Dynamic Bayesian smooth transition autoregressive models applied to hourly electricity load in southern Brazil

Conference or Workshop Item

Dynamic Bayesian smooth transition autoregressive models applied to hourly electricity load in southern Brazil

Álvaro E. Faria
and
Alexandre J. Santos

September 25, 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 polynomial dynamic analytical models suitable for inherently non-stationary time series with non-linear 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. A harmonic version of those models, that accounted for the cyclical component explicitly in a flexible yet parsimonious way, were applied to the well-known series of annual Canadian lynx trappings and showed improved fitting when compared to both the classical STAR and the BSTAR models. Another application to a long series of hourly electricity loading in southern Brazil, covering the period of the South-African Football World Cup in June 2010, illustrates the short-term forecasting accuracy of fast computing harmonic DBSTAR models that account for various characteristics such as periodic behaviour (both within-the-day and within-the-week) and average temperature.

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

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 proposed DBSTAR models 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 STAR models of Chan and Tong (1986) use conveniently chosen smooth transition functions, such as the logistic function

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

adopted here. It has smoothness and location parameters $\gamma \in \Re^+$ and $c \in \Re$, respectively, and transition variable $s_t \in \Re$. Usually, in practice, the transition variable is either a a lagged past value y_{t-d}, where d is a delay parameter, or a chosen exogenous variable. The parameter γ dictates the degree of smoothness of $\pi(s_t; \gamma, c)$ and c is a threshold value 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. For values of γ leaning towards zero the logistic function tends to 1/2 and the *logistic* STAR model is reduced to the average between the combining AR models. As γ increases away from zero,

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 non-linear in form and usually non-linear least square approaches, that approximate the non-linear transition function, are adopted for parametric estimation. We refer to those STAR models as *classical* STAR models throughout.

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 the sampling importance resampling of Gelfand and Smith (1990), for Bayesian inference due to the loss of analytical tractability of posterior distributions as shown in 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 considered known, and a reversible jump MCMC algorithm (Green, 1995) for posterior assessments when p is unknown. Thereafter, the models of Lopes and Salazar (2005) will be referred to as *Bayesian Simulation STAR* (BSTAR) models.

A common characteristic of all those approaches is that they are, without exception, static non-sequential methods for non-linear but stationary AR processes. It is also worth noting that computational Bayesian inference approaches, such as the BSTAR models, are non-parsimonious 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, consequently, not generally appropriate for applications that require fast sequential prior-to-posterior parametric estimation and forecasting.

The DBSTAR models address some of those limitations. Similarly to the classical STAR models, the AR order p and the delay parameter d are fixed a priori in a DBSTAR model. When those parameters are unknown, and initial data is available, a model selection approach can be adopted to determine their optimal values. Also, like BSTAR models, prior distributions must be specified for the state parameters of a DBSTAR model. Those parameters are functions of the AR coefficients and the smoothness parameter as well as the observational variance associated with STAR type models. However, unlike both the classical STAR and the BSTAR models, the observational variance in a DBSTAR model is not constant but allowed to vary in time, albeit slowly and steadily, to account for possible extra variation in the series. The slow changes in the observational variance of a DBSTAR model are determined sequentially from data in an approach based on the Kalman filter.

Another difference from both classical STAR and BSTAR models, is that a DBSTAR model can be parsimoniously formulated to explicitly account for components observed in the underlying time series, that is, level, trend, seasonality and cycle. A DBSTAR model is based on particular formulations of the polynomial *dynamic linear model* (DLM) of West and Harrison (1997) that allows hierarchical component modelling by superposition of models appropriate for specific components.

The main results of a comparative analysis against the classical and the Bayesian STAR models in the Canadian lynx application are shortly described here, while the application to the electricity load in southern Brazil is shown in more detail. More detailed analysis of both applications can be found in Santos (2014) and Faria and Santos (2018).

In both the Canadian lynx trapping and the Brazilian electricity loading applications, harmonic DBSTAR models fitted well the data. The Canadian Lynx data application, apart from showing the improved fitting performances for a harmonic DBSTAR model when compared with both the classical and the BSTAR models, also illustrated how the changing cyclic pattern, amongst others, can be dynamically estimated from the data. In the electricity load application, the formulated harmonic DBSTAR models showed considerably improved fitting over a SARIMA model obtained by SPSS's expert forecasting modeler. The B-splines slightly outperformed the Taylor harmonic DBSTAR model for both fitting and short-term 24-hours-ahead forecasting. The Taylor model performed slightly better for the longer 72-hours-ahead forecasting horizons.

This article is structured as follows. In Section 2, harmonic DBSTAR (HDBSTAR) models are formally defined for the Taylor and the B-splines approximations. Section 3 briefly describes the main results of a comparative fitting analysis between DBSTAR models and both the classical STAR and the BSTAR models when applied to the Canadian lynx series. In Section 4, a more in depth look at the forecasting performances of HDBSTAR models is presented in the Brazilian electricity load series application. The article concludes with some discussion of this work in Section 5.

2 The harmonic DBSTAR model

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. Note that underlined characters are used 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

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

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

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 considered *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 $\underline{\theta}_t$). 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).

2.1 The Taylor DBSTAR model

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 The B-spline DBSTAR model

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 is defined as in Wold (1974).

Computationally, the B-splines are obtained recursively by the Cox-de Boor algorithm (de Boor, 1978). As pointed out by Eilers and Marx (1996), they are rather attractive as base functions for univariate regression in which a linear combination of cubic B-splines gives a curve smooth enough to provide a good fit in many applications. This is the case here for the logistic transition function (2) (or for any of the usual alternative transition functions such the second-order logistic and the exponential). In our case, an appropriate order $r = q + n - 2$ in (5) is usually determined by the number n of knots chosen. In the splines regression context, the choice of n can be a complex task and statistics with penalties for overfitting are used in determining the optimal number. In our context, n (and q) is determined via model selection approach.

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. For the Canadian lynx trappings and electricity loading applications, optimal values of those parameters were obtained via model selection approach.

2.3 The HDBSTAR model

Now, in order to allow the modelling of any observed cyclic behaviour in terms of cyclical components explicitly, we introduce the *Harmonic* DBSTAR (HDBSTAR) model. Similarly to seasonality, the explicit modelling of long term cyclic behaviour allows accounting for changes in that behaviour in a forecasting model. 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 are adopted here.

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

$$\left(Y_t \mid \underline{\theta}_t, \underline{\psi}_t\right) \sim N\left(\underline{F}_{1t}^{'}\underline{\theta}_t + \underline{F}_{2t}^{'}\underline{\psi}_t, \Sigma_t\right) \tag{7}$$

$$\left(\underline{\theta}_t \mid \underline{\theta}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{1t}\underline{\theta}_{t-1}, \mathbf{W}_{1t}\right) \tag{8}$$

$$\left(\underline{\psi}_t \mid \underline{\psi}_{t-1}\right) \sim T_{n_{t-1}}\left(\mathbf{G}_{2t}\underline{\psi}_{t-1}, \mathbf{W}_{2t}\right) \tag{9}$$

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.

Two types of DBSTAR models are defined here, 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. 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) and Eilers and Marx (1996) amongst others for more details on B-splines.

3 Main results of the Canadian lynx application

The Canadian lynx dataset is a yearly series with 114 observations of the number of lynx trapped in the Mackenzie River, district of North-west Canada, from 1821 to 1934. They were collected to improve knowledge about the population dynamics of the ecological system in that area. They have been used in various studies (see e.g. Terasvirta, 1994; Lopes and Salazar, 2005) to analyse and compare the fitting of proposed models. The most famous features of that series are (a) the lack of trend, (b) the presence of irregular changes in the amplitude in time, and (c) the presence of persistent non-regular cyclic oscillations with periods of 10 or 11 years. Those features have been familiar to biologists for a long time and are well known in historical records of trappings of lynx in Canada as described by Elton and Nicholson (1942).

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 presented no evident trend but a clear periodic repetitive behaviour with significant estimated autocorrelations at lags

1, 2 and 11 in the partial auto-correlation function (PACF). However, it may not be appropriate to use the PACF for model order identification in this case as a graphical analysis of the scatterplots of y_{t-u} against y_{t-v} (for $u, v = 1, 2, ..., v > u.$) showed lack of linearity for most lags of the series. Consistently with Elton and Nicholson (1942), a periodogram of the series displayed a spike around the 0.1 frequency indicating cyclical behaviour with a 10-year wavelength.

Following a numerical grid search for optimal values (based on the log-smoothing likelihoods of the models calculated conditionally on values of d, p, r and h) the following three models were selected: a Taylor DBSTAR$(3, 12)$, a cubic B-splines DBSTAR$(6, 12)$ with five knots and a cubic B-splines HDBSTAR$(6, 2, 2)$ also with five knots. Similarly to Terasvirta (1994) and Lopes and Salazar (2005), the optimum delay parameter of the adopted dynamic logistic transition function was found to be $d = 3$ for all models, that is $s_t = y_{t-3}$. However, unlike both Lopes and Salazar (2005) and Terasvirta (1994) that found $p = 11$, the optimum AR order for the non-harmonic Taylor and B-splines DBSTAR models were found to be $p = 12$. The optimal harmonic model though was found to be considerably more parsimonious. In fact, the optimal values for the number of harmonics and AR order were $h = 2$ and $p = 2$ respectively for the B-splines HDBSTAR model. A Taylor expansion of order $r = 3$ was adopted for the Taylor DBSTAR$(3, 12)$ model as higher orders only provided marginal improvements.

Table 1 displays the MAE and the RMSE of the fitted logistic classical STAR(11) from Terasvirta (1994) and the logistic BSTAR(11) from Lopes and Salazar (2005) as well as those of the selected Taylor DBSTAR$(3, 12)$, the B-splines DBSTAR(6, 12) and the B-splines HDBSTAR$(6, 2, 2)$ models. It is clear that the B-splines HDBSTAR(6, 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 B-splines DBSTAR(3, 12), with the largest MAE and RMSE of all DBSTAR models, still outperformed the classical STAR(11) and the BSTAR(11) models, by quite a margin (its MAE and RMSE were 9.8 and 5.7 times lower than the BSTAR(11) of Lopes and Salazar (2005), respectively). A static version of the Taylor DBSTAR(3,12) model (with discount factors set to unity) had an MAE and an RMSE of 0.109 and 0.141, respectively, produced a fitting that was only marginally better than the BSTAR(11)

Model	MAE	RMSE
Lopes and Salazar (2005) - BSTAR(11)	0.118	0.153
Terasvirta (1994) - STAR(11)	0.142	0.179
Taylor DBSTAR$(3, 12)$	0.012	0.015
B-splines DBSTAR$(6, 12)$	0.014	0.027
B-splines HDBSTAR$(6, 2, 2)$	**0.006**	**0.013**

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

As seen above, the sequential dynamic modelling of the Canadian lynx data by the DBSTAR models have allowed much improved fitting to the data compared with the static STAR and BSTAR models. This sequential modelling also allows a better understanding of the lynx population dynamics via analysis of the obtained dynamic parameters estimated from the data that helps to verify how various components varied in time. For example, the estimated cyclic component (obtained from the posterior means of the harmonic components of the HDBSTAR model by Kalman fitering) has shown increased variability in amplitudes from 1824 to 1846, followed by lower oscillations until approximately 1896 after when oscillations increased again almost mirroring the initial period.

It is worth mentioning that a residual analysis of the DBSTAR models above showed residuals to be uncorrelated as well as no significant departures from normality. Please refer to Santos (2014) and Faria and Santos (2018) for more details of the application above.

4 Electricity load forecasting in southern Brazil

The hourly series of electricity load, measured in MegaWatts (MW), and temperature, in degree Celsius (°C), are from the Southeast and Central-West regions of Brazil and span from the first hour on 1 June 2003 to the last hour on 30 June 2010 (that is, 62,088 hourly observations covering the 7-year period). The hourly load data are aggregated for the region while the hourly temperature data are averaged across all the states in those regions. Calendar variables indicating weekdays, national holidays and bridge-holidays (i.e. days between midweek bank holidays that are near a weekend) were used in the models to account for their effects on the load. To measure the short-term forecasting performances of the selected DBSTAR models, the initial 61,368 observations (covering the period from 1 June 2003 to 31 May 2010) and the last 720 observations (corresponding to one month of hourly data from 1 June 2010 to 30 June 2010) were used as in-sample and out-of-sample data respectively.

A preliminary analysis of the data has shown, amongst others, a non-linear S-shaped relationship between load and temperature. In this relationship, differently from the U-shape normally observed for northern hemisphere countries, the load tends to increase with increases in temperature although on a non-linear fashion resembling an S-shape. In fact, 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) when higher rates of increase occur, suggesting that at least two regimes of distinct consumption behaviour is present in the load series that can justify at least in part the use of STAR type of models. Other marked characteristics observed from the data analysis are the long-term positive trend with slowly increasing variability in time, and both within-day and within-week changing periodic behaviour and variability. The within-day variation of load has a general shape that is similar for all weekdays with lower loads in the early hours with minimum levels between 4-8am followed by sharp increases until the middle of the day when they stabilise at an intermediate level until early afternoon. This is followed by peaks in the evening between 7-9pm. Sundays and bridge-holidays show a little shift of consumption to the right in the hour scale as well as lower levels at almost all times.

Similarly, Saturdays and holidays tend to group together with a smaller shift to the right but with larger loads in the early hours. Except for Monday (with slightly lower loads at all times except in the early hours when they are lowest), the patterns for other weekdays are practically undistinguishable from each other. They show the largest levels of load overall with troughs typically at 4-5am, lower peaks at 11am-12pm and 3-4pm and the largest peaks from 7-9pm. The within-week variations are such that daily peaks change their patterns (or shape) along the year according to the season. In particular, evening peaks observed in the winter is not present in the during summer when 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. It is also worth pointing out that due to the changing variability, both within-the-day (that is larger at peak than at off-peak hours) and within-the-week (that is larger at working days than at weekends and holidays), the underlying load series presented intrinsic second-order non-stationarity that no differences and/or transformation in the class of Box-Cox transformations could be found to turn the series stationary. No serious departures from normality were observed.

The Taylor and the B-splines HDBSTAR models that were formulated in this application to account for the above described characteristics of the electricity load series include 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) model and a cubic B-splines HDBSTAR(3, 1, 2) model 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 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 hyper-parameters of the models. 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 during the in-sample period 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).

Both the Taylor HDBSTAR(3, 1, 2) and the cubic B-splines HSDBSTAR(3, 1, 2) models were also used to produce rolling forecasts sequentially from 1 to 72-hours-ahead horizons during the out-of-sample period. 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. Those DBSTAR models were

implemented in R and used prior-to-posterior updating and forecasting routines based on the Kalman filter (and on the DLM and SPLINES packages). 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 models for each forecasting horizon of 1, 12, 24, 48 and 72 hours-ahead. The values in bold represent the lowest values between the two models for each horizon. Note 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.

For illustration, 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. As expected, those results compare unfavourably 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.

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

For the 24-hours-ahead forecasting horizon, that is usually of particular interest to practitioners, plots of observed loads versus forecasts during the out-of-sample period,

displayed most points close to a diagonal line indicating the generally good forecasting accuracy for both the Taylor and B-splines HDBSTAR(3,1,2) models. However, the B-splines model showed a 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 Table 2.

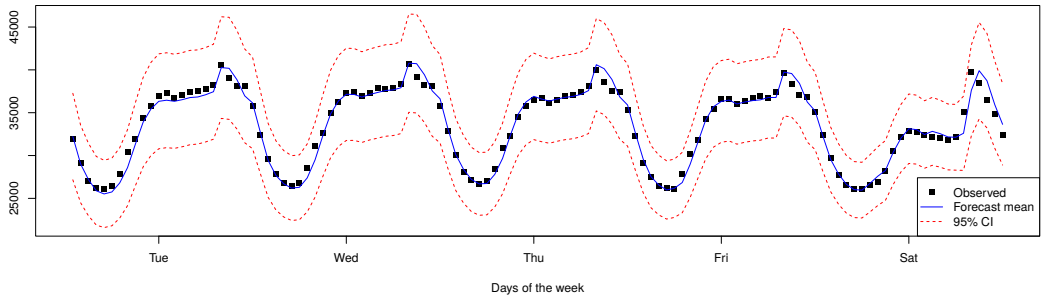
Figure 1 (a)–(d) shows the observed loads (solid square points), their 24-hours-ahead rolling forecasts (solid line) and 95% predictive intervals (dashed lines) during the out-of-sample period by the Taylor HDBSTAR(3, 1, 2) model. At each hour during the out-of-sample period, the forecast at that hour 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, 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 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 1(a), and the South African football world cup when on Tuesday, 15^{th} June, in Figure 1(b), Brazil played North Korea at 15:30 hours as well as at 09:00 on Sunday, 20th June, and at 11:00 on Friday, 25^{th} June, both in Figure 1(c) when Brazil played Ivory Coast and Portugal respectively, followed by Chile at 15:30 on Monday, 28^{th} June, in Figure 1(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.

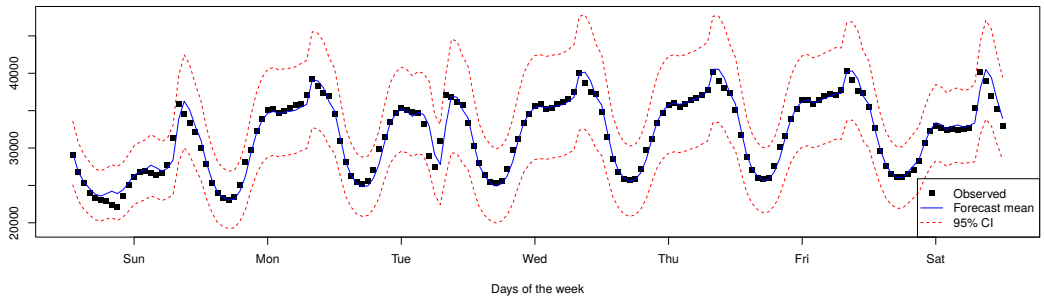
In practice, an expert user could anticipate effects of events like those above in similar occasions and make appropriate interventions in the model (by changing prior hyper-parameters accordingly).

5 Concluding remarks

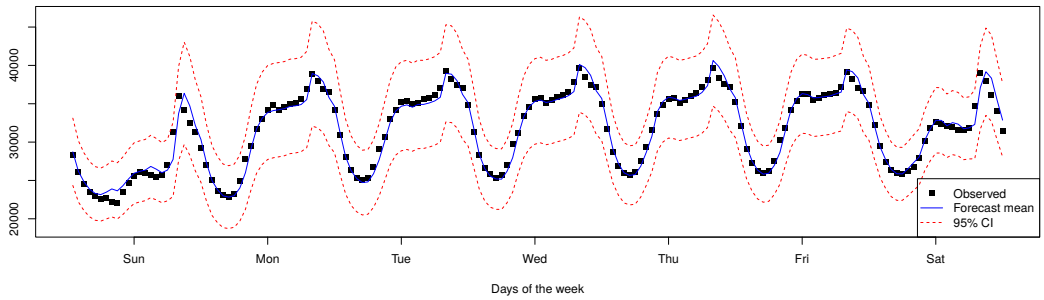
The *dynamic Bayesian smooth transition autoregressive* (DBSTAR) models introduced here are Gaussian analytical approximations of STAR models based on polynomial *dynamic linear models* (DLMs) of West and Harrison (1997). They are appropriate for non-linear and intrinsically non-stationary auto-regressive time series processes such as those exhibiting asymmetric cycles and offer an alternative to both the classical STAR models of Chan and Tong (1986) and the *computational Bayesian STAR* (BSTAR) models of Lopes and Salazar (2005). Two types of DBSTAR models, the Taylor and B-splines DBSTAR models, were defined according to the adopted approximation approach adopted for functional logistic transition functions associated with STAR models. Their unknown



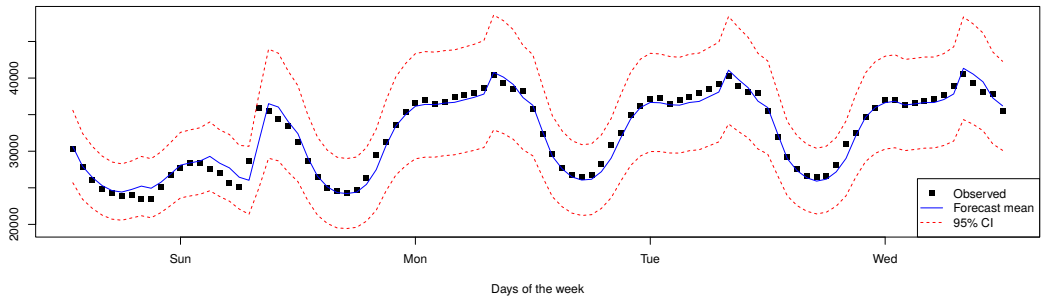
(a) 1-5 June 2010



(b) 13-19 June 2010



(c) 20-26 June 2010



(d) 27-30 June 2010

Figure 1: 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.

parameters are sequentially updated in time analytically via Kalman filtering approach as described in Faria and Santos (2018). Similar to 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 auto-regression 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. 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 auto-regressive 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 repetitive periodic behaviour.

Taylor and the B-splines formulations of HDBSTAR models were applied to a large hourly series of electricity load in a region in Brazil. In this 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.

References

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.

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.

Faria, A. and Santos, A. (2018). Dynamic bayesian smooth transition auto-regressive models for non-linear non-stationary time series. *The Open University, Statistics Technical Report*, 18/04:1–28.

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*.

Santos, A. (2014). *Dynamic Bayesian Smooth Transition Autoregressive models for non-stationary nonlinear time series*. The Open University, PhD Thesis.

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

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

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

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

Wold, S. (1974). Spline functions in data analysis. *Technometrics*, 16:1–11.