Journal of ELECTRICAL ENGINEERING, VOL. 56, NO. 12/s, 2005, 1–5

MULTIVARIATE LSTAR IN GEODESY

Tomáš Bacigál *

Regime-switching models such as threshold autoregressive ones are used to process the data, that are nonlinear in the sense of being piecewise linear. Here we extend the multivariate TAR to LSTAR family of models, where the transition function between regimes is smooth rather then abrupt. LSTAR stands for logistic smooth transition autoregressive. Inclusion of exogenous variables is considered. We briefly outline an application to geodetic data processing.

K e y w o r d s: time series, multivariate analysis, LSTAR, smooth transition regression, regime-switching

2000 *Mathematics Subject Classification*: 37M10 Time series analysis

1 INTRODUCTION

Processes that appear in nature and are subject to observation and analysis in such disciplines as geodesy, hydrology and meteorology, need not be always sufficiently described by linear models like ARMA and the like. There are many types of non-linearities that could "make things turbid", one of them being threshold non-linearity which represents piecewise linear structures and is easily interpretable as we can identify a particular variable (such as temperature or precipitation) that indicates the change in behaviour of other variables (e.g. river flow rate).

Section 2 deals with theoretical background of the models that describe such behaviour (switching between regimes determined by observable variables), particularly the TAR and STAR models. To make the application meaningful, it is essential to test the data for linearity against this particular non-linearity, which is the topic of section 3. Next the theory for model shape identification, its parameters estimation and suitability evaluation is given in section 4, whereas the last two sections contains some practical results of application in geodesy.

2 REGIME–SWITCHING

The most prominent member of the class, which assumes that the regime that occurs at time t can be determined by an observable variable z_t, is the Threshold Autoregressive (TAR) model. For completeness, there is also a class covering the determination by an unobservable process, representative of which is the Markov-Switching model, however it is not of our interest here. The TAR model assumes that the regime is determined by relation between the value of threshold variable z_t and threshold value denoted as r.

Then a 2-regime TAR model assuming an AR(p) in both regimes can be written

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1}\, y_{t-1} + \cdots + \phi_{p,1}\, y_{t-p} + \varepsilon_t & \text{if } z_t \le r, \\ \phi_{0,2} + \phi_{1,2}\, y_{t-1} + \cdots + \phi_{p,2}\, y_{t-p} + \varepsilon_t & \text{if } z_t > r, \end{cases} \tag{1}$$

or alternatively in matrix notation

$$y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t (1 - I[z_t > r]) + \boldsymbol{\phi}_2' \boldsymbol{X}_t I[z_t > r] + \varepsilon_t, \tag{2}$$

where $\boldsymbol{\phi}_j = (\phi_{0,j}, \phi_{1,j}, \ldots, \phi_{p,j})'$ are unknown parameters of j-th regime , $\boldsymbol{X}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, $I[A]$ is an indicator function with $I[A] = 1$ if the event A occurs and $I[A] = 0$ otherwise. M' denotes transposition of M.

A more gradual transition between the different regimes can be obtained by replacing the indicator function $I[z_t > r]$ in (2) by a continuous function $G(z_t, \gamma, r)$ which changes smoothly from 0 to 1 as z_t increases. The resultant model is called a Smooth Transition Autoregressive (STAR) and if rearranged a little it is given by

$$y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t + (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)' \boldsymbol{X}_t G(z_t, \gamma, r) + \varepsilon_t, \tag{3}$$

which is easily extendable to m-regimes version

$$\begin{aligned} y_t = {} & \boldsymbol{\phi}_1' \boldsymbol{X}_t + (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)' \boldsymbol{X}_t G(z_t, \gamma_1, r_1) + \ldots \\ & + (\boldsymbol{\phi}_m - \boldsymbol{\phi}_{m-1})' \boldsymbol{X}_t G(z_t, \gamma_{m-1}, r_{m-1}) + \varepsilon_t, \end{aligned} \tag{4}$$

A popular choice for the so-called transition function G is the logistic function

$$G(z_t, \gamma, r) = \frac{1}{1 + e^{-\gamma(z_t - r)}} \tag{5}$$

which results in the Logistic STAR (LSTAR). The parameter γ determines the smoothness of the transition. Notice that AR model is a special case of the LSTAR

* Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology, Radlinského 11, 813 68 Bratislava, Slovakia, E-mail: bacigal@math.sk

Research supported by VEGA 1/1033/04, VEGA 1/2032/05, VEGA 1/3006/06, APVT - 20 - 003204.

ISSN 1335-3632 © 2005 FEI STU

model in case $\gamma = 0$ and likewise the LSTAR becomes TAR as $\gamma \to \infty$. Alternatively an even transition function can be used, e.g. exponential function $G(z_t, \gamma, r) = 1 - e^{-\gamma(z_t - r)^2}$, so that corresponding model (Exponential STAR) assumes symmetric response of y_t to positive or negative values of $z_t - r$.

For practical purposes it may be more useful to consider the LSTAR model instead of the TAR or ESTAR models since it allows for smooth changes and asymmetric response to shocks.

A more general case considers some explanatory variable x_t to be included in regression and further extension to multivariate space yields

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1 \boldsymbol{X}_t (1 - G(z_t, \gamma, r)) + \boldsymbol{\Phi}_2 \boldsymbol{X}_t G(z_t, \gamma, r) + \boldsymbol{\varepsilon}_t, \quad (6)$$

with

$$\boldsymbol{X}_t = (1, \boldsymbol{y}'_{t-1}, \dots, \boldsymbol{y}'_{t-p}, \boldsymbol{x}'_{t-1}, \dots, \boldsymbol{x}'_{t-q})' \quad (6a)$$

where $\boldsymbol{y}_t = (y_{1t}, \dots, y_{kt})'$ is k-dimensional modelled and $\boldsymbol{x}_t = (x_{1t}, \dots, x_{lt})'$ l-dimensional exogenous variable, p and q are the orders of auto and exogenous regression, respectively, and $\boldsymbol{\Phi}_j$ is the $k \times (1 + pk + ql)$ matrix of unknown parameters corresponding to j-th regime. It is questionable how to denote the resultant general model (6,6a), one possible way is to put the names of all participant model together into the acronym STVARX (Smooth Transition Vector Autoregressive model with eXogenous variables) preferred by some recent authors, or to adopt a simpler version STR (Smooth Transition Regression model) introduced in [8] where contrary to (6a), the model does not incorporate lagged but rather actual value of exogenous variable.

An essential part of the STAR model is surely the threshold (transition) variable z_t which indicates what behaviour to expect at time t. In univariate case it is usual to set $z_t = y_{t-d}$ for an positive integer d, however in more general case it is worth considering other options, e.g. some linear combination of lagged endogenous and exogenous variables included in regression (see [8] for inspiration) or most recent idea of utilizing aggregation operators to reconstruct the threshold variable as shown in [7] (an application to univariate river flow rate time series analysis). In short, for univariate self-exciting model $z_t = A(y_{t-1}, \dots, y_{t-d})$ where A is an aggregation operator (agop). Typical continuous agops on the real line ($R^n \to R$) are arithmetic mean, weighted means or OWA operators $A_{OWA}(a_1, \dots, a_n) = \sum_{i=1}^n w_i a'_i$, where $\boldsymbol{a}'$ denotes a non-decreasing permutation of $\boldsymbol{a} = (a_1, \dots, a_n)$ and $\boldsymbol{w}$ are weights. In class of OWA we can find also MIN ($w_1 = 1$ and $w_i = 0$ otherwise) eventually MAX ($w_n = 1$, $w_i = 0$) operators and all order statistics. A multivariate case with exogenous elements using agops obviously allows for greater variety of construction methods of transition variable z_t, e.g. straightforward nesting $A(A_1(\dots), \dots, A_m(\dots))$ etc.

3 TESTING FOR NONLINEARITY

Before any specific non-linear model is getting started to build up, it is desirable to test the time series for linearity against the suspected non-linearity. There are several methods, one possible way of detection is to compare the in-sample fit of the regime-switching model with that of a linear model (which can be considered as 1-regime model), when the linear model is taken as null and regime-switching one as alternative hypothesis. In the case of 2 regimes it means equality against inequality of the regression parameters in the two regimes.

Tsay in [9] propose one that put threshold nonlinearity (abrupt transition between regimes) against linearity, using a regression rearranged according to the increasing order of threshold variable that effectively transforms a threshold model into a changepoint problem. Another approach utilizes Lagrange Multipliers (LM) statistics and is available for STAR model. Both tests are simple and performs well in finite samples, yet it does not depend on the alternative model, nor does it encounter the problem of unidentified nuisance parameters under the null hypothesis (see discussion in [4], pp.100). As the point of our interest here is the STAR family of models, a reader interested in application of multivariate TAR is referred to our earlier work [2].

Besides equality of the AR parameters in the two regimes, $H_0 : \boldsymbol{\phi}_1 = \boldsymbol{\phi}_2$, the null hypothesis of linearity can alternatively be expressed as $H'_0 : \gamma = 0$. If $\gamma = 0$, the logistic function (5) is equal to 0.5 for all z_t and the STAR model collapse to an AR model with parameters $(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)/2$.

Following [4], rewrite the STAR model (3) as

$$y_t = \frac{1}{2}(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)' \boldsymbol{X}_t + (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)' \boldsymbol{X}_t G^*(z_t, \gamma, r) + \varepsilon_t, \quad (7)$$

where $G^*(z_t, \gamma, r) = G(z_t, \gamma, r) - 1/2$ and approximate the shape function $G^*(z_t, \gamma, r)$ with a third-order Taylor approximation around $\gamma = 0$, that is

$$\begin{aligned} T_3(z_t, \gamma, r) &\approx G^*(z_t, 0, r) + \sum_{i=1}^{3} \frac{1}{i!} \gamma \left(\frac{\partial^i G^*(z_t, \gamma, r)}{\partial \gamma^i} \Big|_{\gamma=0} \right) \\ &= \frac{1}{4}\gamma(z_t - r) + \frac{1}{48}\gamma^3 (z_t - r)^3, \end{aligned} \quad (8)$$

where we have used the fact that $G^*(z_t, \gamma, r)$ and its second derivative with respect to γ evaluated at $\gamma = 0$ equals zero. After substituting $T_3(\cdot)$ for $G(\cdot)$ in (7) and rearranging terms this yields the auxiliary regression

$$y_t = \beta_{0,0} + \boldsymbol{\beta}'_0 \boldsymbol{X}_t + \boldsymbol{\beta}'_1 \boldsymbol{X}_t z_t + \boldsymbol{\beta}'_2 \boldsymbol{X}_t z_t^2 + \boldsymbol{\beta}'_3 \boldsymbol{X}_t z_t^3 + \eta_t, \quad (9)$$

where $\boldsymbol{\beta}_j = (\beta_{j,0}, \dots \boldsymbol{\beta}_{j,p+q})$, $j = 0, 1, 2, 3$, are functions of the parameters $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, γ and r. Inspection of the exact relationships shows that the null hypothesis H'_0: $\gamma = 0$ corresponds to H''_0: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = 0$ (and $\eta_t = \varepsilon_t$), which can be tested by a standard LM-type

test. Note that if z_t is one of the variables included in $\boldsymbol{X}_t$, the terms $\beta_{j,0} z_t^j$, $j = 1, 2, 3$, should be dropped from auxiliary regression to avoid perfect multi-collinearity.

Under the null hypothesis of linearity, the test statistic denoted as LM_3 has an asymptotic χ^2 distribution with $3p$, alternatively $3(pk+ql)$ degrees of freedom if exogenous variable is included and multivariate case (6,6a) considered.

The LM_3 test-statistic based on (9) for multivariate system (6) can be computed as follows:

1. Estimate the model under the null hypothesis of linearity by regressing $\boldsymbol{y}_t$ on $\boldsymbol{X}_t$. Compute the residuals $\hat{\boldsymbol{\varepsilon}}_t$ and the variance-covariance matrix $\boldsymbol{\Sigma}_0 = n^{-1} \sum_{t=h+1}^{n} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$, where $h = \max(p, q, d)$.
2. Estimate the auxiliary regression of $\hat{\boldsymbol{\varepsilon}}_t$ on $\boldsymbol{X}_t$ and $\boldsymbol{X}_t z_t^j$, $j = 1, 2, 3$. Denote the residuals as $\hat{\boldsymbol{e}}_t$, then $\boldsymbol{\Sigma}_0 = n^{-1} \sum_{t=h+1}^{n} \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t'$.
3. $LM_3 = n(\ln|\boldsymbol{\Sigma}_0| - \ln|\boldsymbol{\Sigma}_1|)$.

Under the null hypothesis of linearity, LM_3 has an asymptotic χ^2 distribution with $3(pk + ql + 1)$ degrees of freedom. In small samples it is recommended to use F-version of the LM_3, as it has better size and power properties.

The LM-type test can also be used to select appropriate transition variable by minimizing p-value of the LM_3 computed for several candidates.

4 IDENTIFICATION, ESTIMATION AND EVALUATION

An empirical specification procedure for nonlinear model basically follows these steps: (i) specify an appropriate linear AR (ARX) model of order p (p, q), (ii) test the null hypothesis of linearity against the alternative of regime-switching nonlinearity which includes selecting appropriate variable that determines the regimes, (iii) estimate the parameters in the selected model, (iv) evaluate the model using diagnostic tests, (v) modify the model if necessary and (vi) use the model for descriptive or forecasting purposes.

When selecting orders of linear model (ideally by AIC, BIC), an over-specification of dynamics may be preferred to under-specification as the remaining autocorrelations could affect the outcome linearity test. Transition variable z_t can be sufficiently chosen from the LM-type linearity test minimizing the p-value or directly from estimation of particular models minimizing the sum of squared residuals. To choose the number of regimes, in some applications, past experience and substantial information may help, in others, few procedural techniques are available. One way is to divide the data into subgroups according to the empirical percentiles of z_t and use of linearity test statistic (e.g. LM_3) to detect any model change within each subgroup. Another way is to use an modification of LM-test described above to test a 2-regime STAR model against the alternative of an additive 3-regime model (for details see [4], pp.113).

An important question concerns detecting the appropriate orders p_1, p_2 and q_1, q_2 in general 2-regime model (6), where notation (6a) needs to be respecified to distinguish the regimes. The approach of setting $p_1 = p_2 = p$, $q_1 = q_2 = q$ from linear model can easily be inappropriate and the direct choice of p_j, q_j from nonlinear model based upon information criterion need not be satisfactory either. It seems fair to penalize the inclusion of the additional parameters (p_j, q_j) not for the whole sample size but only for the number of regime-corresponding observations. Such an alternative AIC and BIC proposed in [4] and [9] can be generalized as follows to suit (the s-regimes version of) the model (6):

$$AIC(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{s} \Big(n_j \ln|\hat{\boldsymbol{\Sigma}}_j| + 2k(kp_j + lq_j + 1) \Big),$$

$$BIC(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{s} \Big(n_j \ln|\hat{\boldsymbol{\Sigma}}_j| + (\ln n_j)(kp_j + lq_j + 1) \Big),$$

where $\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n_j} \sum_{t=h+1}^{n} (\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j \boldsymbol{X}_{j,t})(\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j \boldsymbol{X}_{j,t})' \Delta G_{j,t}$ is estimated covariance matrix with $n_j = \sum_{t=h+1}^{n} \Delta G_{j,t}$ and $\Delta G_{j,t} = G_{j-1,t} - G_{j,t}$, where $G_{j,t} = G_j(z_t, \gamma_j, r_j)$ is the transition function corresponding to j-th regime, $G_{0,t} = 1$ and $G_{m,t} = 0$.

Estimation of the parameters $\boldsymbol{\theta} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \gamma, r)'$ in the STAR model (6), where we assign $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$ and $\boldsymbol{X}_t(\gamma, r) = (\boldsymbol{X}_{1,t}'[1 - G(z_t, \gamma, r)], \boldsymbol{X}_{2,t}' G(z_t, \gamma, r))'$, is the problem of minimizing the trace of $\boldsymbol{\Sigma}(\boldsymbol{\Phi}, \gamma, r) = \sum_{t=h+1}^{n} (\boldsymbol{y}_t - \boldsymbol{\Phi} \boldsymbol{X}_t(\gamma, r))(\boldsymbol{y}_t - \boldsymbol{\phi} \boldsymbol{X}_t(\gamma, r))'$. This can be performed directly by nonlinear least squares (NLS) routine $\hat{\boldsymbol{\theta}} = \operatorname{argmin}_{\boldsymbol{\theta}} \operatorname{Tr}(\boldsymbol{\Sigma}(\boldsymbol{\Phi}, \gamma, r))$, for which several iterative optimization algorithms are available in statistical software. Alternatively, for fixed values of γ and r the model is linear in the parameters $\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2$, so that these can be (conditionally upon γ, r) estimated by Ordinary Least Squares (OLS) as

$$\hat{\boldsymbol{\Phi}}(\gamma, r) = \left(\sum_{t=h+1}^{n} \boldsymbol{X}_t(\gamma, r) \boldsymbol{X}_t(\gamma, r)' \right)^{-1} \left(\sum_{t=h+1}^{n} \boldsymbol{X}_t(\gamma, r) \boldsymbol{y}_t' \right)$$

and $(\hat{\gamma}, \hat{r}) = \operatorname{argmin}_{(\gamma, r)} \operatorname{Tr}(\boldsymbol{\Sigma}(\hat{\boldsymbol{\Phi}}, \gamma, r))$.

As the NLS need not always result in global minimum immediately, the conditional OLS grid search can help to define starting values for NLS. However, there is still a notorious problem with parameter γ that converges too slowly so that its estimate is rather imprecise (thus may appears insignificant) unless a large amount of observations (z_t) is available in the neighbourhood of the threshold r. Especially when γ is large, rescaling it becomes important (see [8], pp.123). Also, for ensuring reliable estimates of $\boldsymbol{\Phi}$, each regime should contain at least about 15% of observations, which limits the choice of r.

After an STR model has been estimated, its properties have to be evaluated. A first check is to ensure that the parameter estimates seem reasonable in the light of application (e.g. r outside the range). The next step is to examine residuals for remaining dynamics, that means the

usual tests for autocorrelations, normality and linearity tests as described in [1],[3],[4] or [8] in details. Furthermore, out-of-sample forecasting can also be considered as a way to evaluate estimated regime-switching model, in particular by comparison with forecasts from a benchmark linear model.

5 APPLICATION

Applications in geodesy can be many as the measuring as such is its integral part and the analysis involved plays important role in the interdisciplinary field (statics, geodynamics, geology, hydrology and the like). To try out the above procedures, as input data we use combination of geometric and physical observables, specifically the coordinate variation time series from permanent GPS observations as endogenous (modelled) and meteorologic observations (atmospheric pressure and temperature) as exogenous (explaining) variables realization recorded every 3 hours during 42 days. For any further motivation, see [5].

Before applying to empirical data, the performance was checked on simulations from simple two dimensional and 2-regimes stationary threshold AR model of order 1 with endogenous threshold variable (delayed by 1). Although there has appeared an overspecification of linear AR model order, the linearity test and information criteria gives correct value, and so does the linearity test in choosing the threshold variable (including delay). Estimation of logistic function slope parameter γ yields 100, which in practice corresponds to abrupt transition.

Naturally, real data behaves in much more complicated way and searching for the appropriate representation may appear like looking for a needle in a haystack unless we have good knowledge about the underlying processes and solid experience with fighting the pitfalls of model specification. Number of the variables involved and generality of the model under consideration determines the number of all possible combinations to be treated in the systematic approach.

In our application, firstly the data were inspected in univariate space (transition variable is the function of past values of the data itself), secondly the exogenous variables were included into the regression setup (as transition variable too) and finally we tried to build a vector model (LSTVARX) including all the variables we had at disposal.

As long as only lagged values in transition variable were taken into account, neither reasonable nor uniform result were achieved. The use of aggregation operators in transition improved the models performance in a certain amount, however, the best fit we got by reconsidering the model (6a) and including present value of the exogenous variable. The rationale behind is most probably that time steps of the data is not sufficiently small to contain the effects that are, e.g directly responsible for transition.

It is also essential to extract in advance, or to include into the model, any deterministic component that is believed to be present as could affect the identification correctness significantly, for example forgotten seasonality causes mis-specification of delay in z_t where multiples of season period are primarily preferred.

The best improvements comparing to corresponding linear model was achieved in the third coordinate with temperature as exogenous variable.

6 CONCLUSIONS

The purpose of this paper was to give an overview of one promising method of modelling nonlinear time series, at present largely utilized in econometrics. Introducing the concepts and mediating some empirical suggestions was emphasized instead of detailed description of particular experiment. The next investigation could concern an inspection of aggregation operators and out-of-sample prediction performance.

Acknowledgement

The author gratefully acknowledges many helpful suggestions of Professor Magda Komorníková.

References

[1] ARTLT, J.—ARLTOVÁ, M.: Finanční časové řady, Grada publishing, Praha, 2003.

[2] BACIGÁL, T.: Multivariate threshold autoregressive models in geodesy, Journal of Electrical Engineering **55** No. 12/s (2004).

[3] FRANSES, P.H.: Time Series Models for Business and Economic Forecasting, Cambridge University Press, Cambridge, 1998.

[4] FRANSES, P.H.—van DIJK, D.: Non-linear time series models in empirical finance, Cambridge University Press, Cambridge, 2000.

[5] IGONDOVÁ, M.: Využitie permanentných sietí GPS na modelovanie troposféry a ionosféry, Dizertačná práca, SvF STU, Bratislava, 2004.

[6] KOMORNÍK, J.—KOMORNÍKOVÁ, M.—MESIAROVÁ, J.: Crisp and fuzzy regime-switching models for exchange rates of the Slovak crown to Euro, Proc. of the 10th IFSA World Congress - IFSA 2003, 2003, pp. 438–441.

[7] KOMORNÍK, J.—KOMORNÍKOVÁ, M.—SZÖKEOVÁ, D.: Testing the Adequacy of Regime-Switching Time Series Based on Aggregation Operators, (to appear).

[8] GRANGER, W.J.—TERÄSVIRTA, T.: Modelling Nonlinear Economic relationships, Oxford University Press, Oxford, 1993.

[9] TSAY, R.S.: Testing and modeling multivariate threshold models, Journal of the American Statistical Association **93** (1998), 1188–1202.

Received _ _ 2005

Tomáš Bacigál (Ing) graduated in Geodesy and cartography at the Faculty of Civil Engineering of the Slovak University of Technology, Bratislava, at present he is a PhD student in applied mathematics at the same faculty. His supervisor is Professor Magda Komorníková.