2
The model

Since the articles of Terasvirta and Anderson (1992) and Terasvirta (1994) the Smooth Transition Autoregressive (STAR) model has become one of the most popular non linear time series model in applied econometrics.

It has been employed in modelling several different variables in the economic framework, for example, interest rates in van Dijk and Franses (2000), exchange rates in Taylor, Peel and Sarno (2001) and unemployment in Skalin and Terasvirta (2002) among several others.

In addition, there are some surveys that review characteristics of modelling with the STAR model such as Granger and Terasvirta (1993), Terasvirta (1998), Potter (1999), Franses and van Dijk (2000), and most recently, van Dijk, Terasvirta and Franses (2002).

Consider a common logistic smooth transition autoregressive model of order *p* (STAR(p) model) proposed by Terasvirta (1994):

$$y_t = \phi_{01} + \phi_{11} y_{t-1} + ... + \phi_{p1} y_{t-p} + (\phi_{02} + \phi_{12} y_{t-1} + ... + \phi_{p2} y_{t-p}) \times F(\gamma(y_{t-d} - c)) + \varepsilon_t \qquad (1)$$

where $\varepsilon_t \sim NID(0, \sigma^2)$ and F(.) is the logistic function defined by

$$F(\gamma(y_{t-d} - c)) = (1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1}, \quad \gamma > 0 \qquad (2)$$

where γ determines the smoothness of the change in the value of the logistic function, and thus, the smoothness of the transition from one regime to the other. The parameter *c* indicates the location of the transition (location parameter). The delay parameter *d* of the transition variable can take values in the range of $1 \le d \le p$ or $p < d$. In addition, y_{t-d} is the transition variable. Note that there are other functions that could be used instead of the logistic, each having its own attractive features, for example, the exponential function creating the Exponential Smooth Transition Auto Regressive (ESTAR) (Terasvirta, 1994). However, the logistic has become the most common one used in practice.

When γ is very large, the logistic function F(.) approaches to an indicator function. Hence, the model (1) becomes the Self-Exciting Threshold Autoregressive (SETAR) model with two regimes. See Tong (1983, 1990) for more details about this model.

As pointed in van Dijk et al. (2002), the transition variable can also be an exogenous variable s_t or even a function of lagged endogenous variables. It can also be a linear time trend and so it reduces to a model with smoothly changing parameters, see Lin and Terasvirta (1994).

Another class of models that has been known in the time series literature is the Periodic Auto Regressive model (PAR) used to model seasonal series. In Franses and Paap (2005) they say that before 1996 not many academics and practitioners were considering periodic models for theirs seasonal time series. However, after that, the use of this type of models have increased a great deal as it can be seen in references like Adams and Goodwin (1995), Basawa and Lund (2001, 2000), McLeod (1994) and Ghysels and Osborn (2001).

Consider now a periodic autoregressive model (PAR(p) model) proposed by Herwartz (1999):

$$y_t = \phi_s + \phi_{1s} y_{t-1} + ... + \phi_{ps} y_{t-p} + \varepsilon_t , \tag{3}$$

where ϕ_s is a seasonally varying intercept parameter, ε_t is assumed to be an independent normally distributed series with mean 0 and variance σ^2, although this assumption may be relaxed by allowing ε_t to have seasonal variance σ_s^2. Thus, the autoregressive parameters $\phi_{1s},...\phi_{ps}$ vary according to the periodicity, that is, the number of different seasons within one year which is S. In case of quarterly observed data, S = 4. Hence, S is essentially a function of t. For some more details about periodic models, see Ghysels (1994), Franses and Paap (1994, 1996, 2000, 2002, 2005).

As Franses and Paap (2005) show, in general, the PAR process in (3) can be rewritten as an AR(p) model for the Sx1 vector process $Y_t = (Y_{1,t},...,Y_{S,t})'$, $t = 1,2,...,T$ where $Y_{S,t}$ is the observation of y_t in season s of year T. They refer to this as the vector of quartes (VQ) representation; it is useful for several analyses including, for example, unit roots and forecasting.

In the present thesis, we consider a multiple regime with smooth transition between two periodic autoregressive models, that is the logistic Smooth Transition Periodic Autoregressive model of order *p* (STPAR). It is given by

$$Y_t = G(w_t, s_t; \psi) + \varepsilon_t = \Phi_1 w_t + \Phi_2 w_t F(\gamma_S (s_{t,S} - c_S)) + \varepsilon_t \tag{4}$$

where, G(.) is a non linear function of w_t and s_t defined by the parameters vector ψ. $\varepsilon_t \sim NID(\underline{0}, \Sigma)$, where Σ is diagonal with $\sigma_1^2,...,\sigma_S^2$ and $\underline{0}$ is a vector of zeros with dimension S x S and S x 1, respectively. Note that it is in the VQ representation, s = 1,…,S. $w_t = [1, y_{t-1,1},..., y_{t-p,1},...y_{t-1,S},..., y_{t-p,S}]'$. Φ_j, is a matrix of parameters with dimension S x (SP + 1) and j = 1,2. F(.) is a matrix with dimension S x S containing in the diagonal logistic functions defined by (2) for each season. We assume that the transition variable is an exogenous variable s_t. So far it is just S STAR Models estimated for each s. Note that in the parameter matrix some coefficients are zero. Therefore, we can assume that there is one different nonlinear model for each season if one is working with quarterly data (S = 4).

Furthermore, we remind the reader that, obviously, the number of parameters in a PAR model is four times as many as that in a non-periodic AR, again when one is working with quarterly data. Not to mention that the STAR model has twice the number of parameters as a AR plus two $(\gamma \;\; and \;\; c)$, if one is dealing with a model with only two regimes. In addition, we are interested in using this model in a series of electricity demand which has a daily, weekly and yearly periodicity which would result in a highly parameterised model.

As pointed out in Taylor (2006), in order to overcome this problem a Fourier form proposed by Gallant (1981) was used. The model is then

$$y_t = \phi_1' w_t + \phi_2' w_t \times F(\gamma(s_t - c))\} + \varepsilon_t \tag{5}$$

where $w_t = [1, y_{t-1}, ..., y_{t-p}]'$, $\phi_j = [\phi_{0j}, \phi_{1j}, ..., \phi_{pj}]'$ and

$$\phi_{ij} = \omega_i + \sum_{k=1}^{h} \lambda_{ik} \sin(2k\pi(D(s)/D)) + \kappa_{ik} \cos(2k\pi(D(s)/D)).$$

F(.) is the logistic function and $\varepsilon_t \sim NID(0, \sigma^2)$. j is 1 and 2 because it is restricted to two regimes, however, it is possible to relax in this assumption. i = 0, 1, …, p. D(s)/D is a repeating step function that numbers the periodicity of the series being modelled. Note that the model is not in VQ representation. Now it is two periodic models in the extremes with a smooth transition. It is worth mentioning that it is possible to include more frequencies. h is the number of desired harmonics. Surely, those frequencies can change depending on the application. As we know from the STAR model, we have to restrict γ to be positive in order to identify the model. The transition function is a bounded function of s_t, continuous everywhere in the parameter space for any value of s_t. It is important to remember that when $\gamma = 0$ the model becomes a linear model.

It is straightforward to extend the model for the abrupt transition between the regimes, which is the Self-Exciting Threshold Periodic Autoregressive model (SETPAR), that is

$$y_t = \sum_{j=1}^{r} (\phi_{0j} + \phi_{1j} y_{t-1} + ... + \phi_{pj} y_{t-p} + \sigma_j \varepsilon_t) I(c_{j-1} < y_{t-d} \le c_j) \tag{6}$$

where $\phi_{ij} = \varpi_{ij} + \sum_{k=1}^{h} \lambda_{ik} \sin(2k\pi(D(s)/D)) + \kappa_{ik} \cos(2k\pi(D(s)/D))$

and d > 0 is the delay parameter (integer), $\varpi's$, $\lambda's$, $\kappa's$, are all parameters to be estimated, and j = 1, …, r, k = 1,…, h, and $c_0, c_1, ..., c_r$ are threshold parameters, $c_0 = -\infty, c_r = M < \infty$, and I(E) is an indicator function: I(E) = 1 when event E occurs; zero otherwise. Furthermore, $\varepsilon_t \sim iid(0,1)$, and $\sigma_j > 0$, j = 1,…,r. As in the STPAR, D(s) is repeating step function that numbers the frequencies of the endogenous variable. If one changes the y_{t-d} for just an exogenous variable s_t the model becomes the Threshold Periodic Autoregressive model (TPAR). For more information about the TAR model see, Stevenson, Amaral and Peat (2006).

Needless to say that the model could be extended to allow for exogenous variables $z_{1t},...,z_{kt}$ as additional regressors.

A modeling cycle procedure for the logistic STPAR consisting of the stages of model specification, parameter estimation, evaluation and forecasting is developed in the next section.