Lecture 11: Nonlinear Models

Dr. Joao B. Duarte[1]

[1]Nova School of Business and Economics
University of Cambridge

Masters, Economics: Macroeconometrics
Lisbon
Spring 2017

Lecture Objectives:

- Linear vs Nonlinear models properties.
- Intro to TAR and STAR models.
- Intro to Markov Chains.
- Markov switching models.

Secondary Readings:

- Chapter 7, Applied Econometric Time Series, Enders, Walter, Fourth Edition.
- Chapter 22, Time Series Analysis, Hamilton, James, first edition

Nonlinear Models Motivation

- So far, we have discussed linear time series models. The linear models are at the forefront of time series research.
- Partly, their popularity is related with the simplicity of the models. But we also saw that the Wold decomposition provides solid theoretical basis to take the linear models seriously.
- Moreover, if the time series are Gaussian, then the linear projection is better than any nonlinear forecast!

Nonlinear Models Motivation

- However, often times in real applications, the processes are found to be non-Gaussian.
- And the Wold decomposition, although important, provides a representation only and not a full probabilistic description of time series. Lets recall the Wold Decomposition:
- Wold Decomposition: any weakly stationary stochastic process, z_t, with finite mean, μ, that does not contain deterministic components, can be written as a linear function of uncorrelated random variables, ε_t , as:

$$z_t = \mu + \sum_{i=0}^{\infty} \phi_i \varepsilon_{t-i} \tag{1}$$

where ε_t is a white noise process and $\phi_0 = 1$

Nonlinear Models Motivation

- The Wold decomposition thus only pins down the first and second moments of any process z_t.
- Hence, it is possible for ε to have different third (skewness) and fourth (kurtosis) moments with the same representation as in (1). In other words, many processes could be represented as (1).
- In order, to make the best forecast possible, it is important to take into account the full probabilistic description of z_t.
- Note, that if ε_t is assumed to be Gaussian, then (1) becomes more than a representation, it becomes a full probabilistic description of z_t. In this case, the linear prediction is optimal.

Nonlinear Models Motivation

- What are the behavioral facts that motivate the usage of nonlinear models? Or to put it differently, what are the characteristics in the data that generate non-Gaussian processes?
- Rare events, structural breaks and asymmetries.
- We will cover some popular models that have been useful in dealing with data that present such characteristics: threshold AR (TAR), smooth transition AR models (STAR) and Markov switching models.

TAR model

- The TAR (Tong 1983) is an example of regime switching model that allows the behaviour of a series y_t to depend on an particular state of the system (regime).
- In the business cycle, the behaviour of some aggregates change dramatically during a crises. The volatility increases, but also the persistence of the mean model can also change.
- In such instances, the TAR model can provide an helpful description of the data. One major advantage of this model is that it can be estimated via OLS. This is generally not the case as regime switching models are complex and usually require more advanced estimation techniques.

TAR model

- Suppose that the persistence of y_t is different if the sequence takes positive or negative values. Let a_1 be the persistence of y_t if it takes positive values and a_2 the persistence otherwise. Moreover, let $|a_2| > |a_1|$ such that negative values of y_{t-1} are more persistent. Then we have:

$$y_t = \begin{cases} a_1 y_{t-1} + \varepsilon_{1t} & \text{if} \quad y_{t-1} > 0 \\ a_2 y_{t-1} + \varepsilon_{2t} & \text{if} \quad y_{t-1} \leq 0 \end{cases} \tag{2}$$

Linear vs Nonlinear Models

- Linear models are symmetric and the adjustment process is the same regardless of the initial condition. Consider the homogeneous part of the TAR model:

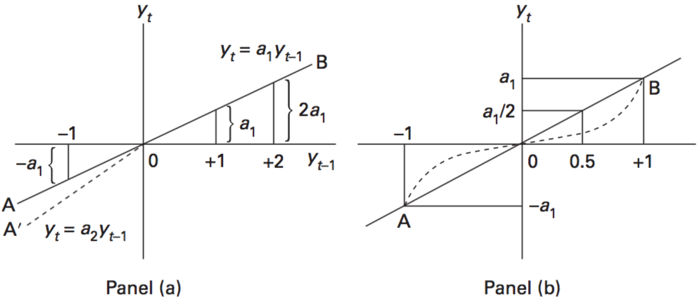


FIGURE 7.1 Two Nonlinear Adjustment Paths

Figure: Enders figure.

TAR Estimation

- The errors are responsible for the regime switching, as in each regime the homogeneous solution would just converge to its respective long-term mean.
- If we also assume that the error variance is the same across the two regimes we can write the model as:

$$y_t = a_1 I_t y_{t-1} + a_2(1 - I_t)y_{t-1} + \varepsilon_t \tag{3}$$

  where $I_t = 1$ if $y_{t-1} > 0$ and $I_t = 0$ if $y_{t-1} \leq 0$.
- I_t is a dummy variable. We use it ti construct the variables coming out of the product of this dummy variable and y_{t-1}. We can then use simple OLS to estimate (3).

General TAR

- Let τ be a threshold that determines when we are in each regime, let p be the number of lags of the AR process in the first regime and r the number of lags in the second regime, a more general version of (3) is given by:

$$y_t = I_t \left[a_{10} + \sum_{i=1}^{p} a_{1i} y_{t-i} \right] + (1 - I_t) \left[a_{20} + \sum_{i=1}^{r} a_{2i} y_{t-i} \right] + \varepsilon_t \quad (4)$$

where $I_t = 1$ if $y_{t-1} > \tau$ and $I_t = 0$ if $y_{t-1} \leq \tau$.

General TAR Estimation

- If τ is known, we can again use OLS. Note that adjustment process does not need to be continuous. Hence, this model can also capture sudden jumps.
- The estimation procedure follows the same variable creation for the posterior application of OLS.
- Example with $\tau = 0$:

Table 7.1 A TAR Model with Regime Dependent Variances

t	1	2	3	4	5	6	7
y_t	0.5	0.3	−0.2	0.0	−0.5	0.4	0.6
y_{t-1}	NA	0.5	0.3	−0.2	0.0	−0.5	0.4
y_{t-2}	NA	NA	0.5	0.3	−0.2	0.0	−0.5
I_t	NA	1	1	0.0	0.0	0.0	1
$I_t y_{t-1}$	NA	0.5	0.3	0.0	0.0	0.0	1
$(1 - I_t) y_{t-1}$	NA	0.0	0.0	−0.2	0.0	−0.5	0
$I_t y_{t-2}$	NA	NA	0.5	0.0	0.0	0.0	−0.5
$(1 - I_t) y_{t-2}$	NA	NA	0.0	0.3	−0.3	0.0	0.0

Figure: Enders TAR data preparation example.

General TAR Estimation

- If the variances of the errors are assumed to be different so that we have:

$$y_t = \begin{cases} a_{10} + a_{11} y_{t-1} + ... + a_{1p} y_{t-p} + \varepsilon_{1t} & \text{if} \quad y_{t-1} > \tau \\ a_{20} + a_{21} y_{t-1} + ... + a_{2r} y_{t-r} + \varepsilon_{2t} & \text{if} \quad y_{t-1} \leq \tau \end{cases} \tag{5}$$

- In this case, we need to separate the data according to the threshold criteria and then estimate each equation by OLS separately.

General TAR Estimation

- For instance, take the previous example and assume once again that $\tau = 0$
- Then, we just need to separate the dataset into two subsets:

Positive		**Negative**	
y_t	y_{t-1}	y_t	y_{t-1}
0.3	0.5	0.0	–0.2
–0.2	0.3	–0.5	0.0
0.6	0.4	0.4	–0.5

Figure: Enders TAR data preparation example 2.

- We apply OLS for each sub-dataset in order to estimate (5).

TAR Estimation with Unknown Threshold

- Often times, in practice we do not know the value of the threshold.
- Chan (1993) shows how to obtain a super-consistent estimate of the threshold τ. This is also called the SETAR model where SE stand for Self-Exciting.
- The idea is to use the value of observations to find the threshold.
- The threshold should be within the data limit points otherwise there is only one regime.
- Finally, take 15% of the highest and lowest observations as possible candidates, and estimate the TAR model for each possible remaining candidate. The estimate with the smallest residual sum of squares contains the consistent estimate of the model.

TAR Estimation with Unknown Threshold

- Example of data trimming in searching for a threshold:

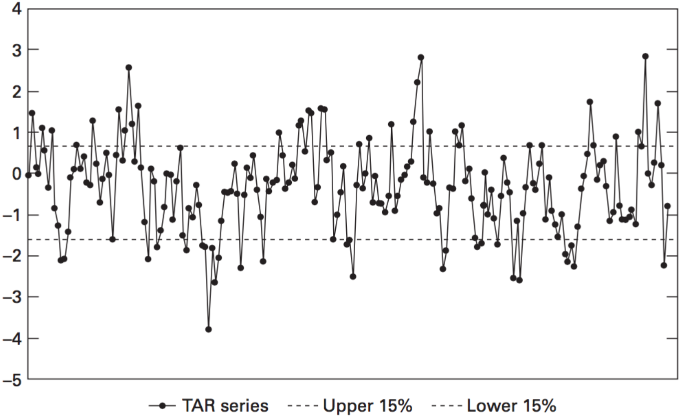


FIGURE 7.3 Estimation of the Threshold

Figure: Enders Figure

Pretesting for TAR model

- In the TAR case, the linear model is nested in the TAR model. Hence, we can use a F-test with the null of $a_{10} = a_{20}$ and $a_{11} = a_{21}$.
- However, under the null hypothesis the nuisance parameter τ is unidentified.
- Hansen (1997) showed how to appropriately obtain the appropriate critical values using a bootstrapping procedure.

$$F^* = \frac{(SSR_r^* - SSR_u^*)/m}{SSR_u^*/(T-2m)}$$

TVAR model

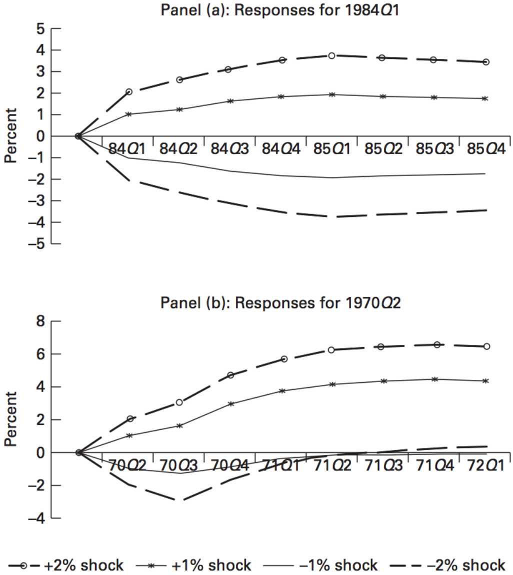
- The TAR can be extended to a multivariate framework:

$$x_t = \begin{cases} A_{10} + A_{11}x_{t-1} + \varepsilon_{1t} & \text{if} \quad q_{t-1} > \tau \\ A_{20} + A_{21}x_{t-1} + \varepsilon_{2t} & \text{if} \quad q_{t-1} \leq \tau \end{cases} \tag{6}$$

- Note, that when we drop the linearity, the Wold decomposition can no longer be used to calculate the IRFs.
- The IRFs now will be different depending on the sign and the regime that we are in. They are history dependent.
- The general point is that the impulse responses from a nonlinear model depend on the sign and magnitude of the shocks as well as the initial state, or history, of the system.

TVAR model

- Example of nonlinear IRFs:



STAR Models

- The TAR model imposes a very sharp transition into another regime. In some cases, the regime switch takes place over a smooth transition.
- Smooth transition autoregressive (STAR) models allow the autoregressive parameters to change slowly. Consider the logistic version of the STAR model called LSTAR:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \theta[\beta_0 + \beta_1 y_{t-1}] + \varepsilon_t \tag{7}$$

where

$$\theta = [1 + \exp\left(-\gamma(y_{t-1} - c)\right)]^{-1} \tag{8}$$

STAR Models

- If we are interested in a symmetric behaviour whereby the regime has more to do with the extremes behaviours, then the exponential STAR can be more appropriate (ESTAR):

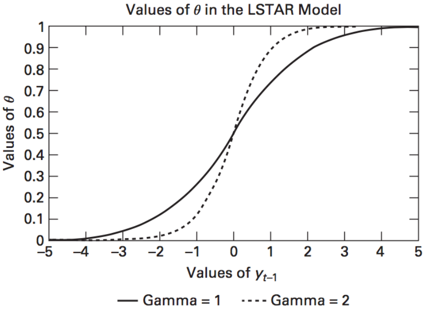
$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \theta[\beta_0 + \beta_1 y_{t-1}] + \varepsilon_t \tag{9}$$

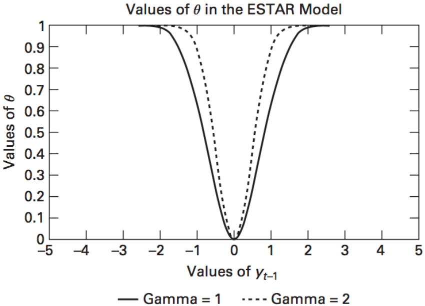
where

$$\theta = 1 - \exp\left(-\gamma(y_{t-1} - c)^2\right) \tag{10}$$

- In both cases, the γ is called the smoothness parameter.

LSTAR vs ESTAR Models





Nonlinear Models in R

- There is an exceptional package in R called `tsDyn`.
- It includes functions to estimate and analyze the nonlinear models we have seen thus far, with the exception of the ESTAR model.
- It also includes specific tests to detect a specific form of nonlinearity associated with a particular nonlinear model.
- The ideal way of testing for nonlinear effects is precisely to do it against a particular form of nonlinearity.

Testing for Nonlinear Effects - Portmanteau Tests

- However, sometimes one does not have a prior idea of the form of nonlinearity.
- Portmanteau tests usually refer to residual-based tests that do not have a specific alternative hypothesis.
- The BDS (Brock, Dechert, Scheinkman, and LaBarron (1996)) is arguably one of the most popular tests for independence.
- The test examines the distance between different pairs of residuals. If the residuals are independent, then the probability that the distance between any pair is less then say d should be the same.
- The only problem with this test with regards to nonlinear effects, is that dependence can be not only associated with nonlinear effects but also with other misspecification problems such as serial correlation, parameter instability, structural breaks and other issues.

Markov Chains Quick Review

- Before introducing the Markov switching models (Hamilton 1989), we will review the basic properties of Markov Chains.
- If we have a discrete finite number of states $s_t \in S$, a Markov chain fully describes the stochastic dynamics of the states s_t.
- The key assumption is that only the current state is need in order to have a full probabilistic description of next period's state. That is:

$$Prob(s_{t+1}|s_t, s_{t-1}, ...) = Prob(s_{t+1}|s_t) \tag{11}$$

Markov Chains Quick Review

- Hence, we can fully describe the systems evolution with a $n \times n$ transition matrix P:

$$
\begin{aligned}
p_{ij} &= Prob(s_{t+1} = j | s_t = i) \quad \forall i, j \\
p_{ij} &\geq 0 \quad \forall i, j \\
\sum_{j=1}^{n} p_{ij} &= 1 \quad \forall i
\end{aligned}
$$

- If s_t has an initial distribution π_0, this distribution together with P is able to characterize the stochastic properties of s_t up to any period T.

Markov Chains Quick Review

- It is also possible to find the unconditional distribution at given time t:

$$\pi'_{t+1} = \pi'_t P \tag{12}$$

Hence,

$$\pi'_t = \pi'_0 P^t \tag{13}$$

Markov Chains Quick Review

- A stationary distribution is one in which:

$$\pi' = \pi' P^t \tag{14}$$

- Every Markov chain has at least one stationary distribution. However, under some mild conditions, it can be shown that is has a unique stationary distribution.

- **Theorem**: If P is both aperiodic and irreducible, then:
  1. P has exactly one stationary distribution π^*
  2. For any initial π_0, we have $||\pi_0' P^t - \pi^*|| \to 0$ as $t \to \infty$

- One simple way to check if the conditions are met is that the all the elements of P must be strictly positive.

Markov Chains Quick Review

- Example: Consider a worker who at any given time t is either unemployed (state 1) or employed (state 2).
- Suppose that an employed worker loses his job with probability β and that an unemployed worker finds a job with probability α.
- In this case $S = \{1, 2\}$ and $p_{12} = \alpha$ and $p_{21} = \beta$. In this case P is given by:

$$P = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

The Markov Switching Model

- The basic TAR model allows the regime to switch depending on the value of y_{t-1}. Hence, the switch is endogenous.
- The Markov switching model developed by Hamilton (1989) proposes that the regime switches are exogenous.
- Suppose, again that the AR process depends on which regime we are in:

$$y_t = \begin{cases} a_{10} + a_{11}y_{t-1} + \varepsilon_{1t} & \text{if we are in regime 1} \\ a_{20} + a_{21}y_{t-1} + \varepsilon_{2t} & \text{if we are in regime 2} \end{cases} \tag{15}$$

- The difference now is that the regime changes are given by a Markov chain instead of by the value of the observed series.

The Markov Switching Model

- In this case, the probability of going from regime 1 to regime 2 is p_{12} and the other way around probability is p_{21}.
- In the Markov switching model, these probabilities will be estimates together with the other parameters of interest.
- The Markov switching models require more advanced estimation techniques. They are typically estimated via quasi-maximum likelihood or Gibbs sampling.

Good Luck!!

Thank you!!