Chapter 2

Non-linearity

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" in practice, but nearly unlimited number of theoretical models, that could be made up on demand. These models, however, should respect some reasonable restrictions to become applicable, mainly, it should be (at least partly) interpretable, not leading to explosive behaviour, and easily usable for forecasting. Still, there are plethora of models possessing these properties ([9]).

At present, the greatest attention is given to regime-switching models and artificial neural network models. The later originates in neurological sciences, where external pulses are filtered through hidden layers such that the initial signals can be properly analysed by brain cells. Artificial neural network (ANN) models achieve increasing popularity for its ability to approximate almost any nonlinear function arbitrarily close. An often-quoted drawback of ANN models is their difficult, if not impossible, interpretability. For this reason they are often considered as "black box" models and constructed mainly for the purpose of pattern recognition and forecasting, although the superior in-sample fit is no guarantee for good out-of-sample performance. For further description see [17], ANNs will not be dealt with in the following. Instead, we focus on the another favourite class of nonlinear models - the regime-switching models- that reflect piecewise linear structures in time series where the transition between different states is indicated by dynamics in certain variable(s). Such a behaviour of a model is easily interpretable and can be found in nature, for example, a change in atmospheric temperature causes sudden change in river flow rate due to the snow melting in mountains.

The chapter is organised as follows. Section 2.1 describes several ways in which the models are designed to give a true picture of switching mechanism between regimes that is determined by some observable variables, particularly TAR and STAR models are foregrounded. To make any application meaningful, it is essential to test the data for linearity against this particular nonlinearity, which is the topic of section 2.2. Next, a theory for model specification – shape identification, parameters estimation and suitability evaluation – is given in section 2.3. Last section brings up a new approach to the regime-switching models, which allows the breaks in linear structure to be indicated by several observables and an aggregation operator. The underling theory is investigated in the framework of multivariate threshold autoregression.

2.1 Description of regime-switching models

The idea behind the regime-switching models quite naturally defines different states of an environment system or regimes, and allows the dynamic behaviour of (observed) variables for possibility to depend on the regime that occurs at any given point in time ([17]). This "state-dependent" (or *regime-switching*) dynamic behaviour of time series means that certain of its properties, such as mean, variance and/or autocorrelations, are different in different regimes. For the example of such a state-dependent behaviour, recall (1.45) or (1.47) in section 1.4 where mean and variance vary through seasons. Hence, every season constitutes a different regime, and, in this interpretation, the regime switch process is deterministic (since we know with certainty in advance, when the regimes occur). In contrast, the focus in this chapter is put on the case when regime switch is stochastic.

Threshold and smooth transition AR

The most prominent member of the regime-switching class of models is the *Threshold Autoregressive* (TAR) model, which assumes that the regime that occurs at time t can be determined by relation between an observable variable z_t and some *threshold value* r. For completeness, there is also a subclass covering the case of determination by an unobservable process, representative of which is the Markov-Switching model. Anyway, a 2-regime TAR model assuming an AR(p) in both regimes can be written as

$$y_t = \begin{cases} \phi_{0,1} + \phi_{1,1} y_{t-1} + \cdots + \phi_{p,1} y_{t-p} + \varepsilon_t & \text{if } z_t \leq r, \\ \phi_{0,2} + \phi_{1,2} y_{t-1} + \cdots + \phi_{p,2} y_{t-p} + \varepsilon_t & \text{if } z_t > r, \end{cases} \tag{2.1}$$

or alternatively in matrix notation

$$y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t (1 - \mathrm{I}[z_t > r]) + \boldsymbol{\phi}_2' \boldsymbol{X}_t \, \mathrm{I}[z_t > r] + \varepsilon_t, \tag{2.2}$$

where $\boldsymbol{\phi}_j = (\phi_{0,j}, \phi_{1,j}, \ldots, \phi_{p,j})'$ are unknown parameters of j-th regime, $\boldsymbol{X}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, $\mathrm{I}[A]$ is an indicator function with $\mathrm{I}[A] = 1$ if the event A occurs and $\mathrm{I}[A] = 0$ otherwise. M' denotes transposition of M. A special case arises when the *threshold variable* z_t is taken to be a lagged value of the time series itself, that is $z_t = y_{t-d}$ for a certain integer $d > 0$. Then the resulting model is called a Self-Exciting TAR (SETAR).

A more gradual transition between the different regimes can be obtained by replacing the indicator function $\mathrm{I}[z_t > r]$ in (2.2) by a continuous function $G(z_t, \gamma, r)$ which changes smoothly from 0 to 1 as z_t increases. The resultant model

$$y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t (1 - G(z_t, \gamma, r)) + \boldsymbol{\phi}_2' \boldsymbol{X}_t G(z_t, \gamma, r) + \varepsilon_t, \tag{2.3}$$

is called a *Smooth Transition Autoregressive* (STAR) and if rearranged a little it is given by

$$y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t + (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)' \boldsymbol{X}_t G(z_t, \gamma, r) + \varepsilon_t, \tag{2.4}$$

which is easily extendable to m-regimes version

$$\begin{aligned} y_t = \boldsymbol{\phi}_1' \boldsymbol{X}_t &+ (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)' \boldsymbol{X}_t G(z_t, \gamma_1, r_1) + \ldots \\ &+ (\boldsymbol{\phi}_m - \boldsymbol{\phi}_{m-1})' \boldsymbol{X}_t G(z_t, \gamma_{m-1}, r_{m-1}) + \varepsilon_t. \end{aligned} \tag{2.5}$$

This is, however, not the only possible generalisation to multiple regime case, other ways are shown in [44], where one can find quite exhaustive and nice illustrative survey of various extensions of STAR models.

Sometimes it is of practical purpose to consider the models (2.2) or (2.3) to allow for exogenous variables $x_{1t}, x_{2t}, \ldots x_{lt}$ as additional regressors, either in their current or time-lagged values. In smooth transition case, the resultant model is simply called smooth transition regression (STR) and is discussed at length in [24]. Some authors denote such extension by adding X (for "eXogenous") to the acronym of a model, e.g., STARX. Nevertheless, henceforth we always point out inclusion of exogenous variables explicitly by word or by expressing the regression matrix $\boldsymbol{X}_t$, if relevant.

Smooth transition functions

Different choices for the *transition function* $G(z_t, \gamma, r)$ give rise to different types of regime-switching behaviour. A popular choice for $G(z_t, \gamma, r)$ is the first-order logistic function

$$G(z_t, \gamma, r) = \frac{1}{1 + e^{-\gamma(z_t - r)}}, \qquad \gamma > 0, \tag{2.6}$$

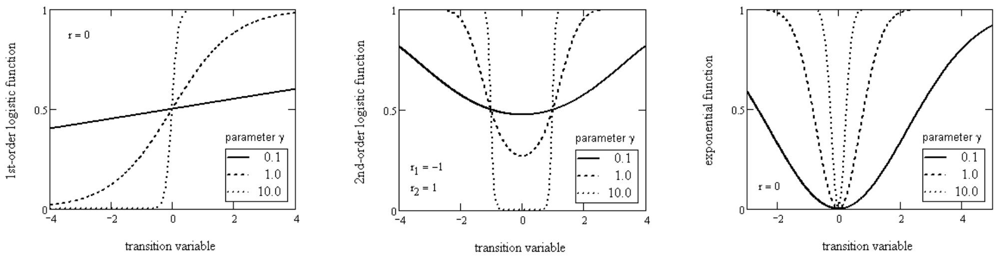

Figure 2.1: Transition functions

which results into the so-called *Logistic STAR* (LSTAR) model. The parameter r can be interpreted as the threshold between the two regimes, in the sense that the logistic function changes monotonically from 0 to 1 as z_t increases, and $G(r, \gamma, r) = 0.5$. The parameter γ determines the smoothness of the transition from one regime to another. Notice that the AR model is a special case of the LSTAR model in case $\gamma = 0$ and likewise the LSTAR becomes TAR as $\gamma \to \infty$. Alternatively, if situation requires to specify the transition function such that the regimes are associated with small and large absolute values of z_t, an even function can be used, e.g. exponential function

$$G(z_t, \gamma, r) = 1 - e^{-\gamma(z_t - r)^2}, \qquad \gamma > 0, \tag{2.7}$$

which has the property that $G(z_t, \gamma, r) \to 1$ both as $z_t \to -\infty$ and $z_t \to \infty$ whereas $G(z_t, \gamma, r) = 0$ for $z_t = r$, so that corresponding model (*Exponential STAR* or ESTAR) assumes symmetric response of y_t to positive or negative values of $z_t - r$. A drawback of the exponential function is that for either $\gamma \to 0$ or $\gamma \to \infty$, the function collapses to a constant (equal to 0 and 1, respectively). Hence the model becomes linear in both cases and the ESTAR does not nest a TAR model as a special case. If this is thought to be desirable, one can instead use the second-order logistic function

$$G(z_t, \gamma, \boldsymbol{r}) = \frac{1}{1 + e^{-\gamma(z_t - r_1)(z_t - r_2)}}, \qquad r_1 < r_2,\ \gamma > 0, \tag{2.8}$$

where now $\boldsymbol{r} = (r_1, r_2)'$. As seen from Figure 2.1, behaviour of this later function causes the STAR to nest a restricted three-regime TAR model, where the restriction is that the outer regimes are identical.

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. Either way, the decision obviously should subordinate to a rationale behind particular application.

Transition 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, or to use an explanatory variable x_t instead, 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 [24] for inspiration) or most recent idea of utilizing aggregation operators to construct the threshold variable (for recent developments on the theory see [41]). In section 2.4 we provide closer look to the aggregation operators being a part of general Threshold VAR model.

Multivariate (S)TAR

Linear Vector AR models, see section 1.6, constitute the most common way of modelling vector time series. In some situations, it could be worthwhile to consider nonlinear models for this purpose. Conceptually it is straightforward to extend the existing univariate regime-switching models to a multivariate context. However, it must be accompanied by a relevant statistical theory. In the following sections, we mainly draw from developments of [45] on this account.

A k-dimensional analogue of the univariate 2-regime STAR model with VAR(p) in both regimes for $(k \times 1)$ vector time series $\boldsymbol{y}_t = (y_{1t}, \dots y_{kt})'$ can be specified as

$$\begin{aligned}\boldsymbol{y}_t = (&\boldsymbol{\Phi}_{1,0} + \boldsymbol{\Phi}_{1,1}\boldsymbol{y}_{t-1} + \dots + \boldsymbol{\Phi}_{1,p}\boldsymbol{y}_{t-p})(1 - G(z_t, \gamma, r)) \\ &+ (\boldsymbol{\Phi}_{2,0} + \boldsymbol{\Phi}_{2,1}\boldsymbol{y}_{t-1} + \dots + \boldsymbol{\Phi}_{2,p}\boldsymbol{y}_{t-p})G(z_t, \gamma, r) + \boldsymbol{\varepsilon}_t, \quad (2.9)\end{aligned}$$

or shortly

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1\boldsymbol{X}_t(1 - G(z_t, \gamma, r)) + \boldsymbol{\Phi}_2\boldsymbol{X}_t G(z_t, \gamma, r) + \boldsymbol{\varepsilon}_t, \qquad (2.10)$$

where $(k \times 1)$ vectors $\boldsymbol{\Phi}_{j,0}$ and $(k \times k)$ matrices $\boldsymbol{\Phi}_{j,i}$, $j = 1, 2$, $i = 1, \dots p$, are stacked to $k \times (1 + kp)$ parameter matrices $\boldsymbol{\Phi}_j = (\boldsymbol{\Phi}_{j,0}, \boldsymbol{\Phi}_{j,1}, \dots \boldsymbol{\Phi}_{j,p})$ corresponding to j-th regime,

$$\boldsymbol{X}_t = (1, \boldsymbol{y}'_{t-1}, \dots \boldsymbol{y}'_{t-p})', \qquad (2.11a)$$

is $(1 + kp)$-dimensional regressor, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots \varepsilon_{kt})'$ is a k-dimensional vector white noise process with mean zero and $(k \times k)$ positive definite covariance matrix $\boldsymbol{\Sigma}$. If an l-dimensional vector time series of exogenous variables,

$\boldsymbol{x}_t = (x_{1t}, \dots x_{lt})'$, is considered to enter the model (2.10), then the regressor will be of form

$$\boldsymbol{X}_t = (1, \boldsymbol{y}'_{t-1}, \dots \boldsymbol{y}'_{t-p}, \boldsymbol{x}'_t)' \tag{2.11b}$$

with dimension equal $(1 + kp + l)$ or alternatively, if lagged values up to q time points are more appropriate,

$$\boldsymbol{X}_t = (1, \boldsymbol{y}'_{t-1}, \dots \boldsymbol{y}'_{t-p}, \boldsymbol{x}'_{t-1}, \dots \boldsymbol{x}'_{t-q})' \tag{2.11c}$$

with $(1 + kp + lq)$-dimensionality. Obviously, the dimension of parameter matrix will change as well. Such models like (2.10) have among many authors adopted an acronym STVAR.

Substitution $G(z_t, \gamma, r) \to \mathrm{I}[z_t > r]$ in (2.10) yields Threshold VAR model

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_1 \boldsymbol{X}_t (1 - \mathrm{I}[z_t > r]) + \boldsymbol{\Phi}_2 \boldsymbol{X}_t \, \mathrm{I}[z_t > r] + \boldsymbol{\varepsilon}_t, \tag{2.12}$$

corresponding to (2.2).

Notice that the regimes are common to the k variables. It is straightforward to generalize the model to incorporate equation-specific transition functions $G_i(z_{it}, \gamma_i, r_i)$ (or $\mathrm{I}[z_{it} > r_i]$), $i = 1, \dots k$, and thereby to allow for equation-specific regime-switching.

In practice, it is many times the case that the parameter matrix of VAR model (in certain regimes) contains unit roots that may indicate common stochastic trend(s). Indeed, it seems that the model of currently most considerable interest among practitioners is the one in which the components of $\boldsymbol{y}_t$ are linked by a linear long-run equilibrium relationship, whereas adjustment towards this equilibrium is nonlinear and can be characterised as regime-switching, with the regimes determined by the size and/or the sign of deviation from equilibrium. In linear time series analysis, this behaviour is captured by cointegration and error-correction models which we will talk about in chapter 3.

2.2 Testing for non-linearity

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 nonlinearity. 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.

However, the testing problem is complicated by the presence of unidentified nuisance parameters under the null hypothesis ([44]). These are the parameters of a model which are not restricted by the null hypothesis, but about which nothing can be learned from the data when the null hypothesis holds true. For example, with the (L)STAR model as alternative, the null hypothesis $H_0: \ \boldsymbol{\phi}_1 = \boldsymbol{\phi}_2$ does not restrict the parameters γ and r in the transition function, but when the null hypothesis is valid, the likelihood is unaffected by the values of γ and r. Likewise, when reformulating the null hypothesis of linearity into $H_0': \ \gamma = 0$ (LSTAR model reduces to AR model with parameters $(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)/2$) then the location parameter r and the parameters $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ are the unidentified parameters ($\boldsymbol{\phi}_j$s can take any value as long as their average remains fixed). The main consequence is that the conventional statistic theory is not available for obtaining the asymptotic null distribution of the test-statistics. Instead, the test-statistics tend to have nonstandard distributions for which analytic expressions are most often not available and the critical values have to be determined by means of simulation. Fortunately, there still exist solutions that lead to applicability of standard asymptotic theory.

Tsay in [45] propose a test that put threshold non-linearity (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, firstly developed by Luukkonen, Saikkonen and Teräsvirta, utilizes Lagrange Multipliers (LM) statistic and is available for STAR models. Both tests are simple and perform well in finite samples, yet they do not depend on the alternative model, nor do they encounter the problem of unidentified nuisance parameters under the null hypothesis.

In the following, we give a brief survey of Tsay's test, which was a priori designed for multivariate time series and the switching between regimes is indicated by lagged values of some threshold variable. The more details we provide in section 2.4 where a more general case with z_t based on aggregation operators is treated. As for the LM-type test, it originated as univariate and later a simple extension to multivariate case was given ([46]). It is described in the second part of this section.

Linearity against threshold nonlinearity

Testing the null hypothesis H_0 of linearity versus the alternative hypothesis H_1 that $\boldsymbol{y}_t$ follows the multivariate TAR model (2.12) with (2.11c) and z_t equal to some lagged variable, ξ_{t-d}, assuming p, q and d are known, is the problem of detecting the change in data behaviour along the increasing z_t.

A linear regression framework $\boldsymbol{y}_t = \boldsymbol{\Phi}\boldsymbol{X}_t + \boldsymbol{\varepsilon}_t$, $t = h+1, \dots n$, where $h = \max(p, q, d)$ and the estimates are biased under H_1, can be rewritten to

$$\boldsymbol{y}_{t(i)+d} = \boldsymbol{\Phi}\boldsymbol{X}_{t(i)+d} + \boldsymbol{\varepsilon}_{t(i)+d}, \qquad i = 1, \dots n-h, \tag{2.13}$$

without any interference to dynamics of the data. The ordering of z_t has been rearranged, so that $z_{(i)}$ denotes the i-th smallest element of $\{z_{h+1-d}, \dots z_{n-d}\}$ and $t(i)$ the time index of $z_{(i)}$. To detect the model change in (2.13), Tsay used recursive least squares method to compute the standardized predictive residuals

$$\hat{\boldsymbol{\eta}}_{t(\tilde{n}+1)+d} = \frac{\boldsymbol{y}_{t(\tilde{n}+1)+d} - \hat{\boldsymbol{\Phi}}_{\tilde{n}}\boldsymbol{X}_{t(\tilde{n}+1)+d}}{\left[1 + \boldsymbol{X}'_{t(\tilde{n}+1)+d}\boldsymbol{V}_{\tilde{n}}\boldsymbol{X}_{t(\tilde{n}+1)+d}\right]^{1/2}}, \tag{2.14}$$

where $\hat{\boldsymbol{\Phi}}_{\tilde{n}}$ is a least squares estimate of $\boldsymbol{\Phi}$ in (2.13) using data points associated with $\tilde{n}$ smallest values of z_{t-d}, and $\boldsymbol{V}_{\tilde{n}} = \left[\sum_{i=1}^{\tilde{n}} \boldsymbol{X}_{t(i)+d}\boldsymbol{X}'_{t(i)+d}\right]^{-1}$. From the regression

$$\hat{\boldsymbol{\eta}}_{t(j)+d} = \boldsymbol{\Psi}\boldsymbol{X}_{t(j)+d} + \boldsymbol{w}_{t(j)+d}, \qquad j = \tilde{n}_0 + 1, \dots n-h, \tag{2.15}$$

where $\tilde{n}_0$ denotes starting point of the recursive least square estimation (usually $\tilde{n}_0 \approx 3\sqrt{n}$), he tested the hypothesis $H_0: \ \boldsymbol{\Psi} = \boldsymbol{0}$ versus $H_1: \ \boldsymbol{\Psi} \neq \boldsymbol{0}$. The test-statistic is defined as

$$C = (n - h - \tilde{n}_0 - K)\big(\ln|\boldsymbol{S}_0| - \ln|\boldsymbol{S}_1|\big), \tag{2.16}$$

where $K = (pk + ql + 1)$ is the length of regressor $\boldsymbol{X}_t$, $|M|$ determinant of M and

$$\boldsymbol{S}_0 = \frac{1}{n-h-\tilde{n}_0}\sum_{j=\tilde{n}_0+1}^{n-h} \hat{\boldsymbol{\eta}}_{t(j)+d}\hat{\boldsymbol{\eta}}'_{t(j)+d}, \ \boldsymbol{S}_1 = \frac{1}{n-h-\tilde{n}_0}\sum_{j=\tilde{n}_0+1}^{n-h} \hat{\boldsymbol{w}}_{t(j)+d}\hat{\boldsymbol{w}}'_{t(j)+d}.$$

Under the null hypothesis of linearity, C is asymptotically a χ^2 random variable with kK degrees of freedom.

Linearity against smooth transition nonlinearity

Before we give a description of general test for multivariate LSTAR model, it comes handy firstly to see background of the test in univariate case.

As already mentioned, besides equality of the AR parameters in 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 (2.6) is equal to 0.5

for all z_t and the LSTAR model collapse to an AR model with parameters $(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)/2$. Following [44], rewrite STAR model (2.4) as

$$y_t = \frac{1}{2}(\boldsymbol{\phi}_1 + \boldsymbol{\phi}_2)'\boldsymbol{X}_t + (\boldsymbol{\phi}_2 - \boldsymbol{\phi}_1)'\boldsymbol{X}_t G^*(z_t, \gamma, r) + \varepsilon_t, \tag{2.17}$$

where $G^*(z_t, \gamma, r) = G(z_t, \gamma, r) - 1/2$ and regressor $\boldsymbol{X}_t$ may contain q lagged values of exogenous variable x_t such that $\boldsymbol{X}_t = (1, y_{t-1}, \ldots y_{t-p}, x_{t-1}, \ldots x_{t-q})'$, for instance. In order to derive a linearity test against (2.17), we approximate the shape function $G^*(z_t, \gamma, r)$ with a third-order Taylor approximation[1] around $\gamma = 0$, that is

$$\begin{aligned} T_3(z_t, \gamma, r) &\approx G^*(z_t, 0, r) + \sum_{i=1}^{3} \frac{1}{i!}\gamma \left(\frac{\partial^i G^*(z_t, \gamma, r)}{\partial \gamma^i} \Big|_{\gamma=0} \right) \\ &= \frac{1}{4}\gamma(z_t - r) + \frac{1}{48}\gamma^3(z_t - r)^3, \end{aligned} \tag{2.18}$$

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 (2.17) and rearranging terms this yields the auxiliary regression

$$y_t = \beta_{0,0} + \boldsymbol{\beta}_0'\boldsymbol{X}_t + \boldsymbol{\beta}_1'\boldsymbol{X}_t z_t + \boldsymbol{\beta}_2'\boldsymbol{X}_t z_t^2 + \boldsymbol{\beta}_3'\boldsymbol{X}_t z_t^3 + e_t, \tag{2.19}$$

where $\boldsymbol{\beta}_i = (\beta_{i,0}, \beta_{i,1}, \ldots \beta_{i,p+q})'$, $i = 0, 1, 2, 3$, are functions of the parameters $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, γ and r. Inspection of the exact relationships show that the null hypothesis $H_0' :\ \gamma = 0$ corresponds to $H_0'' :\ \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{0}$ (and $e_t = \varepsilon_t$), 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 $\beta_{i,0} z_t^i$, $i = 1, 2, 3$, should be dropped from the auxiliary regression to avoid perfect multi-collinearity. This drop off can be simply achieved by omitting the first element of vectors $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\beta}_3$ and $\boldsymbol{X}_t$.

Now consider multivariate system (2.12) with (2.11c). The LM_3 test-statistic based on (2.19) for this multiple equation system 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 $\boldsymbol{\Sigma}_0 = (n-h)^{-1} \sum_{t=h+1}^{n} \hat{\boldsymbol{\varepsilon}}_t \hat{\boldsymbol{\varepsilon}}_t'$.

2. Estimate the auxiliary regression of $\hat{\boldsymbol{\varepsilon}}_t$ on $\boldsymbol{X}_t$ and $\boldsymbol{X}_t z_t^j$, $j = 1, 2, 3$, then $\boldsymbol{\Sigma}_1 = (n-h)^{-1} \sum_{t=h+1}^{n} \hat{\boldsymbol{e}}_t \hat{\boldsymbol{e}}_t'$.

[1]First-order approximation is not sufficient enough in situation, when z_t is identical with one of the variables included in the regressor $\boldsymbol{X}_t$ and the intercept $\phi_{i,0}$ is the only parameter that differs across regimes.

3. The χ^2 version of the LM_3 statistic for testing linearity *equation by equation* can now be computed as

$$LM_{3\,i} = \frac{(n-h)(\Sigma_{0\,i,i} - \Sigma_{1\,i,i})}{\Sigma_{0\,i,i}}, \qquad i = 1, \dots k, \tag{2.20}$$

where $M_{i,i}$ denotes i-th diagonal element of a matrix $\boldsymbol{M}$.

If we again denote the dimension of the regressor $\boldsymbol{X}_t$ as K, in this general case it holds that $K = (pk + ql + 1)$, then under the null hypothesis of linearity, the test-statistic LM_3 has an asymptotic χ^2 distribution with $3K$ degrees of freedom. Note again that n is length of the k-dimensional time series $\boldsymbol{y}_t$ and $h = \max(p, q, d)$.

In small samples it is recommended to use F-version of the LM_3, as it has better size and power properties. Under the null hypothesis, the F-version which can be computed as

$$LM_{3\,i} = \frac{(\Sigma_{0\,i,i} - \Sigma_{1\,i,i})/3K}{\Sigma_{1\;i,i}/(n-h-4K)}, \qquad i = 1, \dots k, \tag{2.21}$$

is approximately F distributed with $3K$ and $(n-h-4K)$ degrees of freedom.

Following [46], the appropriate test for linearity in the *system as a whole* is a log-likelihood test of the null hypothesis $H_0 : \gamma = 0$ in all of the equations. Then the test-statistic

$$LM_3 = (n-h)(\ln|\boldsymbol{\Sigma}_0| - \ln|\boldsymbol{\Sigma}_1|) \tag{2.22}$$

is asymptotically $\chi^2(3kK)$ distributed (compare to (2.16)).

Testing linearity against ESTAR alternative is very similar and uses the auxiliary regression (2.19) with one extra term, $\boldsymbol{\beta}_4 \boldsymbol{X}_t z_t^4$. Then under $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \boldsymbol{\beta}_3 = \boldsymbol{\beta}_4 = \boldsymbol{0}$ the resultant LM-type test-statistic $LM_4 \overset{a}{\sim} \chi^2(4kK)$

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

2.3 Model specification strategy

When building nonlinear time series models, it is strongly recommended to use a "specific-to-general" strategy, which implies starting with a simple or restricted model and proceeding to more complicated ones only if diagnostic tests indicate inadequacy of the maintained model. An empirical specification procedure for nonlinear model basically follows these steps: (i) specify an appropriate linear AR (possibly with included exogenous variables) model of order p, (ii) test the null hypothesis of linearity against the alternative of (STAR or TAR) regime-switching nonlinearity; if linearity is rejected, select the appropriate transition (threshold) variable z_t and the form of the transition function, (iii) estimate the parameters in the selected (S)TAR model, (iv) evaluate the model using diagnostic tests, (v) modify the model if necessary and (vi) use the model for descriptive or forecasting purposes.

In the following, we will go through these steps, more details can be found in [17], [44], [45] and many others. We will try to keep focus on generality to cover the largest area of applicability, yet the most useful and/or simple procedures will be described to provide easiness of practical implementation. Consult also the concepts introduced in chapter 1. For convenience, the TAR model will mostly be treated as special case of LSTAR (when γ is large).

Preliminary specification

When selecting orders of linear model (ideally by AIC and 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 by minimizing the p-value or directly from estimation of particular models by 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 a modification of LM-test to test a 2-regime STAR model against the alternative of an additive 3-regime model. For selecting the transition function G there exist several methods based on LM-type tests.

Selection criteria

An important question concerns detecting the appropriate orders p_1, p_2 and q_1, q_2 in the general 2-regime model (2.10), where notation (2.11c) needs to be respecified, such as

$$\boldsymbol{X}_{j,t} = (1, \boldsymbol{y}'_{t-1}, \ldots \boldsymbol{y}'_{t-p_j}, \boldsymbol{x}'_{t-1}, \ldots \boldsymbol{x}'_{t-q_j})', \qquad j = 1, 2, \tag{2.23}$$

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, n_j.

To our present knowledge, no general formula has been derived and published for the information criteria covering multivariate smooth transition model with regime-varying orders of VAR, yet. Actually, it is an easy exercise and we may start from 2-regime univariate TAR model of order p_1 and p_2, for which $AIC(p_1, p_2) = \sum_{j=1}^{2} \left(n_j \ln \hat{\sigma}_j^2 + 2(p_j + 1)\right)$ given in [17], where the estimated variance $\hat{\sigma}_j^2 = n_j^{-1} \sum_t^{(j)} (y_t - \hat{y}_t)^2$, where $\sum_t^{(j)}$ denotes summing over observations in j-th regime. If we denote the indicator function for regime j as $\mathrm{I}[r_{j-1} < z_t \leq r_j] = \mathrm{I}[z_t > r_{j-1}] - \mathrm{I}[z_t > r_j] = I_{j-1,t} - I_{j,t}$, where $r_0 = -\infty$ and $r_m = \infty$, it is obvious that the n_j will be the number of times when $I_{j-1,t} - I_{j,t}$ becomes 1, and the summation can be rewritten in more explicit form, so that $n_j = \sum_{t=h+1}^{n} (I_{j-1,t} - I_{j,t})$ and $\hat{\sigma}_j^2 = n_j^{-1} \sum_{t=h+1}^{n} (I_{j-1,t} - I_{j,t}) \hat{\varepsilon}_t^2$, where $h = \max(p_1, \ldots p_m, d)$. Now it is quite straightforward to do generalization replacing abrupt transition function for the smooth one and considering multivariate version of commonly used information criteria (see [45]). Then AIC and BIC criteria that suit the m-regimes version of the model (2.10) with (2.23) can be defined as

$$AIC(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{m} \left(n_j \ln |\hat{\boldsymbol{\Sigma}}_j| + 2k(kp_j + lq_j + 1)\right), \tag{2.24}$$

$$BIC(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{m} \left(n_j \ln |\hat{\boldsymbol{\Sigma}}_j| + \ln(n_j) k(kp_j + lq_j + 1)\right), \tag{2.25}$$

where $\boldsymbol{p} = (p_1, \ldots p_m)$, $\boldsymbol{q} = (q_1, \ldots q_m)$, $|\boldsymbol{M}|$ denotes determinant of $\boldsymbol{M}$, and

$$\hat{\boldsymbol{\Sigma}}_j = \frac{1}{n_j} \sum_{t=h+1}^{n} (\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j \boldsymbol{X}_{j,t})(\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j \boldsymbol{X}_{j,t})' \Delta G_{j,t} \tag{2.26}$$

is estimated covariance matrix. Regime-specific number of observations is not necessarily an integer, actually, it is a weight $n_j = \sum_{t=h+1}^{n} \Delta G_{j,t}$ with $\Delta G_{j,t} = G_{j-1,t} - G_{j,t}$, 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 $G_{m,t} = 0$. Recall that $h = \max(\boldsymbol{p}, \boldsymbol{q}, d)$ and d is a time lag associated with transition variable.

Estimation

Estimation of the parameters $\boldsymbol{\theta} = (\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \gamma, r)'$ in the STVAR model (2.10) with regime-specific regressors (2.23) [2], when it comes handy to assign $\boldsymbol{\Phi} =$

[2] Generalisation to m-regimes is straightforward: $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots \boldsymbol{\Phi}_m)$ and $\boldsymbol{X}_t(\boldsymbol{\gamma}, \boldsymbol{r}) = \left(\boldsymbol{X}'_{1,t} \Delta G_{1,t}, \ldots \boldsymbol{X}'_{m,t} \Delta G_{m,t}\right)'$ etc., with notation as by (2.26).

$(\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2)$ and

$$\boldsymbol{X}_t(\gamma, r) = \left(\boldsymbol{X}'_{1,t}[1 - G(z_t, \gamma, r)], \boldsymbol{X}'_{2,t}G(z_t, \gamma, r)\right)',$$

is the problem of minimizing the trace of

$$\boldsymbol{\Sigma}(\boldsymbol{\Phi}, \gamma, r) = \sum_{t=h+1}^{n} \left(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{X}_t(\gamma, r)\right)\left(\boldsymbol{y}_t - \boldsymbol{\Phi}\boldsymbol{X}_t(\gamma, r)\right)'. \tag{2.27}$$

This can be performed directly by nonlinear least squares (NLS) routine

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \operatorname{Tr}\left(\boldsymbol{\Sigma}(\boldsymbol{\Phi}, \gamma, r)\right), \tag{2.28}$$

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) through

$$\hat{\boldsymbol{\Phi}}(\gamma, r)' = \left(\sum_{t=h+1}^{n} \boldsymbol{X}_t(\gamma, r)\boldsymbol{X}_t(\gamma, r)'\right)^{-1}\left(\sum_{t=h+1}^{n} \boldsymbol{X}_t(\gamma, r)\boldsymbol{y}'_t\right) \tag{2.29}$$

and

$$(\hat{\gamma}, \hat{r}) = \underset{(\gamma, r)}{\operatorname{argmin}} \operatorname{Tr}\left(\boldsymbol{\Sigma}(\hat{\boldsymbol{\Phi}}(\gamma, r), \gamma, r)\right). \tag{2.30}$$

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 appear 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 [24], 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.

Evaluation

After a ST(V)AR or T(V)AR 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 specific tests for autocorrelations, normality, parameter constancy and linearity tests as described in [44],[17],[24] or [3] 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.

Point forecasts in univariate case

Consider y_t being described by the general (single equation) nonlinear autoregressive model

$$y_t = F(y_{t-1}; \boldsymbol{\theta}) + \varepsilon_t \tag{2.31}$$

for some nonlinear function $F(y_t; \boldsymbol{\theta})$. The optimal h-step-ahead forecast of y_{t+h} at time t is given by

$$\hat{y}_{t+h|t} = \mathrm{E}[y_{t+h}|\Omega_t], \tag{2.32}$$

where Ω_t again denotes the history of the time series up to and including the observation at time t. Using (2.31) and the fact that $\mathrm{E}[\varepsilon_{t+1}|\Omega_t] = 0$, the optimal 1-step-ahead forecast is $\hat{y}_{t+1|t} = \mathrm{E}[y_{t+1}|\Omega_t] = F(y_{t-1}; \boldsymbol{\theta})$. When the forecast horizon is longer than 1 period, things become more complicated, because in general, the linear conditional expectation operator E cannot be interchanged with the nonlinear operator F, that is $\mathrm{E}[F(\cdot)] \neq F(\mathrm{E}[\cdot])$.

Several methods have been developed to obtain adequate multi-step-ahead forecast. One might attempt to obtain the conditional expectation (2.32) directly by computing

$$\hat{y}_{t+h|t} = \int_{-\infty}^{-\infty} F\left(\hat{y}_{t+h-1|t} + \varepsilon; \boldsymbol{\theta}\right) f(\varepsilon) d\varepsilon, \tag{2.33}$$

where f denotes the density of ε_t. An alternative approach is to approximate the conditional expectation using Monte Carlo or bootstrap methods. The h-step-ahead Monte Carlo forecast is given by

$$\hat{y}^{(mc)}_{t+h|t} = \frac{1}{k} \sum_{i=1}^{k} F\left(\hat{y}_{t+h-1|t} + \varepsilon_i; \boldsymbol{\theta}\right), \tag{2.34}$$

where k is some large number and the ε_i are drawn from the presumed distribution of ε_{t+h-1}. The bootstrap forecast is very similar, the only difference being that the residuals from the estimated model, $\hat{\varepsilon}_t$, $t = 1, \ldots n$, are used,

$$\hat{y}^{(b)}_{t+h|t} = \frac{1}{k} \sum_{i=1}^{k} F\left(\hat{y}_{t+h-1|t} + \hat{\varepsilon}_i; \boldsymbol{\theta}\right). \tag{2.35}$$

2.4 Switching by aggregation operators in threshold VAR

In this section we give detailed theoretical survey on testing and modelling multivariate time series with threshold vector autoregressive model, where threshold variable is constructed using aggregation operators. Firstly, a threshold variable is defined on the basis of aggregation operators and a brief introduction to such operators is given. Then we provide a modification of Tsay's test for linearity against threshold nonlinearity to utilize newly defined threshold variable. For this purpose we also give a modified conditional least square estimation procedure of [45] with corresponding asymptotic properties.

Aggregation operators

In general, the threshold variable z_t can be defined as

$$z_t = \mathcal{A}(a_1, \ldots, a_d) \tag{2.36}$$

where $\mathcal{A}$ is a continuous aggregation operator (agop) and a_i, $i = 1, \ldots d$, is i-th constituting variable that may contribute to the switch between regimes. Usually, for a_i we use lagged values of certain variable, say ξ_{t-i}, which is either endogenous or exogenous variable, so that

$$z_t = \mathcal{A}(\xi_{t-1}, \ldots, \xi_{t-d}). \tag{2.37}$$

Either way, again, it should somehow reflect practical experience.

Typical continuous agops on the real line ($R^d \to R$) are

- arithmetic mean $\mathcal{M}(a_1, \ldots a_d) = \frac{1}{d}\sum_{i=1}^{d} a_i$,

- weighted mean $\mathcal{W}(a_1, \ldots a_d) = \sum_{i=1}^{d} w_i a_i$, where $w_i \in [0,1]$ and $\sum_{i=1}^{d} w_i = 1$,

- OWA operators $\mathcal{OWA}(a_1, \ldots a_d) = \sum_{i=1}^{d} w_i a_i'$ with a_i' as non-decreasing permutation of a_i inputs, i.e. $a_1' \leq \cdots \leq a_d'$,

and the w_i denotes a weight assigned to i-th input. In the class of $\mathcal{OWA}$ we can find also $\mathcal{MIN}$ ($w_1 = 1$ and $w_i = 0$ otherwise) eventually $\mathcal{MAX}$ ($w_d = 1$, $w_i = 0$) operators and all order statistics. Similarly a projection to j-th coordinate, with $w_j = 1$ and $w_i = 0$ otherwise, is a special weighted mean. A convenient way of producing a decreasing sequence $(w_1, \ldots w_d)$ of weight coefficient is based on utilisation of increments of a generating increasing convex bijection φ of $[0,1]$, if we put $w_i = \varphi\left(\frac{d-i+1}{d}\right) - \varphi\left(\frac{d-i}{d}\right)$ for $i = 1, \ldots d$.

Testing for threshold nonlinearity

Given k-dimensional vector $\boldsymbol{y}_t$ of endogenous and l-dimensional vector $\boldsymbol{x}_t$ of exogenous variables, a general m-regimes multivariate threshold model can be defined combining (2.5) and (2.12) as

$$\boldsymbol{y}_t = \sum_{j=1}^{m} \boldsymbol{\Phi}_j \boldsymbol{X}_t \, \mathrm{I}[r_{j-1} < z_t \leq r_j] + \boldsymbol{\varepsilon}_t, \tag{2.38}$$

with $-\infty = r_0 < r_1 < r_2 < \cdots < r_m = \infty$ and with $(pk + ql + 1)$-dimensional regressor $\boldsymbol{X}_t$ given by (2.11c). The error time series $\boldsymbol{\varepsilon}_t = \boldsymbol{\Sigma}_j^{1/2} \boldsymbol{\epsilon}_t$, where j denotes the regime occurred at time t, $\boldsymbol{\Sigma}_j^{1/2}$ is symmetric positive definite matrix and $\{\boldsymbol{\epsilon}_t\}$ is a sequence of serially uncorrelated random vectors with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{I}$, the identity matrix. The threshold variable is assumed to be stationary and have a continuous distribution.

Now consider the null hypothesis (H_0) that time series $\boldsymbol{y}_t$ is linear versus the alternative hypothesis (H_1) that it follows (2.38). The goal is to detect the threshold nonlinearity assuming that p and q are known as well as the threshold variable z_t is defined by (2.37) with known lag d. Let us define a (linear) regression framework

$$\boldsymbol{y}_t = \boldsymbol{\Phi} \boldsymbol{X}_t + \boldsymbol{\varepsilon}_t, \qquad t = h+1, \ldots, n \tag{2.39}$$

where $h = \max(p, q, d)$ and $\boldsymbol{\Phi}$ denotes a $k \times (pk + ql + 1)$ parameter matrix. If the H_0 holds, then the least squares estimates of (2.39) is useful, otherwise the estimates are biased under H_1.

Now, let the ordering of the threshold variable z_t be rearranged increasingly, so that $z_{(i)}$ denotes the i-th smallest value of z_t for $i = 1, \ldots n-h$. Furthermore, let $t(i)$ be the time index of $z_{(i)}$, i.e., $z_{t(i)} = z_{(i)}$. When we rewrite (2.39) in the form

$$\boldsymbol{y}_{t(i)} = \boldsymbol{\Phi} \boldsymbol{X}_{t(i)} + \boldsymbol{\varepsilon}_{t(i)} \qquad i = 1, \ldots n-h \tag{2.40}$$

the dynamic of $\boldsymbol{y}_t$ will not change, $\boldsymbol{X}_t$ remains the independent variable of $\boldsymbol{y}_t$ for all t. What will change is the ordering by which data enter the regression setup. Thus the arranged regression effectively transforms a threshold model into a *changepoint* problem.

One way to detect the changepoint is to use predictive residuals $\hat{\boldsymbol{\varepsilon}}_{t(i+1)}$. If $\boldsymbol{y}_t$ is linear, the recursive least squares estimator of the arranged regression (2.40) is consistent, so that the predictive residuals approach white noise and are uncorrelated with the regressor $\boldsymbol{X}_{t(i+1)}$.

Let $\hat{\mathbf{\Phi}}_{\tilde{n}}$ be a least squares estimate of $\mathbf{\Phi}$ in (2.40) using data points associated with $\tilde{n}$ smallest values of z_t. Let

$$\hat{\boldsymbol{e}}_{t(\tilde{n}+1)} = \boldsymbol{y}_{t(\tilde{n}+1)} - \hat{\mathbf{\Phi}}_{\tilde{n}} \boldsymbol{X}_{t(\tilde{n}+1)} \tag{2.41}$$

be the residual of the 1-step-ahead prediction in the arranged regression, and

$$\hat{\boldsymbol{\eta}}_{t(\tilde{n}+1)} = \frac{\hat{\boldsymbol{e}}_{t(\tilde{n}+1)}}{\left[1 + \boldsymbol{X}'_{t(\tilde{n}+1)} \boldsymbol{V}_{\tilde{n}} \boldsymbol{X}_{t(\tilde{n}+1)}\right]^{1/2}} \tag{2.42}$$

be its standardized version, where

$$\boldsymbol{V}_{\tilde{n}} = \left[\sum_{i=1}^{\tilde{n}} \boldsymbol{X}_{t(i)} \boldsymbol{X}'_{t(i)}\right]^{-1}.$$

Under the H_0, $\hat{\boldsymbol{\eta}}_{(j)}$ and $\boldsymbol{X}_{t(j)}$ should be uncorrelated for all $j = 1, \dots n-h$. We test this hypothesis using the regression

$$\hat{\boldsymbol{\eta}}_{t(j)} = \mathbf{\Psi} \boldsymbol{X}_{t(j)} + \boldsymbol{w}_{t(j)}, \qquad j = \tilde{n}_0 + 1, \dots n-h, \tag{2.43}$$

where $\tilde{n}_0$ is the starting point of recursive regression ($\tilde{n}_0 \approx 3\sqrt{n}$). The problem of interest is then to test the hypothesis $H_0 : \ \mathbf{\Psi} = \mathbf{0}$ against $H_1 : \ \mathbf{\Psi} \neq \mathbf{0}$ with the test-statistic

$$C = [n - h - \tilde{n}_0 - (pk + ql + 1)] \times (\ln|\boldsymbol{S}_0| - \ln|\boldsymbol{S}_1|), \tag{2.44}$$

where

$$\boldsymbol{S}_0 = \frac{1}{n-h-\tilde{n}_0} \sum_{j=\tilde{n}_0+1}^{n-h} \hat{\boldsymbol{\eta}}_{t(j)} \hat{\boldsymbol{\eta}}'_{t(j)}, \qquad \boldsymbol{S}_1 = \frac{1}{n-h-\tilde{n}_0} \sum_{j=\tilde{n}_0+1}^{n-h} \hat{\boldsymbol{w}}_{t(j)} \hat{\boldsymbol{w}}'_{t(j)}.$$

Under the null that $\boldsymbol{y}_t$ is linear, C is asymptotically a χ^2 random variable with $k(pk + ql + 1)$ degrees of freedom.

Remark 1. If $\boldsymbol{\varepsilon}_t$ has conditional heteroscedasticity, then (2.42) no longer holds. The remedy is in modifying the standardization of predictive residuals so that the j-th element

$$\hat{\eta}_{j,t(\tilde{n}+1)} = \hat{e}_{j,t(\tilde{n}+1)} / \left[\hat{\sigma}_j^2 + \boldsymbol{X}'_{t(\tilde{n}+1)} \boldsymbol{V}^*_{\tilde{n}} \boldsymbol{X}_{t(\tilde{n}+1)}\right]^{1/2},$$

where $\hat{\sigma}_j^2 = \sum_{i=1}^{\tilde{n}} \hat{e}^2_{j,t(i)} / (\tilde{n} - pk - ql - 1)$ is the residual mean squared error of the j-th element of $\boldsymbol{y}_t$ and

$$\boldsymbol{V}^*_{\tilde{n}} = \boldsymbol{V}_{\tilde{n}} \left(\sum_{i=1}^{\tilde{n}} \hat{e}^2_{j,t(i)} \boldsymbol{X}_{t(i)} \boldsymbol{X}'_{t(i)}\right) \boldsymbol{V}_{\tilde{n}}.$$

Estimation

Assuming that p, q and m are known and the threshold variable z_t is given like in (2.37), then the parameters of the TAR model (2.38) are delay d, thresholds r_j, linear model parameter matrices $\boldsymbol{\phi}_j$ and covariance matices $\boldsymbol{\Sigma}_j$, $j = 1, \dots m$. Their conditional least squares estimates can be obtained in two steps. First, for given d and $\boldsymbol{r} = (r_1, \dots r_{m-1})$ the model reduces to two separated multivariate linear regressions and the estimates are

$$\hat{\boldsymbol{\Phi}}_j(d, \boldsymbol{r}) = \left(\sum_t^{(j)} \boldsymbol{X}_t \boldsymbol{X}_t' \right)^{-1} \left(\sum_t^{(j)} \boldsymbol{X}_t \boldsymbol{y}_t' \right) \tag{2.45}$$

$$\hat{\boldsymbol{\Sigma}}_j(d, \boldsymbol{r}) = \frac{1}{(n_j - K)} \sum_t^{(j)} \left(\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j(d, \boldsymbol{r}) \boldsymbol{X}_t \right) \left(\boldsymbol{y}_t - \hat{\boldsymbol{\Phi}}_j(d, \boldsymbol{r}) \boldsymbol{X}_t \right)', \tag{2.46}$$

where $\sum_t^{(j)}$ denotes summing over observations in j-th regime, n_j is the number of data points in regime j, and K is the dimension of $\boldsymbol{X}_t$ satisfying $K < n_j$, $j = 1, \dots m$. In the second step, the conditional least squares estimates of d and $\boldsymbol{r}$ are obtained by

$$(\hat{d}, \hat{\boldsymbol{r}}) = \underset{d, \boldsymbol{r}}{\operatorname{argmin}} \operatorname{Tr} \left(\sum_{j=1}^{m} (n_j - K) \hat{\boldsymbol{\Sigma}}_j(d, \boldsymbol{r}) \right), \tag{2.47}$$

where $d \in \{1, 2 \dots d_{max}\}$, thresholds in r lies on a bounded subsets of the real line, say $R_0 \subset R$, and $\operatorname{Tr}(M)$ denotes trace of matrix M. The resulting least squares estimates are

$$\hat{\boldsymbol{\Phi}}_j = \hat{\boldsymbol{\Phi}}_j(\hat{d}, \hat{\boldsymbol{r}}) \qquad \textit{and} \qquad \hat{\boldsymbol{\Sigma}}_j = \hat{\boldsymbol{\Sigma}}_j(\hat{d}, \hat{\boldsymbol{r}}).$$

Now, let us define

$$\begin{aligned} \boldsymbol{D}(r) &= \operatorname{E}[\boldsymbol{X}_t \boldsymbol{X}_t' | z_t = r], \\ D_2(r) &= \operatorname{E}[(\boldsymbol{X}_t' \boldsymbol{X}_t)^2 | z_t = r], \\ \boldsymbol{V}_i(r) &= \operatorname{E}[\boldsymbol{X}_t \boldsymbol{X}_t' \epsilon_{it}^2 | z_t = r], \\ V_{2,i}(r) &= \operatorname{E}[(\boldsymbol{X}_t' \boldsymbol{X}_t)^2 \epsilon_{it}^4 | z_t = r], \end{aligned}$$

and consider the following

Assumptions

1. $(\boldsymbol{X}_t, z_t, \boldsymbol{\epsilon}_t)$ is strictly stationary with β-mixing coefficient $\beta_j = O(j^{-\delta})$, for some $\delta > 4$.

2. $\mathrm{E}[\boldsymbol{\epsilon}_t|F_{t-1}] = \mathbf{0}$, where F_{t-1} is the σ field generated by $(\boldsymbol{X}_{i+1}, z_{i+1}, \boldsymbol{\epsilon}_i)$ for $i \leq t-1$.
3. $\mathrm{E}[|y_{it}|^4] < \infty$, $\mathrm{E}[|x_{jt}|^4] < \infty$ and $\mathrm{E}[|\epsilon_{it}|^4] < \infty$ for all i and j.
4. The density function $f(r)$ of $z_t = \mathcal{A}(\xi_{t-1}, \ldots, \xi_{t-d})$, as defined in (2.37), is positive on $R_0 \subset R$, and $r_1, \ldots, r_{m-1}$ are interior points of R_0.
5. $f(r)$, $\boldsymbol{D}(r)$, $D_2(r)$, $\boldsymbol{V}_i(r)$, $V_{2,i}(r)$ are continuous at $r \in \{r_1, \ldots r_{m-1}\}$.
6. $\boldsymbol{\Delta}_j \equiv \boldsymbol{\Phi}_j - \boldsymbol{\Phi}_{j+1} \neq \mathbf{0}$;
7. $\boldsymbol{\Delta}_{j,i}\boldsymbol{D}(r_j)\boldsymbol{\Delta}'_{j,i} > 0$, $\boldsymbol{\Delta}_{j,i}\boldsymbol{V}_i(r_j)\boldsymbol{\Delta}'_{j,i} > 0$ for $i = 1, \ldots k$ and $j = 1, \ldots m-1$, where $\boldsymbol{\Delta}_{j,i}$ is the i-th row of $\boldsymbol{\Delta}'_j$.

Then, walking the same path as [45], asymptotic properties of the conditional least squares estimates can be established for model (2.38).

Consider model (2.38) and suppose that Assumptions 1–7 hold. We expect that conditional least squares estimators are strongly consistent as the sample size increases. That is, $\hat{\boldsymbol{\Phi}}_j \to \boldsymbol{\Phi}_j$, $\hat{\boldsymbol{r}} \to \boldsymbol{r}$, $\hat{d} \to d$, and $\hat{\boldsymbol{\Sigma}}_j \to \boldsymbol{\Sigma}_j$ almost surely as the sample size n goes to infinity. Furthermore, if $\mathrm{Vec}(\boldsymbol{M})$ is column stacking vector of the matrix $\boldsymbol{M}$, we expect that $\sqrt{n_j}\,\mathrm{Vec}(\hat{\boldsymbol{\Phi}}_j - \boldsymbol{\Phi}_j)$ are asymptotically normal with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Gamma}_j \otimes \boldsymbol{\Sigma}_j$, where $\boldsymbol{\Gamma}_j = \lim_{n\to\infty}\left(\sum_t^{(j)} \boldsymbol{X}_t\boldsymbol{X}'_t\right)/n_j$ and $\otimes$ denotes the Kronecker product.

Note that univariate form of the above hypotheses is proved in [7], [8]. The theory of multivariate aggregation-based switching models is in development only and the verification of the above hypotheses will be the subject of our next investigations.