Lecture 9 - TAR, SETAR and Neural Net

Pedro Valls EESP-FGV & CEQEF-FGV

August 1, 2024

FGV EESP

CENTRO DE ESTUDOS
QUANTITATIVOS EM
ECONOMIA E FINANÇAS

Outline

- Introduction
- Specification
- Estimation
- Testing for STAR and TAR
- Diagnostic tests
- Forecasting - point forecasts
- Forecasting - interval forecasts
- Artificial neural networks
- Example: Forecasting Industrial production growth

Introduction

- Sometimes we can transform a non-linear model in a linear one
- But it is not always possible to transform a non-linear or time-varying parameter model into a linear constant parameter specification.
- For example, the model could be

$$y_t = f(x_t, z_t : \theta_t) + \varepsilon_t \tag{1}$$

- where f is a generic known function linking the explanatory variables x and z to the dependent variable y, according to the time-varying parameters θ_t
- Model specification in the non-linear setting is more complex.
- Parameter estimation is also much more complex and typically analytical solutions are not available.
- The derivation of optimal forecasts can be cumbersome and again analytical expressions are generally unavailable.

Introduction

- We start with the Threshold Autoregressive (TAR) and the Smooth Transition Autoregressive (STAR) Models
- Details on TAR and STAR models can be found [Granger and Teräsvirta, 1993], [Franses and van Dijk, 2000] and [van Dijk et al., 2002].
- We have focused on univariate models but multivariate versions are also feasible see for example [Artis et al., 2007].

Specification

- Let us consider the model

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

- This is an $AR(1)$ but its parameters change depending on whether the variable q_t is above or below the threshold c.
- It is called Threshold AR (TAR).
- If $q_t = y_{t-d}$ then the model is called Self Exciting TAR (SETAR).

Specification

- We can rewrite a SETAR (or TAR) as a single equation, i.e.

$$
\begin{aligned}
y_t &= (\phi_{0,1} + \phi_{1,1} y_{t-1})(1 - I(y_{t-1} > c)) \\
&+ (\phi_{0,2} + \phi_{1,2} y_{t-1}) I(y_{t-1} > c) + \varepsilon_t
\end{aligned} \tag{3}
$$

- where I is the indicator function.
- In the SETAR model the transition between sets of parameter values is abrupt and discontinuous.
- As an alternative to the indicator function I we could use the logistic function G with

$$
G(q_t : \gamma, c) = \frac{1}{1 + \exp(-\gamma \, |q_t - c|)} \tag{4}
$$

- The resulting model is called Logistic Smooth Transition AR (LSTAR or simply STAR).

Specification

- The parameter γ determines the smoothness; when γ is very large the model becomes similar to the TAR model, while for $\gamma = 0$ the model becomes linear.
- The STAR with two regimes can be rewritten as

$$y_t = \phi_1' x_t + (\phi_2 - \phi_1)' x_t G(q_t : \gamma, c) + \varepsilon_t \tag{5}$$

Specification

- Contrary to the linear case, in the non-linear world the specification is from a specific model an then generalize it if necessary.
- In particular [Granger, 1993] suggested the following procedure
  - specify a linear model, e.g. an $AR(p)$
  - test the null hypothesis of linearity against a specific for, of nonlinearity, e.g. SETAR or STAR
  - if linearity is rejected, estimate the specific nonlinear model
  - run diagnostics on the model and modify it of necessary
  - use the model for the required application, e.g. forecasting or computation of impulse response functions

Estimation

- Denote the conditional mean of y_t by $F(x_t : \theta)$ for example F is given by (5) and $\theta = (\phi_1, \phi_2, \gamma, c)$.
- Then we can derive the parameter estimators $\widehat{\theta}$ as the minimizes of the objective function

$$\sum_{t=1}^{T} (y_t - F(x_t : \theta))^2 = \sum_{t=1}^{T} \varepsilon_t^2 \tag{6}$$

- The resulting estimators are called *NLS* and denoted by $\widehat{\theta}_{NLS}$
- Under mild conditions on the functional form F which are satisfied for the case of STAR and TAR models, $\widehat{\theta}_{NLS}$ is consistent and has an asymptotically normal distribution see [Gallant and White, 1988]

Estimation

- In the case of TAR (or SETAR) models a conditional least squares procedure can be used.
- Conditioning on a given threshold variable and the threshold value c, these models are linear since for example $I(y_{t-1} > c)$ is a dummy variable.
- We can obtain the *OLS* estimators for $\boldsymbol{\phi} = (\phi_{0,1}, \phi_{0,2}, \phi_{1,1}, \phi_{1,2})$ and denote this *OLS* estimators by $\widehat{\boldsymbol{\phi}}(c)$.
- For each choice of c and the associated *OLS* estimators $\widehat{\boldsymbol{\phi}}(c)$ we can estimate the variance of the error term $\widehat{\sigma}^2(c)$.
- Choose the one that minimizes $\widehat{\sigma}^2(c)$ say $\widehat{c}$ with associated estimators $\widehat{\boldsymbol{\phi}}(\widehat{c})$ and $\widehat{\sigma}^2(\widehat{c})$

Estimation

- In the case of STAR models the choice of the starting values for the numerical optimization and the estimation of the smoothing parameter γ deserve some attention.
- Can use a grid search
  - First decide a set of values for γ and c to form the grid
  - Fix a value for γ and c say $\overline{\gamma}$ and $\overline{c}$
  - Conditional on these values, the model is linear in ϕ_1, ϕ_2 since $G(y_{t-1} : \gamma, c)$ is just a dummy variable
  - Estimate ϕ_1 and ϕ_2 by *OLS* as

$$\widehat{\boldsymbol{\phi}}(\overline{\gamma}, \overline{c}) = (X_t' X_t)^{-1} X_t' y_t \tag{7}$$

    with $X_t = (x_t'(1 - G(y_{t-1} : \overline{\gamma}, \overline{c})), x_t' G(y_{t-1} : \overline{\gamma}, \overline{c}))$
  - Using the residuals calculate $\widehat{\sigma}_\varepsilon^2(\overline{\gamma}, \overline{c})$
  - Select the values of γ and c as to minimize $\widehat{\sigma}_\varepsilon^2(\overline{\gamma}, \overline{c})$
  - The resulting values, say $\gamma^*, c^*, \widehat{\phi}_1^*, \widehat{\phi}_2^*$ are used as starting values for the numerical optimization to deliver *NLS*

Testing for STAR and TAR

- To test linear versus STAR model under the null hypothesis it can be written as $(\phi_2 - \phi_1)' = 0$ in (5), but the parameters c and γ are unindentified
- An alternative testing procedure is available with an asymptotic $\aleph^2$ distribution by rewriting the model (5) as

$$y_t = \frac{1}{2}(\phi_1 + \phi_2)' x_t + (\phi_2 - \phi_1)' x_t G^*(q_t : \gamma, c) + \varepsilon_t \tag{8}$$

- where $G^*(q_t : \gamma, c) = G(q_t : \gamma, c) - \frac{1}{2}$
- When $\gamma = 0$, $G^* = 0$, then we can approximate $G^*(q_t : \gamma, c)$ with a first order Taylor expansion around zero:

$$\begin{aligned} T_1(q_t \quad : \quad \gamma, c) &= G^*(q_t : 0, c) + \gamma \left. \frac{\partial G^*(q_t : \gamma, c)}{\partial \gamma} \right|_{\gamma=0} \\ &= \frac{\gamma}{4}(q_t - c) \end{aligned}$$

- since $G^*(q_t : 0, c) = 0$

Testing for STAR and TAR

- Substituting T_1 for G^* in (8) and reparametrize the model as:

$$y_t = \alpha + \beta' \widetilde{x}_t + \delta' \widetilde{x}_t q_t + u_t \tag{9}$$

- where $\widetilde{x}_t = y_{t-1}$
- It can be shown that $\gamma = 0$ if and only if $\delta' = 0$ to test linearity versus STAR and the asymptotic distribution is $\aleph^2(p)$ where p is the number of restrictions, e.g. $p = 1$ in the STAR(1) model
- When the alternative is TAR or SETAR we can use an $F-$statistics to test for linearity and we can write it as

$$F(c) = \frac{RSS_1 - RSS_2}{\widehat{\sigma}_2^2(c)} \tag{10}$$

- where RSS_1 and RSS_2 are the residual sum of squares form the linear and the TAR model respectively and $\widehat{\sigma}_2^2(c)$ is the residual variance form the TAR model for the threshold c.

Testing for STAR and TAR

- Since the threshold is unkown [Hansen, 1997] suggested to use the supremum among the $F-$statistics, i.e.

$$F_s = \sup_{c_i \in C} F(c_i) \tag{11}$$

- where C is the set of all possible values of the threshold.
- The distribution of F_s is unkown but can be computed numerically

Diagnostic tests

- Tests for homoskedasticity, no serial correlationa and normality presented before can be used for TAR and STAR models;
- However they only have an asymptotic justification
- Test of interest is whether there exists an additional regime. The same is similar to the test for linearily versus one regime, see [Eitrheim and Teräsvirta, 1996]

- Let us assume for simplicity that y_t only depends on y_{t-1} and write the model in general form as:

$$y_t = F(y_{t-1} : \theta) + \varepsilon_t \tag{12}$$

where $\varepsilon_t \overset{iid}{\sim} (0, \sigma_\varepsilon^2)$ and $t = 1, \ldots, T$

- The optimal in the MSFE sense $h-$steps ahead forecast is

$$\widehat{y}_{T+h} = E(y_{T+h} | I_T) \tag{13}$$

with $I_T = y_T$

- For $h = 1$, the computation of $\widehat{y}_{T+1}$ is straightforward since

$$\widehat{y}_{T+1} = F(y_T : \theta) \tag{14}$$

or

$$\widehat{y}_{T+1} = F(y_T : \widehat{\theta}) \tag{15}$$

if the parameters θ are unkown

Forecasting - point forecasts

- For $h = 2$, assuming that θ is known, we have

$$\widehat{y}_{T+2} = E\left[F(y_{T+1} : \theta)\right] \neq F\left[E(y_{T+1}) : \theta)\right] = F(\widehat{y}_{T+1} : \theta)$$

we can no longer use the simple iterative procedure for $h-$step forecasting

- We could use

$$\widetilde{y}_{T+2} = F(\widehat{y}_{T+1} : \theta)$$

as a two-steps ahead forecast, but it would be biased and not optimal

- The optimal two-steps forecast is

$$\begin{aligned}
\widehat{y}_{T+2} &= E\left[F(F(y_T : \theta) + \varepsilon_{T+1} : \theta) | I_T\right] \\
&= E\left[F(\widehat{y}_{T+1} + \varepsilon_{T+1} : \theta) | I_T\right] \\
&= \int F(\widehat{y}_{T+1} + \varepsilon_{T+1} : \theta) f(\varepsilon) d\varepsilon
\end{aligned} \tag{16}$$

Forecasting - point forecasts

- But it is difficult to compute the integral analytically
- Alternatively can use Monte Carlo or bootstrapping to approximate numerically the integral obtaining

$$\widehat{y}_{T+2}^{(mc)} = \frac{1}{R}\sum_{i=1}^{R} F(\widehat{y}_{T+1} + \varepsilon_i : \theta) \tag{17}$$

  where $\{\varepsilon_i\}_{i=1}^{R}$ are drawm form the entire distribution not the individual ε_i.
- Can drawn from a pre-specified distribution (Monte Carlo) or re-sampled from the in-sample estimated errors (bootstrap)
- For $h > 2$ is procedure is unfeasible.

Forecasting - interval forecasts

- For linear models with normal errors the symmetric forecast intervals knowing thet y_{T+h} ~ $N(\widehat{y}_{T+h} : Var(e_{T+h})$ is given by

$$\left[\widehat{y}_{T+h} - z_{\alpha/2}\sqrt{Var(e_{T+h})} : \widehat{y}_{T+h} + z_{\alpha/2}\sqrt{Var(e_{T+h})}\right] \tag{18}$$

  where $z_{\alpha/2}$ is the $(\ \alpha/2)\%$ critical value for the standard normal density.

- Similarly for non-linear models but the density function of y_{T+h} can be asymmetric and even bimodal.
- The figure below ilustrate this problem.

Forecasting - interval forecasts

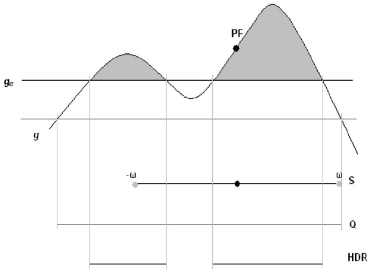


Figure 9.6.1: *Types of interval forecasts with a bimodal density for y_{T+h}: (S) is symmetric, (Q) is quantile based, (HDR) is highest density region*

Artificial neural networks I

- A univariate single layer feed-forward neural network model with n_1 hidden units (and a linear component) is specified as

$$y_t = \zeta_{t-h}\beta_0 + \sum_{i=1}^{n_1} \gamma_{1i} G(\zeta_{t-h}\beta_{1i}) + \varepsilon_t \tag{19}$$

- where y_t is the target variable, G is an activation function (typically the logistic function $G(x) = \frac{1}{(1+\exp(x))}$ and $\zeta_t = (1, y_t, y_{t-1}, \cdots, y_{t-p+1})$.
- The variables ζ_t are the imputs that enter the hidden layer represented by the activation functions $G(\zeta_{t-h}\beta_{1i})$ with connections strengths β_{1i} $i = 1, \ldots, n_1$ and through the weights γ_{1i} determine the output layer y.

Artificial neural networks II

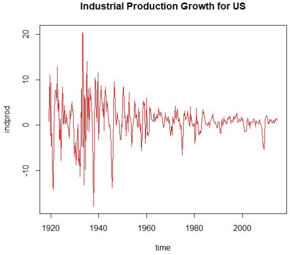
- The non-linear component of (19) namely $\sum_{i=1}^{n_1} \gamma_{1i} G(\zeta_{t-h}\beta_{1i})$ can be interpreted as a set of time-varying intercepts whose time-variation is driven by the evolution of the logistic functions $G(\zeta_{t-h}\beta_{1i})$ $i = 1, \ldots, n_1$
- For this reason, when n os large enough, the model can basically fit very well any type of temporal evolution.
- More flexibility can be obtained with a double layer feed-forward neural network.
- The ANN models in terms of empirical applications in macroeconomics are not very successful for small smaples and "insanity filters", which trim forecasts that are toow different from previously observe values, see e.g. [Stock and Watson, 1999].

Artificial neural networks III

- ANN models are also extensively used in the context of machine learning, for example in deep learning see e.g. [Hinton and Salakhutdinov, 2006].
- Deep learning has been mostly applied in the economic contect for financial applications based in big data.
- For example, [Sirignano et al., 2018] use to analyze mortgage between 1995 and 2014. [Heaton et al., 2016] and [Heaton et al., 2018] employ ANN in the context of portfolio theory.

Example: Forecasting Industrial production growth I

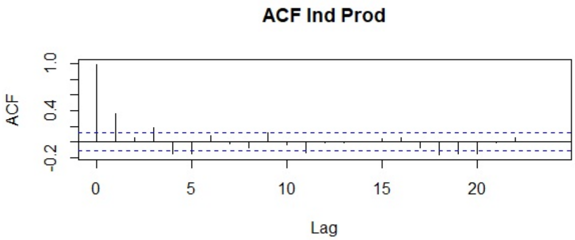
- The growth rate for the US industrial production index from 1930Q1 to 2014Q4
- The estimation sample will be set to start in 1930Q1 and end in 2000Q1, but the series starts in 1919.
- The forecasting sample covers the period from 2000Q2 to 2014Q4

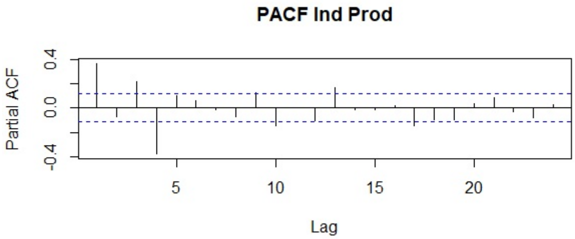


Example: Forecasting Industrial production growth II

- From the correlogram $AR(2)$ specification looks OK

Example: Forecasting Industrial production growth III





Example: Forecasting Industrial production growth IV

- First is to determine whether a TAR model is more appropriate than the $AR(2)$ modelo using Hansen´s sup LR test.
- The Hansen´s sup-LR test produces a value of 58.4956 with 95% confidence interval for the F sup of $[3.6901, 21.0099]$. Therefore we reject linearity
- `Test of linearity against setar(2) and setar(3)`
- ` Test Pval`
- `1vs2 134.9412 0`
- `1vs3 160.9857 0`

1. Diebold-Mariano Test

Model	RMSFE	p-	value
		AR	STAR
AR	1.150700	-	0.227385
TAR	1.108067	0.227385	-
STAR	1.103482	0.284112	0.437089

Reference I

Artis, M., Galvão, A. B., and Marcellino, M. G. (2007).
The transmission mechanism in a changing world.
Journal of Applied Econometrics, 22:39–61.

Eitrheim, Ø. and Teräsvirta, T. (1996).
Testing the adequacy of smooth transition autoregressive models.
Journal of Econometrics, 74(1):59–75.

Franses, P. H. and van Dijk, D. (2000).
Non-linear Time Series Models in Empirical Finance.
Cambridge University Press.

Gallant, A. R. and White, H. (1988).
A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models.
Balckwell.

Reference II

Granger, C. W. J. and Teräsvirta, T. (1993).
Modellind Non-linear Economic Relationships.
OXford University Press.

Granger, Clive, W. J. (1993).
Startegies for modelling nonlinear time-series relationships.
Economics Record, 69:223–238.

Hansen, B. E. (1997).
Inference in tar models.
Studies in Nonlinear Dynamics and Econometrics, 2:1–14.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2016).
Deep learning for finance: deep portfolios.
Applied Stochastic Models in Business and Industry, 33(1):3–12.

Heaton, J. B., Polson, N. G., and Witte, J. H. (2018).
Deep learning in finance.

Reference III

Hinton, G. E. and Salakhutdinov, R. R. (2006).
Reducing the dimensionality of data with neural networks.
Science, 313(5786):504–507.

Sirignano, J., Sadhwani, A., and Giesecke, K. (2018).
Deep learning for mortgage risk.

Stock, J. and Watson, M. (1999).
A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series, pages 1–44.
Oxford University Press, Oxford.

van Dijk, D., Teräsvirta, T., and Franses, P. H. (2002).
Smooth transition autoregressive models - a survey of recent developments.
Econometric Reviews, 21:1–47.