3
Model Implementation

Most of the literature suggests that a “specific-to-general” strategy is strongly recommended for building non-linear time series models, see Granger (1993), for instance. Therefore, one should start with a simple model and then, after evaluation, if the model is not adequate, it should proceed to more complicated ones.

This thesis will follow the steps proposed by Terasvirta (1994) which consist of:

(i) Specify a linear autoregressive model of order p using a model selection criterion as AIC (Akaike, 1974) or SBIC (Schwarz, 1978).

(ii) Select the lags where the periodicity will be counted for using the auto-correlation function.

(iii) Test the null hypothesis of linearity against the model, in this case, STPAR nonlinearity. If linearity is rejected, select the appropriate transition variable s_t;

(iv) Estimate the parameters in the selected STPAR model and select the correct number of h using a selection criterion;

(v) Evaluate the model;

(vi) Finally, obtain the forecasts or use the model for descriptive purposes.

3.1
Model Specification

As said before, one should estimate an auto regression and select the best model according to a model selection criterion. It should be accompanied by an analysis of a proper test for residual autocorrelation, like the portmanteau test of Ljung and Box (1978); in the case of (5) it is also important to analyze the significance of the periodicity involved in the model. In doing so, it is recommended to analyze the estimated autocorrelations for the periodicity.

PUC-Rio - Certificação Digital Nº 0310438/CA

3.2
Linearity Test

Testing linearity is important for two reasons. Firstly, due to the fact that one should use the nonlinear model only if linearity is rejected under the null hypothesis, otherwise a linear model would be suitable to model the underlying time series. Secondly, to select the transition variable s_t. The variable is chosen by running the linearity test, and selecting the lag of the variable that minimizes the p-value of the test (Terasvirta, 1994).

The procedure proposed here is based on a Lagrange Multiplier (LM) test, where auxiliary regressions are estimated to compute the test statistic. It follows the same steps as Terasvirta (1994), but it counts for non linear periodic auto regressive model. Other papers that are extremely important for the understanding of this linearity test are Luukkonen (1990), Tsay (1986), Luukkonen et al (1988) and Saikkonen and Luukkonen (1988).

In order to derive a linearity test against (5), one should approximate the logistic function $F(\gamma(s_t - c)) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}$ with a first order Taylor expansion around γ = 0. The reason for this is simply that a possible null hypothesis for the linearity test would be $H_0 : \gamma = 0$. However, (5) under the null is not identifiable.

Consider that the transition function is $F(\gamma(s_t - c)) = (1 + \exp\{-\gamma(s_t - c)\})^{-1} - 1/2$. The sum of ½ it is helpful to derive the test and it doesn't affect the argument. So, after the expansion the result would be

$$T_1(s_t;\gamma,c) \approx F^*(s_t;0,c) + \gamma \left.\frac{\partial F^*(s_t;\gamma,c)}{\partial \gamma}\right|_{\gamma=0} \tag{7}$$

where it was used the fact that $F^*(s_t;0,c) = 0$. After substituting $T_1(.)$ for $F_1^*(.)$ in (5) and rearranging terms this gives the auxiliary regression model

$$y_t = \beta_{0s}^{'} w_t + \beta_{1s}^{'} w_t s_t + \eta_t \tag{8}$$

PUC-Rio - Certificação Digital Nº 0310438/CA

where $\beta_{js} = (\beta_{js,0}, \beta_{js,1}, ...\beta_{js,p})'$, j = 0,1 and $\eta_t = \varepsilon_t + \phi_{2s}' w_t R_1(s_t; \gamma, c)$, with $R_1(s_t; \gamma, c)$ the remainder of the Taylor expansion and also $\eta_t = \varepsilon_t$ under H_0. Note that s here means only that the parameters follow a Fourier form like in (5). The use of (8) instead of (5) solves the identification problem, and one obtains a simple linearity test. It is easy to see that the parameters β_{js}, in the auxiliary regression (8) are functions of the parameters in the STPAR model (5) and the hypothesis $\gamma = 0$ implies now testing that $\beta_{1si} = 0$ for i = 0,…,p. The test statistic has an asymptotic χ^2 distribution with p+1+2h degrees of freedom under the null hypothesis of linearity.

The reader should be aware that when $s_t = y_{t-d}$ for certain integer$1 \le d \le p$, $\beta_{1s,0} s_t$ should be removed from the regression (8) to avoid perfect multicollinearity. As pointed out by Luukkonen et al. (1988), if only the intercepts differ from one regime to the other and other parameters are the same, the test has no power and to remedy this situation, a third order Taylor expansion of the transition function should be carried out. In this case, the expansion results in

$$T_3(s_t; \gamma, c) \approx \gamma \frac{\partial F^*(s_t; \gamma, c)}{\partial \gamma}\bigg|_{\gamma=0} + \frac{1}{6}\gamma^3 \frac{\partial F^*(s_t, \gamma, c)}{\partial \gamma^3}\bigg|_{\gamma=0} = \frac{1}{4}\gamma(s_t - c) + \frac{1}{48}\gamma^3 (s_t - c)^3 \qquad (8)$$

Where it was used the fact that the second derivative of $F^*(s_t; \gamma, c)$ with respect to γ evaluated at $\gamma = 0$ equals zero. Using this approximation yields the auxiliary model

$$y_t = \beta_{0s}' w_t + \beta_{1s}' w_t s_t + \beta_{2s}' w_t s_t^2 + \beta_{3s}' w_t s_t^3 + \eta_t \qquad (9)$$

where $\eta_t = \varepsilon_t + \phi_{2s}' w_t R_3(s_t; \gamma, c)$, and the null hypothesis is now defined as $\beta_{isj} = 0$ for i = 1,…,p, and j = 1, 2, 3. Again, note that $\eta_t = \varepsilon_t$ when the null hypothesis is true. It has an asymptotic χ^2 distribution with 3(p+1+2h) degrees of freedom. In small or medium size samples χ^2 distribution has a poor approximation to the actual small sample distribution and to circumvent that

PUC-Rio - Certificação Digital Nº 0310438/CA

problem Granger and Terasvirta (1993 – ch.7) suggest the use of an F distribution. The regression can become very large and Luukkonen et al. (1988) suggest that the regression (9) should be augmented only by the terms that depend on the intercept $\phi_{1,0}$ and $\phi_{2,0}$, which are $\beta_{1s,0}, \beta_{2s,0}$ and $\beta_{3s,0}$ creating then a more parsimonious or "economic" version of the test. Hence, the final regression is

$$y_t = \beta_{0s}^{'} w_t + \beta_{1s}^{'} w_t s_t + \beta_{2s,0} s_t^2 + \beta_{3s,0} s_t^3 + \eta_t \tag{10}$$

Now, it has a χ^2 distribution with p+3+2h degrees of freedom.

The test can be carried out like this:

1) Regress y_t on w_t with the Fourier terms and compute $SSR_1 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2$;
2) Regress $\hat{\varepsilon}_t$ on w_t and on the others nonlinear regressors of (9) for the full version or (10) for the "economic" version and compute the residual sum of squares $SSR_2 = \sum_{t=1}^{T} \hat{\eta}_t^2$;
3) Calculate a χ^2 statistic as $LM_{\chi^2} = T \dfrac{SSR_1 - SSR_2}{SSR_1}$.

Or the F version of the test which is:

$$LM_F = \frac{SSR_1 - SSR_2 / 3(p+1+2h)}{SSR_2 / (T - 4p - 4 - 6h)}$$

where T is the number of observations.

3.3
Estimation

Once the variables and the transition variable have been selected, the next stage in the modeling cycle involves estimation of the parameters in the STPAR model.

PUC-Rio - Certificação Digital Nº 0310438/CA

The parameters in (5) can be estimated by conditional maximum likelihood or non linear least squares. When $\varepsilon_t \sim NID(0,\sigma^2)$ the methods are coincident. Then, the vector of parameters ψ can be estimated as

$$\hat{\psi} = \arg\min_{\psi} Q_T(\psi) = \arg\min_{\psi} \sum_{t=1}^{T} (y_t - G(w_t, s_t; \psi))^2 \tag{11}$$

It is easy to see that the difference between (5) and (1) is just that w_t is "bigger" in the former. It incorporates the Fourier forms into the matrix as other variables. Under certain regularity conditions, which are discussed in Wooldrige (1994) and Potscher and Prucha (1997), among others, the NLS estimates are consistent and asymptotically normal, i. e.

$$\sqrt{T}(\hat{\psi} - \psi_0) \rightarrow N(0, C), \tag{12}$$

where ψ_0 denotes the true parameters value. The asymptotic covariance matrix C of the estimators can be estimated consistently. For further details see Dijk et al. (2002).

In order to reduce the dimensionality of the NLS estimation problem, as suggested by Leybourne, Newbold and Vougas (1998) and Medeiros and Veiga (2005), the concentrated least squares were used. When the parameters γ and c are known the STPAR reduces to a linear model. Therefore, one can compute the parameters $\varpi's$, $\lambda's$, $\kappa's$, conditional on γ and c by Ordinary Least Squares (OLS).

Another important issue in this type of model is the selection of starting values for the optimization routine. It has proven to be very sensitive to those values. In this thesis it was based in the idea of van Dijk et al. (2002), which is a two dimensional grid search over γ and c. Replacing the transition function (2) by

$$F(\gamma(s_t - c)) = (1 + \exp\{-\gamma / \hat{\sigma}_{s_t} (s_t - c)\})^{-1}, \tag{13}$$

PUC-Rio - Certificação Digital Nº 0310438/CA

Where $\hat{\sigma}_{s_t}$ is the sample standard deviation of s_t, makes γ approximately normalized. Then, a set of grid values for c was defined as sample percentiles of the transition variable.

Finally, some comments about the estimation of the slope parameter γ are necessary. As pointed out by many other authors, it is rather difficult to obtain precise estimates of this parameter. One needs a large amount of data around c to have an accurate estimate of γ. This is due to the fact that when the value of γ is large, even large changes in this parameter will generate only small effects in the transition function which will be close to a step function. Therefore, the reader should not take into account the insignificance of the parameter if judged by its t-statistic against the hypothesis of non linearity.

However, in this thesis, the analysis with real data (NSW electricity demand) will use a long time series with more than fifty thousands observations. This should be taken into account for small and medium size samples.

PUC-Rio - Certificação Digital Nº 0310438/CA

3.4
Model Evaluation

After estimating the parameters one needs some tools to evaluate the model in order to verify if it was correctly specified. The tests proposed here are Lagrange Multiplier (LM) tests for the hypothesis of no error autocorrelation and LM-type test for the hypothesis of no remaining nonlinearity. These tests were implemented based on the work of Eitrheim and Terasvirta (1996) and Medeiros and Veiga (2003). They are constructed in the same manner as the linearity test described previously that is, using Taylor expansions to linearize the problem and then constructing auxiliary regressions to compute the statistic of the test.

3.4.1
Test of Serial Independence

Consider the following nonlinear model of order p with autocorrelated errors:

$$y_t = G(w_t, s_t; \psi) + \varepsilon_t \tag{14}$$

$$\varepsilon_t = \theta' \upsilon_t + \mu_t = \left(\sum_{j=1}^{q} \theta_j L^j \right) \varepsilon_t + \mu_t \qquad \mu_t \sim iid(0, \sigma^2) \tag{15}$$

Where L is lag operator, w_t = (1, y_{t-1}, …, y_{t-p})', $\psi = (\psi_1, ..., \psi_r)'$, $\upsilon_t = (\varepsilon_{t-1}, ..., \varepsilon_{t-q})'$, and $\theta = (\theta_1, ..., \theta_q)'$, $\theta_q \neq 0$. We need to assume that the roots of the polynomial in the error equation lies inside the unit circle, i.e. ε_t is stationary, and furthermore, that under the assumption of ε_t following an identically and independent distribution with mean 0 and variance σ^2, implying that $\theta = 0$, the process y_t is stationary and ergodic such that the parameters of (14) can be consistently estimated by nonlinear least squares. So the hypothesis of serial independence of errors ε_t in (14) would be $H_0 : \theta = 0$.

Assuming the necessary starting values fixed, the conditional normal log-likelihood function for observation t (t = 1, …, T) takes the form

$$l_t = c - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left\{ y_t - \sum_{j=1}^{q} \theta_j y_{t-j} - G(w_t, s_t; \psi) + \sum_{j=1}^{q} \theta_j G(w_{t-j}, s_{t-j}; \psi) \right\}^2 \tag{16}$$

As pointed out in Eitrheim and Terasvirta (1996), the information matrix is block diagonal such that the second partial derivative with respect to σ^2 forms its own block. Considering σ^2 fixed in (14) to derive the test statistic, the first partial derivatives of the normal log-likelihood with respect to θ *and* ψ are

$$\begin{aligned} \frac{\partial l_t}{\partial \theta_j} &= \left(\frac{u_t}{\sigma^2} \right) \{ y_{t-j} - G(w_{t-j}, s_{t-j}; \psi) \}, \qquad j = 1, ..., q, \\ \frac{\partial l_t}{\partial \psi} &= \left(\frac{u_t}{\sigma^2} \right) \left\{ \frac{\partial G(w_t, s_t; \psi)}{\partial \psi} \right\} - \sum_{j=1}^{q} \theta_j \frac{\partial G(w_{t-j}, s_{t-j}; \psi)}{\partial \psi} \end{aligned} \tag{17}$$

Under H_0, the consistent estimators of (16) are

PUC-Rio - Certificação Digital Nº 0310438/CA

$$\left.\frac{\partial \hat{l}_t}{\partial \theta}\right|_{H_0} = \left(\frac{1}{\sigma^2}\right)\hat{\varepsilon}_t\hat{\upsilon}_t \quad \textit{and} \quad \left.\frac{\partial \hat{l}_t}{\partial \psi}\right|_{H_0} = -\left(\frac{1}{\sigma^2}\right)\hat{\varepsilon}_t\hat{z}_t,$$

Where $\hat{\upsilon}_t = (\hat{\varepsilon}_{t-1},...,\hat{\varepsilon}_{t-q})'$, $\hat{\varepsilon}_{t-j} = y_{t-j} - G(w_{t-j}, s_{t-j}; \hat{\psi})$, j = 1,…,q, $\hat{z}_t = \partial G(w_t, s_t, \hat{\psi})/\partial\psi$ and $\hat{\sigma}^2 = 1/T\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. Finally, the LM statistic can be written as

$$LM = \sigma^{-2}\left(\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\upsilon}_t'\right)\left\{\sum_{t=1}^{T}\hat{\upsilon}_t\hat{\upsilon}_t' - \sum_{t=1}^{T}\hat{\upsilon}_t\hat{z}_t'\left(\sum_{t=1}^{T}\hat{z}_t\hat{z}_t'\right)^{-1}\sum_{t=1}^{T}\hat{z}_t\hat{\upsilon}_t'\right\}\sum_{t=1}^{T}\hat{\upsilon}_t\hat{\varepsilon}_t' \tag{18}$$

Under the conditions that the moments implied by (17) exist, LM has an asymptotic $\chi^2(q)$ distribution.

The STPAR model in (5) can be written like (14), therefore, the components of $\hat{z}_t$ in the case when the transition function is a logistic function are as follows:

$$\frac{\partial \hat{G}}{\partial \phi_1} = w_t \tag{19}$$

$$\frac{\partial \hat{G}}{\partial \phi_2} = w_t F(s_t; \hat{\gamma}, \hat{c}) = w_t \hat{F}_t \tag{20}$$

$$\frac{\partial \hat{G}}{\partial \gamma} = \hat{g}_\gamma(t) = (1+\exp\{-\hat{\gamma}(s_t-\hat{c})\})^{-2}\exp\{-\hat{\gamma}(s_t-\hat{c})\}(s_t-\hat{c})\hat{\phi}_2' w_t \tag{21}$$

$$\frac{\partial \hat{G}}{\partial c} = \hat{g}_c(t) = \hat{\gamma}(1+\exp\{-\hat{\gamma}(s_t-\hat{c})\})^{-2}\exp\{-\hat{\gamma}(s_t-\hat{c})\}\hat{\phi}_2' w_t \tag{22}$$

For more details about the test, see Eitrheim and Terasvirta (1996). Hence, the test can be performed in four stages as follows:

1) Estimate the STPAR under the assumption of uncorrelated errors and compute the residual $\hat{\varepsilon}_t$.

PUC-Rio - Certificação Digital Nº 0310438/CA

2) Regress $\hat{\varepsilon}_t$ on $w_t, w_t\hat{F}_t$, $\hat{g}_\gamma(t)$ and $\hat{g}_c(t)$; Compute the residuals sum of squares $SSR_1 = \sum_{t=1}^{T} \tilde{\varepsilon}_t^2$.

3) Regress $\tilde{\varepsilon}$ on υ_t, w_t, $w_t\hat{F}_t$, $\hat{g}_\gamma(t)$ and $\hat{g}_c(t)$; Compute the residuals sum of squares SSR_2.

4) Compute the χ^2 statistic

$$LM_{\chi^2} = T\frac{SSR_1 - SSR_2}{SSR_1}$$

Or the F version of the test which is preferable when one is dealing with a small or moderate sample.

$$LM_F = \frac{\dfrac{SSR_1 - SSR_2}{q}}{\dfrac{SSR_2}{T - n - q}}$$

Where T is the number of observations and n is the dimension of the gradient vector $\hat{z}_t$. It is worth mentioning that step 2 is done just to make sure that errors are ortogonalized in relation to the regression.

PUC-Rio - Certificação Digital Nº 0310438/CA

3.4.2
Test of no remaining nonlinearity

As pointed out in Eitrheim and Terasvirta (1996) a non-linear model can be misspecified in a lot of different ways. In this thesis we will concentrate only in one specific way to determine whether there is remaining nonlinearity or not. In fact, what we will consider testing is the hypothesis of no additional nonlinear structure. For that we shall consider an additive STPAR with three regimes:

$$y_t = \phi_1' w_t + \phi_2' w_t F_1(\gamma_1(s_t - c_1)) + \phi_3' w_t F_2(\gamma_2(h_t - c_2)) + \varepsilon_t \qquad (23)$$

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

$$\varepsilon_t \sim iid(0, \sigma^2).$$

where i = 1, 2, 3 and F_2 is the same logistic function as F_1 and h_t is another transition variable or the same with a different lag. What we need to consider is that the model has been estimated without the third regime which is the second nonlinear component, and the null hypothesis H_0: $\gamma_2 = 0$ is tested against (23). For that, assuming the same as before, that is, under this null hypothesis the parameters $\phi_{1s}, \phi_{2s}, \gamma_1$ and c_1 can be consistently estimated by NLS.

Following the same idea as the test showed previously, we need to replace F_2 in (23) by its third order Taylor series approximation around $\gamma_2 = 0$ because under this null hypothesis the model is not identified. This equals $T_2 = f_{20} + f_{21}h_t + f_{22}h_t^2 + + f_{23}h_t^3$, where f_{2j}, j = 0,1,2,3, are functions of γ_2 such that $f_{20} = f_{21} = f_{22} = f_{23} = 0$ for $\gamma_2 = 0$. The result of (23) after some reparameterization is

$$y_t = \phi_1' w_t + \phi_2' w_t F_1(\gamma_1(s_t - c_1)) + \beta_2' w_t h_t + \beta_3' w_t h_t^2 + \beta_4' w_t h_t^3 + \mu_t \quad (24)$$

where $w_t = [y_{t-1}, ..., y_{t-p}]'$. Under H_0 which is equivalent now to testing H_0: $\beta_2 = \beta_3 = \beta_4 = 0$, $\mu_t = \varepsilon_t$ under H_0. The dimension of the null hypothesis is different now, increased from p to 3p. In fact one is testing against an additive STPAR component. The asymptotic theory behind the test is really similar to that one presented for the test with no autocorrelation. The test statistic is the same as (18) with $\hat{z}_t$ as before, however, $\hat{\upsilon}_t$ is different. It is now equal to $(w_t' h_t, w_t' h_t^2, w_t' h_t^3)$'. The test can be performed in the same steps as shown before in the test of no autocorrelation. The difference is just the new definition of $\hat{\upsilon}_t$ at stage 3 and the degrees of freedom in the F test, 3p and T – n – 3p, respectively.

PUC-Rio - Certificação Digital Nº 0310438/CA

3.5
Forecasting

The forecasting is one of the most important issues in the scope of this work. Forecasting with a periodic auto regression is roughly the same as an AR model. Therefore, one can actually forecast with the STPAR in the same way as a STAR.

The non-linearity of the models STPAR make the forecast with more than one step ahead much more complicated than in the linear models. Considering one step ahead forecast one could say that it is the same as an auto regressive model. For a complete discussion about forecasting see Lundbergh and Teräsvirta (2000). The next discussion will be based in a general procedure to obtain forecasts for more than one step ahead forecast.

Consider the model STPAR in (5). It follows that $E(y_{t+1} \mid Y_t) = f(w_{t+1};\psi)$, which is the unbiased forecast of y_{t+1} made at time t conditional to all previous observations Y_t up until that time. In this case, all the relevant information is included in $w_{t+1} = (1, y_t, y_{t-1},..., y_{t-(p-1)})'$; the notation of the forecast is then $y^f_{t+1|t}$. To forecast for more than one step ahead is not straightforward because to obtain $E(y_{t+2} \mid Y_t)$ is somehow more difficult. It is then

$$y^f_{t+2|t} = E(y_{t+2} \mid Y_t) = E\{[g(w^f_{t+2|t};\psi) + \varepsilon_{t+2}] \mid Y_t\} = E\{g(w^f_{t+2|t};\psi) \mid Y_t \tag{25}$$

where $w^f_{t+2|t} = (1, y^f_{t+1|t} + \varepsilon_{t+1}, y_t, y_{t-1},..., y_{t-(p-2)})'$; The exact expression for (14) would be

$$y^f_{t+2|t} = E\{g(w^f_{t+2|t};\psi) \mid Y_t\} = \int_{-\infty}^{\infty} g(w^f_{t+2|t};\psi) d\Phi(z) dz \tag{26}$$

where $\Phi(z)$ is the cumulative distribution function of ε_{t+1}. Numerical integration would be needed to obtain the forecast, and for longer time horizons multiple integration is necessary.

PUC-Rio - Certificação Digital Nº 0310438/CA

Computationally, it is more feasible to obtain the forecasts recursively without numerical integration. As pointed out in Granger and Teräsvirta (1993) another way to do it would by simulating. One could use Monte Carlo, relying on the error distribution, or bootstrap, applying resample to obtain the forecast. An advantage of these numerical approximations to the true expectations relies on the fact that instead of having one point forecast, one has a number of point forecasts for each period to be predicted. Hence, what is available is a density forecast, and confidence intervals may be constructed on the basis of them. The reader needs to remember that in doing so one of these procedures, it doesn't account for sampling uncertainty and, therefore, the intervals are somewhat narrow.

In this thesis, Monte Carlo simulation was used. For k steps ahead forecast horizon it follows that

$$y^{f}_{t+k,i|t} = 1/N \sum_{i=1}^{N} g(w^{f}_{t+k,i};\psi) \tag{27}$$

where each of the N values of $\varepsilon_{t+(k-1)}$ in $w^{f}_{t+k,,i}$ is drawn independently from a normal error distribution of the estimated model, that is with the same mean and variance. By the weak law of large numbers, the forecast is asymptotically unbiased as $N \to \infty$.

PUC-Rio - Certificação Digital Nº 0310438/CA