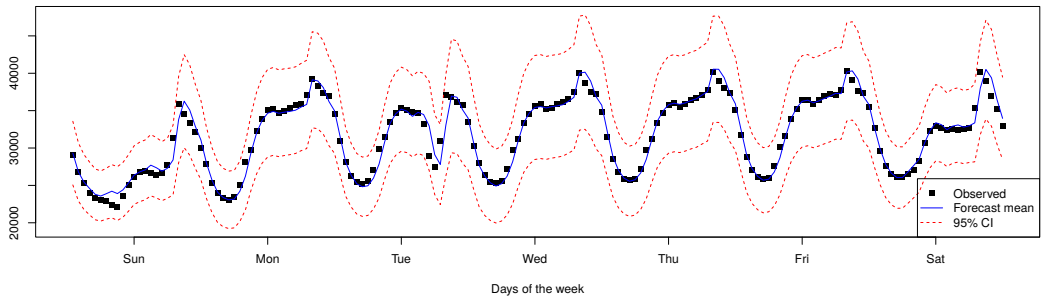
Certainly that an expert user could anticipate effects of events like those above in similar occasions and make appropriate interventions in the model (by changing prior hyperparameters accordingly).



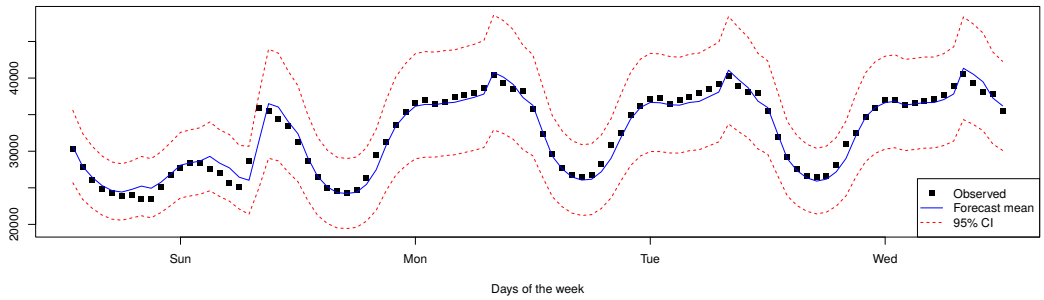
(a) 1-5 June 2010



(b) 13-19 June 2010



(c) 20-26 June 2010



(d) 27-30 June 2010

Figure 11: Out-of-sample observed values (dots), the 24-hours-ahead forecasts (solid line) and the corresponding 95% prediction intervals (dashed lines) of the hourly load by the Taylor HDBSTAR(2, 1, 2) model from the first to the last hour in the periods (a) 1-5 June 2010; (b) 13-19 June 2010; (c) 20-26 June 2010; and (d) 27-30 June 2010.

5 Discussion

In this paper, Gaussian *dynamic Bayesian smooth transition autoregressive* (DBSTAR) models for nonlinear autoregressive time series processes are proposed to address some limitations of existing STAR type models. Unlike those, DBSTAR models allow parameters to adapt to data dynamically in time. More specifically, the DBSTAR models consider analytical approximations for STAR models based on polynomial *dynamic linear models* (DLMs) of West and Harrison (1997) as alternative to both the classical STAR models of Chan and Tong (1986) and the *computational Bayesian STAR* (BSTAR) models of Lopes and Salazar (2005). Taylor series and B-splines approximations were adopted for functional logistic transition functions associated with STAR models.

Classical STAR models have static parameters (including observational variance) and adopt an approach similar to the Taylor approximation but at the parameter estimation stage. Only parameters associated with the autoregressive components, the observational variance and the transition function smoothing constant are estimated from data while other parameters such as autoregressive order, transition function threshold value and delay constant are assumed fixed a priori. BSTAR models, on the other hand, also have static parameters all assumed unknown (including those the classical STAR models assume fixed) that are estimated from data via full Bayesian approach. Due to loss of analytical tractability, BSTAR models rely on MCMC algorithms for approximate posterior assessments of the unknown parameters. Reliance on MCMC approaches makes applicability to high-frequency time series impractical. Both classical STAR and BSTAR models are applicable to intrinsically stationary time series and are generally non-parsimonious models requiring large AR orders to capture long term effects such as seasonality and cycle.

DBSTAR models, however, sequentially update their unknown parameters sequentially in time analytically via Kalman filtering. Similar to the classical STAR models some parameters such as the AR order and transfer function location values are assumed fixed a priori. However, unlike both the classical STAR and the BSTAR the unknown parameters (AR coefficients, observational variance and transition smoothing variable) are all dynamic. In applications where initial data are available optimisation approaches can be used to determine suitable values for the fixed parameters. The DBSTAR model formulation is thus useful in applications demanding sequential parametric change in time (including observational variance) and fast computing. However, a complicating factor in the DBSTAR model formulation is that the solution of a set of polynomial equations is required at each time step if the first two moments of the unknown parameters associated with a STAR model are to be determined. This is because, at each time step, a DBSTAR model produces posterior distributions for the parameters of the polynomial DLM that results from approximating the STAR transition function by Taylor expansion or B-splines. Those parameters are polynomial functions of the original STAR parameters. Thus, a DBSTAR model naturally estimate parameters that have polynomial autoregression interpretability of their own but will demand extra computational processing for parametric interpretability associated with parameters of a STAR model.

Similarly to DLMs, DBSTAR models can be formulated to account for components

of the underlying process without need for transformations. Trend, seasonality and cycle components are easily accounted for in a parsimonious way. Heteroskedasticity can also be accounted for by either incorporating a variance law in the model or with the use of variance discounting techniques when slow but steady changes in the unknown observational variance are allowed. The parameters associated with with the slow changes are estimated sequentially from data.

To model observed cyclical behaviour in the data that are not accounted for by seasonal components, *harmonic* DBSTAR (HDBSTAR) models have been defined that explicitly include components for cycles. Fourier form representations of cycles with combinations of sine/cosine waves provide an economic parametric characterisation and facilitate their interpretation. In general, lower autoregressive orders are required by a HDBSTAR model comparatively to a DBSTAR (and a STAR) model. The parsimony of a HDBSTAR model is balanced by larger amplitudes in the autoregressive coefficients. This is an advantage of the HDBSTAR models over the DBSTAR, classical STAR and BSTAR, for modelling time series in the presence of cyclical behaviour.

Taylor and the B-splines formulations of HDBSTAR models were applied to the well-known annual Canadian lynx data and to an hourly series of electricity load in a region in Brazil. In both applications, the HDBSTAR models fitted well the data. The Canadian Lynx data application showed improved fitting performances for the HDBSTAR models when compared with both the classical and the BSTAR models. In the Brazilian electricity load application, the formulated HDBSTAR models showed considerably improved fitting over a SARIMA model obtained by SPSS's expert forecasting modeler. The B-splines slightly outperformed the Taylor HDBSTAR model for both fitting and short-term forecasting. The Taylor model performed slightly better for longer forecasting horizons.

Finally, it is worth mentioning that some of the main limitations of the proposed DBSTAR models such as the ad-hoc determination of the AR order p (and the harmonic order h in HDBSTAR models) could be addressed by a fully Bayesian sequential approach, such as particle filtering, that treated those parameters as unknown at a hopefully low computational cost.

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.

Broemeling, L. D. (1985). *Bayesian analysis of linear models*. Decker, New York.

Campbell, E. P. (2004). Bayesian selection of threshold autoregressive models. *Journal of Time Series Analysis*, 25:467–482.

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

Chen, C. and Lee, J. (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis*, 16:483–492.

Chen, C. W. (1998). A bayesian analysis of generalized threshold autoregressive models. *Statistics & Probability Letters*, 40:15–22.

de Boor, C. (1978). *A Practical Guide to Splines*. Springer, Berlin.

Dierckx, P. (1993). *Curve and surface fitting with splines*. Clarendon, Oxford.

Dordonnat, V., Koopman, S., Ooms, M., Dessertaine, A., and Collet, J. (2008). An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24:566–587.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with b-splines and penalties (with discussions). *Statistical Science*, 11:89–121.

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.

Geweke, J. and Terui, N. (1993). Bayesian threshold autoregressive models for nonlinear time series. *Journal of Time Series Analysis*, 14:441–454.

Green, P. J. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732.

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

Lubrano, M. (2000). Bayesian analysis of nonlinear time-series models with a threshold. *Nonlinear Econometric Modelling in Time Series: Proceedings of the Eleventh International Symposium in Economic Theory*.

Pardo, A., Meneu, V., and Valor, E. (2002). Temperature and seasonality influences on spanish electricity load. *Energy Economics*, 24:55–70.

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.

Tong, H. (2011). Threshold models in time series analysis - 30 years on (with discussions). *Statistics and Its Interface*, 4:107–136.

Tong, H. and Lim, K. (1980). Threshold autoregressions, limit cycles, and data. *Journal of the Royal Statistical Society B*, 42:245–292.

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 the Royal Statistical Society, Series B*, 48:70–78.

West, M. (1995). Bayesian inference in cyclical component dynamic linear models. *Journal of the American Statistical Association*, 90:1301–1312.

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