

# **Thresholds and Smooth Transitions in Vector Autoregressive Models**

**Kirstin Hubrich and Timo Teräsvirta**

**CREATES Research Paper 2013-18**

# Thresholds and Smooth Transitions in Vector Autoregressive Models

Kirstin Hubrich\* and Timo Teräsvirta†

\*European Central Bank, Frankfurt am Main

†CREATES, Aarhus University

June 6, 2013

## Abstract

This survey focuses on two families of nonlinear vector time series models, the family of Vector Threshold Regression models and that of Vector Smooth Transition Regression models. These two model classes contain incomplete models in the sense that strongly exogenous variables are allowed in the equations. The emphasis is on stationary models, but the considerations also include nonstationary Vector Threshold Regression and Vector Smooth Transition Regression models with cointegrated variables. Model specification, estimation and evaluation is considered, and the use of the models illustrated by macroeconomic examples from the literature.

**Keywords:** common nonlinearity, impulse response analysis, linearity testing, multivariate nonlinear model, nonlinear cointegration, threshold estimation

**JEL Classification Codes:** C32, C51, C52, C53

**Acknowledgements:** This research has been supported by CREATES - Center for Research in Econometric Analysis of Time Series (DNR78), funded by the Danish National Research Foundation. The views expressed in this article are those of the authors and should not be interpreted as reflecting the views of the European Central Bank.

# 1 Introduction

Since its appearance as a macroeconomic modelling tool more than thirty years ago (Sims, 1980), the basic linear vector autoregressive (VAR) model has been applied in various economic modelling situations and developed further in many ways. For a concise overview of the linear VAR model covering the basic statistical inference, interpretation of the estimated model through the impulse response analysis, and forecasting, see Lütkepohl (2006).

Although linear VAR models have served economists well, there are situations where a nonlinear model would be appropriate. Economic theory may suggest employing a nonlinear model. Alternatively, observed time series may indicate that the variables of interest could be nonlinearly related. For example, the recent financial crisis and the Great Recession suggest that describing quantitative relationships between the financial sector and the macroeconomy would require a nonlinear model. Nonlinear models may also be useful in analysing monetary policy. Positive and negative monetary policy shocks may have asymmetric effects, and the effect may depend nonlinearly on the size of the shock as opposed to linear models in which this effect is proportional to the size of the shock. Nonlinear models may also have a role to play in studies of effectiveness of fiscal policy. The effects of certain fiscal policy measures may depend on the phase of the business cycle, and investigating fiscal multiplier effects may require nonlinear tools.

There are different ways of modelling nonlinear relationships in economics. The leading examples of parametric nonlinear VAR models are the vector threshold autoregressive (VTAR), the vector smooth transition autoregressive (VSTAR) and the vector Markov-switching autoregressive (VMSAR) model. The first two models differ from the third one in that in the former all variables of the model are observable, whereas in the VMSAR model an unobservable, 'latent', variable plays a central rôle. In this survey the focus will be on the VTAR and VSTAR models. Incomplete models that contain strongly exogenous variables are considered as well: these are called the vector switching regression (VSR) and vector smooth transition regression (VSTR) model, respectively. Statistical theory behind these models is discussed and their use illustrated by empirical examples selected from the literature. The focus is on stationary models, but statistical inference in non-stationary models with cointegrated variables also receives attention. The important class of VMSAR models and their applications deserve a survey of their own and are not included in this review.

The choice between VTAR and VSTAR models depends on the economic problem to be investigated. The VTAR model is designed to describe situations in which the dynamic behaviour of a set of random variables can be

modelled by defining a limited number of linear states or regimes that the process can visit. VSTAR models can be viewed as models with a continuum of states, that is, the dynamic behaviour of the variables changes smoothly between a number (often two) extreme states or regimes. An observable switch or transition variable determines the regime that generates the next observation.

The plan of this chapter is as follows. Single-equation switching regression and smooth transition regression models are introduced in Section 2. The VSR and VSTR models, including some special cases, are defined in Section 3. Parameter estimation by least squares is considered in Section 4. Section 5 is devoted to building adequate VSR and VSTR models: a modelling cycle consisting of specification, estimation and evaluation of stationary models of this kind is presented. Approaches to inference in nonstationary models with cointegrated variables are considered as well. Section 6 contains a discussion of the possibility that nonlinearity in these models can be eliminated by a linear transformation, that is, nonlinearity in the individual equations of the model is a common feature. Generalized impulse response functions are considered in Section 7, and forecasting with VSR and VSTR models in Section 8. The purpose of Section 9 is to illustrate the use of the models by presenting applications to macroeconomic time series. Final remarks and suggestions for further reading can be found in Section 10.

## 2 Background: single equation models

### 2.1 Switching regression model

The switching regression (SR) model was introduced by Quandt (1958), who also discussed parameter estimation. Subsequently, Quandt (1960) considered linearity testing but did not yet solve the problem of the asymptotic null distribution of his test statistic. A dynamic SR model is defined as follows:

$$y_t = \sum_{i=1}^q (\mu_i + \sum_{j=1}^p \phi_{ij} y_{t-j} + \gamma_i' \mathbf{x}_t + \varepsilon_{it}) I(c_{i-1} < s_t \leq c_i) \quad (1)$$

where  $y_t$  and  $s_t$  are stationary random variables,  $\mathbf{x}_t$  is a vector of strongly exogenous variables,  $\{\varepsilon_{it}\} \sim \text{iid}(0, \sigma_i^2)$ ,  $\mu_i$  and  $\phi_{ij}$  are parameters,  $i = 1, \dots, q$ , and  $j = 1, \dots, p$ ,  $\gamma_i$ ,  $i = 1, \dots, q$ , are  $n \times 1$  parameter vectors,  $c_0 = -\infty$ , and  $c_q = \infty$ .  $I(A)$  is an indicator function:  $I(A) = 1$  when  $A$  occurs, zero otherwise.

Later, Howell Tong introduced an autoregressive version of (1),  $\gamma_i = \mathbf{0}$  for all  $i$ , and called it the self-exciting threshold autoregressive (SETAR) model.

In the SETAR model,  $s_t = y_{t-d}$ ,  $d > 0$ , hence the term 'self-exciting'. It quickly became the most popular nonlinear univariate time series model in many areas of application. Chan (1993) worked out the asymptotic theory for least squares estimators of the SETAR model with a single threshold, and Qian (1998) did the same for maximum likelihood estimators. In particular, they showed that under certain conditions, the least squares estimator of  $c_1$  is super consistent. For an excellent review of SETAR models, see Tong (1990). For further developments of asymptotic inference, see Hansen (1997, 1999, 2000).

As an economic example, Pfann, Schotman and Tschernig (1996) used a single-threshold SETAR model to describe the dynamic behaviour of the three-month US T-bill interest rate. The time series extends from 1962(1) to 1990(6). The results show that when the interest rate is sufficiently low it behaves like a random walk. At high values, the rate is mean-reverting. The switch extends to the variance as in (11). The estimated variance is clearly higher in the high rate than in the low rate regime.

One obtains another form of a SR or threshold model by assuming that switching in (1) is controlled by a latent variable instead of an observable one. This is done by defining a latent discrete random variable  $s_t$  with  $q$  possible values and replacing the indicator in (1) by  $I(s_t = k)$ . In economic applications it is often assumed that  $s_t$  has memory. The most common case is the one where  $\{s_t\}$  is a (first-order) Markov chain; see Lindgren (1978) for this type of model.

## 2.2 Smooth transition regression model

The origins of smooth transition models can be traced back to Bacon and Watts (1971) who suggested that instead of fitting two linear regression lines to a bivariate point cloud, one could have a smooth transition from one line to the other as a function of the explanatory variable. The smooth transition regression (STR) model has the form

$$y_t = \phi_0 + \sum_{j=1}^p \phi_j y_{t-j} + \boldsymbol{\gamma}' \mathbf{x}_t + (\psi_0 + \sum_{j=1}^p \psi_j y_{t-j} + \boldsymbol{\eta}' \mathbf{x}_t) G(\boldsymbol{\gamma}, \mathbf{c}; s_t) + \varepsilon_t \quad (2)$$

where  $\phi_j$  and  $\psi_j$ ,  $j = 0, 1, \dots, p$ , are parameters,  $\boldsymbol{\gamma}$  and  $\boldsymbol{\eta}$  are  $n \times 1$  parameter vectors, and  $G(\boldsymbol{\gamma}, \mathbf{c}; s_t)$  is a bounded continuous function of  $s_t$ , called the transition function. The error sequence  $\{\varepsilon_t\} \sim \text{iid}(0, \sigma^2)$ . The transition function of Bacon and Watts (1971) was a hyperbolic tangent function,

and the parameters of the ensuing nonlinear model were estimated using Bayesian methods. Maddala (1977, p. 396) suggested a logistic function for the purpose, and that has become a standard choice. For  $\mathbf{c} = c$  (scalar),

$$G(\gamma, c; s_t) = (1 + \exp\{-\gamma(s_t - c)\})^{-1}, \quad \gamma > 0 \quad (3)$$

where  $\gamma > 0$  is an identifying restriction. Equations (2) and (3) define a single-transition logistic STR model. A pure time series smooth transition autoregressive (STAR) model,  $\boldsymbol{\gamma} = \boldsymbol{\eta} = \mathbf{0}$  and  $s_t = y_{t-d}$ ,  $d > 0$ , in (2), was introduced by Chan and Tong (1986). An alternative to the logistic function (3) is the exponential transition function:

$$G(\gamma, c; s_t) = 1 - \exp\{-\gamma(s_t - c)^2\} \quad (4)$$

and (2) with (4) define an exponential STR (ESTR) model. Haggan and Ozaki (1981) already considered a univariate version of this model in the special case  $c = 0$  in (4). Teräsvirta (1994) constructed a model building strategy for STAR models that Eitrheim and Teräsvirta (1996) completed.

Smooth transition type models can describe aggregate dynamics in situations where individual agents respond differently to a stimulus. LSTAR models are often an alternative to single-threshold TAR models. ESTR models have been used for modelling real exchange rates in cases where the mean reversion is symmetric around an equilibrium and the strength of the attraction depends nonlinearly on the size of the deviation from this equilibrium. See Taylor, Peel and Sarno (2001) for discussion and examples. More discussion on all these models and modelling time series with them can be found in Teräsvirta, Tjøstheim and Granger (2010). For forecasting with STAR models, see also Lundbergh and Teräsvirta (2002), Teräsvirta (2006) and Kock and Teräsvirta (2011).

### 2.3 Switching mean autoregressive model

Lanne and Saikkonen (2002) introduced a univariate threshold autoregressive (TAR) model in which only the intercept is switching, whereas the autoregressive structure remains unchanged. Their model can be written as follows:

$$y_t = \sum_{j=1}^r \phi_{0j} I(c_{j-1} < y_{t-d} \leq c_j) + \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t \quad (5)$$

where  $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ . This specification is used for describing 'near unit root' behaviour. That means modelling series which may 'look' nonstationary but are stationary and fluctuate within bounds. Interest rate or unemployment

rate series serve as examples. The switching intercept in (5) causes level shifts in realisations, although the sequence  $\{y_t\}$  itself is stationary and ergodic when the roots of the lag polynomial  $1 - \sum_{j=1}^p \phi_j z^j$  lie outside the unit circle.

One of the two examples in Lanne and Saikkonen (2002) consists of fitting the intercept-switching TAR model with one and two thresholds to the monthly UK Treasury bill series, 1964(1) to 1997(9). The results show that the persistence (the sum of the autoregressive coefficients) considerably decreases when one moves from a pure linear autoregressive model, for which the unit root hypothesis is not rejected, to the intercept-switching TAR model with two thresholds. When the TAR model is tested against a linear non-stationary autoregressive model containing a unit root, the null hypothesis is not rejected. For the test and the asymptotic theory behind it, see Lanne and Saikkonen (2002).

The switching-mean autoregressive (SM-AR) model, see González and Teräsvirta (2008) or González, Hubrich and Teräsvirta (2009), is a variant of (5). In the SM-AR model the switching intercept has been replaced by an explicitly nonstationary component. The model has the following form:

$$y_t = \delta(t) + \sum_{j=1}^p \phi_j y_{t-j} + \varepsilon_t. \quad (6)$$

where  $\varepsilon_t \sim \text{iid}(0, \sigma^2)$ . The deterministic time-varying intercept is defined as

$$\delta(t) = \delta_0 + \sum_{j=1}^q \delta_j G(\gamma_j, c_j, t/T)$$

where  $\delta_j$ ,  $j = 0, 1, \dots, q$ , are parameters and  $G(\gamma_j, c_j, t/T)$  logistic functions. The roots of  $1 - \sum_{j=1}^p \phi_j z^j$  are assumed to lie outside the unit circle, so  $\{y_t - \delta(t)\}$  is a stationary sequence. The unconditional (switching) mean is

$$\mathbf{E}_t y_t = (1 - \sum_{j=1}^p \phi_j \mathbf{L}^j)^{-1} \delta(t).$$

where  $\mathbf{E}_t$  indicates that the expectation is time-varying, and  $\mathbf{L}$  is the lag operator.

It is also possible to reparameterise (6) as follows:

$$y_t = \delta(t) + \sum_{j=1}^p \phi_j \{y_{t-j} - \delta(t-j)\} + \varepsilon_t. \quad (7)$$

In that case,  $\mathbf{E}_t y_t = \delta(t)$ . This parameterisation is analogous to that in the Markov-switching model Hamilton (1989) introduced. In that model the deterministic  $\delta(t)$  in (7) is replaced by a latent random variable  $\mu(s_t)$ , where  $\mu(s_t)$  represents the local (switching) mean of  $y_t$ . The latent variable  $s_t$  is assumed to have a (first-order) Markov structure, which implies that the switching mean obtains  $q$  different values.

### 3 Vector models

#### 3.1 General nonlinear dynamic vector model

A general nonlinear dynamic vector model can be written as follows:

$$\mathbf{y}_t = \mathbf{f}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}) + \boldsymbol{\varepsilon}_t \quad (8)$$

where  $\mathbf{y}_t$  is a stochastic  $m \times 1$  vector,  $\mathbf{x}_t$  is an  $n \times 1$  vector of exogenous variables,  $\boldsymbol{\varepsilon}_t$  is an  $m \times 1$  error vector with  $\mathbf{E}\boldsymbol{\varepsilon}_t = \mathbf{0}$  and  $\text{cov}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Sigma} > 0$ , and  $\boldsymbol{\theta}$  is a parameter vector. The  $m \times 1$  vector  $\mathbf{f}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta})$  has the form

$$\mathbf{f}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}) = (f_1(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}), \dots, f_m(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}))'. \quad (9)$$

The  $n$ -vector  $\mathbf{x}_t$  may contain an intercept, other deterministic components and strongly exogenous variables. In some cases the whole vector, excluding the deterministic components, may be lagged. All variables in (8) are observable and all parameters non-random. The sequence  $\{\mathbf{y}_t\}$  is often assumed weakly stationary, but sometimes it is only assumed that the sequence of first differences  $\{\Delta \mathbf{y}_t\}$  is weakly stationary.

This survey concentrates on special cases of (8). A useful simplification is the one in which nonlinearity is additive. This means that the function  $\mathbf{f}$  defined in (9) can be decomposed as follows:

$$\mathbf{f}(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}) = \boldsymbol{\mu} + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\Gamma} \mathbf{x}_t + \mathbf{f}_1(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}_1) \quad (10)$$

where  $\boldsymbol{\mu}$  is an  $m \times 1$  intercept vector (other deterministic components are assumed away for notational simplicity),  $\boldsymbol{\Phi}_j$  is an  $m \times m$  parameter matrix,  $j = 1, \dots, p$ , and  $\boldsymbol{\Gamma}$  is an  $m \times n$  parameter matrix. The models to be considered below have this or a similar structure. The function  $\mathbf{f}_1$  can sometimes be linear in parameters, if not in variables.



## 3.2 Vector threshold autoregressive model

### 3.2.1 Definition

The vector switching regression (VSR) or threshold autoregressive (VTAR) model is a straightforward generalisation of the corresponding single-equation model. The former term applies to incomplete models that contain strongly exogenous regressors, whereas the VTAR model generalises the linear vector autoregressive model. The stationary VSR model is defined as follows:

$$\mathbf{y}_t = \sum_{i=1}^q (\boldsymbol{\mu}_i + \sum_{j=1}^p \boldsymbol{\Phi}_{ij} \mathbf{y}_{t-j} + \boldsymbol{\Gamma}_i \mathbf{x}_t + \boldsymbol{\varepsilon}_{it}) I(c_{i-1} < s_t \leq c_i) \quad (11)$$

where  $\mathbf{y}_t$  and  $\boldsymbol{\varepsilon}_{it}$ ,  $j = 1, \dots, q$ , are stochastic  $m \times 1$  vectors, and  $\boldsymbol{\mu}_i$  is an  $m \times 1$  vector of intercepts,  $i = 1, \dots, q$ . Furthermore,  $\boldsymbol{\Phi}_{ij}$  are  $m \times m$  coefficient matrices,  $i = 1, \dots, q$ , and  $j = 1, \dots, p$ ,  $\boldsymbol{\Gamma}_i$  are  $m \times n$  coefficient matrices,  $i = 1, \dots, q$ ,  $c_0 = -\infty$ , and  $c_q = \infty$ . When  $\boldsymbol{\Gamma}_i = \mathbf{0}$  for all  $i$ , (11) becomes a VTAR model. The errors  $\boldsymbol{\varepsilon}_{it}$  are serially uncorrelated with mean  $\mathbf{0}$  and positive definite covariance matrix  $\boldsymbol{\Sigma}_i$ ,  $i = 1, \dots, q$ . The vector  $\mathbf{x}_t$  contains  $n$  stationary variables that are strongly exogenous for the parameters of the model. If the stationary variable  $s_t$  is replaced by  $t/T$ , where  $T$  is the number of observations, (11) becomes a linear VR model with  $q - 1$  breaks.

In (11), a single stationary and continuous switch-variable  $s_t$  determines the regime of the whole system. This assumption has been relaxed by Tena and Tremayne (2009) such that each equation may have a separate switch-variable and some equations may even be linear. The authors argue that there are good economic reasons for that, but in economic applications other than theirs, see Section 9.1, this has not been the case. A systematic modelling strategy for VSR or VTAR models with applications can be found in Tsay (1998) and is discussed in Section 5.2.

Finally, it may be noted that substituting a latent indicator for the observable  $s_t$  in (11) yields the following vector regression model:

$$\mathbf{y}_t = \sum_{i=1}^q (\boldsymbol{\mu}_i + \sum_{j=1}^p \boldsymbol{\Phi}_{ij} \mathbf{y}_{t-j} + \boldsymbol{\Gamma}_i \mathbf{x}_t + \boldsymbol{\varepsilon}_{it}) I(s_t = \nu_i) \quad (12)$$

in which the indicator variable  $s_t$  is defined similarly to the single-equation model of Lindgren (1978). Typically, the switches in all equations are controlled by the same indicator function, as is the case in (12). This assumption can be relaxed, however, see Sims, Waggoner and Zha (2008). Recent economic applications include Sims and Zha (2006) and Hubrich and Tetlow (2012). Their approach is Bayesian; for more details and discussion in

the classical framework, see Krolzig (1997) and Jacobson, Lindh and Warne (2002).

### 3.2.2 Nested vector threshold autoregressive model

The VTAR model (11) may be generalised in various ways. It is possible to add thresholds and threshold variables such that the switches are controlled by more than variable. One of these models is a multivariate generalisation of the univariate Nested TAR (NTAR) model by Astatkie, Watts and Watt (1997). It is a four-regime model with the following form:

$$\begin{aligned}
\mathbf{y}_t = & \left\{ (\boldsymbol{\mu}_{11} + \sum_{j=1}^p \boldsymbol{\Phi}_{11j} \mathbf{y}_{t-j}) I(c_1 < s_{1t}) + (\boldsymbol{\mu}_{12} + \sum_{j=1}^p \boldsymbol{\Phi}_{12j} \mathbf{y}_{t-j}) I(c_1 \geq s_{1t}) \right\} \\
& \times I(c_2 < s_{2t}) + \left\{ (\boldsymbol{\mu}_{21} + \sum_{j=1}^p \boldsymbol{\Phi}_{21j} \mathbf{y}_{t-j}) I(c_1 < s_{1t}) \right. \\
& \left. + (\boldsymbol{\mu}_{22} + \sum_{j=1}^p \boldsymbol{\Phi}_{22j} \mathbf{y}_{t-j}) I(c_1 \geq s_{1t}) \right\} I(c_2 \geq s_{2t}) + \boldsymbol{\varepsilon}_t
\end{aligned} \tag{13}$$

where the error process  $\{\boldsymbol{\varepsilon}_t\} \sim \text{iid}(\mathbf{0}, \boldsymbol{\Sigma})$  and thus does not switch. In this model, the first threshold variable  $s_{1t}$  controls switching as in (11). But then, the two states between which the process switches according to  $s_{1t}$  can change due to the other threshold variable  $s_{2t}$ . The standard VTAR model is thus nested in the more general Vector NTAR (VNTAR) model.

There may not be any applications of the VNTAR model (13) with stochastic stationary threshold variables  $s_{1t}$  and  $s_{2t}$  to economic time series. A special case, however, in which  $s_{2t} = t$ , has been considered by Galvão (2006). In this model, the two threshold regimes determined by  $s_{1t}$  change once during the observation period: the second threshold indicates a structural break in the VTAR process. The author calls this model the Structural Break Threshold Vector Autoregressive (SBTVAR) model. She discusses specification and estimation issues and applies the model to predicting recessions, the 2001 recession in the US in particular.

### 3.2.3 Threshold vector error correction

Following the way linear cointegrated VAR models are parameterised, there is an extension of the VTAR model applicable to nonstationary time series. Balke and Fomby (1997) introduced this extension to allow discontinuous changes in the attraction towards the equilibrium characterised by the cointegrating relationships. To define the model, assume for simplicity that  $\boldsymbol{\Gamma}_i = \mathbf{0}$

for all  $i$  and that  $q = 3$ . The model (11) is then rewritten as follows:

$$\Delta \mathbf{y}_t = \sum_{j=1}^3 (\boldsymbol{\mu}_j + \boldsymbol{\Pi}_j \mathbf{y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\Psi}_{jk} \Delta \mathbf{y}_{t-k} + \boldsymbol{\varepsilon}_{jt}) I(c_{j-1} < s_t \leq c_j) \quad (14)$$

for  $p \geq 2$ . When  $p = 1$ , the weighted sum of lagged first differences vanishes. Assume that in each regime,  $\mathbf{y}_t$  is integrated of order one ( $\Delta \mathbf{y}_t$  stationary in the mean) and the  $m \times m$  matrix  $\boldsymbol{\Pi}_j = \mathbf{A}_j \mathbf{B}'$ , where  $\mathbf{A}_j$ ,  $j = 1, \dots, q$ , and  $\mathbf{B}$  are  $m \times r$  matrices with full column rank. Furthermore,  $s_t$  is continuous and stationary as before. If  $m = 2$ , say, then  $\mathbf{A}_j = (\alpha_{1j}, \alpha_{2j})'$  and  $\mathbf{B} = (1, \beta_2)'$ , and the variables  $y_{1t}$  and  $y_{2t}$  are cointegrated with the cointegrating vector  $\mathbf{B}$ . This model is called the threshold vector error correction (TVEC) model. The difference between the linear cointegrated VAR model and (14) with  $q = 3$  is that strength of the attraction varies in the three regimes according to  $\mathbf{A}_j$ . The cointegrating relationships defined by the matrix  $\mathbf{B}$  remain linear.

A special case with economic applications has  $q = 3$  and  $\mathbf{A}_2 = \mathbf{0}$ . This allows modelling relationships in which error correction operates outside an interval, but is incomplete in the sense that there is a band within which no adjustment takes place. The intercept is restricted in the cointegrating relationship, so that

$$\begin{aligned} \Delta \mathbf{y}_t &= \sum_{j=1}^3 \{ \mathbf{A}_j (\mathbf{B}' \mathbf{y}_{t-1} - \boldsymbol{\mu}_j) \\ &\quad + \sum_{k=1}^{p-1} \boldsymbol{\Psi}_{jk} \Delta \mathbf{y}_{t-k} + \boldsymbol{\varepsilon}_{jt} \} I(c_{j-1} < s_t \leq c_j) \end{aligned} \quad (15)$$

where  $\boldsymbol{\mu}_j$  is the regime-specific mean vector of  $\mathbf{B}' \mathbf{y}_{t-1}$ . Model (15) is called the band-TVEC model; see Lo and Zivot (2001) for discussion. The corresponding univariate model is called the band-TAR model.

In applications of (15), the interval  $(c_1, c_2]$  often contains zero. For example, when considering trade with identical commodities, transportation costs help sustain a difference in the price of this commodity in two regions. Price convergence can only occur when the absolute difference between the prices is sufficiently large, so profitable export or import is possible. In this example, the transition variable  $s_t$  in (15) is the difference between the prices.

### 3.2.4 Stationarity in threshold vector error correction models

Stationarity and ergodicity of the equilibrium component  $\mathbf{B}' \mathbf{y}_{t-1}$  and the differences  $\Delta \mathbf{y}_t$  defined by the TVEC process (14) where are considered in

Bec and Rahbek (2004). Let  $\mathbf{z}_t = (\mathbf{y}'_t \mathbf{B}, \Delta \mathbf{y}'_{t-1}, \dots, \Delta \mathbf{y}'_{t-p+2})'$ . The focus is on the case where the threshold variable  $s_t = \|\mathbf{z}_{t-1}\|$  and the indicator has the form  $I(\|\mathbf{z}_{t-1}\| \geq c)$  in (14). This implies that the resulting three-regime model is 'symmetric': the mid-regime is different from the identical outer regimes,  $\mathbf{A}_1 = \mathbf{A}_3 \neq \mathbf{A}_2$ , and the two thresholds separating them are  $c_1 = -c$  and  $c_2 = c > 0$ . Under the conditions given by Bec and Rahbek (2004), see their Theorem 2,  $\mathbf{B}'\mathbf{y}_t$  and  $\Delta \mathbf{y}_t$  are geometrically ergodic, and assuming the process starts at  $t = 0$ , they can be given initial distributions such that they are stationary. Stationarity is seen to depend only on the parameters of the outer regimes. For stationarity, the characteristic polynomial matrix for those regimes or regime

$$C_j(z) = \mathbf{I}_m(1 - z) - \mathbf{A}_j \mathbf{B}' z - \sum_{k=1}^{p-1} \Psi_{jk}(1 - z)z^k, \quad j = 1 \text{ or } 3$$

must have  $m - r$  roots equal to unity, while the remaining roots  $|z| > 1$  (Johansen, 1995, Theorem 4.2). Bec and Rahbek (2004) also look at the case in which  $s_t = \|\mathbf{B}'\mathbf{y}_{t-1}\|$  but then they assume that  $\Psi_{jk} = \Psi_k$ ,  $k = 1, \dots, p$ , that is, the short-run coefficients do not switch.

Saikkonen (2005) studies a rather general class of nonlinear error correction models with the error correction component  $\mathbf{B}'\mathbf{y}_t$  as before. Conditional heteroskedasticity in the errors is also allowed. A three-regime VTAR model with  $\mathbf{A}_1 = \mathbf{A}_3 \neq \mathbf{A}_2$  is a member of this class. The threshold variable is a function of  $\mathbf{z}_{t-1}$  defined as before. Under the conditions given in the paper and using a transformation of the general model, Saikkonen (2005) is able to show that given a proper choice of initial distributions,  $\mathbf{B}'\mathbf{y}_t$  and  $\Delta \mathbf{y}_t$  are both strictly and weakly stationary.

It should be mentioned, however, that the results in Bec and Rahbek (2004) are not restricted to mere VTAR models either but apply to a more general class containing, among other things, the case in which the threshold is not completely observed but contains an unobserved random element. The results in Saikkonen (2005) cover this extension as well. See also Section 3.3.3, where extensions to smooth transition VEC models are discussed.

### 3.3 Vector smooth transition autoregressive model

#### 3.3.1 Definition

When the stationary smooth transition regression model is generalised to the vector case, one obtains a logistic vector STR (LVSTR) model of order  $p$ . This model first appeared in Anderson and Vahid (1998) who were interested

in testing common nonlinearities in vector models. Their approach to this testing problem is discussed in Section 6.3. For notational simplicity, a single-transition (two extreme regimes) version of the LVSTR model is considered here. For more general representations, see Teräsvirta and Yang (2013). The model may be defined as follows:

$$\begin{aligned}
\mathbf{y}_t &= \boldsymbol{\mu}_0 + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\Gamma} \mathbf{x}_t + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \\
&\quad \times \left\{ \boldsymbol{\mu}_1 + \sum_{j=1}^p (\boldsymbol{\Psi}_j \mathbf{y}_{t-j} + \boldsymbol{\Xi} \mathbf{x}_t) \right\} + \boldsymbol{\varepsilon}_t \\
&= \boldsymbol{\mu}_0 + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \boldsymbol{\mu}_1 + \sum_{j=1}^p \{ \boldsymbol{\Phi}_j + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \boldsymbol{\Psi}_j \} \mathbf{y}_{t-j} \\
&\quad + (\boldsymbol{\Gamma} + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \boldsymbol{\Xi}) \mathbf{x}_t + \boldsymbol{\varepsilon}_t
\end{aligned} \tag{16}$$

where  $\mathbf{y}_t$  is an  $m \times 1$  vector,  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\mu}_1$  are  $m \times 1$  intercept vectors,  $\boldsymbol{\Phi}_j$  and  $\boldsymbol{\Psi}_j$ ,  $j = 1, \dots, p$ , are  $m \times m$  parameter matrices,  $\boldsymbol{\Gamma}$  and  $\boldsymbol{\Xi}$  are  $m \times n$  parameter matrices, and

$$\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) = \text{diag}\{G_1(\gamma_1, \mathbf{c}_1, s_{1t}), \dots, G_m(\gamma_m, \mathbf{c}_m, s_{mt})\} \tag{17}$$

is an  $m \times m$  diagonal matrix of bounded logistic transition functions. It is usually assumed that  $\{\mathbf{y}_t\}$  defined by (16) and (17) is a stationary and ergodic sequence.

Equation (16) defines quite a general LVSTR model. One way of simplifying it is to assume that the transition variables as well as the parameters of the transition functions are the same for all equations. As already noted, a similar assumption is made in VTAR models. In that case,  $\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) = G(\gamma, \mathbf{c}; s_t) \mathbf{I}_m$ . The LVSTAR model suggested by Weise (1999) contains this assumption. When  $\gamma \rightarrow \infty$ , this model becomes a two-regime VTAR model. Setting  $s_{jt} = t$ ,  $j = 1, \dots, m$ , leads to an LVSTAR model that defines the alternative model in testing parameter constancy of linear VAR models by LM-type tests; see He, Teräsvirta and González (2009).

Another simplification is to assume that the LVSTAR model is intercept-switching, that is,  $\boldsymbol{\Psi}_j = \mathbf{0}$ ,  $j = 1, \dots, p$ , and  $\boldsymbol{\Xi} = \mathbf{0}$  in (16). The intercept-switching model is already fairly flexible specification. Besides, as the intercept  $\boldsymbol{\mu}_0 + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \boldsymbol{\mu}_1$  is bounded for all values of  $\mathbf{s}_t$ , the model has the same stability condition as the linear VAR model. The expectation

$$\mathbf{E} \mathbf{y}_t = (\mathbf{I}_m - \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{L}^j)^{-1} \{ \boldsymbol{\mu}_0 + \mathbf{E} \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) \boldsymbol{\mu}_1 \}$$

can only be computed numerically. An empirical example of the intercept-switching LVSTAR model can be found in Section 9.2. It would also be possible to generalise the intercept-switching AR model of Lanne and Saikkonen (2002) to the vector case, but it seems that this has not been done in the literature on vector TAR models.

It would be possible to define smooth transition counterparts to the VN-TAR model (13) and its special case, the SBTVAR model. The latter in fact exists as a univariate model and is called the Time-Varying STAR (TV-STAR) model, see Lundbergh, Teräsvirta and van Dijk (2003).

### 3.3.2 Smooth transition vector error correction

A Smooth Transition Vector Error Correction (STVEC) model can be defined along the same lines as the TVEC model (14). As an example, a single-transition STVEC model (without exogenous variables) can be written as follows:

$$\begin{aligned} \Delta \mathbf{y}_t = & \boldsymbol{\mu}_1 + \mathbf{A}_1 \mathbf{B}' \mathbf{y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\Phi}_k \Delta \mathbf{y}_{t-k} \\ & + \mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) (\boldsymbol{\mu}_2 + \mathbf{A}_2 \mathbf{B}' \mathbf{y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\Psi}_k \Delta \mathbf{y}_{t-k}) + \boldsymbol{\varepsilon}_t \end{aligned} \quad (18)$$

where  $\Delta \mathbf{y}_t$  is  $\mathbf{I}(0)$ ,  $\mathbf{A}_j \mathbf{B}'$ ,  $j = 1, 2$ , are reduced rank  $m \times m$  matrices as in (14) and  $\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t)$  is a diagonal matrix defined in (17). Rothman, van Dijk and Franses (2001) defined the STVEC model assuming  $\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) = G(\boldsymbol{\gamma}, \mathbf{c}; s_t) \mathbf{I}_m$ . Camacho (2004) constructed a two-dimensional ( $m = 2$ ) STVEC model without this assumption. His application is discussed in Section 9.1. Goodwin, Holt and Prestemon (2012) also specified a two-dimensional STVEC model without assuming identical transition functions and applied it to modelling the exchange rate pass-through in timber products crossing the border between Canada and the US. Their model could be viewed as being of type (16) in the sense that it contains strongly exogenous variables. The authors, however, closed the system by constructing separate linear equations for these variables.

### 3.3.3 Stationarity in smooth transition vector error correction models

Conditions for stationarity and ergodicity of nonstationary STVEC models can be found as special cases of the general nonstationary model introduced

in Saikkonen (2008). The general model in that paper is defined by

$$\begin{aligned} \Delta \mathbf{y}_t &= \sum_{j=1}^q h_j(\mathbf{S}'\mathbf{z}_{t-1}, \eta_t) (\boldsymbol{\mu}_j^* + \mathbf{A}_j^* \mathbf{B}' \mathbf{y}_{t-1} + \sum_{k=1}^{p-1} \boldsymbol{\Phi}_{jk}^* \Delta \mathbf{y}_{t-j}) \\ &\quad + \left( \sum_{j=1}^q h_j(\mathbf{S}'\mathbf{z}_{t-1}, \eta_t) \boldsymbol{\Sigma}_j \right)^{1/2} \boldsymbol{\varepsilon}_t \end{aligned} \quad (19)$$

where  $B$  and  $\mathbf{A}_j^*$ ,  $j = 1, \dots, q$ , are  $m \times r$  matrices with full column rank,  $\boldsymbol{\varepsilon}_t \sim \text{iid}(\mathbf{0}, \boldsymbol{\Sigma})$ , and  $\mathbf{S}$  is typically a selection matrix or vector, selecting the threshold or transition variable or a linear combination of variables from the elements of  $\mathbf{z}_{t-1}$  defined as before. For example, setting  $\mathbf{S}' = [\mathbf{I}_r \ \mathbf{0}]$  gives  $\mathbf{S}'\mathbf{z}_{t-1} = \mathbf{B}'\mathbf{y}_{t-1}$ , and one can for example choose  $h_j(\mathbf{S}'\mathbf{z}_{t-1}, \eta_t) = I(\|\mathbf{B}'\mathbf{y}_{t-1}\| > c)$  as in Bec and Rahbek (2004). Furthermore,  $\eta_t$  is a scalar random variable making the threshold or the location parameter in the logistic transition function unobservable: here we set  $\eta_t = 0$ . It is assumed that  $\sum_{j=1}^q h_j(\mathbf{S}'\mathbf{z}_{t-1}, \eta_t) = 1$ . Since we are interested in STVEC models, we assume that  $h_j(\mathbf{s}'\mathbf{z}_{t-1})$ ,  $j = 1, \dots, q$ , where  $\mathbf{s}$  is a selection vector, are functions of logistic functions. For example, for  $q = 2$  we have

$$h_1(\mathbf{s}'\mathbf{z}_{t-1}) = G(\gamma, c, \mathbf{s}'\mathbf{z}_{t-1}) = (1 + \exp\{-\gamma(\mathbf{s}'\mathbf{z}_{t-1} - c)\})^{-1}, \quad \gamma > 0$$

so  $h_2(\mathbf{s}'\mathbf{z}_{t-1}) = 1 - G(\gamma, c, \mathbf{s}'\mathbf{z}_{t-1})$ . This yields a reparameterised version of (18) with a common transition function for all equations. Selecting  $q = 3$ , one obtains

$$h_j(\mathbf{s}'\mathbf{z}_{t-1}) = \begin{cases} 1 - G(\gamma, c_1; \mathbf{s}'\mathbf{z}_{t-1}) & \text{for } j = 1 \\ G(\gamma, c_1; \mathbf{s}'\mathbf{z}_{t-1}) - G(\gamma, c_2; \mathbf{s}'\mathbf{z}_{t-1}) & \text{for } j = 2 \\ G(\gamma, c_2; \mathbf{s}'\mathbf{z}_{t-1}) & \text{for } j = 3 \end{cases} \quad (20)$$

where  $c_1 < c_2$ . The positive slope parameter  $\gamma$  is the same for both logistic functions. The argument can also be a function of  $\mathbf{s}'\mathbf{z}_{t-1}$  such as  $\|\mathbf{s}'\mathbf{z}_{t-1}\|$ . When  $\gamma \rightarrow \infty$ , assuming  $\mathbf{A}_1 = \mathbf{A}_3$  and that the transition variable equals  $\|\mathbf{B}'\mathbf{y}_{t-1}\|$  the stationarity results for  $\mathbf{B}'\mathbf{y}_t$  and  $\Delta \mathbf{y}_t$  in models with conditional heteroskedasticity are obtained from Saikkonen (2005). When  $\mathbf{A}_1 \neq \mathbf{A}_3$ , this is no longer the case, and the corresponding results are worked out by Saikkonen (2008).

In proving stationarity and ergodicity of the long-run equilibrium relations and the differences is that the relevant matrix polynomials

$$C_j(z) = \mathbf{I}_m(1 - z) - \mathbf{A}_j \mathbf{B}' z - \sum_{k=1}^{p-1} \boldsymbol{\Phi}_{jk}^* (1 - z) z^k \quad (21)$$

have  $m - r$  unit roots, whereas  $|z| > 1$  for the rest. In our example with the logistic transition function and  $q = 2$  this concerns both  $C_1(z)$  and  $C_2(z)$ . When  $h_j(\mathbf{s}'\mathbf{z}_{t-1})$ ,  $j = 1, 2, 3$ , are defined by (20), the relevant polynomials are  $C_1(z)$  and  $C_3(z)$ . The roots of  $C_2(z)$  are not restricted.

The proof is based on a model transformation similar to the one in Saikkonen (2005) but requires additional conditions and is technically more complex than its predecessor. For details, see Saikkonen (2008). These results apply to the case in which all equations of the STVEC model have the same transition function. As yet, no results seem to be available for the more general situation characterised by (16) and (17).

### 3.4 Vector shifting mean autoregressive model

Holt and Teräsvirta (2012) introduced a multivariate version of the shifting mean autoregressive model. It is a straightforward generalisation of the univariate model in Section 2.3. Accordingly, it can be defined in two ways, either

$$\mathbf{y}_t = \boldsymbol{\delta}(t) + \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \quad (22)$$

where  $\boldsymbol{\delta}(t) = (\delta_1(t), \dots, \delta_m(t))'$  with  $\delta_j(t) = \delta_{j0} + \sum_{i=1}^{q_j} \delta_{ji} G(\gamma_{ji}, c_{ji}, t/T)$ ,  $j = 1, \dots, m$ ,  $\boldsymbol{\varepsilon}_t \sim \text{iid}(\mathbf{0}, \boldsymbol{\Sigma})$ ,  $\boldsymbol{\Sigma} > 0$ , or

$$\mathbf{y}_t = \boldsymbol{\delta}(t) + \sum_{j=1}^p \boldsymbol{\Phi}_j \{\mathbf{y}_{t-j} - \boldsymbol{\delta}(t-j)\} + \boldsymbol{\varepsilon}_t. \quad (23)$$

In both cases,  $|\mathbf{I} - \sum_{j=1}^p \boldsymbol{\Phi}_j z^j| \neq 0$  for  $z \leq 1$ , that is,  $\mathbf{y}_t - \boldsymbol{\delta}(t)$  is a weakly stationary sequence. The transition functions are logistic:

$$G(\gamma_{ji}, c_{ji}, t/T) = (1 + \exp\{-\gamma_{ji}(t/T - c_{ji})\})^{-1}, \quad \gamma_{ji} > 0$$

for all  $i$  and  $j$ . In (22), the shifting mean equals

$$\mathbf{E}_t \mathbf{y}_t = (\mathbf{I} - \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{L}^j)^{-1} \boldsymbol{\delta}(t) \quad (24)$$

whereas for (23),  $\mathbf{E}_t \mathbf{y}_t = \boldsymbol{\delta}(t)$ . As in single-equation SM-AR models, the intercept can be a very flexible function of time, and in some situations the VSM-AR model can be viewed as a reasonable alternative to models with unit roots and stochastic trends. Since  $\boldsymbol{\delta}(t)$  is bounded, the expectation (24), even if time-varying, is also bounded.



### 3.5 Vector contemporaneous-threshold model

Dueker, Psaradakis, Sola and Spagnolo (2011) introduced the multivariate (or vector) contemporaneous-threshold (C-MSTAR) model which they write belongs to the class of vector STAR models. It is actually an  $m$ -dimensional mixture model consisting of  $q^2$  basic linear equations that are weighted together by time-varying probabilities. The first-order (for simplicity) model can be written as follows:

$$\mathbf{y}_t = \sum_{i=1}^{q^2} G_i(\mathbf{z}_{t-1})(\boldsymbol{\mu}_i + \mathbf{A}_{i1}\mathbf{y}_{t-1} + \boldsymbol{\Sigma}_i^{1/2}\boldsymbol{\varepsilon}_t)$$

where  $\boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \mathbf{I}_m)$ ,  $\mathbf{z}_{t-1}$  is a linear function of  $\mathbf{y}_{t-1}$ ,

$$G_i(\mathbf{z}_{t-1}) = \frac{\Phi(\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{c}_i - \boldsymbol{\mu}_i - \mathbf{A}_{i1}\mathbf{y}_{t-1}\})}{\sum_{j=1}^{m^2} \Phi(\boldsymbol{\Sigma}_i^{-1/2}\{\mathbf{c}_i - \boldsymbol{\mu}_i - \mathbf{A}_{i1}\mathbf{y}_{t-1}\})}$$

$i = 1, \dots, q^2$ , is bounded between zero and one,  $\Phi(\cdot)$  is the cumulative distribution function of the standard normal distribution, and  $\mathbf{c}_i = (c_{i1}, c_{i2})'$ ,  $i = 1, \dots, q^2$ , are the location vectors. For example, if  $q = 2$ , then  $\mathbf{c}_1 = (c_1, c_2)'$ ,  $\mathbf{c}_2 = (c_1, -c_2)'$ ,  $\mathbf{c}_3 = (-c_1, c_2)'$ , and  $\mathbf{c}_4 = (-c_1, -c_2)'$ . The authors discuss linearity testing ( $q = 1$ ), stability, and maximum likelihood estimation of the parameters, and apply their model to consider the relationship between stock prices and interest rates using a long annual time series from 1900 to 2000. Since the number of parameters increases rapidly as a function of  $q$  and the lag length, the applications of the C-MSTAR model are likely to be low-dimensional.

## 4 Estimation

### 4.1 Estimation of vector switching regression models

Parameters of the VSR model (11) can be estimated using least squares. This includes estimating the thresholds  $c_i$ ,  $i = 1, \dots, r - 1$ . Consider the case in which  $q = 2$ , so there is one threshold,  $c_1$ . First set the threshold at  $c^{(i)}$  where  $c^{(i)}$  is the  $i$ th value of the order statistic of the transition variable  $s_t$ . Then form two sets of observations:  $\mathcal{T}_1$  containing the  $n_1$  observations for which  $s_t \leq c^{(i)}$  and  $\mathcal{T}_2$  comprising the remaining  $T - n_1$  ones, and the sum of squares function

$$Q(c^{(i)}) = \text{tr} \sum_{t \in \mathcal{T}_1} \boldsymbol{\varepsilon}_t'(c^{(i)})\boldsymbol{\varepsilon}_t(c^{(i)}) + \text{tr} \sum_{t \in \mathcal{T}_2} \boldsymbol{\varepsilon}_t'(c^{(i)})\boldsymbol{\varepsilon}_t(c^{(i)})$$

where  $\boldsymbol{\varepsilon}_t(c^{(i)}) = \mathbf{y}_t - (\boldsymbol{\mu}_i + \sum_{j=1}^p \boldsymbol{\Phi}_{ij} \mathbf{y}_{t-j} + \boldsymbol{\Gamma}_i \mathbf{x}_t)$  for  $t \in \mathcal{T}_i$ ,  $i = 1, 2$ . The parameters  $\boldsymbol{\mu}_i$ ,  $\boldsymbol{\Phi}_{ij}$ ,  $j = 1, \dots, p$ ,  $\boldsymbol{\Gamma}_i$ ,  $i = 1, 2$ , are estimated by minimising  $Q(c^{(i)})$  with respect to these parameters. The final estimates, including that of  $c_1$ , are obtained by repeating this for all  $c^{(i)} \in \mathcal{C} = \{c^{(i)} : c(q) \leq c^{(i)} \leq c(1-q)\}$  where  $c(q)$  is the  $q$ th quantile of the order statistic of  $s_t$  and selecting the parameter estimates including  $c_1$  that minimise  $Q(c_1)$ :

$$Q(\hat{c}_1) = \arg \min_{c^{(i)} \in \mathcal{C}} Q(c^{(i)}).$$

If  $T$  is very large,  $\mathcal{C}$  can be replaced by a grid to save computations. Tsay (1998) showed that under regularity conditions including the existence of fourth moments for all variables and the error process, the least squares estimators, including that  $c_1$ , are consistent. As already mentioned, in the univariate TAR case Chan (1993) and Qian (1998) showed that  $\hat{c}_1$  is super consistent and that the estimators other than  $\hat{c}_1$  are asymptotically normal. These results hold when the conditional mean  $\mathbf{E}\{\mathbf{y}_t | \mathcal{F}_{t-1}\}$  is discontinuous at the threshold, see Tsay (1998) for discussion. Chan and Qian also showed that if the transition variable is selected from a set  $\mathcal{S} = \{s_t^{(1)}, \dots, s_t^{(D)}\}$  of stationary switch-variables, then minimising  $Q(c_1, \mathcal{S})$  even leads to consistent selection of  $s_t$ , provided  $\mathcal{S}$  contains the true switch-variable. The covariance matrix is then estimated as

$$\hat{\boldsymbol{\Sigma}} = n_1^{-1} \sum_{t \in \mathcal{T}_1} \boldsymbol{\varepsilon}_t(\hat{c}_1) \boldsymbol{\varepsilon}_t(\hat{c}_1)' + (T - n_1)^{-1} \sum_{t \in \mathcal{T}_2} \boldsymbol{\varepsilon}_t(\hat{c}_1) \boldsymbol{\varepsilon}_t(\hat{c}_1)'.$$

Seo (2011b) considers estimation of nonlinear error correction models. One of his objects of interest is the TVEC model with one cointegrating relationship such that the cointegrating vector is the transition variable denoted by  $s_t$  in (14). This was the case in the original TVEC model by Balke and Fomby (1997). The model contains a single threshold, but as Seo (2011b) points out, replacing  $s_t$  by  $|s_t|$ , a symmetric three-regime model (the band-TVEC model (15) with  $c_1 = -c_2$ ) is also included in his considerations. The interesting result is that under assumptions given in the paper, the least squares estimator of the cointegrating vector is  $T^{3/2}$ -consistent (the convergence is extremely rapid), and the estimator of the threshold parameter (there is one threshold if the transition variable is an absolute-valued) is super consistent. Seo (2011b) suggests that estimation of these parameters be carried out using a grid search; see also Hansen and Seo (2002).

VTAR models may also be estimated using heuristic (derivative-free) methods such as simulated annealing and genetic algorithms. These methods can be useful in estimating TVEC models, in which the threshold and cointegration parameters have to be estimated jointly for efficiency. This is

true in particular if the number of cointegrating vectors exceeds one. Hansen and Seo (2002) have solved this estimation problem by a grid search, because their cointegrating vector only contained one parameter and the model had just a single threshold. In higher-dimensional models, the grid search could be very slow. Yang, Tian and Yuan (2007) suggested a hybrid genetic-simulated annealing algorithm to this problem. El-Shagi (2011) compared various genetic algorithms applied to estimating TVEC models. For a good overview of heuristic methods, including genetic algorithms and simulated annealing, see Winker and Gilli (2004).

## 4.2 Estimation of vector smooth transition regression models

The single-transition VLSTAR model (16) can be estimated by nonlinear least squares. If the transition functions are not assumed to be the same, estimation can proceed equation by equation as in single-equation STR models; see for example Teräsvirta et al. (2010, Section 16.3). In the present section the focus is on the case where all equations have the same transition function. The model contains the parameters  $\theta = \{\Phi_1, \dots, \Phi_p; \Psi_1, \dots, \Psi_p; \Gamma, \Xi; \gamma, c\}$ . Set

$$\mathbf{B} = (\boldsymbol{\mu}_0, \Phi_1, \dots, \Phi_p, \boldsymbol{\mu}_1, \Psi_1, \dots, \Psi_p, \Gamma, \Xi)$$

and

$$\mathbf{z}_t(\gamma, c) = (\mathbf{1}', \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, G(\gamma, c; s_t)\mathbf{1}', G(\gamma, c; s_t)\mathbf{y}'_{t-1}, \dots, G(\gamma, c; s_t)\mathbf{y}'_{t-p}, \mathbf{x}'_t, G(\gamma, c; s_t)\mathbf{x}'_t)'$$

The NLS estimator of  $\theta$  is obtained by solving the optimization problem

$$\hat{\theta}_{NLS} = \arg \min_{\theta} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{Bz}_t(\gamma, c))' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \mathbf{Bz}_t(\gamma, c)) \quad (25)$$

assuming  $\boldsymbol{\Sigma} = \mathbf{I}$ . When  $\gamma$  and  $c$  are known the solution for  $\mathbf{B}$  is analytic. This offers a key to simplifying the nonlinear optimisation problem. In general, finding the optimum may be numerically demanding in the sense that convergence may be slow and the algorithm may converge to some local minimum. This is due to the shape of the sum of squares function in (25): the function can be rather 'flat' in some directions and possess many local optima.

For this reason, finding a suitable starting-value of  $\theta$  for the algorithm is crucial. The search to find one can be carried out by means of a grid search

similar to the one in the previous section. The basic idea is to construct a discrete grid in the parameter space of  $\gamma$  and  $c$  and estimate the parameters in  $\mathbf{B}$  conditionally on each pair  $(\gamma, c)$  in the grid. Choosing the pair  $(\gamma, c)$  which produces the smallest residuals sum of squares yields a starting-value for the nonlinear optimization. More specifically, the corresponding average residual sum of squares function is concentrated with respect to  $\mathbf{B}$ . Therefore the optimization is only performed with respect to  $\gamma$  and  $c$ .

The same procedure is employed in solving the nonlinear optimisation problem (25). After estimating  $\mathbf{B}$  conditionally on  $\gamma$  and  $c$ , one retains the estimated  $\mathbf{B}$  and solves the two-dimensional nonlinear estimation problem to obtain new values for  $\gamma$  and  $c$ . The matrix  $\mathbf{B}$  is then re-estimated conditionally on these new values of  $\gamma$  and  $c$ , and this is continued to convergence. Dividing each iteration into these two components reduces the dimension of the nonlinear estimation problem and thus saves computation time. If the grid is dense, the initial step-length of any derivative-based optimisation algorithm must be sufficiently short so that optimisation with a high probability leads to the local maximum closest to the value found using the grid approach and not to a lower local maximum further away.

The covariance matrix  $\Sigma$  can be estimated from the residual matrices

$$\widehat{\mathbf{E}}'\widehat{\mathbf{E}} = \sum_{t=p+1}^T \left( \mathbf{y}_t - \widehat{\mathbf{B}}\mathbf{z}_t(\widehat{\gamma}, \widehat{c}) \right) \left( \mathbf{y}_t - \widehat{\mathbf{B}}\mathbf{z}_t(\widehat{\gamma}, \widehat{c}) \right)'$$

as  $\widehat{\Sigma} = (T - p)^{-1}\widehat{\mathbf{E}}'\widehat{\mathbf{E}}$ . If a more efficient estimator is desired, one has to iterate the solution (25) by plugging in  $\widehat{\Sigma}^{-1}$  and applying generalised least squares to solve the optimisation problem. In practice, the estimates of  $\mathbf{B}$ ,  $\gamma$  and  $c$  are not likely to change much by these iterations, however.

It is also possible to apply heuristic methods such as simulated annealing or genetic algorithms mentioned in Section 4.1 to solving the optimisation problem (25). These methods may also be used for obtaining starting-values for an appropriate derivative-based algorithm. There is not yet much empirical evidence available of how well they perform in LVSTAR models. Auerbach and Gorodnichenko (2012) estimated a three-dimensional LVSTAR model with a single transition function with  $c = 0$  by the Monte Carlo Markov Chain method. A complication is that the error covariance matrix is assumed time-varying and controlled by the same transition function as the conditional mean. In our notation:

$$\Sigma_t = \Sigma_0 + G(\gamma, c; s_t)\Sigma_1$$

where  $\Sigma_0$  and  $\Sigma_0 + \Sigma_1 > 0$ . The authors simply assume this form for the covariance matrix and do not test the hypothesis  $\Sigma_t \equiv \Sigma$ .

## 5 Building nonlinear vector models

### 5.1 General considerations

In a few modelling situations, economic theory may offer guidance concerning the form of the relationship the researcher wants to use for characterising dynamic behaviour of a number of economic variables. For example, given the model classes considered in this chapter, an appropriate economic or other theory may suggest thresholds rather than smooth transitions, or the other way round. It may also be clear already at the outset what the switch or transition variable should be.

Nevertheless, there will almost always be choices to be made before settling for a final specification. The choice between linearity and nonlinearity is a very important one. It is possible that the linear model adequately describes the relationship considered, in which case a nonlinear model is not needed. In fact, the first step of a proper modelling strategy is to test linearity. This and other specification issues will be discussed below.

After the model has been specified, it has to be estimated, but that is not enough. Before an estimated model can be put to use, it has to be evaluated. The purpose of the evaluation is to find out whether the assumptions made when estimating the model appear satisfied, given the estimates. A battery of misspecification tests are available for this purpose, and some of them are discussed in Section 5.3.2. A complete modelling cycle thus consists of specifying the model, estimating its parameters and evaluating the estimated model. An early example is the cycle Box and Jenkins (1970) constructed for building linear ARMA or ARIMA models. Tsay (1989) considered specification and estimation of univariate TAR models. A modelling technique for STAR models was designed in Teräsvirta (1994) and Eitrheim and Teräsvirta (1996), and an analogous cycle for STR models discussed in Teräsvirta (1998).

### 5.2 Building stationary vector switching regression models

#### 5.2.1 Testing linearity

It is not advisable to fit a VSR model to any set of stationary time series without first testing linearity. An obvious practical reason is that the linear VR model is easier to work with than a VSR model, so linearity should be checked first. There is also an important statistical reason: the VSR model is not identified when the data-generating process is linear. If the

model contains only one regime, the time series are not informative about the switch-point which therefore cannot be estimated consistently. This lack of identification invalidates the standard testing approach: the asymptotic null distribution of the likelihood ratio statistic is unknown and does not have an analytic expression. This difficulty and how to deal with it was first discussed by Davies (1977). For a recent overview, see Teräsvirta et al. (2010, Chapter 5).

In VSR models, this identification problem can be handled in different ways. Following Tsay (1998), let

$$\mathbf{y}_t = \Theta \mathbf{z}_t + \boldsymbol{\varepsilon}_t \quad (26)$$

where  $\Theta = (\boldsymbol{\mu} \ \Phi_1 \ \dots \ \Phi_p \ \Gamma)$  and  $\mathbf{z}_t = (\mathbf{1}', \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, \mathbf{x}'_t)'$  be a linear VR model (only one regime in (11)) with independent errors. In order to test this VR model against a VSR model with one threshold and the threshold or switch variable  $s_t$ , arrange the observations in the ascending order according to the values of  $s_t$ . Since the error vectors  $\boldsymbol{\varepsilon}_t$  are independent, this can be done without affecting the dynamic properties of the model. Let the arranged model be

$$\mathbf{y}_{t(i)} = \Theta \mathbf{z}_{t(i)} + \boldsymbol{\varepsilon}_{t(i)}, \quad i = 1, \dots, T. \quad (27)$$

Next, estimate  $\Theta$  from (27) using the first  $K$  arranged observations, call the least squares estimator  $\tilde{\Theta}^{(K)}$ , and compute the residuals

$$\tilde{\boldsymbol{\varepsilon}}_{t(i)} = \mathbf{y}_{t(i)} - \tilde{\Theta}^{(K)} \mathbf{z}_{t(i)}, \quad i = 1, \dots, K$$

and the predictive residuals

$$\tilde{\boldsymbol{\eta}}_{t(i)} = \tilde{\boldsymbol{\varepsilon}}_{t(i)} (1 + \mathbf{z}'_{t(i)} (\mathbf{V}^{(K)})^{-1} \mathbf{z}_{t(i)})^{-1}, \quad i = 1, \dots, K$$

where  $\mathbf{V}^{(K)} = (1/K) \sum_{i=1}^K \mathbf{z}_{t(i)} \mathbf{z}'_{t(i)}$ . Do this for  $K = T_0 + 1, \dots, T$ , where  $T_0 \geq \dim(\mathbf{z}_t)$  and form the regression

$$\tilde{\boldsymbol{\eta}}_{t(i)} = \boldsymbol{\Psi} \mathbf{z}_{t(i)} + \mathbf{v}_{t(i)}, \quad i = T_0 + 1, \dots, T.$$

If (26) is the true model, the predictive residuals are asymptotically vector white noise and uncorrelated with  $\mathbf{z}_{t(i)}$ . Testing  $H_0: \boldsymbol{\Psi} = \mathbf{0}$  against  $H_1: \boldsymbol{\Psi} \neq \mathbf{0}$  thus reveals misspecification in (26). Under regularity conditions, the test statistic, based on the residuals  $\tilde{\boldsymbol{\eta}}_{t(i)}$  and  $\tilde{\mathbf{v}}_{t(i)}$  and given in Tsay (1998), has an asymptotic  $\chi^2$ -distribution under  $H_0$ .

This test does not explicitly have the VSR model as the alternative hypothesis. Another possibility would be to generalise the approach of Hansen (1996) to VSR models. The test would be carried out as follows:

1. Estimate the VR (or VAR) model (26) and a VSR (or a VTAR) model with a single threshold (the threshold variable is assumed known). Test the null hypothesis that (26) is the true model. For example, this can be done using a Wald statistic for testing the hypothesis that the coefficients in the two regimes are identical. The problem is that the resulting statistic is a function of the unknown nuisance parameter  $c$  (the threshold) and thus not operational. To solve the problem, carry out the test for a set of values of  $c \in \mathcal{C}$ . Typically,  $\mathcal{C}$  contains the values between the  $\beta$  and  $1 - \beta$  quantiles of the empirical distribution of the assumed threshold variable. Often  $\beta = 0.15$ . Calculate the maximum or the average value of these statistics over  $c \in \mathcal{C}$ . This yields two nuisance parameter free statistics, call them  $\sup W(c)$  and  $\text{ave}W(c)$ . The null distribution of these statistics is obtained numerically as follows:
2. Generate  $T$  observations from the estimated VR model either by simulation (drawing the error vectors from a distribution) or by a bootstrap using the estimated residuals, estimate a VR model and a VSR model from these observations and calculate the value of  $\sup W(c)$  or  $\text{ave}W(c)$  as in the previous step.
3. Repeat the second step  $N$  times. Arrange the  $N$  values of the test statistic obtained this way to form its empirical cumulative distribution function.
4. Compare the original value of the statistic with the distribution. If it exceeds the  $(1 - \alpha)$ th quantile of the distribution (large values are critical), reject the null hypothesis at the significance level  $\alpha$ .

Balke (2000) and Atanasova (2003) may be mentioned as examples of the use of this method. They applied the Wald test but the likelihood ratio test would also be a possibility.

A third, and so far untested, possibility would be to generalise the approach of Strikholm and Teräsvirta (2006) to vector models. This would require testing the linear VR model against a vector STR model by a test described in Section 5.3.1. Since the alternative hypothesis only contains the VSR model as a special case, this method must have lower power than that of Hansen (1996). Its main advantage would be computational simplicity.

Specification of lag lengths, exogenous variables and the switch- or threshold variable requires estimation. This has already been discussed in Section 4.1 in the case where  $q = 2$ . Generalising this to a larger number of regimes is obvious. However, the single-threshold model seems to be the most popular one in applications to stationary time series.

### 5.2.2 Testing for threshold cointegration

Cointegration is a major issue in the analysis of vectors of nonstationary time series. Consequently, testing the null hypothesis of no cointegration is an important problem in building TVEC models presented in Section 3.2.3. There exist tests testing linear against threshold cointegration and testing the null of no cointegration against threshold cointegration. The early tests assume a two-regime (single threshold) model as the alternative. Hansen and Seo (2002) developed a supremum LM-test of testing a VAR with linear cointegration against a TVEC model in the case where both the null model and the alternative contain a single cointegrating vector and the cointegrating vector (the error correction term) is the threshold variable. The LM approach makes it possible to use the cointegrating vector estimated under the null hypothesis in the test. Seo (2006) constructed a supremum Wald test of no cointegration against a TVEC model in this framework. He also showed how to approximate the small-sample distribution of the test using a residual-based bootstrap in order to reduce the size distortion present in the asymptotic test.

Gonzalo and Pitarakis (2006) generalised the approach of Hansen and Seo (2002) for testing linear against threshold cointegration such that the cointegration rank is not restricted to one. Furthermore, the threshold variable can be other than the error correction term. It has to be strictly stationary and ergodic, however. An additional restriction is that the threshold variable is what the authors call 'external': it is independent of the iid error vector at all lags and leads. Under these and other assumptions Gonzalo and Pitarakis (2006) derived the limiting null distribution of their supremum Wald statistic that they tabulated for some values of  $m$ . Recently, Krishnakumar and Neto (2012) constructed a supremum Wald test of no cointegration against the band-TVEC model. In their case, the threshold variable is assumed stationary and ergodic with a continuous and symmetric distribution. A set of critical values for the test can be found in the paper.

Finally, it may be mentioned that the tests are derived for alternatives in which the short-run coefficients of the model do not switch and the error covariance matrices remain constant. These models are therefore somewhat less general than the ones defined by equations (14) and (15) in Section 3.2.3.

### 5.2.3 Evaluation

As already mentioned, the estimated VSR model has to be evaluated before putting it into use. This implies testing the model by appropriate misspecification tests. Tsay (1998) mentions testing for serial correlation



and conditional heteroskedasticity, without providing any details. Testing is straightforward in cases where the objective function is discontinuous at the threshold, because the estimator of the threshold is super or  $T$ -consistent instead of being  $\sqrt{T}$ -consistent like the other estimators. This does not, however, concern tests of the switch parameter. The existing tests relate to single-equation models; see Hansen (1997, 2000) and Section 6.4 for discussion. The faster convergence rate means that in testing the threshold may be assumed known, which in turn implies that the standard test of no error autocorrelation for linear VAR or VR models is available in this case as well.

Other available tests include tests of constancy of the error covariance matrix. The test by Eklund and Teräsvirta (2007) makes use of the decomposition of the conditional covariance matrix by Bollerslev (1990) and contains several parametric alternatives to the null hypothesis. There is also a recent test by Yang (2013) that relies on the spectral decomposition of the covariance matrix. Under the alternatives the eigenvalues of the covariance matrix vary over time. These tests, however, require the errors to be independent and normal.

As Strikholm and Teräsvirta (2006) pointed out, super consistency of the threshold parameter estimator also allows using the linearity test against STAR to select the number of regimes in the SR model. This generalises to VSR models, although at this moment small sample properties of this technique are not known. It may be expected, however, that the LM-type test discussed in Section 5.3.1 is size-distorted in small samples and has to be modified to correct the bias.

One can also check stability of the estimated AR model. In the absence of analytical stability results, this may be done by simulation. One then switches off the noise and generates observations from the estimated model using different initial values or 'histories'. If the generated realisations converge to the same point, the conclusion is that the simulations do not contradict the stability assumption. They cannot confirm it either, but then a single realisation converging to a different point suffices to invalidate this assumption.

## **5.3 Building stationary vector smooth transition regression models**

### **5.3.1 Specification and estimation**

Camacho (2004) was the first to develop a modelling strategy for stationary vector STR models. He considered bivariate single-transition models. Teräsvirta and Yang (2013) generalised this to multitransition multivari-

ate LVSTR models. As already discussed, the modelling strategy has three stages: specification, estimation, and evaluation. Specification involves testing linearity and, if the null hypothesis is rejected, finding the number of transitions. Furthermore, possibilities to reduce the size of the model, i.e., imposing parameter restrictions, are considered, which already requires estimation.

**Testing linearity** As already mentioned, the first step of the specification stage is to test linearity of the model against VLSTAR. The arguments for doing so are the same as in the VSR case. The linear VR model is easier to handle than the VLSTAR one, and the latter is not identified when the data-generating process is linear.

The way linearity is tested depends on how the model is defined. If the model is (16), linearity can be tested equation by equation. Lagrange multiplier tests are preferred, because they only require estimating the model under the linearity hypothesis. A computationally easy way to perform the test is to approximate the transition function by a Taylor series expansion around the null hypothesis  $\gamma_j = 0$  (the  $j$ th equation) and base the test on this approximation, as Luukkonen, Saikkonen and Teräsvirta (1988) suggested. This LM-type test has an asymptotic  $\chi^2$ -distribution under the null hypothesis. Alternatively, one may construct the empirical distribution of the test statistic as in Hansen (1996). Doing so is computationally more demanding than performing the Taylor series alternative, but the advantage is that the test has higher power than the corresponding LM-type test if the true alternative is indeed a STAR equation. See Teräsvirta et al. (2010, Chapter 5) for discussion.

If it is assumed that the transition function is the same for the whole system, one can carry out a systemwise test as discussed in Yang and Teräsvirta (2013). Consider the  $m$ -dimensional two-regime logistic VSTR model

$$\mathbf{y}_t = \mathbf{B}'_1 \mathbf{z}_t + \mathbf{G}_t \mathbf{B}'_2 \mathbf{z}_t + \varepsilon_t \quad (28)$$

where  $\mathbf{B}'_1 = (\boldsymbol{\mu}_0 \ \boldsymbol{\Phi}_1 \ \dots \ \boldsymbol{\Phi}_p \ \boldsymbol{\Gamma})$ ,  $\mathbf{B}'_2 = (\boldsymbol{\mu}_1 \ \boldsymbol{\Psi}_1 \ \dots \ \boldsymbol{\Psi}_p \ \boldsymbol{\Xi})$ ,  $\mathbf{z}_t = (\mathbf{1}', \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p}, \mathbf{x}'_t)'$ , the sequence  $\{\varepsilon_t\} \sim \text{iid}\mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$ , and  $\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) = G(\boldsymbol{\gamma}, \mathbf{c}; s_t) \mathbf{I}_m$ . The null hypothesis of linearity can be written as  $H_0: \boldsymbol{\gamma} = \mathbf{0}$ . When the null is true,  $\mathbf{G}_t \equiv (1/2) \mathbf{I}_m$  and (28) becomes linear. Neither the location parameter  $c$  in the logistic function nor the parameters in the linear combination  $\mathbf{B}_1 + (1/2)\mathbf{B}_2$  are identified. The alternative hypothesis  $H_1: \boldsymbol{\gamma} > \mathbf{0}$ . Using a third-order Taylor expansion approximation of the logistic transition function  $G(\boldsymbol{\gamma}, \mathbf{c}; s_t)$  around  $\boldsymbol{\gamma} = \mathbf{0}$  helps circumvent the identification problem and

yields

$$\begin{aligned}\mathbf{G}_t &= \text{diag}\{a_{01} + a_{11}(\gamma)s_t + a_{21}(\gamma^2)s_t^2 + a_{31}(\gamma^3)s_t^3 + r_{1t}, \dots, a_{0m} + a_{1m}(\gamma)s_t \\ &\quad + a_{2m}(\gamma^2)s_t^2 + a_{3m}(\gamma^3)s_t^3 + r_{mt}\} \\ &= \mathbf{A}_0 + \mathbf{A}_1s_t + \mathbf{A}_2s_t^2 + \mathbf{A}_3s_t^3 + \mathbf{R}_t\end{aligned}\quad (29)$$

where  $\mathbf{A}_0 = \text{diag}(a_{01}, \dots, a_{0m})$ ,  $\mathbf{A}_j = \text{diag}(a_{j1}(\gamma^j), \dots, a_{jm}(\gamma^j))$ ,  $j = 1, 2, 3$ , are functions of  $\gamma$  such that  $\gamma = 0$  implies  $a_{ji}(\gamma^j) = 0$ , and  $\mathbf{R}_t = \text{diag}(r_{1t}, \dots, r_{mt})$ . Inserting (29) into (28) gives

$$\begin{aligned}\mathbf{y}_t &= \mathbf{B}'_1\mathbf{z}_t + (\mathbf{A}_0 + \mathbf{A}_1s_t + \mathbf{A}_2s_t^2 + \mathbf{A}_3s_t^3 + \mathbf{R}_t)\mathbf{B}'_2\mathbf{z}_t + \boldsymbol{\varepsilon}_t \\ &= (\mathbf{B}'_1 + \mathbf{A}_0\mathbf{B}'_2)\mathbf{z}_t + \mathbf{A}_1\mathbf{B}'_2\mathbf{z}_ts_t + \mathbf{A}_2\mathbf{B}'_2\mathbf{z}_ts_t^2 + \mathbf{A}_3\mathbf{B}'_2\mathbf{z}_ts_t^3 + \mathbf{R}_t\mathbf{B}'_2\mathbf{z}_t + \boldsymbol{\varepsilon}_t \\ &= \boldsymbol{\Theta}'_0\mathbf{z}_t + \boldsymbol{\Theta}'_1\mathbf{z}_ts_t + \boldsymbol{\Theta}'_2\mathbf{z}_ts_t^2 + \boldsymbol{\Theta}'_3\mathbf{z}_ts_t^3 + \boldsymbol{\varepsilon}_t^*\end{aligned}\quad (30)$$

where  $\boldsymbol{\Theta}_0 = \mathbf{B}_1 + \mathbf{B}_2\mathbf{A}_0$ ,  $\boldsymbol{\Theta}_j = \mathbf{B}_2\mathbf{A}_j$ ,  $j = 1, 2, 3$ , and  $\boldsymbol{\varepsilon}_t^* = \mathbf{R}_t\mathbf{B}'_2\mathbf{z}_t + \boldsymbol{\varepsilon}_t$ . The null hypothesis implies  $\mathbf{A}_0 = (1/2)\mathbf{I}_m$ ,  $\mathbf{A}_j = \mathbf{0}$ ,  $j = 1, 2, 3$ , and  $\mathbf{R}_t = \mathbf{0}$  in (30).

Thus, under the null hypothesis, the model (28) is linear, with  $\boldsymbol{\Theta}_0 = \mathbf{B}_1 + (1/2)\mathbf{B}_2$  and  $\boldsymbol{\Theta}_j = \mathbf{0}$ ,  $j = 1, 2, 3$ . The linearity test is therefore equivalent to the test of  $H_0 : \boldsymbol{\Theta}_j = \mathbf{0}$ ,  $j = 1, 2, 3$ , in (30). Moreover, due to the fact that  $\mathbf{R}_t = \mathbf{0}$  under the null hypothesis, the error term  $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t$  when  $H_0$  holds. Since the Lagrange multiplier test only requires estimating the model under the null hypothesis, the remainder term does not affect the normality of the errors or the standard asymptotic inference.

The corresponding Lagrange multiplier test under the null is derived from the score

$$\frac{\partial \log L(\tilde{\theta})}{\partial \boldsymbol{\Theta}_1} = \sum_{t=1}^T (\mathbf{z}_ts_t \ \mathbf{z}_ts_t^2 \ \mathbf{z}_ts_t^3) (\mathbf{y}_t - \tilde{\mathbf{B}}'_1\mathbf{z}_t)' \tilde{\boldsymbol{\Omega}}^{-1} = \mathbf{Z}'_3 (\mathbf{Y} - \mathbf{Z}_0\tilde{\mathbf{B}}_1) \tilde{\boldsymbol{\Omega}}^{-1}\quad (31)$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_T \end{bmatrix}, \mathbf{Z}_0 = \begin{bmatrix} \mathbf{z}'_1 \\ \mathbf{z}'_2 \\ \vdots \\ \mathbf{z}'_T \end{bmatrix}, \mathbf{Z}_3 = \begin{bmatrix} \mathbf{z}'_1s_1 & \mathbf{z}'_1s_1^2 & \mathbf{z}'_1s_1^3 \\ \mathbf{z}'_2s_2 & \mathbf{z}'_2s_2^2 & \mathbf{z}'_2s_2^3 \\ \vdots & \vdots & \vdots \\ \mathbf{z}'_Ts_T & \mathbf{z}'_Ts_T^2 & \mathbf{z}'_Ts_T^3 \end{bmatrix}$$

and  $\tilde{\mathbf{B}}_1$  and  $\tilde{\boldsymbol{\Omega}}$  are estimates from the restricted model. Under regularity conditions, the score converges to a matricvariate normal distribution with zero mean and variance  $\mathbf{Z}'_3(\mathbf{I}_T - \mathbf{P}_Z)\mathbf{Z}_3 \otimes \boldsymbol{\Omega}^{-1}$ , where  $\mathbf{P}_Z \equiv \mathbf{Z}_0(\mathbf{Z}'_0\mathbf{Z}_0)^{-1}\mathbf{Z}'_0$  is the projection matrix.

Luukkonen et al. (1988) pointed out that the third-order Taylor expansion is strictly needed only when the only nonlinear term in the equation is

the intercept, i.e.,  $\mathbf{B}_2 = (\boldsymbol{\beta}_{21} \mathbf{0})$  and  $s_t$  equals one of the remaining stochastic elements in  $\mathbf{z}_t$ . In that case the test based on the first-order Taylor expansion has only trivial power against the alternative. The second-order Taylor expansion does not remedy this problem, as  $a_{2j}(\gamma^2) = 0$  whenever  $c_j = 0$ .

The test can be performed in two vector regressions as follows:

1. Estimate the restricted model: regress  $\mathbf{Y}$  on  $\mathbf{Z}_0$ . Collect the residuals  $\tilde{\mathbf{E}} = (\mathbf{I}_T - \mathbf{P}_Z)\mathbf{Y}$ , and compute the matrix residual sum of squares  $\mathbf{RSS}_0 = \tilde{\mathbf{E}}'\tilde{\mathbf{E}}$ .
2. Run an auxiliary regression of  $\tilde{\mathbf{E}}$  on  $(\mathbf{Z}_0, \mathbf{Z}_3)$ . Compute the residuals  $\hat{\mathbf{E}}$  and the matrix residual sum of squares  $\mathbf{RSS}_1 = \hat{\mathbf{E}}'\hat{\mathbf{E}}$ .
3. Compute the test statistic

$$\begin{aligned} LM &= T \operatorname{tr}\{\mathbf{RSS}_0^{-1}(\mathbf{RSS}_0 - \mathbf{RSS}_1)\} \\ &= T(m - \operatorname{tr}\{\mathbf{RSS}_0^{-1}\mathbf{RSS}_1\}). \end{aligned} \quad (32)$$

The joint test statistic defined in (31) and (32) collapses into the univariate LM-type linearity test statistic when  $m = 1$ . This joint test can also be applied to any subset of equations in the system, for instance, to check whether some equations in the system are nonlinear with a common transition variable. It is also possible to test subhypotheses. For example, it may be assumed that the exogenous variables  $\mathbf{x}_t$  only enter (28) linearly, i.e.,  $\boldsymbol{\Xi} = \mathbf{0}$ .

In small samples, the empirical size of this LM-type test is sensitive to the dimension of the model. The size distortion increases with  $m$ . If the errors are assumed independent and identically distributed, Rao's F-test (Rao, 1973, Section 8c.5) effectively corrects the size. If the errors contain higher-order dependence, the problem can be solved by applying the wild bootstrap. These alternatives are discussed in Yang and Teräsvirta (2013).

It should be noted, however, that a rejection of the null hypothesis does not automatically imply that all  $m$  equations of the alternative VLSTAR model have the same transition function. This is because the same approximation (30) can also be obtained by assuming that all equations have the same transition variable but not necessarily the same transition function.

Harvill and Ray (1999) generalised the univariate linearity test of Tsay (1986) to vector series. The test is based on the following model:

$$\mathbf{y}_t = \sum_{j=1}^p \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\Psi} \operatorname{vech}(\mathbf{z}_t \mathbf{z}_t') + \boldsymbol{\varepsilon}_t$$

where  $\mathbf{z}_t = (\mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$  is an  $mp \times 1$  vector,  $\Psi$  is an  $m \times mp(mp+1)/2$  parameter matrix, and  $\varepsilon_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \Sigma)$ . The null hypothesis is  $\Psi = \mathbf{0}$ . It is tested using Wilks's  $\Lambda$ :

$$\Lambda = \frac{|\mathbf{RSS}_0|}{|\mathbf{RSS}_0 + \mathbf{RSS}_1|} \quad (33)$$

where, as previously,  $\mathbf{RSS}_0$  is the residual sum of squares matrix from estimating the null model, and  $\mathbf{RSS}_1$  is obtained by regressing the residuals from that regression on  $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$  and  $\text{vech}(\mathbf{z}_t \mathbf{z}'_t)$ . Statistic (33) has an approximate F-distribution under  $H_0$ , see Harvill and Ray (1999) for details. One could also use Rao's F-statistic instead of (33). Simulations in Yang and Teräsvirta (2013) suggest that although Wilks's  $\Lambda$  is less size-distorted in small samples than most other statistics, Rao's F-test is the best performer.

Weise (1999) has yet another solution. Take a linear VAR model ( $\Xi = \mathbf{0}$  in (16) but a generalisation is possible) and test linearity equation by equation as in Teräsvirta (1998) using the Taylor expansion approach and the same transition variable. Let  $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{1t}, \dots, \tilde{\varepsilon}_{mt})'$  be the residual vector of the estimated linear VAR model and  $\tilde{\eta}_t = (\tilde{\eta}_{1t}, \dots, \tilde{\eta}_{mt})'$  the corresponding vector from the  $m$  auxiliary regressions. Form the covariance matrices  $\tilde{\Sigma}_\varepsilon = (T-p)^{-1} \sum_{j=p+1}^T \tilde{\varepsilon}_t \tilde{\varepsilon}'_t$  and  $\tilde{\Sigma}_\eta = (T-p)^{-1} \sum_{j=p+1}^T \tilde{\eta}_t \tilde{\eta}'_t$  and the likelihood ratio-type test statistic

$$LR = T(\ln |\tilde{\Sigma}_\varepsilon| - \ln |\tilde{\Sigma}_\eta|)$$

that has an asymptotic  $\chi^2$ -distribution with  $pm^2$  degrees of freedom when the linearity hypothesis holds. This approach provides results both for individual equations and for the whole system. Weise (1999) simulates the statistic and finds its small-sample behaviour satisfactory. For an application, see Section 9.2.

**Finding the number of transitions** When linearity is rejected, the VL-STAR model with a single transition is estimated. The next step is to test one transition against two. This is either done equation by equation or for the whole system, depending on the situation. In the former case, the relevant single-equation test of no additive nonlinearity is described in Teräsvirta (1998). If it is assumed that one transition function controls the whole system, the same test is designed for the whole system, see Yang and Teräsvirta (2013).

**Specifying the lag structure** Linearity testing is, when the number of observations allows it, preferably carried out with a null model that is free of error autocorrelation, because it may distort the empirical size of the test.

It is then desirable to find and impose restrictions on the parameters of the VLSTAR model. There are three types of inraequation restrictions. As an example, consider the parameter matrices  $\Phi_j = [\phi_{ij}]$  and  $\Psi_j = [\psi_{ij}]$  in (16). First, the  $j$ th lag  $y_{i,t-j}$  may only enter linearly:  $\psi_{ij} = 0$ . Second, this lag may be completely redundant:  $\phi_{ij} = \psi_{ij} = 0$ . Third, the effect of  $y_{i,t-j}$  is nonlinear and zero at low values of  $s_t$ :  $\phi_{ij} = 0$ . Finally, the effect of  $y_{i,t-j}$  is nonlinear and approaches zero at high values of  $s_t$ :  $\phi_{ij} = -\psi_{ij}$ . Testing these restrictions and imposing them requires estimation of the VLSTAR model.

### 5.3.2 Evaluation

**Misspecification tests** Misspecification tests are the most important and useful tool for evaluating VLSTR models. Out-of-sample forecasting may also be used to select between different models. The available tests include the test of no serial correlation, no remaining nonlinearity, and the test of parameter constancy. For space reasons it is not possible to consider all these tests in detail. Suffice it so say that the estimated model can be tested equation by equation see Teräsvirta (1998), or as a system. As an example of the latter, the test of no error autocorrelation can be carried out by assuming that under the alternative, the error sequence  $\{\varepsilon_t\}$  in (16) has the following structure:

$$\varepsilon_t = \sum_{j=1}^q \Lambda_j \varepsilon_{t-j} + \eta_t$$

where  $\eta_t \sim \text{iid}\mathcal{N}(0, \Sigma_\eta)$ ; see Camacho (2004) or Yang and Teräsvirta (2013). Under the null hypothesis,  $\Lambda_j = \mathbf{0}$ ,  $j = 1, \dots, q$ . If the corresponding hypothesis is tested equation by equation, the contemporaneous cross-correlations between errors are ignored. The tests of no additive nonlinearity and parameter constancy can likewise be conducted either equation by equation or jointly for the whole system.

These tests have been discussed in the situation in which the equations contain a single transition, but they can be generalised to multitransition equations. As in the case of VTAR models, it is possible to test the constancy of the error covariance matrix against various parametric alternatives using the tests of Eklund and Teräsvirta (2007) and Yang (2013). The stability of a VLSTAR model may be checked numerically in the same way as that of the VTAR model.

### 5.3.3 Bayesian approach to building vector smooth transition autoregressive models

It is possible to construct VSTAR models using Bayesian techniques, as shown by Gefang and Strachan (2010). The authors consider a bivariate ( $m = 2$ ) variant of the logistic (16) with  $\mathbf{x} = \mathbf{0}$  and  $\mathbf{G}(\gamma, \mathbf{c}; \mathbf{s}_t) = G(\gamma, c; s_t)\mathbf{I}_2$ . Determination of prior distributions contains one detail worth mentioning. As  $\gamma = 0$  in the logistic transition function implies linearity and an unidentified LVSTAR model, this value is excluded from the support of the prior distribution of  $\gamma$ . For a general discussion of this identification problem in the Bayesian framework, see Bauwens, Lubrano and Richard (1999, pp. 241-242).

Specification of the model consists of estimating a large number of models with varying transition functions and lag lengths (these are not fixed in advance) and using Bayes factors to choose between models. Since linearity cannot be excluded a priori, the set of models compared also contains linear VAR models. Posterior distributions of the parameters are estimated using Gibbs sampling. The aim of the application in Gefang and Strachan (2010) is to study (pairwise) business cycle linkages between the UK, the US and Germany. They find that nonlinear VLSTAR models are superior to linear VARs in describing these relationships.

## 5.4 Inference in nonstationary vector smooth transition autoregressive models

The statistical inference discussed in the preceding section builds on the assumption that  $\{\mathbf{y}_t\}$  is stationary. The situation becomes more complex when that is no longer true. Inference in STVEC models is studied in Kristensen and Rahbek (2010, in press). In the former paper the authors consider maximum likelihood estimation of an  $m$ -dimensional nonlinear vector error correction model in which the adjustment is symmetric and the lags of  $\Delta\mathbf{y}_t$  enter linearly. The model is written as follows:

$$\Delta\mathbf{y}_t = \mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \boldsymbol{\theta}) + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j \Delta\mathbf{y}_{t-j} + \boldsymbol{\varepsilon}_t \quad (34)$$

where  $\{\boldsymbol{\varepsilon}_t\} \sim \text{iid}(\mathbf{0}, \boldsymbol{\Sigma})$ . The symmetry arises from the fact that

$$\mathbf{h}(\mathbf{B}'\mathbf{y}, \boldsymbol{\theta}) = \mathbf{A}_1\mathbf{B}'\mathbf{y} + o(\|\mathbf{B}'\mathbf{y}\|)$$

for some  $m \times r$  matrix  $\mathbf{A}_1$  as  $\|\mathbf{B}'\mathbf{y}\| \rightarrow \infty$ . The characteristic polynomial

$$C_1(z) = \mathbf{I}_m(1 - z) - \mathbf{A}_1\mathbf{B}'z - \sum_{k=1}^{p-1} \Phi_k(1 - z)z^k$$

has  $m - r$  unit roots, whereas  $|z| > 1$  for the remaining ones. The 'symmetric' STVEC model is obtained as a special case of (34) by defining

$$\mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \boldsymbol{\theta}) = \mathbf{A}_1\mathbf{B}' + \mathbf{A}_2\mathbf{B}'[1 + \exp\{(\mathbf{B}'\mathbf{y}_{t-1} - \boldsymbol{\omega})'\mathbf{C}(\mathbf{B}'\mathbf{y}_{t-1} - \boldsymbol{\omega})\}^{-1}] \quad (35)$$

where  $\mathbf{C}$  is a positive definite matrix, and  $\boldsymbol{\theta} = (\boldsymbol{\omega}', \text{vech}(\mathbf{C})', \text{vec}(\mathbf{A}_1)', \text{vec}(\mathbf{A}_2)')'$ . A single transition function controls nonlinearity in all equations, and the transition variable is a function of the  $r$  error-correcting vectors. Under regularity conditions for  $\mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \boldsymbol{\theta})$  that exclude VTAR models, and other conditions given in the paper, Kristensen and Rahbek (2010) derive consistency and the asymptotic distribution of the (quasi) maximum likelihood estimators of the parameters of (the normalized version of) (34). The short-run parameter vector estimators  $\hat{\boldsymbol{\theta}}$  and  $\text{vec}(\hat{\Phi}_k)$ ,  $k = 1, \dots, p - 1$ , are  $\sqrt{T}$ -consistent, whereas the estimator of (normalized)  $\mathbf{B}$  is super or  $T$ -consistent. The estimators have nonstandard distributions, so confidence intervals have to be constructed numerically.

Kristensen and Rahbek (in press) consider the same model as in Kristensen and Rahbek (2010) but allow what they call asymmetric adjustment: the transition function in (34) equals (Saikkonen, 2008)

$$G_t(\mathbf{y}_{t-1}, \boldsymbol{\theta}) = [1 + \exp\{\mathbf{s}'(\mathbf{B}'\mathbf{y}_{t-1} - \boldsymbol{\omega})\}]^{-1} \quad (36)$$

where  $\mathbf{s}$  is an  $r \times 1$  vector. In the former case (symmetric adjustment),  $\|\mathbf{B}'\mathbf{y}_{t-1} - \boldsymbol{\omega}\| \rightarrow 0$  as  $\|\mathbf{y}_{t-1}\| \rightarrow \infty$ , whereas in the latter, the nonlinear component will contribute to  $\Delta\mathbf{y}_t$  even when  $\|\mathbf{y}_{t-1}\| \rightarrow \infty$ . The authors derive tests for testing hypothesis about the short-run parameters  $\boldsymbol{\theta} = (\boldsymbol{\omega}', \mathbf{s}', \text{vec}(\mathbf{A}_1)', \text{vec}(\mathbf{A}_2)')'$  and  $\text{vec}(\Phi_j)$ ,  $j = 1, \dots, p$ , and the long-run parameters  $\text{vec}(\mathbf{B})$ . Note that the hypothesis  $\text{vec}(\mathbf{A}_2) = \mathbf{0}$  is a special case because it is a linearity hypothesis. As in the stationary case, testing it directly, that is, without approximating the alternative (see below), leads to nonstandard asymptotic inference because  $\boldsymbol{\omega}$  and  $\text{vech}(\mathbf{C})$  in (35) or  $\boldsymbol{\omega}$  and  $\mathbf{s}$  in (36) are not identified when the null hypothesis holds. Testing linear against nonlinear cointegration will be discussed next.

## 5.5 Testing for nonlinear cointegration in smooth transition error correction models

Kristensen and Rahbek (in press) discuss the standard solution to testing  $H_0: \text{vec}(\mathbf{A}_2) = \mathbf{0}$ ; see for example Hansen (1996). First define a set  $\Theta_1$  of values of



$\boldsymbol{\theta}_1 = (\boldsymbol{\omega}', \text{vech}(\mathbf{C})')'$  (symmetric model) or  $\boldsymbol{\theta}_1 = (\boldsymbol{\omega}', \mathbf{s}')'$  (asymmetric model) and compute the test statistic using these values in turn. These statistics are not operational because they are functions of  $\boldsymbol{\theta}_1$ . The proposed test statistic is the supremum statistic, for example the LM statistic  $LM_{\text{NL}}$  is defined by

$$LM_{\text{NL}} = \sup_{\boldsymbol{\theta}_1 \in \Theta_1} LM(\boldsymbol{\theta}_1)$$

where  $LM(\boldsymbol{\theta}_1)$  is an inoperational statistic. The empirical null distribution of  $LM_{\text{NL}}$  is computed by simulation as discussed in Section 5.2.1. The power of the test strongly depends on the number of values in  $\Theta_1$  and the selected values of  $\boldsymbol{\theta}_1$ .

Seo (2011a) also tests linear cointegration against nonlinear one. Furthermore, he presents a test of linearity of the short-run component. In testing linear against nonlinear cointegration, the model is

$$\Delta \mathbf{y}_t = \mathbf{A}\{\mathbf{B}'\mathbf{y}_{t-1} + \mathbf{g}(\mathbf{y}_{t-1})\} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j \Delta \mathbf{y}_{t-j} + \mathbf{u}_t \quad (37)$$

where  $\{\mathbf{u}_t\}$  is a martingale difference sequence with respect to  $\mathbf{y}_{t-j}$ ,  $j \geq 1$ , and no specific parametric structure is assumed for the  $r \times 1$  vector  $\mathbf{g}(\mathbf{y}_{t-1})$ . The null hypothesis is that  $\mathbf{g}(\mathbf{y}_{t-1}) = \mathbf{0}$ . When the null hypothesis concerns the short-run parameters, the alternative model is

$$\Delta \mathbf{y}_t = \mathbf{A}\mathbf{B}'\mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j \Delta \mathbf{y}_{t-j} + \mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \Delta \bar{\mathbf{y}}_{t-1}) + \mathbf{u}_t \quad (38)$$

where  $\Delta \bar{\mathbf{y}}_{t-1} = \text{vec}(\Delta \mathbf{y}_{t-1}, \dots, \Delta \mathbf{y}_{t-p+1})$ . The null hypothesis is that  $\mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \Delta \bar{\mathbf{y}}_{t-1}) = \mathbf{0}$ . The argument is that under the null hypothesis,  $\mathbf{u}_t$  is neither dependent on  $\mathbf{y}_{t-1}$  in (37), nor on  $\mathbf{B}'\mathbf{y}_{t-1}$  or  $\Delta \bar{\mathbf{y}}_{t-1}$  in (38). The idea of the test is to approximate the alternative, which is done by replacing the nonlinear component by a polynomial approximation. Thus in (37),

$$\mathbf{g}(\mathbf{y}_t) \approx \mathbf{Z}'_{1t} \boldsymbol{\delta}_1$$

where  $\mathbf{Z}_{1t}$  is an  $s_1 \times r$  matrix, whose elements are powers and cross-products of the elements of  $\mathbf{y}_t$ , and  $\boldsymbol{\delta}$  is an  $s_1 \times 1$  parameter vector. The null hypothesis is  $\boldsymbol{\delta}_1 = \mathbf{0}$ . Analogously, in (38), it is assumed that

$$\mathbf{h}(\mathbf{B}'\mathbf{y}_{t-1}, \Delta \bar{\mathbf{y}}_{t-1}) \approx \mathbf{Z}'_{2t} \boldsymbol{\delta}_2$$

where the  $s_2 \times p$  matrix  $\mathbf{Z}_{2t}$  consists of powers and cross-products of the elements of  $\mathbf{B}'\mathbf{y}_{t-1}$  and  $\Delta \bar{\mathbf{y}}_{t-1}$ , and the null hypothesis is  $\boldsymbol{\delta}_2 = \mathbf{0}$ . The resulting

tests have the advantage that assuming  $\mathbf{u}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$  the test statistics have a  $\chi^2$ -distribution under  $H_0$ ; for details see Seo (2011a). The test of  $\delta_2 = \mathbf{0}$  can be viewed as a generalisation of the univariate linearity test by Luukkonen et al. (1988) or the one in Teräsvirta, Lin and Granger (1993), see also Section 5.3.1.

Kapetanios, Shin and Snell (2006) test the null hypothesis of no cointegration against smooth transition cointegration in VAR models assuming that under the alternative the model contains exactly one cointegrating relationship. This implies that the  $\mathbf{A}_1, \mathbf{A}_2$ , and  $\mathbf{B}$  in (18) are  $m \times 1$  vectors. Compared to (18), the model of Kapetanios et al. (2006) does not contain intercepts, and the lags of  $\Delta \mathbf{w}_t$  enter the model linearly. Writing the (normalized) cointegrating relationship as

$$u_t = y_{1t} - \mathbf{B}'\mathbf{w}_t$$

where  $\mathbf{w}_t = (y_{2t}, \dots, y_{mt})'$ , the alternative can be expressed as the following transformed system:

$$\begin{aligned} \Delta y_{1t} &= \alpha_1 u_{t-1} + \sum_{j=1}^{p-1} \phi'_j \Delta \mathbf{y}_{t-j} + \phi'_0 \Delta \mathbf{w}_t \\ &+ \{ \alpha_2 u_{t-1} \exp\{-\gamma(u_{t-1} - c)^2\} + \sum_{j=1}^{p-1} \psi'_j \Delta \mathbf{y}_{t-j} + \varepsilon_{1t}, \gamma > 0 \\ \Delta \mathbf{w}_t &= \sum_{j=1}^{p-1} \Phi'_{2j} \Delta \mathbf{y}_{t-j} + \varepsilon_{2t} \end{aligned} \quad (39)$$

Several things are worth noticing in (39). First,  $\Delta \mathbf{w}_t$  is assumed exogenous to the parameters of interest in the equation of  $\Delta y_{1t}$ . Second, the nonlinearity is ESTAR or 'symmetric' type. Third, the transition variable is the lagged error correction term itself. Finally, nonlinearity is not extended to the short-run dynamic structure of the model.

Testing is carried out by first regressing  $y_{1t}$  on  $\mathbf{w}_t$  assuming that  $y_{2t}, \dots, y_{mt}$  are weakly exogenous to the parameters of interest in this equation. The residuals are  $\tilde{u}_t = y_{1t} - \tilde{\beta}' \mathbf{w}_t$ ,  $t = 1, \dots, T$ . Then, following Luukkonen et al. (1988),  $\exp\{-\gamma(u_{t-1} - c)^2\}$  is approximated with its first-order Taylor expansion at  $\gamma = 0$  and the auxiliary regression

$$\Delta y_{1t} = \delta_1 \tilde{u}_{t-1} + \delta_2 \tilde{u}_{t-1}^2 + \delta_3 \tilde{u}_{t-1}^3 + \sum_{j=1}^{p-1} \phi'_j \Delta \mathbf{y}_{t-j} + \phi'_0 \Delta \mathbf{w}_t + \varepsilon_{1t}^*$$

formed to test the null hypothesis  $\delta_1 = \delta_2 = \delta_3 = 0$ . The resulting tests are LM-type tests.

## 5.6 Bayesian approach to building smooth transition vector error correction models

Building smooth transition vector error correction models using Bayesian techniques is discussed in Gefang (2012). The purpose of the paper is to study Granger causality between money and output using monthly US time series from 1959(1) to 2006(12) using a four-dimensional STVEC model. The model can be written as follows:

$$\begin{aligned} \Delta \mathbf{y}_t = & \boldsymbol{\mu}_1 + \mathbf{A}_1 \mathbf{B}'_1 \mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Phi}_j \Delta \mathbf{y}_{t-j} \\ & + \{ \boldsymbol{\mu}_2 + \mathbf{A}_2 \mathbf{B}'_2 \mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \boldsymbol{\Psi}_j \Delta \mathbf{y}_{t-j} \} G(\gamma, c; s_t) + \boldsymbol{\varepsilon}_t, \quad \gamma > 0 \end{aligned} \quad (40)$$

where  $\mathbf{y}_t$  is a  $m \times 1$  vector of nonstationary variables such that  $\Delta \mathbf{y}_t$  is  $I(0)$ , and  $\boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ . The matrices  $\mathbf{A}_1$  and  $\mathbf{B}_1$  are both  $m \times r_1$ ,  $r_1 < m$ , whereas  $\mathbf{A}_2$  and  $\mathbf{B}_2$  are  $m \times r_2$ ,  $r_2 < m$ . The ranks  $r_1$  and  $r_2$  are not fixed in advance, and they are not assumed equal. This means that the extreme regimes corresponding to  $G(\gamma, c; s_t) = 0$  and  $G(\gamma, c; s_t) = 1$ , respectively, can contain a different number of cointegrating relationships. In the application,  $m = 4$ , so the ranks can be between one and three, which implies nine possible combinations for  $r_1$  and  $r_2$ . How to interpret the changing ranks (and the 'fractional' ranks when  $G(\gamma, c; s_t) \neq 0$ ) is not discussed. Models without cointegration in (40) are not considered separately, as the author argues that they are special cases of models with ranks between one and three,  $\mathbf{A}_1 = \mathbf{0}$  and  $\mathbf{A}_2 = \mathbf{0}$ . In that case, the ranks are not identified, however. The transition variable  $s_t$  is not predetermined, and the set of possible candidates is defined. Some of the stochastic ones used in the study are nonstationary, which means that the local dynamic structure of the corresponding model mostly changes in one direction and does not fluctuate between the two extreme regimes.

As in the stationary case in Section 5.3.3, the posterior distributions are obtained using Gibbs sampling. Model selection is carried out analogously to the stationary case using Bayes factors. The most frequently selected models support the assumption of nonlinear Granger causality from money to output. They are in fact ones in which  $r_1 \neq r_2$ .

## 6 Common nonlinearities

### 6.1 Definitions

In a system of nonlinear equations such as the VTAR or VLSTAR model in which nonlinearity is additive, common nonlinearities may occur. Anderson and Vahid (1998) were the first to consider this possibility. They discussed both the general case in which the nonlinear model (8) is additive as in (10), and the common nonlinearity in the VLSTAR and VTAR models. The concept of common nonlinearity is defined as follows. Consider the model (8) with (10) assuming that it contains  $m$  equations. Suppose now that there is an  $m \times r$  matrix  $\mathbf{M}$ ,  $r < m$ , such that

$$\mathbf{M}'\mathbf{f}_1(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}_1) = \mathbf{0}.$$

Consequently, nonlinearity in the relationship in (8) with (10) can be eliminated by defining a linear combination  $\mathbf{z}_t = \mathbf{M}'\mathbf{y}_t$  which removes the nonlinear component  $\mathbf{f}_1(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}; \mathbf{x}_t; \boldsymbol{\theta}_1)$  from the system. Using the terminology of Engle and Kozicki (1993), nonlinearity is a feature, and if it is eliminated by left-multiplying the equation (10) by  $\mathbf{M}'$ , it is a common feature.

Apply now this concept to the VLSTAR model (16). A necessary but not sufficient condition for common nonlinearity is that the transition function matrix

$$\mathbf{G}(\boldsymbol{\gamma}, \mathbf{c}; \mathbf{s}_t) = \text{diag}\{G_1(\gamma_1, \mathbf{c}_1, s_{1t}), \dots, G_m(\gamma_m, \mathbf{c}_m, s_{mt})\} = G(\boldsymbol{\gamma}, \mathbf{c}; s_t)\mathbf{I}_m \quad (41)$$

that is, that the transition is common to all equations in the model. In (16), define the  $m \times ((p+1)m + n)$  coefficient matrix  $\boldsymbol{\Pi} = (\boldsymbol{\mu}_1 \ \boldsymbol{\Psi}_1 \ \dots \ \boldsymbol{\Psi}_p \ \boldsymbol{\Xi})$ . For common nonlinearity it is required that  $\mathbf{M}'\boldsymbol{\Pi} = \mathbf{0}$ , that is, the  $m \times r$  matrix  $\mathbf{M}$ ,  $r < m$ ,  $\text{rk}(\mathbf{M}) = r$ , eliminates the nonlinear component from (16). This is a very strong condition if  $m$ ,  $n$ , and  $p$  are large, as it requires that  $\boldsymbol{\Pi}$  has a reduced rank. For this reason it is useful to define partial common nonlinearity such that the orthogonality condition only applies to a subset of columns of  $\boldsymbol{\Pi}$ . For example, consider the following  $m$ -dimensional first-order VLSTAR model

$$\mathbf{y}_t = \boldsymbol{\mu}_0 + \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + (\boldsymbol{\mu}_1 + \boldsymbol{\Psi}_1\mathbf{y}_{t-1})G(\boldsymbol{\gamma}, \mathbf{c}; s_t) + \boldsymbol{\varepsilon}_t \quad (42)$$

where  $\boldsymbol{\varepsilon}_t \sim \text{iid}\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ . The interest may now be focussed on finding out whether the model may be transformed into a model in which there are  $r < m$  linear combinations of the elements of  $\mathbf{y}_t$  defined by  $\mathbf{M}$  such that the resulting  $r$ -dimensional model is intercept-switching, i.e.,  $\mathbf{M}'\boldsymbol{\Psi}_1 = \mathbf{0}$  in (42).

The definition of common nonlinearity is similar to that of co-breaking. As defined in Hendry and Massmann (2007), the matrix  $\mathbf{M}$  is co-breaking of order  $q$  if left-multiplying a linear vector model with breaks with  $\mathbf{M}'$  eliminates the breaks. A similar definition can be developed for a vector TAR model with a common threshold. Consider the following two-regime generalisation of the VTAR model (11):

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\mu}_1 + \sum_{j=1}^p \boldsymbol{\Phi}_{1j} \mathbf{y}_{t-j} + \boldsymbol{\Gamma}_1 \mathbf{x}_t \\ &+ \mathbf{I}(\mathbf{s}_t > \mathbf{c}_1) (\boldsymbol{\mu}_2 + \sum_{j=1}^p \boldsymbol{\Phi}_{2j} \mathbf{y}_{t-j} + \boldsymbol{\Gamma}_2 \mathbf{x}_t) + \boldsymbol{\varepsilon}_t \end{aligned} \quad (43)$$

where the  $m \times m$  indicator matrix

$$\mathbf{I}(\mathbf{s}_t > \mathbf{c}_1) = \text{diag}(I(s_{1t} > c_1), \dots, I(s_{mt} > c_m)). \quad (44)$$

A necessary condition for a common threshold is

$$\mathbf{I}(\mathbf{s}_t > \mathbf{c}_1) = I(s_t > c_1) \mathbf{I}_m \quad (45)$$

that is,  $s_{1t} = \dots = s_{mt}$  and  $c_{11} = \dots = c_{1m}$  in (44). This may be completed by a condition analogous to the one for vector STAR or STR models. As already discussed, the condition (45) is satisfied in the original VTAR or VSR model (11) of Tsay (1998).

## 6.2 Co-shifting

Common nonlinearity in shifting-mean vector autoregressive models is called co-shifting. In this case, the transition functions have the same transition variable, and so a necessary condition for strong co-shifting is

$$\mathbf{G}_j(\boldsymbol{\gamma}_j, \mathbf{c}_j; t/T) = G(\boldsymbol{\gamma}_j, c_j; t/T) \mathbf{I}_m, \quad j = 1, \dots, q. \quad (46)$$

This requires that all equations of the model have the same number of shifts. If  $q$ , the number of shifts, is less than  $m$ , the number of equations, the condition (46) is also sufficient. If  $q \geq m$ , then the necessary condition has to be completed by the condition

$$\mathbf{M}'(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \dots, \boldsymbol{\delta}_q) = \mathbf{0}$$

where  $\mathbf{M}$  is again an  $m \times q$  matrix,  $\text{rk}(\mathbf{M}) = q$ . Partial co-shifting may be defined analogously to partial common nonlinearity. There may then

be a subset of diagonal matrices  $\mathbf{G}_j(\boldsymbol{\gamma}_j, \mathbf{c}_j; t/T)$  satisfying (46), and if their number  $q_1 < m$ , then this is sufficient for weak co-shifting. For an application of co-shifting see Holt and Teräsvirta (2012) who fit an VSM-AR model to a pair of annual hemispheric temperature series and discuss comovements in them.

### 6.3 Testing common nonlinearity in VSTR models

Anderson and Vahid (1998) consider testing common nonlinearity in a nonlinear model of additive type (10). In their notation,  $\mathbf{z}_t$  is the  $n \times 1$  vector of nonlinear regressors. It is implicitly assumed that the nonlinear model is linear in parameters. As an example, suppose that one is testing linearity against STAR equation by equation assuming that the transition variable is  $s_t$  using the technique of Luukkonen et al. (1988). Consider the  $i$ th equation of (42). Then, if  $s_t$  is not an element in  $\mathbf{z}_t$ , the auxiliary regression has the following form:

$$y_{it} = \boldsymbol{\theta}'_i \mathbf{z}_t + \boldsymbol{\theta}'_i \mathbf{z}_t s_t + \boldsymbol{\theta}'_i \mathbf{z}_t s_t^2 + \boldsymbol{\theta}'_i \mathbf{z}_t s_t^3 + \varepsilon_{it}^*.$$

and  $\mathbf{z}_t = (\mathbf{z}'_t s_t, \mathbf{z}'_t s_t^2, \mathbf{z}'_t s_t^3)'$  is a  $3(p+1+n) \times 1$  vector. Let  $\mathbf{y}_t^\dagger$  and  $\mathbf{z}_t^\dagger$  be the corresponding variables after the linear influence of these variables have been eliminated. Then, there is common nonlinearity of at least dimension one, if there exists an  $m \times 1$  vector  $\mathbf{m}_1$  such that

$$\mathbf{E}(\mathbf{m}'_1 \mathbf{y}_t^\dagger \otimes \mathbf{z}_t^\dagger) = 0.$$

This vector can be estimated using the generalised method of moments and the normalisation constraint  $\mathbf{m}'_1 \mathbf{Y}' \mathbf{Y} \mathbf{m}_1 = 1$ , where  $\mathbf{Y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_T)'$ . Anderson and Vahid (1998) show that the solution is the eigenvector of  $\mathbf{y}_t^\dagger$  that corresponds to the smallest squared canonical correlation  $\lambda_1^2$  between  $\mathbf{y}_t^\dagger$  and  $\mathbf{z}_t^\dagger$ . Under regularity conditions, the test statistic  $Q_T = T\lambda_1^2$  has a  $\chi^2$ -distribution with  $3(p+1+n) \times 1$  degrees of freedom when the null hypothesis of a single common nonlinearity vector is valid. If one tests the null hypothesis of  $q < m$  common nonlinearities, the statistic is  $Q_T = T \sum_{i=1}^q \lambda_i^2$ , where  $\lambda_i^2$ ,  $i = 1, \dots, q$ , are the  $q$  smallest squared canonical correlations between  $\mathbf{y}_t^\dagger$  and  $\mathbf{z}_t^\dagger$ .

For a full information maximum likelihood approach to this testing problem using techniques originally developed for cointegration analysis, see Hungnes (2012).

### 6.4 Testing common nonlinearity in VSR models

Anderson and Vahid (1998) base a test of common nonlinearity on the

arranged regression of Tsay (1998) considered in Section 5.2.1; see their paper for details. Testing the hypothesis (45) directly against the alternative that the equations have the same switch variable but different switch-points or thresholds would be difficult, because there is no statistical theory for testing the hypothesis  $c_{11} = \dots = c_{1m}$  in the VTAR model (43). There is, however, asymptotic distribution theory for testing the null hypothesis  $c_1 = c_1^0$  in a univariate two-regime TAR model,  $q = 2$  in (1). The likelihood ratio test statistic is

$$LR_T(c_1^0) = T \frac{Q(c_1^0) - Q(\hat{c}_1)}{Q(\hat{c}_1)}$$

where  $Q(c_1^0)$  is the sum of squared residuals from the model estimated under the null hypothesis, and  $Q(\hat{c}_1)$  is the same from the estimated alternative. Under a set of assumptions including  $\varepsilon_{it} = \varepsilon_t \sim \text{iid}\mathcal{N}(0, \sigma^2)$ ; see Hansen (1997, 2000),  $LR_T(c_1^0)$  has an analytic (nonstandard) asymptotic null distribution, and Hansen tabulates critical values for customary significance levels.

Tena and Tremayne (2009) use this result to test (45). They estimate a VTAR model equation by equation and choose an estimated threshold from one of the equations to represent  $c_1^0$  and test this null hypothesis in the other equations. This is repeated for all equations. They also suggest testing the hypothesis  $c_{1i} = 0$  equation by equation for  $i = 1, \dots, m$  and comparing the results, because  $c_1^0 = 0$  is a sensible value in the application the authors are working with. Doing so only provides indirect evidence against the null hypothesis, but in some cases the results may be sufficiently convincing for the researcher to reject the null hypothesis, albeit without a formal significance level to support the decision.

## 7 Generalised impulse response functions

Interpreting the coefficient estimates of a VLSTAR model is not possible, except possibly for  $\hat{\gamma}$  and  $\hat{c}$  that indicate the shape and location of the transition function. A similar remark is valid for VTAR models. In order to understand the dynamic behaviour of the model, impulse response functions are most helpful. As Koop, Pesaran and Potter (1996) have pointed out, estimating impulse response functions for nonlinear models is not as easy as it may be for linear VAR models. The reason is that the impulse response is no longer proportional to the size of the shock, nor is it independent of the 'history', that is, the starting-values. These authors thus defined the so-called generalized impulse response (GIR) function which is a random variable and a function of both the size of the shock and the history. It is defined as

follows:

$$GIRF(h, \boldsymbol{\varepsilon}_t, \Omega_{t-1}) = \mathbf{E}\{\mathbf{y}_{t+h} | \boldsymbol{\varepsilon}_t^\delta, \Omega_{t-1}\} - \mathbf{E}\{\mathbf{y}_{t+h} | \Omega_{t-1}\} \quad (47)$$

where  $\boldsymbol{\varepsilon}_t^\delta$  is the vector of shocks, and  $\Omega_{t-1} = \{\omega_{t-j} : j \geq 1\}$  is the set of possible histories but can be replaced by an appropriate subset. The conditioning variables in (47),  $\boldsymbol{\varepsilon}_t^\delta$  and  $\Omega_{t-1}$ , are assumed random. That  $\Omega_{t-1}$  is random means that its values in computing densities of the random GIR function are drawn from the set  $\{\omega_{t-j} : j \geq 1\}$  or a subset.

As is the case for standard impulse responses from VAR models, estimating the GIR function requires ordering the variables in the vector  $\mathbf{y}_t$ . This is equivalent to making weak exogeneity assumptions in the model. Once this has been done, the GIR function can be estimated by simulation or bootstrapping the residuals of the estimated model. How this is done is described in Tena and Tremayne (2009). First draw a history from the set of histories. Second, form the Cholesky decomposition of the residual covariance matrix:  $\widehat{\boldsymbol{\Sigma}} = \widehat{\mathbf{C}}\widehat{\mathbf{C}}'$ , where  $\widehat{\mathbf{C}}$  is a lower triangular matrix defined by the appropriate weak exogeneity assumptions. Then orthogonalize the residuals:  $\mathbf{e}_s = \widehat{\mathbf{C}}^{-1}\widehat{\boldsymbol{\varepsilon}}_s$ ,  $s = p + 1, \dots, T$ . Next, either sample from the distribution of  $\mathbf{e}_s$  or draw the residuals  $\{\mathbf{e}_t^*, \mathbf{e}_{t+1}^*, \dots, \mathbf{e}_{t+h}^*\}$  from this set with replacement. Suppose one wants to shock the  $i$ th variable. Then form another set by replacing the  $i$ th element of  $\mathbf{e}_t^*$  by a shock  $e_i^\delta$  drawn from a set of shocks. For example, one may be interested in shocks that are at least as large as the 75% quantile of the (empirical) distribution of the transformed residual. This yields another sequence of residuals  $\{\mathbf{e}_t^\delta, \mathbf{e}_{t+1}^*, \dots, \mathbf{e}_{t+h}^*\}$ , where  $\mathbf{e}_t^\delta = (e_{1t}, \dots, e_{i-1,t}, e_i^\delta, e_{i+1,t}, \dots, e_{mt})'$ . Transform back to the original residuals:  $\widehat{\boldsymbol{\varepsilon}}_{t+j}^* = \widehat{\mathbf{C}}\mathbf{e}_{t+j}^*$ ,  $j = 1, \dots, h$ , and  $\widehat{\boldsymbol{\varepsilon}}_t^\delta = \widehat{\mathbf{C}}\mathbf{e}_t^\delta$ . Generate two sequences of  $\mathbf{y}_{t+j}$ ,  $j = 0, 1, \dots, h$ , from the estimated models using the two sequences of residuals and compute the difference element by element. This gives one observation of (47) for periods  $t, t + 1, \dots, t + h$ , when the system is shocked at time  $t$ .

An estimate of the conditional distribution of the random GIR function given  $\Omega_{t-1}$  is generated by repeating these steps by new draws for  $\mathbf{e}_t^\delta$  and thus generating new observations of (47). Repeating everything up until this phase with histories drawn from the whole set yields an unconditional empirical distribution of (47). This distribution may also be conditional if the size of the shock and/or a subset of histories is restricted. For example, the interest may lie in finding out whether certain shocks affect the output differently when the economy is in a recession than when it is in expansion. One may then estimate a GIR function for both cases by selecting appropriate subsets of histories and shocks and compare the results.

If one computes GIR functions for several horizons, as is normally the case, the question is how to present them. The highest density regions



(HDR) for characterizing distributions, see Hyndman (1996) or Teräsvirta et al. (2010, Ch. 15), can be recommended for the purpose. They may be conveniently graphed using fan charts, see for example Wallis (1999) or Lundbergh and Teräsvirta (2002). Teräsvirta and Yang (2013) apply box and whiskers graphs. The latter have the drawback that they miss potential multimodality of the HDR, so the user of those has to make sure that the estimated HDR is unimodal.

## 8 Forecasting with vector threshold autoregressive and smooth transition models

The forecasts considered here are conditional means: the forecast of  $\mathbf{y}_{T+h}$  given the conditioning information  $\mathcal{F}_T$  at time  $T$ , denoted  $\mathbf{y}_{T+h|T}$ , in theory equals

$$\mathbf{y}_{T+h|T} = \mathbf{E}\{\mathbf{y}_{T+h}|\mathcal{F}_T\}. \quad (48)$$

Supposing that there is a nonlinear model, such as the VTAR or VSTAR model,  $\mathcal{F}_T = \Omega_T$ , the history of  $\{\mathbf{y}_t\}$  up until  $T$ . This means that forecasting  $\mathbf{y}_{T+h}$  has computationally much in common with computing GIR functions. Forecasting is in fact simpler as history is not a random variable. Furthermore, the residuals need not be orthogonalised if the error sequences  $\{\boldsymbol{\varepsilon}_{T+1}^{(i)}, \dots, \boldsymbol{\varepsilon}_{T+h}^{(i)}\}$ ,  $i = 1, \dots, B$ , are generated by a bootstrap, i.e., by  $B$  independent draws from the set of residuals  $\{\widehat{\boldsymbol{\varepsilon}}_{p+1}, \dots, \widehat{\boldsymbol{\varepsilon}}_T\}$ . Some researchers construct the forecast by assuming  $\boldsymbol{\varepsilon}_{T+1} = \dots = \boldsymbol{\varepsilon}_{T+h} = \mathbf{0}$ , but this simplification leads to biased forecasts. For discussion, see for example Teräsvirta (2006), Kock and Teräsvirta (2011) or Teräsvirta et al. (2010, Chapter 14). The forecast, an unbiased approximation of the conditional mean in (48), becomes

$$\mathbf{y}_{T+h|T} = (1/B) \sum_{j=1}^B \mathbf{y}_{T+h}^{(j)}$$

where  $\mathbf{y}_{T+h}^{(j)}$  is the  $j$ th realization of  $\mathbf{y}_{T+h}$  obtained from the model, given the history  $\Omega_T$ , and the error sequence  $\{\boldsymbol{\varepsilon}_{T+1}^{(j)}, \dots, \boldsymbol{\varepsilon}_{T+h}^{(j)}\}$ . Note that if one makes an assumption of the multivariate distribution of  $\boldsymbol{\varepsilon}_t$  it becomes possible to also draw the error sequence from this distribution using the estimated error covariance matrix  $\widehat{\boldsymbol{\Sigma}}$ . This requires that the errors are homoskedastic, whereas the bootstrap allows unconditional heteroskedasticity.

It seems that most applications focus on describing macroeconomic relationships rather than forecasting. The paper by Galvão (2006) mentioned in Section 3.2.2 constitutes an exception. This may be a reason for the fact

that as yet there are few studies (Galvão’s is one) comparing forecasting performance of nonlinear VAR models to that of linear VARs. Consequently, usefulness of VTAR and VSTAR models in macroeconomic forecasting is still an open question.

## 9 Applications

Vector TAR and STAR models have been fitted to macroeconomic and financial time series. These models have been used to describe asymmetries, for example in the response of real variables to changes in monetary policy. A relevant question for monetary policy decisions and their evaluation is whether or not the response of these variables to monetary policy shocks varies nonlinearly according to the business cycle. The recent financial crisis has made the dependence of the interaction between the financial sector and the macroeconomy on the financial conditions an interesting research topic. In the following we consider four macroeconomic applications of vector TAR and STAR models that illustrate how these models can be used to address a variety of relevant economic questions. In addition to the empirical results, we highlight several model-building issues and choices.

### 9.1 Applications of Vector TAR models

Both stationary VTAR and nonstationary TVEC models have been considered in the empirical literature. In this section we discuss two examples to illustrate both economic issues that might be investigated with this model class and different modelling approaches. We begin by considering Balke (2000) who discusses the possibility that credit plays a role of a nonlinear propagator of shocks to US macroeconomic variables. The author constructs a four-dimensional TVAR model in which the variables appear in the following order: output growth, inflation, the Fed funds rate and an indicator of credit conditions (measured in a number of alternative ways). A noteworthy detail is that the two-regime TVAR model is in a triangular form:

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_t + \mathbf{B}(\mathbf{L})\mathbf{y}_{t-1} + (\mathbf{C}\mathbf{y}_t + \mathbf{D}(\mathbf{L})\mathbf{y}_{t-1})I(s_{t-d} < c) + \mathbf{v}_t \quad (49)$$

where  $\mathbf{A}$  and  $\mathbf{C}$  are  $4 \times 4$  upper triangular matrices with zeroes on the main diagonal,  $\mathbf{B}(\mathbf{L}) = \sum_{j=1}^{p_b} \mathbf{B}_j \mathbf{L}^j$ ,  $\mathbf{D}(\mathbf{L}) = \sum_{j=1}^{p_d} \mathbf{D}_j \mathbf{L}^j$ ,  $\mathbf{v}_t$  is the error vector with uncorrelated elements, and  $s_t$  is the regime indicator. The upper triangular matrices determine the contemporaneous ‘structural’ relationships between the four variables and thus identify the shocks.

Since one does not know which one of the potential credit condition variables, if any, would be the switch or threshold variable, the analysis begins by testing linearity against (49). This is done by applying the method of Hansen (1996) and generating the null distribution of the test statistic by simulation. Balke considers both the supremum and the average Wald tests, as discussed in Section 5.2.1. His tests reject the linearity hypothesis for all three credit condition variables: the spread of the 4–6 month commercial paper and the six-month Treasury bill, the mix of bank loans and commercial paper in total firm external finance, and difference of growth rate in short-term debt in small and large manufacturing firms. Balke then estimates a triangular TVAR model using each of these indicators as the threshold variable. He notes that the three indicators split the sample into two regimes in more or less the same way.

Balke (2000) does not report any evaluation results for the estimated model(s) but concentrates on generalized impulse responses. The set of histories is divided into periods of 'normal' and 'tight' credit conditions, and the shock is plus/minus one or two standard deviations computed from the distribution of the corresponding residuals. Results for both regimes are reported separately for the four different shocks assuming that the shocked variable is any of the four included in  $\mathbf{y}_t$ . What is graphed is not the density of the estimated GIR function but rather its mean. The shocks are identified as the model is triangular with properly ordered equations, and the errors independent. It is found that the shocks generally have a larger (average) effect on output growth in when the economy is in the tight credit regime than when the opposite is true. Evidence of asymmetry in the response to shocks is also found.

Tena and Tremayne (2009) apply the VTAR model to studying the transmission of monetary policy to industrial output in the UK. Their paper contains both a general (aggregate output) and a sectoral analysis, of which only the former is considered here. The reason why we review this particular application is that in setting up the model the authors do not assume a common threshold a priori, and modelling proceeds equation by equation. The seasonally adjusted quarterly time series are the industrial production index, the retail price index, an interest rate, and the Pound sterling/US Dollar exchange rate. The models are built on first differences of these series, and the period of observation is 1972(1)–2007(1). The variables appear in the vector to be modelled in the aforementioned order that is based on relevant weak exogeneity assumptions.

Linearity is tested using the lagged industrial production growth rate as the threshold variable. This is due to the fact that the interest lies in the asymmetries in the transmission mechanism which depends on the state of

the economy. It turns out that the inflation (growth rate of the retail price index) equation is linear. Parameters are estimated equation by equation following Hansen (1997). The hypothesis of a common threshold for the three threshold equations is tested as discussed in Section 6.4 and rejected. Evaluation involves testing for another threshold (Hansen, 2000), but the null hypothesis of a single threshold is not rejected.

The sectoral analysis also contains estimating GIR functions. Due to the focus on monetary policy, the interest rate is shocked. The size of the shock comprises two values, and a separate GIR function is estimated for them. Two sets of histories are considered: expansionary and recessionary episodes (measured using the industrial output). The whole HDR is not reported, but 95% confidence intervals from the estimated densities are given for horizons extending from one to ten quarters. The main conclusion is that negative interest rate shocks have a bigger impact on output than positive ones.

Most applications of the VTAR or TVEC model have been to macroeconomic series. But then, validity of the Law of One Price has been investigated by a TVEC model using prices of individual commodities in different cities as in Lo and Zivot (2001). Further, an empirical example in Tsay (1998) concerns the high-frequency (one-minute) S&P 500 stock index and the corresponding index futures data. These time series have also been modelled using the TVEC model; see Martens, Kofman and Vorst (1998). A comprehensive list of applications of TAR models, both vector and univariate ones, can be found in Hansen (2011).

## 9.2 Applications of Vector STAR and STR models

Camacho (2004) makes good use of the modelling strategy described in Section 5.3 when building a VSTAR model for describing the relationship between the quarterly US GDP ( $y_{1t}$ ) and the Composite Leading Indicator (CLI) of the Conference Board ( $y_{2t}$ ). The time period is 1959(1)–2002(1). The latter variable is monthly, and the last month of every quarter is taken to construct a quarterly series to match the quarterly GDP series. Choosing the STVEC model as the starting-point allows for the possibility that these two nonstationary variables be cointegrated. Cointegration is tested for using the nonparametric cointegration test of Bierens (1997), the result being that the two variables are not cointegrated. This leads to eliminating the cointegration terms from (18).

The next step is to test linearity, which is done equation by equation. This is done using the Taylor expansion approach LM-type tests in Luukkonen et al. (1988). Camacho (2004) also consider the Exponential VSTAR

(EVSTAR) model, which is (16) with the exponential transition function

$$G_j(\gamma_j, \mathbf{c}_j; s_{jt}) = 1 - \exp\{-\gamma_j(s_{jt} - c_j)^2\}, \quad \gamma_j > 0$$

in (17). Even a third VSTAR model, the so-called LVSTAR-D model is considered but not discussed in detail here. The linearity tests reject the null hypothesis for most transition variables, and test sequences as in Teräsvirta (1994) are used to select the transition function. This leaves the model builder with three alternatives: an LVSTAR, an EVSTAR and an LVSTAR-D model. A final choice is left to the evaluation stage. In the LVSTAR model,  $s_t = \Delta y_{1,t-2}$  and in the EVSTAR one  $s_t = \Delta y_{2,t-2}$  (in both cases the same transition variable for both equations). The estimated equations only contain one lag of each variable, except for the transition variables. A noteworthy detail is that the estimated models are intercept-switching STAR models; see Section 3.3.1.

The results of the evaluation stage indicates that the LVSTAR model is superior to the other two VSTAR variants. It has the best fit and passes most misspecification tests that include testing the hypotheses of no error autocorrelation, no additive nonlinearity, and parameter constancy. Furthermore, one-quarter-ahead real-time forecasts for the GDP growth from this model are more accurate than the corresponding ones from the other two models.

Camacho (2004) also carries out tests of equal forecasting ability with forecasts from the LVSTAR model as the benchmark. The idea is to construct the one-period-ahead forecasts from 1978(1) to 2002(2) in such a way that they as accurately as possible reflect the forecasting situation, that is, the information available for the forecasters at the time of forecasting. The details of this can be found in the paper. The main outcome is that all tests reject the null hypothesis of equal forecastability in favour of the intercept-switching LVSTAR forecasts.

The goals of Weise (1999) are to investigate whether (a) positive and negative monetary shocks have asymmetric effects on output and prices; (b) whether the effects of monetary shocks vary over the business cycle; and (c) whether the effects of monetary shocks vary disproportionately with the size of the shock. After first developing a simple structural model incorporating asymmetric nominal rigidities and discussing its potential implications for the nonlinear econometric model, the author chooses an LVSTAR framework (16) to pursue his goals. The quarterly US time series of first differences of the logarithmic industrial production, the consumer price index, and M1 cover the period from 1960(2) to 1995(2). Before proceeding, the series are filtered by regressing them separately on a constant, seasonal dummies, two

intercept dummies accounting for presumed structural breaks, a linear trend, and two breaking linear trends in which the breaks are synchronized with the intercept dummies. This makes the filtered series stationary.

A linear trivariate VAR model with four lags forms the basis of the analysis. Linearity is tested using the likelihood ratio test approach discussed in Section 5.3.1. It is found that linearity is never rejected when the transition variable is money (M1). Lags of the other two variables produce rejections, some of which are very strong. Selection of final models based on the results of the tests yields three LVSTAR models with the first lag of industrial production, the second lag of the consumption prices and the first lag of their differences as transition variables.

The models are estimated by assuming a single transition function and fixing the location parameter so it is not estimated. Estimation is equation by equation as discussed in Section 4.2. No evaluation of the estimated models is carried out, but the dynamic properties of the estimated models are investigated by impulse response analysis. The interest lies in the response of output and prices to monetary shocks. As discussed in Section 7, the functions are generated by drawing from different sets of histories and shocks. Monetary shocks are identified in the customary way by appropriately ordering the equations and applying the Cholesky decomposition in generating the shocks. Randomness of the GIR functions is not taken fully into account, however, as only averages and not complete estimated densities are reported. Shocks to the money supply are found to have stronger output effects and weaker price effects when output growth is initially low than when it is high. Positive and negative monetary shocks have nearly symmetric effects to these variables. Only shocks of  $\pm 1$  and  $\pm 2$  standard deviations have been considered.

Vector STAR and STR models have been applied less frequently than their threshold counterparts. Rothman et al. (2001) specify and estimate an LVSTAR model to investigate the direction of nonlinear Granger causality in the relationship between money and output in the United States. Rahman and Serletis (2010) apply the LVSTAR model to studying asymmetries in the effects of oil price and monetary policy shocks to output growth in the US. Also using US time series, Auerbach and Gorodnichenko (2012) investigate whether fiscal policy is more effective in recessions than in expansions. Differences across regimes are allowed in the variance-covariances of the shocks and in the transmission. The authors find that fiscal policy is more effective in recessions. They also investigate fiscal multipliers of disaggregate spending variables and find that military spending has the largest multiplier of the ones considered.

## 10 Final remarks

Nonlinear models have become a useful tool in modelling economic relationships given the experiences gained and conclusions drawn from the financial crisis and the Great Recession. The purpose of this chapter is to provide an overview of vector threshold autoregressive and smooth transition models. This includes modelling nonstationary cointegrated variables with models of this kind. Model building is emphasised, and various testing situations described. Space limitations have made it necessary to restrict the discussion of applications to a few interesting examples.

Readers interested in more details about modelling strategies and issues of vector threshold autoregressive and smooth transition models discussed in this survey are referred to Tsay (1998) for vector TAR models and to Camacho (2004), Teräsvirta and Yang (2013) and Yang and Teräsvirta (2013) for their smooth transition counterparts. Hansen (2011) provides a comprehensive review of both univariate and vector TAR models and contains a large number of references. Gonzalo and Pitarakis (2006) present a useful survey of testing for cointegration in the TVEC framework; for an update see Gonzalo and Pitarakis (2013). For a more general discussion on vector nonlinear models, see Teräsvirta et al. (2010, Chapter 11).

## References

- Anderson, H. M. and Vahid, F.: 1998, Testing multiple equation systems for common nonlinear components, *Journal of Econometrics* **84**, 1–36.
- Astatkie, T., Watts, D. G. and Watt, W. E.: 1997, Nested threshold autoregressive (NeTAR) models, *International Journal of Forecasting* **13**, 105–116.
- Atanasova, C.: 2003, Credit market imperfections and business cycle dynamics: A nonlinear approach, *Studies in Nonlinear Dynamics & Econometrics* **7**, Issue 4, Article 5.
- Auerbach, A. J. and Gorodnichenko, Y.: 2012, Measuring the output responses to fiscal policy, *American Economic Review: Economic Policy* **4**, 1–27.
- Bacon, D. W. and Watts, D. G.: 1971, Estimating the transition between two intersecting straight lines, *Biometrika* **58**, 525–534.

- Balke, N.: 2000, Credit and economic activity: Credit regimes and nonlinear propagation of shocks, *Review of Economics and Statistics* **82**, 344–349.
- Balke, N. S. and Fomby, T. B.: 1997, Threshold cointegration, *International Economic Review* **38**, 627–645.
- Bauwens, L., Lubrano, M. and Richard, J.-F.: 2000, *Bayesian Inference in Dynamic Econometric Models*, Oxford University Press, Oxford.
- Bec, F. and Rahbek, A.: 2004, Vector equilibrium correction models with non-linear discontinuous adjustments, *Econometrics Journal* **7**, 628–651.
- Bierens, H. J.: 1997, Nonparametric cointegration analysis, *Journal of Econometrics* **77**, 379–404.
- Bollerslev, T.: 1990, Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model, *Review of Economics and Statistics* **72**, 498–505.
- Box, G. E. P. and Jenkins, G. M.: 1970, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.
- Camacho, M.: 2004, Vector smooth transition regression models for US GDP and the composite index of leading indicators, *Journal of Forecasting* **23**, 173–196.
- Chan, K. S.: 1993, Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model, *Annals of Statistics* **21**, 521–533.
- Chan, K. S. and Tong, H.: 1986, On estimating thresholds in autoregressive models, *Journal of Time Series Analysis* **7**, 178–190.
- Davies, R. B.: 1977, Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrika* **64**, 247–254.
- Dueker, M. J., Psaradakis, Z., Sola, M. and Spagnolo, F.: 2011, Multivariate contemporaneous-threshold smooth transition autoregressive models, *Journal of Econometrics* **160**, 311–325.
- Eitrheim, Ø. and Teräsvirta, T.: 1996, Testing the adequacy of smooth transition autoregressive models, *Journal of Econometrics* **74**, 59–75.
- Eklund, B. and Teräsvirta, T.: 2007, Testing constancy of the error covariance matrix in vector models, *Journal of Econometrics* **140**, 753–780.



- El-Shagi, M.: 2011, An evolutionary algorithm for the estimation of threshold vector error correction models, *International Economics and Economic Policy* **8**, 341–362.
- Engle, R. F. and Kozicki, S.: 1993, Testing for common features, *Journal of Business and Economic Statistics* **11**, 369–386.
- Galvão, A. B. C.: 2006, Structural break threshold VARs for predicting US recessions using the spread, *Journal of Applied Econometrics* **21**, 463–487.
- Gefang, D.: 2012, Money-output causality revisited - A Bayesian logistic smooth transition VECM perspective, *Oxford Bulletin in Economics and Statistics* **74**, 131–151.
- Gefang, D. and Strachan, R.: 2010, Nonlinear impacts of business cycles on the U.K. - A Bayesian smooth transition VAR approach, *Studies in Nonlinear Dynamics and Econometrics* **14**, Issue 1, Article 2.
- González, A., Hubrich, K. and Teräsvirta, T.: 2009, Forecasting inflation with gradual regime shifts and exogenous information, *CREATES Research Paper 2009-3*, Aarhus University.
- González, A. and Teräsvirta, T.: 2008, Modelling autoregressive processes with a shifting mean, *Studies in Nonlinear Dynamics and Econometrics* **12**, No. 1, Article 1.
- Gonzalo, J. and Pitarakis, J.-Y.: 2006, Threshold effects in multivariate error correction models, in T. C. Mills and K. Patterson (eds), *Palgrave Handbooks of Econometrics*, Vol. 1, Econometric Theory, Palgrave Macmillan, Basingstoke, pp. 578–609.
- Gonzalo, J. and Pitarakis, J.-Y.: 2013, Estimation and inference in threshold type regime switching models, in N. Hashimzade and M. A. Thornton (eds), *Handbook of research methods and applications in empirical macroeconomics*, Edward Elgar, Cheltenham, pp. 189–204.
- Goodwin, B. K., Holt, M. T. and Prestemon, J. P.: 2012, Exchange rate pass-through in timber products: The case of oriented strand board in Canada and the United States, *Technical report*, University of Alabama at Tuscaloosa.
- Haggan, V. and Ozaki, T.: 1981, Modelling non-linear random vibrations using an amplitude-dependent autoregressive time series model, *Biometrika* **68**, 189–196.

- Hamilton, J. D.: 1989, A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica* **57**, 357–384.
- Hansen, B. E.: 1996, Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**, 413–430.
- Hansen, B. E.: 1997, Inference in TAR models, *Studies in Nonlinear Dynamics and Econometrics* **2**, 1–14.
- Hansen, B. E.: 1999, Testing for linearity, *Journal of Economic Surveys* **13**, 551–576.
- Hansen, B. E.: 2000, Sample splitting and threshold estimation, *Econometrica* **68**, 575–603.
- Hansen, B. E.: 2011, Threshold autoregression in economics, *Statistics and Its Interface* **4**, 123–127.
- Hansen, B. E. and Seo, B.: 2002, Testing for two-regime threshold cointegration in vector error-correction models, *Journal of Econometrics* **110**, 293–318.
- Harvill, J. L. and Ray, B. K.: 1999, A note on tests for nonlinearity in a vector time series, *Biometrika* **86**, 728–734.
- He, C., Teräsvirta, T. and González, A.: 2009, Testing parameter constancy in vector autoregressive models against continuous change, *Econometric Reviews* **28**, 225–245.
- Hendry, D. F. and Massmann, M.: 2007, Co-breaking: Recent advances and a synopsis of the literature, *Journal of Business and Economic Statistics* **25**, 33–51.
- Holt, M. T. and Teräsvirta, T.: 2012, Global hemispheric temperature trends and co-trending: A shifting mean vector autoregressive analysis, *CREATES Research Paper 2012-54*, Aarhus University.
- Hubrich, K. and Tetlow, R. J.: 2012, Financial stress and economic dynamics: The transmission of crises, *Finance and economics discussion series 2012-82*, The Federal Reserve Board.
- Hungnes, H.: 2012, Testing for co-non-linearity, *Discussion Paper 699*, Statistics Norway, Research Department.

- Hyndman, R. J.: 1996, Computing and graphing highest density regions, *American Statistician* **50**, 120–126.
- Jacobson, T., Lindh, T. and Warne, A.: 2002, Growth, saving, financial markets and Markov switching regimes, *Studies in Nonlinear Dynamics and Econometrics* **5**, Issue 4, Article 1.
- Johansen, S.: 1995, *Likelihood-based inference in cointegrated vector autoregressive models*, Oxford University Press, Oxford.
- Kapetanios, G., Shin, Y. and Snell, A.: 2006, Testing for cointegration in nonlinear smooth transition error correction models, *Econometric Theory* **22**, 279–303.
- Kock, A. B. and Teräsvirta, T.: 2011, Forecasting with nonlinear time series models, in M. P. Clements and D. F. Hendry (eds), *The Oxford Handbook of Economic Forecasting*, Oxford University Press, Oxford, pp. 61–87.
- Koop, G., Pesaran, M. H. and Potter, S. M.: 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics* **74**, 119–147.
- Krishnakumar, J. and Neto, D.: 2012, Testing uncovered interest rate parity and term structure using a three-regime threshold unit root VECM An application to the Swiss 'isle' of interest rates, *Oxford Bulletin of Economics and Statistics* **74**, 180–202.
- Kristensen, D. and Rahbek, A.: 2010, Likelihood-based inference for cointegration with nonlinear error-correction, *Journal of Econometrics* **158**, 78–94.
- Kristensen, D. and Rahbek, A.: in press, Testing and inference in nonlinear cointegrating vector error correction models, *Econometric Theory* .
- Krolzig, H.-M.: 1997, *Markov-Switching Vector Autoregressions Modelling, Statistical Inference and Applications to Business Cycle Analysis*, Springer, Berlin.
- Lanne, M. and Saikkonen, P.: 2002, Threshold autoregressions for strongly autocorrelated time series, *Journal of Business and Economic Statistics* **20**, 282–289.
- Lindgren, G.: 1978, Markov regime models for mixed distributions and switching regressions, *Scandinavian Journal of Statistics* **5**, 81–91.

- Lo, M. C. and Zivot, E.: 2001, Threshold cointegration and nonlinear adjustment to the law of one price, *Macroeconomic Dynamics* **5**, 533–576.
- Lundbergh, S. and Teräsvirta, T.: 2002, Forecasting with smooth transition autoregressive models, in M. P. Clements and D. F. Hendry (eds), *A Companion to Economic Forecasting*, Blackwell, Oxford, pp. 485–509.
- Lundbergh, S., Teräsvirta, T. and van Dijk, D.: 2003, Time-varying smooth transition autoregressive models, *Journal of Business and Economic Statistics* **21**, 104–121.
- Luukkonen, R., Saikkonen, P. and Teräsvirta, T.: 1988, Testing linearity against smooth transition autoregressive models, *Biometrika* **75**, 491–499.
- Lütkepohl, H.: 2006, Vector autoregressive models, in T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics*, Vol. 1, Econometric Theory, Palgrave Macmillan, Basingstoke, pp. 477–510.
- Maddala, D. S.: 1977, *Econometrics*, McGraw-Hill, New York.
- Martens, M., Kofman, P. and Vorst, T. C. F.: 1998, A threshold error-correction models for intraday futures and index returns, *Journal of Applied Econometrics* **13**, 245–263.
- Pfann, G. A., Schotman, P. C. and Tschernig, R.: 1996, Nonlinear interest rate dynamics and implications for the term structure, *Journal of Econometrics* **74**, 149–176.
- Qian, L.: 1998, On maximum likelihood estimators for a threshold autoregression, *Journal of Statistical Planning and Inference* **75**, 21–46.
- Quandt, R. E.: 1958, The estimation of parameters of a linear regression system obeying two separate regimes, *Journal of the American Statistical Association* **53**, 873–880.
- Quandt, R. E.: 1960, Tests of the hypothesis that a linear regression system obeys two separate regimes, *Journal of the American Statistical Association* **55**, 324–330.
- Rahman, S. and Serletis, A.: 2010, The asymmetric effects of oil price and monetary policy shocks: A nonlinear VAR approach, *Energy Economics* **32**, 1460–1466.

- Rao, C. R.: 1973, *Linear Statistical Inference and its Applications*, Wiley, New York.
- Rothman, P., van Dijk, D. and Franses, P. H.: 2001, A multivariate STAR analysis of the relationship between money and output, *Macroeconomic Dynamics* **5**, 506–532.
- Saikkonen, P.: 2005, Stability results for nonlinear error correction models, *Journal of Econometrics* **127**, 69–81.
- Saikkonen, P.: 2008, Stability of regime switching error correction models under linear cointegration, *Econometric Theory* **24**, 294–318.
- Seo, B.: 2011a, Nonparametric testing for linearity in cointegrated error-correction models, *Studies in Nonlinear Dynamics and Econometrics* **15**, Issue 2, Article 6.
- Seo, M.: 2006, Bootstrap testing for the null of no cointegration in a threshold vector error correction model, *Journal of Econometrics* **134**, 129–150.
- Seo, M. H.: 2011b, Estimation of nonlinear error correction models, *Econometric Theory* **27**, 201–234.
- Sims, C. A.: 1980, Macroeconomics and reality, *Econometrica* **48**, 1–48.
- Sims, C. A., Waggoner, D. F. and Zha, T.: 2008, Methods for inference in large multiple equation Markov-switching models, *Journal of Econometrics* **146**, 255–274.
- Sims, C. and Zha, T.: 2006, Were there regimes shifts in U.S. monetary policy?, *American Economic Review* **96**, 54–81.
- Strikholm, B. and Teräsvirta, T.: 2006, A sequential procedure for determining the number of regimes in a threshold autoregressive model, *Econometrics Journal* **9**, 472–491.
- Taylor, M. P., Peel, D. A. and Sarno, L.: 2001, Nonlinear mean-reversion in real exchange rates: Towards a solution to the Purchasing Power Parity puzzles, *International Economic Review* **42**, 1015–1042.
- Tena, J. D. and Tremayne, A. R.: 2009, Modelling monetary transmission in UK manufacturing industry, *Economic Modelling* **26**, 1053–1066.
- Teräsvirta, T.: 1994, Specification, estimation, and evaluation of smooth transition autoregressive models, *Journal of the American Statistical Association* **89**, 208–218.

- Teräsvirta, T.: 1998, Modeling economic relationships with smooth transition regressions, in A. Ullah and D. E. Giles (eds), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552.
- Teräsvirta, T.: 2006, Forecasting economic variables with nonlinear models, in G. Elliott, C. W. J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Vol. 1, Elsevier, Amsterdam, pp. 413–457.
- Teräsvirta, T., Lin, C.-F. and Granger, C. W. J.: 1993, Power of the neural network linearity test, *Journal of Time Series Analysis* **14**, 309–323.
- Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J.: 2010, *Modelling Non-linear Economic Time Series*, Oxford University Press, Oxford.
- Teräsvirta, T. and Yang, Y.: 2013, Specification, estimation and evaluation of vector smooth transition autoregressive models with applications, *CREATES Research Paper*, Aarhus University.
- Tong, H.: 1990, *Non-Linear Time Series. A Dynamical System Approach*, Oxford University Press, Oxford.
- Tsay, R. S.: 1986, Nonlinearity tests for time series, *Biometrika* **73**, 461–466.
- Tsay, R. S.: 1989, Testing and modeling threshold autoregressive processes, *Journal of the American Statistical Association* **84**, 231–240.
- Tsay, R. S.: 1998, Testing and modeling multivariate threshold models, *Journal of the American Statistical Association* **93**, 1188–1202.
- Wallis, K. F.: 1999, Asymmetric density forecasts of inflation and the Bank of England’s fan chart, *National Institute Economic Review* (January 1999), 106–112.
- Weise, C. L.: 1999, The asymmetric effect of monetary policy: A nonlinear vector autoregression approach, *Journal of Money, Credit, and Banking* **31**, 85–108.
- Winker, P. and Gilli, M.: 2004, Applications of optimization heuristics to estimation and modelling problems, *Computational Statistics and Data Analysis* **47**, 211–223.
- Yang, Y.: 2013, Testing constancy of the error covariance matrix in vector models against parametric alternatives using a spectral decomposition, *CREATES Research Paper*, Aarhus University.

- Yang, Y. and Teräsvirta, T.: 2013, Linearity and misspecification tests for vector smooth transition regression models, *CREATES Research Paper*, Aarhus University.
- Yang, Z., Tian, Z. and Yuan, Z.: 2007, GSA-based maximum likelihood estimation for threshold vector error correction model, *Computational Statistics and Data Analysis* **52**, 109–120.

# Research Papers 2013



- 2013-01: Mikko S. Pakkanen: Limit theorems for power variations of ambit fields driven by white noise
- 2013-02: Almut E. D. Veraart and Luitgard A. M. Veraart: Risk premia in energy markets
- 2013-03: Stefano Grassi and Paolo Santucci de Magistris: It's all about volatility (of volatility): evidence from a two-factor stochastic volatility model
- 2013-04: Tom Engsted and Thomas Q. Pedersen: Housing market volatility in the OECD area: Evidence from VAR based return decompositions
- 2013-05: Søren Johansen and Bent Nielsen: Asymptotic analysis of the Forward Search
- 2013-06: Debopam Bhattacharya, Pascaline Dupasand Shin Kanaya: Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets
- 2013-07: Sílvia Gonçalves, Ulrich Hounyo and Nour Meddahi: Bootstrap inference for pre-averaged realized volatility based on non-overlapping returns
- 2013-08: Katarzyna Lasak and Carlos Velasco: Fractional cointegration rank estimation
- 2013-09: Roberto Casarin, Stefano Grassi, Francesco Ravazzolo and Herman K. van Dijk: Parallel Sequential Monte Carlo for Efficient Density Combination: The Deco Matlab Toolbox
- 2013-10: Hendrik Kaufmann and Robinson Kruse: Bias-corrected estimation in potentially mildly explosive autoregressive models
- 2013-11: Robinson Kruse, Daniel Ventosa-Santaulària and Antonio E. Noriega: Changes in persistence, spurious regressions and the Fisher hypothesis
- 2013-12: Martin M. Andreasen, Jesús Fernández-Villaverde and Juan F. Rubio-Ramírez: The Pruned State-Space System for Non-Linear DSGE Models: Theory and Empirical Applications
- 2013-13: Tom Engsted, Stig V. Møller and Magnus Sander: Bond return predictability in expansions and recessions
- 2013-14: Charlotte Christiansen, Jonas Nygaard Eriksen and Stig V. Møller: Forecasting US Recessions: The Role of Sentiments
- 2013-15: Ole E. Barndorff-Nielsen, Mikko S. Pakkanen and Jürgen Schmiegel: Assessing Relative Volatility/Intermittency/Energy Dissipation
- 2013-16: Peter Exterkate, Patrick J.F. Groenen, Christiaan Heij and Dick van Dijk: Nonlinear Forecasting With Many Predictors Using Kernel Ridge Regression
- 2013-17: Daniela Osterrieder: Interest Rates with Long Memory: A Generalized Affine Term-Structure Model
- 2013-18: Kirstin Hubrich and Timo Teräsvirta: Thresholds and Smooth Transitions in Vector Autoregressive Models