

# Simple nonlinear time series models for returns\*

José Maria Gaspar

September 5, 2014

Econometrics II - Doctoral programme in Economics - FEP

## 1 A brief overview

Several time series in Economics and Finance present statistical properties that are empirically time varying, i.e., the statistical law that drives the data in each regime need not be the same. As a result, the possibility of having structural breaks may pose as a serious problem of misspecification resulting in erroneous estimation, misleading inference and poor predictions. To account for these issues, a few models have been developed to be able to deal with these nonlinearities. A first class of models pertains to the *threshold autoregressive* (TAR) models. The TAR is an  $AR(p)$  type with discontinuities. The sudden shift in regime occurs when an observed variable jumps above a certain threshold denoted as  $c$ . If the threshold variable is a lagged value of the time series itself of order  $d$  and  $d = 1$ , then a self-exciting TAR (SETAR)<sup>1</sup> with 2 regimes and an  $AR(1)$  in both regimes is given by:

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}y_{t-1} + \varepsilon_t & \text{if } y_{t-1} \leq c \\ \phi_{0,2} + \phi_{1,2}y_{t-1} + \varepsilon_t & \text{if } y_{t-1} > c, \end{cases} \quad (1)$$

---

\*This assignment was developed for the course of Econometrics II of the PhD programme in Economics at the School of Economics and Management, University of Porto.

<sup>1</sup>The regime is determined by the time series itself.

which can also be represented as:

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1}) I[1 - y_{t-1} > c] + (\phi_{0,2} + \phi_{1,2}y_{t-1}) (I[y_{t-1} > c]) + \varepsilon_t.$$

Generalizations of this model include changing the threshold variable or replacing the indicator function  $I[y_{t-1} > c]$  with some continuous function. We thus obtain a more gradual transition between any regime. Any such resulting model is called a *smooth transition* (STAR) model. For instance, a fairly popular choice is to use the logistic function:

$$G(y_{t-1}; \gamma, c) = \frac{1}{1 + \exp(1 - \gamma[y_{t-1} - c])},$$

and replace it in the indicator function. We refer to this as the Logistic STAR (LSTAR) model.

## 1.1 Estimation

The 2-regime STAR is given by:

$$y_t = (\phi_{0,1} + \phi_{1,1}y_{t-1}) I_t + (\phi_{0,2} + \phi_{1,2}y_{t-1}) (1 - I_t) + \varepsilon_t, \quad (2)$$

where  $I_t = 1(q_{t-1} > c)$  is a known function of the data and  $q_t$  is the threshold variable. The model in (2) is a non linear regression and is discontinuous. To estimate it, we can use sequential conditional least squares: (i) estimate  $\phi$ 's using LS for a given  $c$ ; and (ii) the LS estimate of  $c$  minimizes:

$$\hat{c} = \arg \min_{c \in [\underline{c}, \bar{c}]} \hat{\sigma}_\varepsilon^2(c).$$

Moreover, under  $\varepsilon \sim iidN(0, \sigma_\varepsilon^2)$ , then LS is equivalent to maximum likelihood estimation. For further estimation and testing procedures, see Terasvirta (1994, JASA). For either of the procedures, nonlinear estimation relies heavily on initial values.

For estimation of a LSTAR, there are problems with  $c$  and  $\gamma$ . When the latter is large, we need many observations in the neighborhood of  $c$  to estimate  $\gamma$  accurately.

## 1.2 Testing and model selection

Testing for linearity in time series pertains to testing equality in the two regimes. Under the null, we have  $H_0 = \phi_{i,1} = \phi_{i,2}$ . Alternatively, we can express the null as  $\gamma = 0$  which

reduces the logistic function to 0.5 and the STAR model to an AR. Hansen (1996) suggests using the following test statistic:

$$F = \sup_{c \in [\underline{c}, \bar{c}]} F[c],$$

where

$$F(c) = T \frac{(\tilde{\sigma}_\varepsilon^2(c) - \hat{\sigma}_\varepsilon^2(c))}{\hat{\sigma}_\varepsilon^2(c)},$$

with  $\tilde{\sigma}_\varepsilon^2(c) = (1/T) \sum (y_t - y_{t-1} \tilde{\phi})^2$ . If one rejects the null hypothesis of linearity, then one should proceed to choose  $G(\cdot)$  and estimate the STAR model. Model selection and parameter choice is usually based on the AIC, SBC, or the HQ criteria. A detailed numerical example shall be illustrated in the next section.

### 1.3 Forecasting

We evaluate (2) at  $t = T + 1$  and do recursive one-step predictions to obtain  $t = T + 1, T + 2, \dots$ , where  $\theta_{i,j}, c$  is replaced by its point estimates. Consider the case where  $y_t$  is described by the general nonlinear autoregressive model of order 1,

$$y_t = F(y_{t-1}; \theta) + \varepsilon,$$

for some nonlinear function  $F(\cdot)$ . When using a LS criterion, the optimal point forecasts of future values of the time series are given by their conditional expectations and the one-step ahead forecast is  $\hat{y}_{t+1|t} = F(y_t; \theta)$ . Two or more step ahead point forecasts is far more complicated. For a well referenced documentation on this point, see Franses and van Dijk (2000).

## 2 Numerical application

In this section I shall present a numerical illustration of the models discussed in the previous section, along with the procedures required for the usual tasks in time series analysis. For this purpose, I use the `tsDyn` package for nonlinear time series modelling available in R-Project. For all the available linear and nonlinear models in the package one needs only to run the command `availableModels()` once the package is loaded.

## 2.1 Data

The simulation is based on the univariate data set ( $UsUnemp$ ) from Caner and Hansen (2001) containing monthly data for US adult male unemployment rate spanning from 1956:M1 to 1999:M8.

## 2.2 Nonlinear models for US Unemployment

Figure 1 plots the time series  $UsUnemp$ , and the ACF plus the PACF for the time series. Although it is not clear cut from the picture, it already seems to evidence the existence

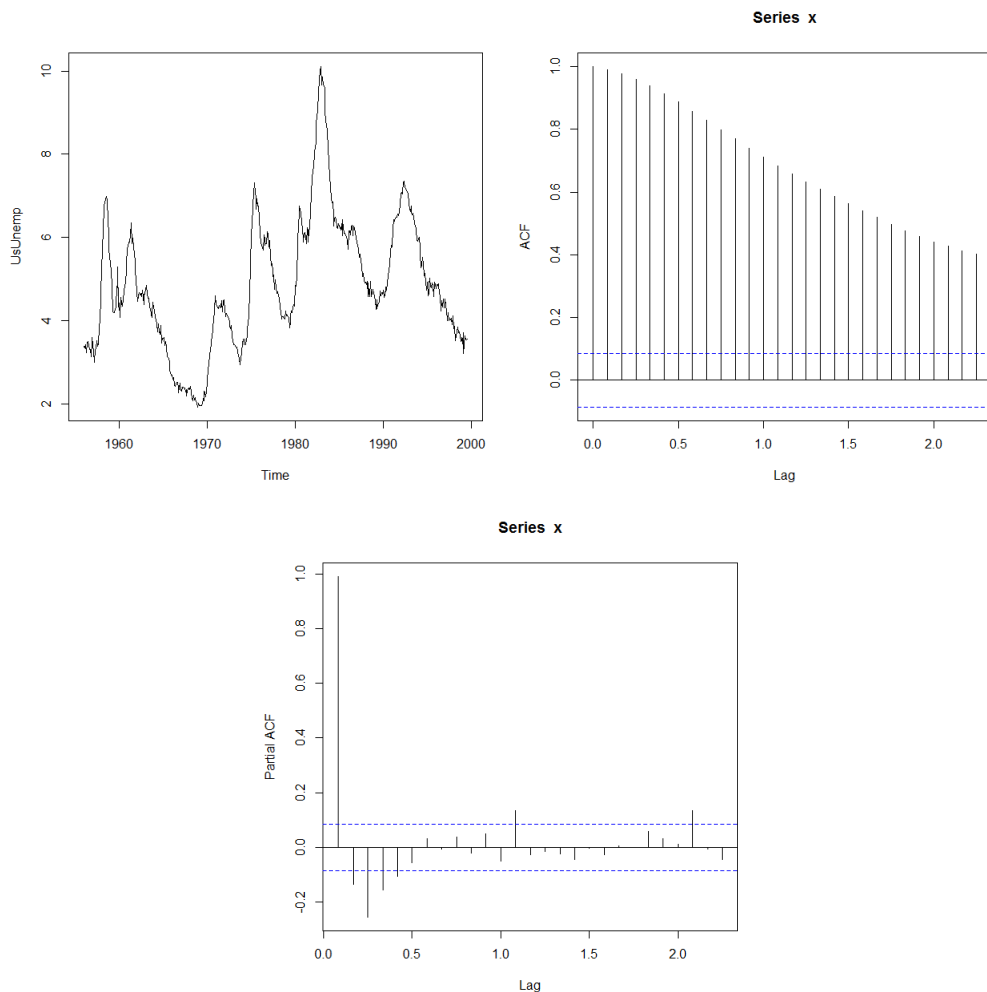


Figure 1 – Plot, ACF and PACF of  $y_t = UsUnemp_t$ .

of some structural breaks. Before proceeding with estimation, the next step will be to test for linearity in the series. This can be done with the `delta.lin.test` command

in the `tsDyn` package available in R. Assuming embedding dimensions  $m$  from 2 up to 5, the linearity test results are reported in table 1. The results are ambiguous as the

$m$	$eps$			
	0.7979	1.5958	2.3938	2.3938
2	0.30	0.12	0.22	0.52
3	0.02	0.02	0.58	0.12
4	0.02	0.08	0.04	0.22
5	0.18	0.02	0.26	0.64

Table 1 – Linearity test results. Middle values are reported p-values.  $eps$  represents different length scales and  $m$  are different dimensions.

p-values differ greatly depending on the embedding dimension and length scales. But for the sake of our purposes, and knowing in advance the results from model selection criteria, let us reject the null of linearity at the 5% significance level.

Next, I estimate different STAR models. For the sake of comparison, I shall begin by estimating a simple AR(2) model. The R estimation output is as follows:

```
Non linear autoregressive model
```

```
AR model
```

```
Coefficients:
```

```
      const      phi.1      phi.2
0.04863972  1.14281890 -0.15273731
```

Next, we estimate a SETAR(2,2) with time delay  $d = 1$ , whose reported results are given by the following output:

```
Non linear autoregressive model
```

```
SETAR model ( 2 regimes)
```

```
Coefficients:
```

```
Low regime:
```

```
      const.L      phiL.1      phiL.2
-0.0006888229  0.9560012261  0.0486719125
```

```
High regime:
```

```

      const.H      phiH.1      phiH.2
0.03465109  1.28923101 -0.29769290

```

Threshold:

-Variable:  $Z(t) = + (0) X(t) + (1)X(t-1)$

-Value: 4.978

Proportion of points in low regime: 60.15%                      High regime: 39.85%

Finally, we estimate a LSTAR with the same configurations. The output can be easily checked by running the commands in the script file. If we assign *mod* to be a list containing each of the three estimated models, the `sapply(mod,AIC)` command computes the AIC criterion for each one of the models. The reported values are given in table 2. Though the AIC is close for the SETAR and LSTAR models, I opt for the

	linear	SETAR	LSTAR
AIC	-1660.848	-1669.549	-1667.143

Table 2 – Model selection using AIC. The SETAR model seems the more appropriate.

SETAR as a more suitable model based on the results of table 2.

After choosing the model we are going to work with, the `tsDyn` package contains a very useful function that helps choosing the best parameters for the model, such as, e.g., the number of autoregressive polynomials for each regime, the *threshold delay*  $d$  for the threshold variable, and the threshold  $c$ . Let us assume there are at most 3 AR polynomials in each regime ( $m \leq 3$ ) and that  $d \leq 2$ . Inserting this in the function `selectSETAR()`, we get the following results:

```

Using maximum autoregressive order for low regime: mL = 3
Using maximum autoregressive order for high regime: mH = 3
Searching on 363 possible threshold values within regimes with
sufficient ( 15% ) number of observations
Searching on 9801 combinations of thresholds ( 363 ),
thDelay ( 3 ), mL ( 3 ) and mH ( 3 )
Results of the grid search for 1 threshold
      thDelay mL mH      th pooled-AIC
1          0  1  3 3.845823 -245.4812

```

2	0	1	3	3.844265	-244.8682
3	0	1	3	3.671880	-244.5442
4	0	1	3	3.828556	-244.2263
5	0	1	3	3.836013	-244.0534
6	0	2	3	3.845823	-243.9030
7	0	1	3	3.863612	-243.8684
8	0	1	3	3.856234	-243.6822
9	0	1	3	3.661763	-243.6590
10	0	1	3	3.817562	-243.6098

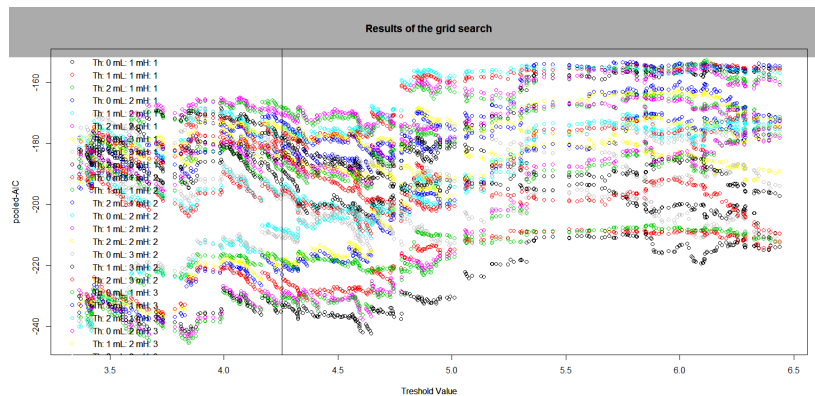


Figure 2 – Pooled AIC for each  $c$  after a grid search.

From figure 2 and the results above we choose to estimate a SETAR with  $m = 3$ ,  $mL = 1, mH = 3, d = 0$ , and  $c = 3.845823$ , based on the pooled-AIC, where  $mL$  and  $mH$  are the number of AR polynomials for the low and high regimes, respectively.

The estimation results of our new SETAR model are as follows:

Non linear autoregressive model

SETAR model ( 2 regimes)

Coefficients:

Low regime:

const.L      phiL.1  
0.01124622   1.00085091

High regime:

```

      const.H      phiH.1      phiH.2      phiH.3
0.09456131  1.10669723  0.21194097 -0.33680573

```

Threshold:

-Variable:  $Z(t) = + (1) X(t) + (0)X(t-1) + (0)X(t-2)$

-Value: 3.846 (fixed)

Proportion of points in low regime: 25.91%

High regime: 74.09%

Residuals:

```

      Min      1Q      Median      3Q      Max
-1.022366 -0.129020 -0.014286  0.107983  0.760483

```

Fit:

residuals variance = 0.03721, AIC = -1713, MAPE = 3.21%

Coefficient(s):

```

      Estimate  Std. Error  t value  Pr(>|t|)
const.L  0.011246   0.084584   0.1330  0.894277
phiL.1   1.000851   0.027611  36.2489 < 2.2e-16 ***
const.H  0.094561   0.043720   2.1629  0.031006 *
phiH.1   1.106697   0.045341  24.4083 < 2.2e-16 ***
phiH.2   0.211941   0.069233   3.0613  0.002318 **
phiH.3  -0.336806   0.044952  -7.4925  2.94e-13 ***

```

---

Threshold

Variable:  $Z(t) = + (1) X(t) + (0) X(t-1) + (0) X(t-2)$

Value: 3.846 (fixed)

We also re-estimate the linear and LSTAR models with the same specifications and reapply the `sapply(mod,AIC)` command just to arrive at the conclusion that the SETAR is still a better fit.



I have also ran a forecasting examplly using the data, which can be applied running the script file in R. An interesting task can be inspecting the fitted model skeleton. This can be achieved by computing the forecasting results under each model. Figure 3 contains fitted model skeletons for the AR, SETAR and LSTAR models. The skeleton

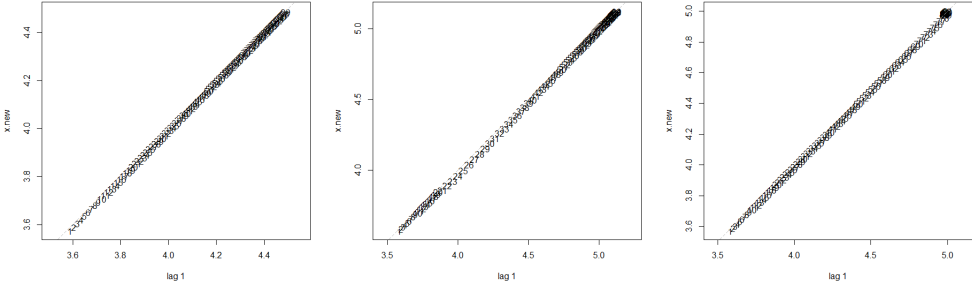


Figure 3 – Fitted model skeletons. From left to right: AR, SETAR and LSTAR models.

models show no evidence of a stationary solution. In fact, they all show divergence. This could, however, be a consequence of the strong evidence of presence of unit roots in the AR polynomials, something which I have conveniently dismissed for the sake of brevity in this discussion.

### 3 Concluding remarks

I have briefly described the main models used to fit time series that evidence strong non linearity. This strong non linearity stems mainly from discontinuities in different regimes. The models discussed fall into the class of the Threshold Autoregressive (TAR) models. A change in a regime occurs when a certain variable falls or jumps above a certain threshold. If that variable is the observed variable itself, we have the particular case of self exciting TAR (SETAR) models. If the threshold variable is some continuous function, we call the model a smooth transition (STAR) model. Nested in the latter is the logistic LSTAR (model), the case in which the function is of the logistic type. We have seen that testing for linearity is tantamount for testing equality of parameters under different regimes. We have also seen that forecasting for more than two periods ahead can be quite difficult for nonlinear models.

We have completed this assignment by running a simulation in R using data on US Unemployment from 1956 to 1999. We have compared linear models to SETAR and

LSTAR models and found that the SETAR fits the data comparably better.

## References

- [1] Caner and Hansen B. (2001), "Threshold autoregression with a unit root *Econometrica*", 2001, 69, 1555-1596.
- [2] Franses, P.H. and Dick v. D. (2000) "*Nonlinear Time Series Models in Empirical Finance*", Cambridge: Cambridge University Press.
- [3] Hansen, B. (1996) "Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis," *Econometrica*, 64(2), 413-30.
- [4] Terasvirta (1994), "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models", *JASA*, vol 89, pp 208-218