Econometric Reviews

ISSN: 0747-4938 (Print) 1532-4168 (Online) Journal homepage: www.tandfonline.com/journals/lecr20

A sequential test procedure for the choice of the number of regimes in multivariate nonlinear models

Andrea Bucci

To cite this article: Andrea Bucci (25 Dec 2025): A sequential test procedure for the choice of the number of regimes in multivariate nonlinear models, Econometric Reviews, DOI: 10.1080/07474938.2025.2598416

To link to this article: https://doi.org/10.1080/07474938.2025.2598416

ECONOMETRIC REVIEWS
https://doi.org/10.1080/07474938.2025.2598416

Taylor & Francis
Taylor & Francis Group

A sequential test procedure for the choice of the number of regimes in multivariate nonlinear models

Andrea Bucci

Department of Economics and Law, University of Macerata, Macerata, Italy

ABSTRACT

This article proposes a sequential test procedure for determining the number of regimes in nonlinear multivariate autoregressive models. The procedure relies on linearity and no additional nonlinearity tests for multivariate smooth transition and threshold autoregressive models. We conduct a simulation study to evaluate the finite-sample properties of the proposed test in small samples. Our findings indicate that the sequential procedure is capable of correctly identifying the number of regimes in the presence of both smooth and abrupt regime changes. The sequential procedure is also applied to real-world data. In the analysis of US monthly interest rates, the test identifies multiple regimes, confirming that the adjustment of interest rates toward their long-run equilibrium is state-dependent. Finally, when applied to Icelandic river flows, the procedure identifies a three-regime structure consistent with hydrological cycles.

ARTICLE HISTORY
Received 11 March 2025
Accepted 26 November 2025

KEYWORDS
Multivariate time series; nonlinear model; regime Identification

1. Introduction

Since the seminal work by Sims (1980), vector autoregressive (VAR) models have been a cornerstone of multivariate time series analysis. However, these models often fail to capture the complex, nonlinear interactions and regime shifts that characterize real-world economic and financial data (Hamilton, 2016), such as abrupt market crashes, sudden policy shifts, and asymmetric responses to external shocks.

In order to overcome these limitations, linear models have been extended to include a regime-switching mechanism. This can be done either using an underlying observable transition variable, a Markov-switching process, or using machine learning methods. This article focuses on a specific class of such models: those where the nonlinearity in the conditional mean is governed by an observable transition variable. Among the models based on a transition variable, there are the vector logistic smooth transition regression (VLSTR) and the vector threshold regression (VTR). These models resemble the nonlinear behaviors and structural breaks often observed in real-world data (see Hubrich and Teräsvirta, 2013, for a comprehensive review) through a transition function that leads the regime changes. In particular, vector smooth transition models extend the VAR model by allowing for a smooth transition between regimes, while vector threshold models involve an abrupt regime change. In both models, each regime is characterized by a distinct linear specification, and each observation is a weighted average of the conditional means of the regimes (Lanne and Virolainen, 2024). A key advantage of VLSTAR models over alternatives like Markov-switching VARs is that the regime shifts are driven by an observable variable, which is often endogenous to the system. This provides a direct economic interpretation for the switching mechanism and accommodates both gradual and abrupt dynamic changes. Despite their increasing popularity (Caggiano, Castelnuovo, and Figueres, 2017; Jiang et al., 2021; Lanne and

CONTACT Andrea Bucci andrea.bucci@unimc.it Department of Economics and Law, University of Macerata, Macerata, Italy.

Virolainen, 2024), their application in practice is partly limited due to challenges in model specification, such as selecting the appropriate number of regimes and identifying suitable transition variables.

Proper specification tests are crucial for these models, especially because they may not be identified under the null hypothesis of linearity or when a lower-regime model is the data-generating process (Davies, 1987). In the univariate context, a linearity test has been developed by Luukkonen, Saikkonen, and Teräsvirta (1988), while Eitrheim and Teräsvirta (1996) construct misspecification tests for smooth transition autoregressive (STAR) models, including tests for error autocorrelation, no additional nonlinearity, and parameter constancy. For multivariate models, Camacho (2004) and Teräsvirta and Yang (2014a) extended these tests. The former has developed a modelling strategy for a bivariate VLSTAR model along with several misspecification tests. The latter builds upon Camacho's approach by extending the linearity and misspecification tests, generalizing the tests beyond two time series.

Nevertheless, determining the number of regimes in multivariate models is unexplored. Several attempts to identify the number of regimes have been made in the univariate framework, including those by Hansen (1999), Gonzalo and Pitarakis (2002), and Strikholm and Teräsvirta (2006). Inspired by the sequential testing procedure for structural breaks proposed by Bai and Perron (1998), the latter two approaches recommend starting with a linear model (*i.e.*, a single-regime model) and iteratively testing between models with m and $m+1$ regimes until the null hypothesis of m regimes cannot be rejected.

This article aims to fill this critical gap in the multivariate context by proposing an easy-to-implement sequential procedure for selecting the number of regimes in VLSTAR and VTAR models. This task is a crucial prerequisite for any subsequent analysis, since without a correctly specified m, practitioners cannot reliably identify the current state of the system, interpret the economic meaning of different regimes, or generate accurate forecasts. Building upon the univariate approach of Strikholm and Teräsvirta (2006), we extend the methodology to multivariate models. Specifically, we employ and give a more formal context to the linearity and nonadditive nonlinearity tests proposed by Teräsvirta and Yang (2014a), and we use them in a sequential procedure for the identification of the number of regimes. By employing a Taylor series approximation around the null hypothesis and using Lagrange multiplier (LM) tests, we address the identification problems that typically arise under the null. This approach enables testing for linearity and additional nonlinearity in a stepwise fashion, allowing practitioners to identify the appropriate number of regimes without excessive computational burden. Moreover, it allows to handle cross-sectional heterogeneity in the transition function, so that the regimes do not need to be synchronized and can be in different phases at a given point in time. In addition, the proposed methodology is flexible enough to allow handling vector threshold models. Finally, the proposed methodology facilitates regime-aware forecasting. Once the number of regimes is determined, the value of the transition variable s_t allows for a straightforward identification of the current regime, enabling the use of regime-specific parameters for more accurate out-of-sample predictions.

It is well known that sequential testing procedures, such as those used for covariate selection in regression or for determining the order of a polynomial in partly linear models, suffer from several limitations (Leeb and Pötscher, 2005). These include cumulative testing errors, potential overfitting, and issues with controlling the overall type I error rate. In our application, the sequential procedure for selecting the number of regimes somehow shares these drawbacks. However, despite these limitations, alternative methods for determining the number of regimes in multivariate nonlinear models remain almost unexplored. While approaches based on information criteria (*e.g.*, AIC or BIC) or global tests have been proposed in univariate contexts, extending them to the multivariate nonlinear setting poses significant challenges. This is primarily due to the nonstandard asymptotic properties arising from unidentified nuisance parameters under the null and the possibility that different series may exhibit regime switches at distinct points. Consequently, the sequential procedure introduced here, despite its known limitations, represents one of the few practical methods currently available for this problem.

To evaluate the finite-sample properties of the proposed tests, we conduct a simulation study. The results demonstrate that the tests exhibit good empirical sizes under various specifications, highlighting their robustness and practical applicability. Furthermore, we apply the sequential procedure to two

real-world datasets. The purpose of the empirical application is to benchmark our procedure against established findings. We re-examine two well-studied datasets to compare the number of regimes identified by our method with those reported in the literature. For both time series, consistent with previous research, we assume that the conditional mean is subject to regime changes driven by an observable transition variable. On the one hand, we try to detect the number of regimes in US monthly interest rates (Tsay, 1998). On the other hand, the sequential procedure is applied to daily Icelandic river flow data, which have been shown to be nonlinear in several former applications (Tong, Thanoon, and Gudmundsson, 1985; Tsay, 1998; Teräsvirta and Yang, 2014b; Livingston Jr and Nur, 2020). In both cases, the number of regimes detected overlaps with what was found in the related literature.

The article is organized as follows. Section 2 describes the vector logistic smooth transition autoregressive model. In Section 3, we define the linearity test, while the sequential test procedure is introduced in Section 4. The tests are then applied to simulated data in Section 5 to compute their empirical sizes and selection frequencies, and to real data in Section 6. Section 7 concludes.

2. The VLSTAR model

A specification for the general vector logistic smooth transition autoregressive (VLSTAR) model can be found in Teräsvirta and Yang (2014b). For ease of notation, in this study we do not include exogenous variables in the model. Let $\mathbf{y}_t$ be an $n \times 1$ vector of dependent variables, the VLSTAR model with m regimes can be defined as follows:

$$\begin{aligned}\mathbf{y}_t = \boldsymbol{\mu}_0 + \sum_{j=1}^{p} \Phi_{0,j}\mathbf{y}_{t-j} + \mathbf{G}_t^{(1)}\left(\mathbf{s}_t; \boldsymbol{\gamma}_1, \mathbf{c}_1\right)\left[\boldsymbol{\mu}_1 + \sum_{j=1}^{p} \Phi_{1,j}\mathbf{y}_{t-j}\right] + \dots \\ + \mathbf{G}_t^{(m-1)}\left(\mathbf{s}_t; \boldsymbol{\gamma}_{m-1}, \mathbf{c}_{m-1}\right)\left[\boldsymbol{\mu}_{m-1} + \sum_{j=1}^{p} \Phi_{m-1,j}\mathbf{y}_{t-j}\right] + \boldsymbol{\varepsilon}_t \end{aligned} \tag{1}$$

where $\boldsymbol{\mu}_d$ is an $n \times 1$ vector of intercepts, for $d = 0, \dots, m-1$, $\Phi_{d,j}$ is an $n \times n$ matrix of parameters for the j-th lag and $\mathbf{G}_t^{(d)}\left(\mathbf{s}_t; \boldsymbol{\gamma}_d, \mathbf{c}_d\right)$ is a diagonal matrix of transition functions such that

$$\mathbf{G}_t^{(d)}\left(\mathbf{s}_t; \boldsymbol{\gamma}_d, \mathbf{c}_d\right) = \operatorname{diag}\left\{g_{1,t}^{(d)}\left(s_{1,t}; \gamma_{1d}, c_{1d}\right), \dots, g_{n,t}^{(d)}\left(s_{n,t}; \gamma_{nd}, c_{nd}\right)\right\} \tag{2}$$

where $s_{i,t}$, for $i = 1, \dots, n$, is a weakly stationary transition variable for the i-th equation, while γ_{id} and c_{id} are, respectively, the slope parameter and the location parameter where the transitions occur for the d-th regime.

The elements of $\mathbf{G}_t^{(d)}$ in Eq. (2) are usually specified as standard logistic functions[1]

$$g_{i,t}^{(d)}\left(s_{i,t}; \gamma_{id}, c_{id}\right) = \left[1 + \exp\left\{-\gamma_{id}\left(s_{i,t} - c_{id}\right)\right\}\right]^{-1}, \quad \gamma_{id} > 0.$$

This specification is extremely flexible, since for $\gamma_{id} \to \infty, \forall d$, the diagonal elements of $\mathbf{G}_t^{(d)}\left(\mathbf{s}_t; \boldsymbol{\gamma}_d, \mathbf{c}_d\right)$ (for ease of notation, we will refer to this function as $\mathbf{G}_t^{(d)}$) approach the indicator function, $\mathbb{1}(s_{i,t} > c_{id})$, thus the model becomes a vector threshold autoregressive (VTAR) model as the one introduced by Tsay (1998), while for $\gamma_{id} \to 0$, the model becomes a simple VAR. This means that the approach proposed in this study based on a smooth transition model can also be implemented for the selection of the number of regimes in a VTAR, for γ_{id} sufficiently large (see Section 4.1 for a discussion). The lagged values of one of the dependent variables in $\mathbf{y}_t$, or a combination of them (Camacho, 2004; Kheifets and Saikkonen, 2020), are usually chosen as $s_{i,t}$ for smooth transition models. However, a stationary exogenous variable can also be used, and, according to He, Teräsvirta, and González (2008), a temporal trend such as $s_{i,t} = t/T$

[1] We use standard logistic functions because of their simplicity, but a more general version of the logistic function can also be used (He, Teräsvirta, and González, 2008).

can be employed as well without violating the asymptotic theory. In this case, the VLSTAR model can be considered a special case of a time-varying autoregressive (TV-VAR) model and, for $\gamma_{i,d} \to \infty$, the changes of regimes identify structural breaks in the model. This could provide a good alternative to already existing methods for the identification of co-shifting in multivariate time series (Hendry and Mizon, 1998).

Model (1) can be reparametrized in the following form

$$\mathbf{y}_t = \left\{ \sum_{d=1}^{m} \mathbf{G}_t^{(d-1)} \mathbf{B}_r' \right\} \mathbf{x}_t + \boldsymbol{\varepsilon}_t = \tilde{\mathbf{G}}_t' \mathbf{B}' \mathbf{x}_t + \boldsymbol{\varepsilon}_t \tag{3}$$

where $\tilde{\mathbf{G}}_t = \left(\mathbf{I}_n, \mathbf{G}_t^{(1)}, \ldots, \mathbf{G}_t^{(m-1)}\right)'$ is a $mn \times n$ matrix, $\mathbf{I}_n$ is an $n \times n$ identity matrix, $\mathbf{x}_t = \left[1, \mathbf{y}_{t-1}', \mathbf{y}_{t-2}', \ldots, \mathbf{y}_{t-p}'\right]'$ is a $(1 + pn) \times 1$ vector and $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_m)$ is a $(1 + pn) \times mn$ matrix of parameters, where $\mathbf{B}_d = \left(\boldsymbol{\mu}_d', \Phi_{d,1}', \ldots, \Phi_{d,p}'\right)'$. Setting $\mathbf{G}_t^{(0)} = \mathbf{I}_n$ indicates that no transitions are allowed before the first change of regime. The set of parameters to be estimated is $\boldsymbol{\theta} = \{\mathbf{B}, \boldsymbol{\Gamma}, \mathbf{C}\}$, where $\boldsymbol{\Gamma}$ and $\mathbf{C}$ are $n \times m$ matrices of parameters of the transition functions defined as follows:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nm} \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nm} \end{bmatrix}.$$

The linearity and additive nonlinearity testing problems in model (1) concern testing the additive $(m-1)$-th component, therefore the null hypothesis is that $\boldsymbol{\mu}_{m-1} = \mathbf{0}$, $\Phi_{m-1,j} = \mathbf{0}$, $j = 1, \ldots, p$, in which case $\mathbf{G}_t^{(m-1)}$ is not identified for the presence of unidentified nuisance parameters. This problem, firstly studied by Davies (1987) and Watson and Engle (1985), has the direct consequence that the standard asymptotic inference does not hold as the asymptotic distribution of the test is not known under the null. To overcome it, Hansen (1996) has provided an empirical null distribution by simulation and has given the asymptotic theory for inference. Nevertheless, this method is computationally demanding and applies only in the case of a common transition function among all the equations, $\mathbf{G}_t = g(s_t|\gamma, c)\mathbf{I}_n$, so the number of nuisance parameters is restricted to two. Alternatively, the use of a Taylor series approximation around the null and a Lagrange multiplier (LM) test has been used to circumvent the identification problem, see Luukkonen, Saikkonen, and Teräsvirta (1988), Teräsvirta (1994) for the univariate smooth transition model. More recently, Seong, Cho, and Teräsvirta (2022) consider testing both the null hypotheses in a univariate smooth transition model and combining the results in a single quasi-likelihood ratio test statistic (Cho, Ishida, and White, 2011; White and Cho, 2012). Following Luukkonen, Saikkonen, and Teräsvirta (1988) and Strikholm and Teräsvirta (2006), we propose to approximate the logistic function in the alternative hypothesis through a L-order Taylor approximation around $\gamma_i = 0$, as further discussed in Section 3. It should be recalled that, although the true parameter value is strictly positive under the alternative, the expansion is performed around the point of nonidentification, so we can use an approximation around $\gamma_i = 0$.

3. Linearity test

In our sequential procedure, the linearity test is the first step since the smooth transition model is not identified if the linear model is the true data-generating process. When the system foresees a different transition variable for each equation, linearity can be tested equation-by-equation through the test introduced by Luukkonen, Saikkonen, and Teräsvirta (1988). Otherwise, the joint linearity test introduced in Teräsvirta and Yang (2014a) can be performed when a single transition variable is used. In the next sections, we deepen the theory behind the linearity and no additional nonlinearity tests already proposed in Teräsvirta and Yang (2014a).

By considering a 2-regime model (*i.e.*, $m = 2$), Eq. (3) becomes

$$\mathbf{y}_t = \mathbf{B}_1'\mathbf{x}_t + \mathbf{G}_t\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t. \tag{4}$$

Testing linearity in Eq. (4) equals testing the null hypothesis $\mathrm{H}_0 : \gamma_i = 0,\ i = 1, \ldots, n$. Under the null, we have that $\mathbf{G}_t = (1/2)\,\mathbf{I}_n$ and that Eq. (4) is linear, meaning that the null hypothesis creates an identification problem for the parameters in the linear combination $\mathbf{B}_1 + (1/2)\mathbf{B}_2$ and for the location parameter, c_i. As already pointed out above, this identification problem can be overcome by approximating the logistic function through an L-order Taylor approximation around $\gamma_i = 0$, such that

$$g_{i,t}\left(s_t \mid \gamma_i, c_i\right) \approx \sum_{l=0}^{L} \upsilon_{i,l} s_t^l + r_{i,t}$$

where $\upsilon_{i,0}, \ldots, \upsilon_{i,L}$ are the coefficients and $r_{i,t}$ is the reminder term. This means that $\mathbf{G}_t$ can be written as follows:

$$\mathbf{G}_t \approx \operatorname{diag}\left\{\sum_{l=0}^{L} \upsilon_{1,l} s_t^l + r_{1,t}, \ldots, \sum_{l=0}^{L} \upsilon_{n,l} s_t^l + r_{n,t}\right\} \approx \sum_{l=0}^{L} \Upsilon_l s_t^l + \mathbf{R}_t \tag{5}$$

where $\Upsilon_l = \operatorname{diag}\left(\upsilon_{1,l}, \ldots, \upsilon_{n,l}\right)$ and $\mathbf{R}_t = \operatorname{diag}\left(r_{1,t}, \ldots, r_{n,t}\right)$. Inserting Eq. (5) in (4) yields:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{B}_1'\mathbf{x}_t + \left(\sum_{l=0}^{L} \Upsilon_l s_t^l + \mathbf{R}_t\right)\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \\ &= \mathbf{D}_0'\mathbf{x}_t + \sum_{l=1}^{L} \mathbf{D}_l'\mathbf{x}_t s_t^l + \boldsymbol{\varepsilon}_t^*\end{aligned} \tag{6}$$

where $\mathbf{D}_0 = \mathbf{B}_1 + \mathbf{B}_2\Upsilon_0'$, $\mathbf{D}_l = \mathbf{B}_2\Upsilon_l'$ and $\boldsymbol{\varepsilon}_t^* = \mathbf{R}_t\mathbf{B}_2'\mathbf{x}_t + \boldsymbol{\varepsilon}_t$. In the auxiliary VAR in Eq. (6), testing linearity is equal to testing the null hypothesis $\mathrm{H}_0\colon \mathbf{D}_1 = \cdots = \mathbf{D}_L = \mathbf{0}$. Under the null hypothesis $\mathbf{R}_t = \mathbf{0}$, therefore the error term is $\boldsymbol{\varepsilon}_t^* = \boldsymbol{\varepsilon}_t$, so that the distributional properties of the error process are not affected by the Taylor approximation under the null hypothesis.

A special case of a VLSTAR with a common transition variable is the one with a transition function that is common to all the equations, *i.e.*, $\mathbf{G}_t = g_t(s_t; \gamma, c)\mathbf{I}_n$ with $g(s_t; \gamma, c)$ being a scalar. In such a case, we have that

$$g_t(s_t; \gamma, c) = \sum_{l=0}^{L} \upsilon_l s_t^l + r_t$$

which leads to $\Upsilon_l = \upsilon_l\mathbf{I}_n$ and $\mathbf{R}_t = r_t\mathbf{I}_n$. Inserting these elements in Eq. (6), the construction of the LM-type statistic remains the same as below.

Denoting $\mathbf{Y} = \left(\mathbf{y}_1, \ldots, \mathbf{y}_T\right)'$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)'$, $\mathbf{E}^* = \left(\boldsymbol{\varepsilon}_1^*, \ldots, \boldsymbol{\varepsilon}_T^*\right)'$, $\widetilde{\mathbf{D}}_L = \left(\mathbf{D}_1', \ldots, \mathbf{D}_L'\right)'$, and $\mathbf{Z}_L = \left(\mathbf{z}_1', \ldots, \mathbf{z}_T'\right)'$, where $\mathbf{z}_t = (\mathbf{x}_t' s_t, \mathbf{x}_t' s_t^2, \ldots, \mathbf{x}_t' s_t^L)'$, Eq. (6) can be written as

$$\mathbf{Y} = \mathbf{X}\mathbf{D}_0 + \mathbf{Z}_L\widetilde{\mathbf{D}}_L + \mathbf{E}^*. \tag{7}$$

The null hypothesis is $\widetilde{\mathbf{D}}_L = \mathbf{0}$, while the subscript in $\mathbf{Z}$ and $\widetilde{\mathbf{D}}$ indicates the order of the Taylor expansion.

Let $\boldsymbol{\theta} = \left(\mathbf{d}_0', \mathbf{d}_1'\right)' \in \Theta$ be the unknown parameters of the model (7) with the true values $\boldsymbol{\theta}_0$, where $\mathbf{d}_0 = \operatorname{vec}(\mathbf{D}_0)$, $\mathbf{d}_1 = \operatorname{vec}(\widetilde{\mathbf{D}}_L)$, $\Theta = \Theta_{\mathbf{d}_0} \times \Theta_{\mathbf{d}_1}$ is the parametric space with $\Theta_0 \in \mathbb{R}^{\tau_0}$ and $\Theta_1 \in \mathbb{R}^{\tau_1}$, with $\tau_0 = (1 + pn)n$ and $\tau_1 = (1 + pn)n + 2n$. Below, we assume that Θ_0 and Θ_1 are compact and $\boldsymbol{\theta}_0$ is an interior point of Θ. To compute a test for the null hypothesis, the log-likelihood of model (7) for T observations can be approximated by the conditional log-likelihood as follows

$$\ell_T(\boldsymbol{\theta}; \mathbf{Y}|\mathcal{F}_{t-1}) = \sum_{t=1}^{T} \ell_t(\boldsymbol{\theta}) = k - \frac{1}{2}\sum_{t=1}^{T} \log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T} \boldsymbol{\varepsilon}_t'{\Omega_{\mathbf{t}}}^{-1}\boldsymbol{\varepsilon}_t \tag{8}$$

where $k = -Tn\log(2\pi)/2$ and $\boldsymbol{\varepsilon}_t = \mathbf{y}_t - \mathbf{D}_0'\mathbf{x}_t - \sum_{l=1}^{L}\mathbf{D}_l'\mathbf{x}_t s_t^l = \mathbf{y}_t - \mathbf{D}_0'\mathbf{x}_t - \widetilde{\mathbf{D}}_L'\mathbf{z}_t$, with $\mathbf{z}_t = \left(\mathbf{x}_t' s_t, \mathbf{x}_t' s_t^2, \ldots, \mathbf{x}_t' s_t^l\right)'$ and $E\left\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\right\} = \Omega_t$ is a positive definite covariance matrix, with $\lim_{T\to\infty}(1/T)\sum_{t=1}^{T}\Omega_t = \Omega$ (see the following Assumption 3 for further details). Consequently, the limiting covariance matrix can be estimated from $(1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$ and can be used in the construction of the test statistic.

We need to specify the following assumptions in order to define an LM test.

Assumption 1. The log-likelihood $\ell_T(\boldsymbol{\theta};\mathbf{Y})$, defined as in Eq. (8), is twice continuously differentiable with respect to $\boldsymbol{\theta}$ in an open neighborhood of $\widetilde{\mathbf{D}}_L = \mathbf{0}$.

Assumption 2. The maximum likelihood estimators of the parameters $\mathbf{D}_0$ are consistent under the null hypothesis $\widetilde{\mathbf{D}}_L = 0$.

Assumption 3. The stochastic sequence $\{\boldsymbol{\varepsilon}_t\}$ is a martingale difference sequence with respect to an increasing sequence of σ-fields, $\mathcal{F}_t$ with

$$\sup_t E\left\{|\varepsilon_{i,t}|^{2+\alpha}|\mathcal{F}_{t-1}\right\} < \infty \qquad \text{a.s.}$$

for some $\alpha > 0$ and $i = 1, \ldots, n$, with $E\left\{\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'|\mathcal{F}_{t-1}\right\} = \Omega_t$, where Ω_t is a positive definite matrix with the following asymptotic limit

$$\lim_{T\to\infty}(1/T)\sum_{t=1}^{T}\Omega_t = \Omega \qquad \text{a.s.}$$

for some positive definite matrix Ω.

Assumption 4. $\mathbf{X}'\mathbf{X}$ and $\mathbf{Z}_L'(\mathbf{I} - \mathbf{P}_\mathbf{X})\mathbf{Z}_L$, where $\mathbf{P}_X$ is the limiting projection matrix of $\mathbf{X}$, $\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, are positive definite matrices.

Assumption 2 is a high-level assumption, while Assumption 3 guarantees the existence of the second moments for $\mathbf{y}_t$ and the convergence of the sample moments to their true values (He, Teräsvirta, and González, 2008) and permits the use of asymptotic theory for a martingale difference sequence (MDS), even when the assumption of i.i.d. errors is not valid, *e.g.*, in the case of conditionally heteroskedastic errors (Wang, Zhu, and Shao, 2022). Assumption 4 ensures that the matrices are invertible to compute the test statistic.

The block of the score vector involving the parameters under test, $\tilde{\boldsymbol{\theta}}$, can be written as follows

$$\begin{aligned}\frac{\partial\ell_T(\tilde{\boldsymbol{\theta}};\mathbf{Y})}{\partial\widetilde{\mathbf{D}}_L} &= -\frac{\partial}{\partial\widetilde{\mathbf{D}}_L}(1/2)\sum_{t=1}^{T}\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t = \frac{\partial}{\partial\widetilde{\mathbf{D}}_L}\sum_{t=1}^{T}\mathbf{z}_t'\widetilde{\mathbf{D}}_L\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}_t \\ &= \sum_{t=1}^{T}\mathbf{z}_t\boldsymbol{\varepsilon}_t'\boldsymbol{\Omega}^{-1} = \mathbf{Z}_L'\mathbf{E}\boldsymbol{\Omega}^{-1}, \end{aligned} \tag{9}$$

see for example Lütkepohl (1996) and Appendix A. Evaluated under H_0, the score obtained in Eq. (9) becomes

$$\frac{\partial\ell_T(\tilde{\boldsymbol{\theta}};\mathbf{Y})}{\partial\widetilde{\mathbf{D}}_L}\,|_{H_0} = \sum_{t=1}^{T}\mathbf{z}_t\hat{\boldsymbol{\varepsilon}}_t'\hat{\boldsymbol{\Omega}}^{-1} = \mathbf{Z}_L'\hat{\mathbf{E}}\hat{\boldsymbol{\Omega}}^{-1}$$

where $\hat{\mathbf{E}} = \left(\hat{\boldsymbol{\varepsilon}}_1, \hat{\boldsymbol{\varepsilon}}_2, \ldots, \hat{\boldsymbol{\varepsilon}}_T\right)'$, $\hat{\boldsymbol{\varepsilon}}_t = \mathbf{y}_t - \hat{\mathbf{D}}_0'\mathbf{x}_t$, and $\hat{\boldsymbol{\Omega}} = (1/T)\sum_{t=1}^{T}\hat{\boldsymbol{\varepsilon}}_t\hat{\boldsymbol{\varepsilon}}_t'$. The matrix $\hat{\mathbf{D}}_0$ is the maximum likelihood (ML) estimator of $\mathbf{D}_0$ under the null hypothesis. The consistency of the ML estimator is guaranteed under the stationarity conditions provided by Kheifets and Saikkonen (2020).

Theorem 1. *Under Assumptions 1–3 and under the regularity conditions provided by Basawa, Feigin, and Heyde (1976), the score vector has an asymptotic Normal distribution (Breusch and Pagan, 1980). As the score is normal and* $\mathbf{Z}_L'(\mathbf{I}_T - \mathbf{P}_\mathbf{X})\mathbf{Z}_L$ *is positive definite, the vectorized LM test statistic*

$$LM_L = vec\left(\hat{\mathbf{E}}'\mathbf{Z}_L\right)' \left\{\left(\mathbf{Z}_L'(\mathbf{I}_T - \mathbf{P}_\mathbf{X})\mathbf{Z}_L\right) \otimes \hat{\mathbf{\Omega}}\right\}^{-1} vec\left(\hat{\mathbf{E}}'\mathbf{Z}_L\right) \tag{10}$$

has an asymptotic χ^2*-distribution with* $Ln(1 + np)$ *degrees of freedom when the null hypothesis holds (Teräsvirta and Yang, 2014a).*

Proof. See Appendix B. □

Through some algebra, the test statistic in (10) can be written avoiding vectorization and Kronecker products as follows:

$$LM_L = \text{tr}\left\{\hat{\mathbf{\Omega}}^{-1}\hat{\mathbf{E}}'\mathbf{Z}_L\left[\mathbf{Z}_L'(\mathbf{I}_T - \mathbf{P}_\mathbf{X})\mathbf{Z}_L\right]^{-1}\mathbf{Z}_L'\hat{\mathbf{E}}\right\}. \tag{11}$$

The test can also be performed using the TR^2-form as explained in the Appendix C.

In this setting, the choice of L is somewhat arbitrary. Nevertheless, Teräsvirta and Yang (2014b) state that choosing L equal to 3 helps balancing the trade-off between the size properties of the tests. Accordingly, a third-order Taylor expansion is performed in this article.

4. Determining the number of regimes

Once rejected the null of linearity, the practitioner should account for the possible presence of some nonlinearity not gathered from a 2-regime model. This means that there may exist an additional nonlinear component that enters the model additively. In this article, we build upon the additive nonlinearity test introduced by Teräsvirta and Yang (2014a) which can also be used in a sequential procedure for the detection of the number of regimes. Following the findings by Bai and Perron (1998), Strikholm and Teräsvirta (2006) suggest the use of a sequential testing procedure for additive nonlinearity in the univariate framework. We here extend such a procedure in the multivariate framework to specify the number of regimes, m.

If the equations do not share the same transition variable, identifying the number of regimes is not straightforward and a suitable choice would be to select the minimum number of regimes identified in an equation-by-equation test. When a common transition variable is assumed throughout the system, the number of regimes can be identified from the following procedure which somehow extends in the multivariate framework the sequential test proposed by Strikholm and Teräsvirta (2006). We further discuss in Section 4.1 how this procedure can be applied also in the case of a VTAR model as the data-generating process. It should be recalled that the stationarity conditions provided in Kheifets and Saikkonen (2020) are generally valid for a single transition, but they are valid only if all the equations share the same shape parameter, γ, in the case of multiple transitions. This means that the consistency of the ML estimator, and the stability of the LM test results are guaranteed only for $H_0\colon\ m = 2$. Consequently, the tests can be only used to suggest the presence of at least three regimes.

The sequential testing procedure can start directly from the case of linearity testing against a 2-regime model. Hence, the first step of the procedure foresees the implementation of the linearity test shown in Eq. (11) to test the null hypothesis of $m = 1$ against $m = 2$. If H_0 is rejected at a given level, α, there could exist additive nonlinearity in the model. Therefore, the purpose of the practitioner may be sequentially testing for $m - 1$ versus m regimes until a nonrejection.

If we write Eq. (1) for $m = 3$ regimes as follows

$$\mathbf{y}_t = \mathbf{B}_1'\mathbf{x}_t + \mathbf{G}_t^{(1)}\mathbf{B}_2'\mathbf{x}_t + \mathbf{G}_t^{(2)}\mathbf{B}_3'\mathbf{x}_t + \boldsymbol{\varepsilon}_t, \tag{12}$$

testing for nonadditive nonlinearity equals to test $\text{H}_0\colon\ \gamma_{2,i} = 0,\ i = 1, \ldots, n$, against the alternative $H_1\colon\ \exists\gamma_{2,i} > 0$. Clearly, the test can be extended to a generic number of m regimes.

As for the linearity test in Section 3, the alternative model is not identified under the null hypothesis. Once again, Taylor's approximation of $\mathbf{G}_t^{(2)}$ allows us to overcome this problem and obtain a feasible test statistic. Using an L-order Taylor approximation, Eq. (12) becomes

$$\mathbf{y}_t = \mathbf{B}_1'\mathbf{x}_t + \mathbf{G}_t^{(1)}\mathbf{B}_2'\mathbf{x}_t + \left(\sum_{l=0}^{L} \Upsilon_l^{(2)} s_t^l + \mathbf{R}_t^{(2)}\right)\mathbf{B}_3'\mathbf{x}_t + \boldsymbol{\varepsilon}_t \tag{13}$$

where $\Upsilon_l^{(2)}$ is the diagonal matrix of coefficients of the L-order Taylor expansion of $g_{i,t}^{(2)}$. As for the linearity test, the null hypothesis implies $\Upsilon_l^{(2)} = \mathbf{0}$ for $l = 1, \ldots, L$. By reparametrizing, Eq. (13) can be written as

$$\mathbf{y}_t = \tilde{\mathbf{G}}_0'\mathbf{x}_t + \sum_{l=1}^{L} \tilde{\mathbf{G}}_l'\mathbf{x}_t s_t^l + \boldsymbol{\varepsilon}_t^* \tag{14}$$

where $\tilde{\mathbf{G}}_0 = \left(\mathbf{B}_1' + \mathbf{G}_t^{(1)}\mathbf{B}_2' + \Upsilon_0^{(2)}\mathbf{B}_3'\right)'$, $\tilde{\mathbf{G}}_l = \left(\Upsilon_l^{(2)}\mathbf{B}_3'\right)'$ and $\boldsymbol{\varepsilon}_t^* = \mathbf{R}_t^{(2)}\mathbf{B}_3'\mathbf{x}_t + \boldsymbol{\varepsilon}_t$.

The null hypothesis in the VAR in Eq. (14) is $\mathrm{H}_0\colon \tilde{\mathbf{G}}_1 = \ldots = \tilde{\mathbf{G}}_L = 0$. Let be $\mathbf{Y} = \left(\mathbf{y}_1', \ldots, \mathbf{y}_T'\right)'$, $\mathbf{X} = \left(\mathbf{x}_1', \ldots, \mathbf{x}_T'\right)'$, $\mathbf{E}$ the $T \times n$ matrix of residuals from Eq. (14), and $\mathbf{Z}_L$ specified as before, then model (14) can be written as

$$\mathbf{Y} = \mathbf{X}\tilde{\mathbf{G}}_0 + \mathbf{Z}_L\tilde{\mathbf{G}}_L + \mathbf{E}. \tag{15}$$

Let suppose that $\mathbf{K} = \left[\mathrm{vec}\left(\partial\tilde{\mathbf{G}}_1'\mathbf{B}'\mathbf{x}_1/\partial\boldsymbol{\theta}\right)' \ldots \mathrm{vec}\left(\partial\tilde{\mathbf{G}}_T'\mathbf{B}'\mathbf{x}_T/\partial\boldsymbol{\theta}\right)'\right]$ is the vectorized first order derivatives of $\tilde{\mathbf{G}}_t'\mathbf{B}'\mathbf{x}_t$ w.r.t to the vector of parameters $\boldsymbol{\theta}$ and that $\mathbf{P}_\mathbf{K} = \mathbf{K}(\mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'$, the test statistic can be computed similarly to the one in Section 3. Therefore, we have that the test statistic

$$LM_L = \mathrm{tr}\left\{\hat{\Omega}^{-1}\hat{\mathbf{E}}'\mathbf{Z}_L\left[\mathbf{Z}_L'\left(\mathbf{I}_T - \mathbf{P}_\mathbf{K}\right)\mathbf{Z}_L\right]^{-1}\mathbf{Z}_L'\hat{\mathbf{E}}\right\} \tag{16}$$

has an asymptotic χ^2-distribution with $Ln(1 + np)$ degrees of freedom under Assumptions 1–3 from Section 3, and under the assumption that $\mathbf{K}'\mathbf{K}$ and $\mathbf{Z}_L'(\mathbf{I}_T - \mathbf{P}_\mathbf{K})\mathbf{Z}_L$ are positive definite matrices (see Theorem 3 in Teräsvirta and Yang, 2014a, for further details).

The asymptotic distribution of the LM statistic has the desired null distribution only when $m = 2$ in testing $\mathbf{G}_t^{(m-1)} = (1/2)\mathbf{I}_n$. There are moment conditions for the asymptotic distribution theory to be valid (Eitrheim and Teräsvirta, 1996; Teräsvirta and Yang, 2014a). In the univariate case, a STAR model with logistic-type transition functions must satisfy the condition $E(\varepsilon_t^8) < \infty$. A sufficient condition in the multivariate case is $E(\varepsilon_{i,t}^8) < \infty$, for $i = 1, \ldots, n$. To compute K as stated before, the vector of first-order partial derivatives of $\tilde{\mathbf{G}}_t'\mathbf{B}'\mathbf{x}_t$ is necessary, see Appendix D in Teräsvirta and Yang (2014a) for further details. As for the linearity test, the no-additive nonlinearity test can also be performed using the TR^2-form in a multi-step regression problem as before, see Appendix C.

The LM test statistics can be used in a top-down sequential testing procedure that foresees testing the null hypothesis of $\gamma_{m-1,i} = 0$ for a growing number of regimes until nonrejection. Therefore, the number of regimes to be included in the model is the minimum for which the null hypothesis of no-additive nonlinearity cannot be rejected.

4.1. Applying the sequential procedure for a vector threshold autoregressive model

The tests presented in this study are also valid if the practitioner is analysing a VTAR model. Since the VLSTAR nests the VTAR for γ_{id} sufficiently large, the idea is to apply the tests directly on a VLSTAR approximation of the VTAR. This should also solve the drawback of finding the first derivative of the indicator function in the VTAR model or using a bootstrap procedure as proposed for the univariate framework in Giannerini, Goracci, and Rahbek (2024). Let's suppose a 2-regime VTAR model with a

single transition variable[2], a common threshold among the dependent variables and a not-switching error term, defined as follows

$$\mathbf{y}_t = \left(\boldsymbol{\mu}_1 + \sum_{j=1}^{p} \Phi_{1,j}\mathbf{y}_{t-j}\right) \mathbb{1}\left(s_t \leq c\right) + \left(\boldsymbol{\mu}_2 + \sum_{j=1}^{p} \Phi_{2,j}\mathbf{y}_{t-j}\right) \mathbb{1}\left(s_t > c\right) + \boldsymbol{\varepsilon}_t$$

which can alternatively be written as

$$\mathbf{y}_t = \boldsymbol{\mu}_1^* + \sum_{j=1}^{p} \Phi_{1,j}^*\mathbf{y}_{t-j} + \left(\mu_2^* + \sum_{j=1}^{p} \Phi_{2,j}^*\mathbf{y}_{t-j}\right) \mathbb{1}(s_t > c) + \boldsymbol{\varepsilon}_t. \tag{17}$$

It follows that the indicator function $\mathbb{1}(\cdot)$ in Eq. (17) can be approximated by a logistic function where the slope parameter γ is fixed and equal to a sufficiently large value. Consequently, the estimated parameters of the approximation are consistent under the same assumptions of the VLSTAR model (see also Luukkonen, Saikkonen, and Teräsvirta, 1988). This means that the aforementioned tests for a VLSTAR model can also be applied when the data-generating process is a VTAR.

Alternatively, the number of regimes in a VTAR model can be found extending to the multivariate the procedure proposed in Strikholm and Teräsvirta (2006), which will be denoted as the ST approach so on. Consider a single-lag three-regime VTAR model (where, for simplicity, we imposed $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}_3 = \mathbf{0}$)

$$\mathbf{y}_t = \left(\Phi_1\mathbf{y}_{t-1}\right) \mathbb{1}(s_t \leq c_1) + \left(\Phi_2\mathbf{y}_{t-1}\right) \mathbb{1}(c_1 < s_t \leq c_2) + \left(\Phi_3\mathbf{y}_{t-1}\right) \mathbb{1}(s_t > c_2) + \boldsymbol{\varepsilon}_t$$

which can be reformulated as

$$\mathbf{y}_t = \Phi_1^*\mathbf{y}_{t-1} + \left(\Phi_2^*\mathbf{y}_{t-1}\right) \mathbb{1}(s_t > c_1) + \left(\Phi_3^*\mathbf{y}_{t-1}\right) \mathbb{1}(s_t > c_2) + \boldsymbol{\varepsilon}_t \tag{18}$$

with $c_1 < c_2$. The sequential ST test procedure can be summarized in the following steps:

1. Set $\Phi_3^* = \mathbf{0}$ in Eq. (18) and approximate the indicator function $\mathbb{1}(s_t > c_1)$ with $\mathbf{G}_t^{(1)}(s_t; \gamma_1, c_1) = g_t^{(1)}(s_t; \gamma_1, c_1)\mathbf{I}_n$, where $g_t^{(1)}$ is a standard logistic function, such that

   $$\mathbf{y}_t = \Phi_1^*\mathbf{y}_{t-1} + \left(\Phi_2^*\mathbf{y}_{t-1}\right) \mathbf{G}_t^{(1)}(s_t; \gamma_1, c_1) + \boldsymbol{\varepsilon}_t.$$

   Then, apply the linearity test presented in Section 3 for a VLSTAR model, imposing $H_0\colon \gamma_1 = 0$.
2. If the null hypothesis is rejected at a given significance level α, estimate the coefficients in (18) model imposing $\Phi_3^* = \mathbf{0}$. As further detailed in Chan (1993) and Gonzalo and Pitarakis (2002), the threshold estimator, $\hat{c}_1$, is super consistent.
3. Use the super consistent (Seo, 2011) estimator, $\hat{c}_1$, in Eq. (18) and test the linearity of the following model

   $$\mathbf{y}_t = \Phi_1^*\mathbf{y}_{t-1} + \left(\Phi_2^*\mathbf{y}_{t-1}\right) \mathbb{1}(s_t > \hat{c}_1) + \left(\Phi_3^*\mathbf{y}_{t-1}\right) \mathbf{G}_t^{(2)}(s_t; \gamma_2, c_2) + \boldsymbol{\varepsilon}_t$$

   where $\mathbf{G}_t^{(2)}(s_t; \gamma_2, c_2) = g_t^{(2)}(s_t; \gamma_2, c_2)\mathbf{I}_n$.
4. If the null hypothesis, $H_0\colon \gamma_2 = 0$, is rejected at a significance level, estimate the parameters in Eq. (18) and use the super consistent estimates, $\hat{c}_1, \hat{c}_2$, to test the linearity of (18) against a four-regime model.
5. Continue the procedure until a nonrejection.

Since the convergence rates of the c_d, for $d = 1, \ldots, m-1$, are faster than the rates of the estimates of the other parameters (Seo, 2011), we can treat the thresholds as known and assume the model in Eq. (18) as a linear model. Therefore, this second approach for VTAR models only relies on the linearity test presented in Section 3. This, indeed, extends in the multivariate the sequential procedure for the definition of the number of thresholds proposed by Strikholm and Teräsvirta (2006).

[2] In our application to VTAR models, we restrict the specification to a single transition variable. This is in line with the standard definition of VTAR models in the literature (see Tsay, 1998), where a common threshold variable is used to capture abrupt regime switches across all equations.

5. Simulation study

In vector models, the standard LM-type tests can be strongly oversized when the null hypothesis foresees the estimation of a large set of parameters and when the size of the sample is not large. In practice, the nominal size of the test tends to overestimate the true probability of type I error in finite samples, see also Honda (1988). As in Laitinen (1978) and Meisner (1979), to overcome this limitation, we use a Bartlett-type correction that allows to rescale the degrees of freedom of the test and apply an F-statistic. In a Monte Carlo simulation study conducted by Bera, Byron, and Jarque (1981), the authors show that this correction is able to correct the oversize of LM.

The Laitinen-Meisner correction consists of a degree of freedom rescaling of the form $(nT - S)/(W \times nT)$, where n and T are defined as before, S is the number of parameters, and W is the number of restrictions, see Laitinen (1978) and Meisner (1979). The F-type LM test statistic, or rescaled LM test statistic, can be computed as

$$LM_L^{\text{resc}} = LM_L \cdot \frac{nT - S}{W \times nT} \tag{19}$$

and follows an $F(W, nT - S)$ distribution.

We carry out some simulation experiments for the finite-sample performance of the test procedure. Specifically, we investigate the empirical size and the selection frequencies for the sequential procedure. We also consider a Wilks' lambda test statistic based on Wilks' Λ-distribution (Anderson, 2003). In Appendix B of Teräsvirta and Yang (2014b), the authors show that Wilks' Λ is applicable in our testing situation and how the test is performed in this framework. To provide a benchmark for our sequential testing procedure, we also compute selection frequencies using a common alternative approach, the Akaike Information Criterion (AIC). This comparison allows us to evaluate the performance of our proposed method against a standard model selection tool used in practice.

5.1. Simulating from a VAR model

We first simulate from the following VAR(1) model with $n = 3$

$$\mathbf{y} = \mathbf{B}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{20}$$

where the parameter matrix, $\mathbf{B}$, has 0.1 entries and diagonal values ρ either equal to 0.5, 0.6, and 0.7, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$. A linear model would cause the sequential procedure to stop at the first step. To compute the empirical sizes, we generate 1000 replications from the VAR(1) model, using three sample sizes ($T = 400, 600, 1000$) and three different nominal sizes, $\alpha = 0.10, 0.05, 0.01$. To compute the TR^2-form of the test statistics, we use a third-order Taylor expansion, therefore $L = 3$, and we assume that the transition variable in the alternative model is the lag of the first dependent variable, $y_{1,t-1}$. In the model in Eq. (20), the vector of intercepts is equal to zero, but in practice one does not know this, therefore we estimate a model with intercepts to compute the test statistics.

Analysing the empirical sizes of the test in Table 1, it can be noticed that these are close to the nominal values in all of the three tests. When simulating $T = 1000$ observations, the empirical size of the LM test statistics (first three columns), the rescaled LM (columns four to six) and the Wilks' statistics (last three columns) almost coincides with the nominal size. The persistence level of the dependent variables seems to be slightly relevant, especially when $\rho = 0.7$, since a larger sample size is needed to approach the nominal sizes.

5.2. Simulating from a VLSTAR model

To apply the sequential procedure, we then generate 1000 replications from the model specified in Eq. (12). We do not include any explanatory variable and we select a single lag for the simulation of $n = 3$

Table 1. Empirical size of linearity test. The empirical size is in per cent based on 1000 replications from a VAR(1) model

	LM_3			LM_3^{resc}			Wilks		
	10%	5%	1%	10%	5%	1%	10%	5%	1%
$T = 400$									
$\rho = 0.5$	9.5	5.3	1.4	7.7	4.2	1.0	9.4	5.3	1.4
$\rho = 0.6$	10.9	5.6	1.0	8.7	4.2	0.7	10.6	5.5	1.0
$\rho = 0.7$	12.0	7.1	1.9	9.9	5.8	1.5	11.8	7.1	2.0
$T = 600$									
$\rho = 0.5$	9.9	5.3	1.4	8.6	4.7	1.1	9.6	5.3	1.4
$\rho = 0.6$	10.6	5.3	1.1	9.6	4.8	0.7	10.6	5.3	1.1
$\rho = 0.7$	11.5	5.8	1.5	10.5	4.8	1.2	11.5	5.7	1.6
$T = 1000$									
$\rho = 0.5$	11.6	6.8	0.9	10.2	6.0	0.7	11.5	6.8	1.0
$\rho = 0.6$	11.5	5.2	1.2	10.6	4.2	1.1	11.5	5.1	1.2
$\rho = 0.7$	11.1	5.4	1.0	10.4	4.7	0.9	10.9	5.3	1.1

Note: Data are simulated from a VAR(1) model and we test the null of $H_0: m = 1$ against $H_1: m = 2$. We simulate using different values of $\mathbf{B}$.

dependent variables, such that:

$$\mathbf{y}_t = \mathbf{B}_1'\mathbf{y}_{t-1} + \mathbf{G}_t^{(1)}\mathbf{B}_2'\mathbf{y}_{t-1} + \mathbf{G}_t^{(2)}\mathbf{B}_3'\mathbf{y}_{t-1} + \ldots + \mathbf{G}_t^{(m-1)}\mathbf{B}_{m-1}'\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{21}$$

where $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$, thus we are supposing uncorrelated errors. For each realization, we use the searching grid mechanism for the initial values of γ and c, we estimate the VLSTAR model and compute the residuals matrix. Since the transition function parameters in the VLSTAR model are estimated numerically, relatively large samples are required for a reasonable estimation accuracy, therefore we choose $T = 400, 600, 1000$.

In a first attempt, we compute the selection frequencies by simulating $\mathbf{y}_t$ from Eq. (21) with $m = 2$ and $\mathbf{G}_t^{(1)} = g_t(y_{1,t-1}; \gamma = 2, c = 0) \cdot \mathbf{I}_3$. The data-generating process (DGP) relies on a parameter matrix $\mathbf{B}_1$ with 0.1 entries and diagonal values equal to 0.7, while $\mathbf{B}_2 = (0.2 \cdot \mathbf{I}_3) - \mathbf{B}_1$ guarantees that the joint spectral radius of $\mathbf{B}_1$ and $\mathbf{B}_1 + \mathbf{B}_2$ (Kheifets and Saikkonen, 2020) is lower than 1. Moreover, we use $y_{1,t-1}$ as a transition variable. In the application of the sequential procedure, we always estimate a model with a vector of intercepts, even if they are not present in the DGP, and we let the model estimate possible different values of γ and c for each equation. We decided to use selection frequencies to evaluate the performance of our approach with simulated data for a threefold reason: first, their use aligns with the approach taken by Strikholm and Teräsvirta (2006) which is the work that inspired our study; second, a sequential test only proceeds if the null hypothesis is rejected in the first step, which means that if the procedure fails to reject the null of linearity when the DGP is the VLSTAR model with two regimes, empirical sizes and powers are irrelevant; third, selection frequencies capture not only the rate at which the correct number of regimes is identified but also the frequency with which the procedure tends to overestimate or underestimate the true number of regimes. This dual perspective is especially important in sequential testing, where errors in early stages can propagate through later steps.

The procedure starts with a linearity test against a two-regime model (using $y_{1,t-1}$ as a transition variable for the alternative model), then foresees testing $m = 2$ vs $m = 3$ regimes and continue until a nonrejection. As before, to compute the TR^2-form of the test statistics, we use a third-order Taylor expansion, therefore $L = 3$. The results are reported in Table 2.

For any sample size, the selection frequencies indicate a high tendency to identify two regimes correctly. As expected, increasing the sample size enhances the accuracy of the tests in identifying the correct number of regimes. In fact, the selection frequencies of $\hat{m} = 2$ are all above 99% for all the tests when $T = 1000$.

The selection frequencies based on the AIC are reported in Table 3. While the AIC correctly identifies two regimes as the most frequent choice, its accuracy is consistently lower than that of our sequential tests. For instance, at $T = 1000$, the selected number of regimes is equal to 2 in all the replications, whereas the AIC selects the correct model in only 91.9% of replications. Notably, the AIC's performance

Table 2. Selection frequencies for a VLSTAR model. The frequency is in per cent based on 1000 replications from model (21) with $m = 2$ regimes

	LM_3			LM_3^{resc}			Wilks		
	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
$T = 400$									
$\alpha = 0.10$	3.2	96.8	0.0	4.1	95.9	0.0	3.2	96.8	0.0
$\alpha = 0.05$	5.6	94.4	0.0	7.1	92.9	0.0	5.5	94.5	0.0
$\alpha = 0.01$	16.3	83.7	0.0	20.7	79.3	0.0	15.2	84.8	0.0
$T = 600$									
$\alpha = 0.10$	0.2	99.6	0.2	0.2	99.6	0.2	0.2	99.6	0.2
$\alpha = 0.05$	0.2	99.6	0.2	0.2	99.7	0.1	0.2	99.6	0.2
$\alpha = 0.01$	0.9	99.1	0.0	0.2	98.6	0.0	0.2	99.2	0.0
$T = 1000$									
$\alpha = 0.10$	0.0	99.9	0.1	0.0	99.9	0.1	0.0	99.9	0.1
$\alpha = 0.05$	0.0	99.9	0.1	0.0	99.9	0.1	0.0	99.9	0.1
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0

Note: Data are simulated from a model with $m = 2$ regimes. We simulate using $\mathbf{B}_1$ specified as a matrix with 0.1 entries and 0.7 diagonal values, and for $\mathbf{B}_2 = 0.2 \cdot \mathrm{I}_3 - \mathbf{B}_1$. Underlined values denote the identified number of regimes.

Table 3. Selection frequencies using AIC.

	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
$T = 400$	0.0	93.2	6.8
$T = 600$	0.0	94.3	5.7
$T = 1000$	0.0	91.9	8.1

Note: The frequency is in per cent based on 1000 replications from the VLSTAR model in Eq. (21) with $m = 2$ regimes.

does not improve with the sample size; in fact, its tendency to incorrectly select three regimes increases as T grows.

We then simulate from a VLSTAR model with $m = 3$ regimes and $n = 3$, using common $c_1 = -2$, $c_2 = 0.5$, $\gamma_1 = 2$ and $\gamma_2 = 40$ among the equations. In our simulation, $\mathbf{B}_1$ is a 3×3 parameter matrix with 0.1 entries and diagonal values equal to 0.7, while $\mathbf{B}_2 = (0.1 \cdot \mathbf{I}_3) - \mathbf{B}_1$ and $\mathbf{B}_3 = -0.7 \cdot \mathbf{I}_3$ guarantee that the joint spectral radius of $\mathbf{B}_1$, $\mathbf{B}_1 + \mathbf{B}_2$ and $\mathbf{B}_1 + \mathbf{B}_3$ is lower than 1 (Kheifets and Saikkonen, 2020). Moreover, we use $y_{1,t-1}$ as a transition variable, and we include the following constant vector terms: $\boldsymbol{\mu}_0 = \boldsymbol{\iota}$, $\boldsymbol{\mu}_1 = -2\boldsymbol{\iota}$ and $\boldsymbol{\mu}_2 = 2\boldsymbol{\iota}$, where $\boldsymbol{\iota}$ is a 3×1 vector of ones.

The results of the selection frequencies of the sequential procedure are reported in Table 4. In the case of a DGP with three regimes, the sample size strongly affects the results in terms of selection frequencies. In fact, when $T = 400$, none of the tests is capable of correctly identifying the third regime at any significance level. Conversely, when $T = 1000$, all the tests exhibit a selection frequency of three or more regimes higher than 85%. Recall that we do not go ahead with the sequential procedure to test $m = 3$ against $m = 4$ regimes, since the results for the stationarity of a VLSTAR model provided in Kheifets and Saikkonen (2020) are valid for two or more transitions only if the equations share a common γ.

Table 5 presents the AIC's selection frequencies when the true DGP has three regimes. In contrast to our sequential procedure, the AIC wrongly selects two regimes in all the scenarios. Even with $T = 1000$ observations, the AIC fails to identify the correct number of regimes in 88.6% of replications, revealing a strong tendency to underfit the model. This result highlights the potential pitfalls of relying on information criteria for this problem and underscores the value of the formal testing procedure proposed in this article.

5.3. Simulating from a VTAR model

Since we want to understand whether the sequential procedure introduced here is also valid for the case of an abrupt regime change, we compute the selection frequencies when data are simulated from a VTAR

Table 4. Selection frequencies for a VLSTAR model. The frequency is in per cent based on 1000 replications from model (21) with $m = 3$ regimes

	LM_3			LM_3^{resc}			Wilks		
	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
$T = 400$									
$\alpha = 0.10$	0.0	66.5	33.5	0.0	74.2	25.8	0.0	67.0	33.0
$\alpha = 0.05$	0.0	77.1	22.9	0.0	82.0	18.0	0.0	77.5	22.5
$\alpha = 0.01$	0.0	88.8	11.2	0.0	92.7	7.3	0.0	88.4	11.6
$T = 600$									
$\alpha = 0.10$	0.0	26.5	73.5	0.0	30.9	69.1	0.0	26.4	73.6
$\alpha = 0.05$	0.0	37.8	62.2	0.0	44.0	56.0	0.0	37.8	62.2
$\alpha = 0.01$	0.0	60.7	39.3	0.0	67.5	32.5	0.0	60.4	39.6
$T = 1000$									
$\alpha = 0.10$	0.0	1.7	98.3	0.0	2.1	97.9	0.0	1.7	98.3
$\alpha = 0.05$	0.0	3.1	96.9	0.0	3.3	96.7	0.0	3.1	96.9
$\alpha = 0.01$	0.0	9.4	90.6	0.0	10.8	89.2	0.0	9.3	90.7

Note: Data are simulated from a model with $m = 3$ regimes. We simulate using $\mathbf{B}_1$ specified as a matrix with 0.1 entries and 0.7 diagonal values, for $\mathbf{B}_2 = 0.1 \cdot \mathbf{I}_3 - \mathbf{B}_1$ and $\mathbf{B}_3 = -0.7 \cdot \mathbf{I}_3$. Underlined values denote the identified number of regimes.

Table 5. Selection frequencies using AIC.

	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
$T = 400$	0.0	94.4	5.6
$T = 600$	0.0	93.3	6.7
$T = 1000$	0.0	88.6	11.4

Note: The frequency is in per cent based on 1000 replications from the VLSTAR model in Eq. (21) with $m = 3$ regimes.

model with 2 regimes and $n = 3$ dependent variables, specified as follows

$$\mathbf{y}_t = \left(\Phi_1 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t < c) + \left(\Phi_2 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t \geq c) + \boldsymbol{\varepsilon}_t \tag{22}$$

where s_t is the lag of the first dependent variable, $y_{t-1,1}$. We set $\Phi_1 = \mathbf{B}_1$, where $\mathbf{B}_1$ is specified as in the previous section, we also set $\Phi_2 = (0.2 \cdot \mathbf{I}_3) - \boldsymbol{\Phi}_1$, the threshold c is set equal to 0, while $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$.

In Table 6, we compare the selection frequencies obtained using the sequential procedure based on the approximation of the indicator function through the smooth transition function (Panel A) with those obtained from the extension of the approach proposed by Strikholm and Teräsvirta (2006) to the multivariate (denoted by ST in Panel B of the table). The results in Panel A are consistent with those of the VLSTAR model, with even higher selection frequency rates at each combination of sample and nominal sizes, confirming that the sequential procedure is equally applicable and reliable for VTAR models. As with the VLSTAR model, the accuracy of the test procedure increases with sample size and significance levels. The extension of the ST approach to the multivariate in Panel B leads to even higher selection frequencies, which are equal to 100 for any sample and nominal size.

In line with these findings, when using the AIC computed from the VLSTAR model (panel A of Table 7), the selection frequencies of the right number of regimes are higher than 99% for any sample size. However, the AIC computed from the VTAR models wrongly indicates 3 regimes in almost 25% of the replications (panel B) with respect to the perfect percentages observed in Table 6.

Finally, we simulate the data from the following DGP with $m = 3$ regimes:

$$\mathbf{y}_t = \left(\boldsymbol{\mu}_1 + \Phi_1 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t \leq c_1) + \left(\boldsymbol{\mu}_2 + \Phi_2 \mathbf{y}_{t-1}\right) \mathbb{1}(c_1 < s_t \leq c_2) + \left(\boldsymbol{\mu}_3 + \Phi_3 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t > c_2) + \boldsymbol{\varepsilon}_t \tag{23}$$

where $c_1 = -2$, $c_2 = 0.5$, $\boldsymbol{\mu}_1 = \boldsymbol{\iota}$, $\boldsymbol{\mu}_2 = -2\boldsymbol{\iota}$, $\boldsymbol{\mu}_3 = 2\boldsymbol{\iota}$, with $\boldsymbol{\iota}$ being a 3×1 vector of ones, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$. As for the case of the VLSTAR model, we specify $\boldsymbol{\Phi}_1$ as a matrix with 0.1 entries and 0.7 diagonal values, we set $\boldsymbol{\Phi}_2 = (0.1 \cdot \mathbf{I}_3) - \boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_3 = -0.7 \cdot \mathbf{I}_3$.

The findings in Table 8 suggest that the approach using the VLSTAR model to test no-additive nonlinearity is never capable of identifying the real number of regimes correctly. Conversely, the ST approach shows a percentage selection rate of three regimes higher than 95% for any sample size,

Table 6. Selection frequencies for a VTAR model. The frequency is in per cent based on 1000 replications from model (22) with $m = 2$ regimes

	LM_3			LM_3^{resc}			Wilks		
	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} \geq 3$
Panel A: Indicator function approximation									
	$T = 400$								
$\alpha = 0.10$	0.0	99.6	0.4	0.0	99.8	0.2	0.0	99.6	0.4
$\alpha = 0.05$	0.0	99.9	0.1	0.0	100.0	0.0	0.0	99.9	0.1
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$T = 600$									
$\alpha = 0.10$	0.0	99.7	0.3	0.0	99.9	0.1	0.0	99.7	0.3
$\alpha = 0.05$	0.0	99.9	0.1	0.0	99.9	0.1	0.0	99.9	0.1
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$T = 1000$									
$\alpha = 0.10$	0.0	99.9	0.1	0.0	99.9	0.1	0.0	99.9	0.1
$\alpha = 0.05$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
Panel B: ST approach									
	$T = 400$								
$\alpha = 0.10$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.05$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
	$T = 600$								
$\alpha = 0.10$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.05$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
	$T = 1000$								
$\alpha = 0.10$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.05$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0
$\alpha = 0.01$	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0

Note: Data are simulated from a model with $m = 2$ regimes. We simulate using $\mathbf{\Phi}_1$ specified as a matrix with 0.1 entries and 0.7 diagonal values, and for $\mathbf{\Phi}_2 = 0.2 \cdot \mathbf{I}_3 - \mathbf{\Phi}_1$. Underlined values denote the identified number of regimes.

Table 7. Selection frequencies using AIC.

	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
Panel A: VLSTAR model			
$T = 400$	0.0	99.6	0.4
$T = 600$	0.0	99.3	0.7
$T = 1000$	0.0	99.2	0.8
Panel B: VTAR model			
$T = 400$	0.0	74.9	25.1
$T = 600$	0.0	77.5	22.5
$T = 1000$	0.0	75.6	24.4

Note: The frequency is in per cent based on 1000 replications from the VTAR model in Eq. (22) with $m = 2$ regimes.

nominal level and type of test. The explanation of these mixed results may lie in the fact that identifying multiple regimes requires accurately estimating multiple transition functions and their parameters. When the true DGP involves abrupt transitions as in a VTAR model, the smooth transition model may struggle with identification due to the overlap and interaction of the smooth transitions. This can lead to the estimation procedure failing to converge to the true parameters or not capturing the additional regimes. Finally, Table 9 reports the selection frequencies through AIC using the same 3-regime VTAR model. Conversely to what happens in the 2-regime DGP, the number of regimes is correctly selected using the AIC from the VTAR model (panel B) in all the replications. Instead, using the AIC from the VLSTAR model suggests a 2-regime model in most cases. Once again, the methodology introduced in this article seems to be more reliable and less volatile to the sample size and the number of regimes.

Table 8. Selection frequencies for a VTAR model. The frequency is in per cent based on 1000 replications from model (24) with $m = 3$ regimes

	LM_3			LM_3^{resc}			Wilks		
	$\hat{m}=1$	$\hat{m}=2$	$\hat{m}\geq 3$	$\hat{m}=1$	$\hat{m}=2$	$\hat{m}\geq 3$	$\hat{m}=1$	$\hat{m}=2$	$\hat{m}\geq 3$
Panel A: Indicator function approximation									
	$T=400$								
$\alpha=0.10$	0.0	92.6	7.4	0.0	93.5	6.5	0.0	92.8	7.2
$\alpha=0.05$	0.0	94.0	6.0	0.0	94.3	5.7	0.0	94.0	6.0
$\alpha=0.01$	0.0	94.5	5.5	0.0	94.7	5.3	0.0	94.5	5.5
$T=600$									
$\alpha=0.10$	0.0	94.0	6.0	0.0	95.0	5.0	0.0	94.0	6.0
$\alpha=0.05$	0.0	95.4	4.6	0.0	95.6	4.4	0.0	95.5	4.5
$\alpha=0.01$	0.0	96.6	3.4	0.0	96.7	3.3	0.0	96.6	3.4
$T=1000$									
$\alpha=0.10$	0.0	91.6	8.4	0.0	92.7	7.3	0.0	91.6	8.4
$\alpha=0.05$	0.0	94.3	5.7	0.0	95.5	4.5	0.0	94.5	5.5
$\alpha=0.01$	0.0	96.8	3.2	0.0	97.2	2.8	0.0	96.8	3.2
Panel B: ST approach									
	$T=400$								
$\alpha=0.10$	0.0	0.4	99.6	0.0	0.5	99.5	0.0	0.4	99.6
$\alpha=0.05$	0.0	0.7	99.3	0.0	0.8	99.2	0.0	0.7	99.3
$\alpha=0.01$	0.0	2.4	97.6	0.0	3.0	97.0	0.0	2.1	97.9
	$T=600$								
$\alpha=0.10$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0
$\alpha=0.05$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0
$\alpha=0.01$	0.0	0.3	99.7	0.0	0.3	99.7	0.0	0.2	99.8
	$T=1000$								
$\alpha=0.10$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0
$\alpha=0.05$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0
$\alpha=0.01$	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0

Note: Data are simulated from a model with $m = 3$ regimes. We simulate using $\mathbf{\Phi}_1$ specified as a matrix with 0.1 entries and 0.7 diagonal values, for $\mathbf{\Phi}_2 = 0.1 \cdot \mathbf{I}_3 - \mathbf{\Phi}_1$ and $\mathbf{\Phi}_3 = -0.7 \cdot \mathbf{I}_3$. Underlined values denote the identified number of regimes.

Table 9. Selection frequencies using AIC.

	$\hat{m}=1$	$\hat{m}=2$	$\hat{m}=3$
Panel A: VLSTAR model			
$T=400$	0.0	59.1	40.9
$T=600$	0.0	65.2	34.8
$T=1000$	0.0	71.0	29.0
Panel B: VTAR model			
$T=400$	0.0	0.0	100.0
$T=600$	0.0	0.0	100.0
$T=1000$	0.0	0.0	100.0

Note: The frequency is in per cent based on 1000 replications from the VTAR model in Eq. (22) with $m = 3$ regimes.

5.4. Using the sequential procedure for predictions

To further highlight the usefulness of properly selecting the number of regimes in nonlinear models, we also report a forecasting exercise with simulated data. Here, we simulate $T = 1000$ observations from the 3-regime VTAR model, and we split the observations into a training sample of $T_{train} = 800$ observations and a testing sample of $T_{test} = 200$ observations. We then use a rolling window scheme to produce one-step ahead predictions through a VAR, a 2-regime VTAR, and an m-regime VTAR model, where m is the number of regimes selected through the ST approach. At each temporal step, the models are re-estimated and the sequential procedure is used to select the ideal number of regimes. The idea of this section is to provide evidence that when the number of regimes is properly selected, the predictive accuracy is improved with respect to wrong (like the VTAR with two regimes, VTAR_2) or simpler models (as the VAR).

Table 10. Statistical evaluation of the predictive accuracy among the different predicting models.

Model	MSE_{y_1}	MSE_{y_2}	MSE_{y_3}	$\overline{MSE}$	Euclidean	GW
VAR	1.7324	3.0369	3.170942	2.6468	2.0060	-
$VTAR_2$	1.2150	1.7127	2.005671	1.6445	1.6172	-7.8373*
$VTAR_{ST}$	1.0883	0.9052	1.103506	1.0324	2.5478	-11.7155*

Note: $VTAR_2$ denotes a prediction from a 2-regime model, while $VTAR_{ST}$ refers to predictions from an m-regime VTAR model, where m is selected using the ST sequential procedure. $\overline{MSE}$ is the average among the three dependent variables for each forecasting method. The asterisk * indicates a significant result at a 1% significance level.

As in the previous section, we simulate the data for 3 variables from the following model:

$$\begin{aligned}\mathbf{y}_t = &\left(\boldsymbol{\mu}_1 + \Phi_1 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t \leq c_1) + \left(\boldsymbol{\mu}_2 + \Phi_2 \mathbf{y}_{t-1}\right) \mathbb{1}(c_1 < s_t \leq c_2) \\ &+ \left(\boldsymbol{\mu}_3 + \Phi_3 \mathbf{y}_{t-1}\right) \mathbb{1}(s_t > c_2) + \boldsymbol{\varepsilon}_t \end{aligned} \tag{24}$$

where $c_1 = -2$, $c_2 = 0.5$, $\boldsymbol{\mu}_1 = \boldsymbol{\iota}$, $\boldsymbol{\mu}_2 = -2\boldsymbol{\iota}$, $\boldsymbol{\mu}_3 = 2\boldsymbol{\iota}$, with $\boldsymbol{\iota}$ being a 3×1 vector of ones, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_3)$. Again, $\boldsymbol{\Phi}_1$ is specified as a matrix with 0.1 entries and 0.7 diagonal values, while $\boldsymbol{\Phi}_2 = (0.1 \cdot \mathbf{I}_3) - \boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_3 = -0.7 \cdot \mathbf{I}_3$.

The mean-squared errors (MSE) among the three variables, along with the average Euclidean distances, are reported in Table 10. In the table, we also report the Giacomini-White (GW) test (Giacomini and White, 2006) using the Euclidean distance as a loss function to assess statistical significance in the predictive accuracy and considering the predictions from the VAR as a reference. Recall that a negative and significant value indicates that the model in the row significantly outperforms the VAR in terms of predictive accuracy. It can be noticed that the lowest MSE for each of the time series in $\mathbf{y}_t$ is given by the VTAR model, whose regimes are selected using the sequential procedure presented in this article. When the DGP is nonlinear, the predictions from a linear model are the worst, as shown by the first row of Table 10. If we look at the GW test in the last column, we can observe that the VTAR models, both the 2-regime and the m-regime one, strongly and significantly outperform the VAR model in terms of predictive accuracy.

6. Empirical applications

6.1. Interest rate term structure

As an empirical application, we first apply the sequential test procedure to the U.S. monthly interest rates data already used in Tsay (1998). The dataset contains the 3-month treasury bill rates ($Y_{1,t}$) and 3-year treasury notes ($Y_{2,t}$) for the period from June 1953 to September 2022 ($T = 832$). These represent the short-term and intermediate-term series in the term structure of the interest rates. The theory of the term structure (Campbell and Shiller, 1991; Diebold and Li, 2006) suggests that these rates should move together over the long run, even if they diverge in the short term. Deviations from this long-run equilibrium, represented by the yield spread, are believed to contain crucial information about the future state of the economy.

Therefore, we begin our analysis by testing for cointegration between $\ln(Y_{1,t})$ and $\ln(Y_{2,t})$ using the Johansen test. This strongly rejects the null of no cointegrating vectors ($\lambda = 34.78 > 14.90$, where λ is the test statistics and 14.90 is the 5% critical value), and cannot reject the null of $r = 1$ cointegrating vector ($\lambda = 5.23 < 8.18$). The resulting estimated cointegrating vector is statistically indistinguishable from (1, -1), which implies that the yield spread, $\ln(Y_{1,t}) - \ln(Y_{2,t})$, represents the deviations from the long-run equilibrium. This provides a strong theoretical and empirical foundation for using the spread as a transition variable in a regime-switching model.

The key economic hypothesis of this application is that the dynamic adjustment of interest rates back to their long-run equilibrium is not linear. Instead, the speed and nature of this adjustment depend on the state of the economy, which is signaled by the yield spread itself. For instance, a positive spread is associated with economic expansions, where adjustment dynamics may be stable. An inverted spread is

a well-known (Harvey, 1988; Estrella and Mishkin, 1998) predictor of recessions, a state in which market participants may react more forcefully, leading to a faster and different adjustment dynamic.

This state-dependent behavior is a form of nonlinearity in the conditional mean that cannot be captured by a standard linear Vector Error Correction Model (VECM). The idea in the paper of Tsay (1998) was to see if the nonlinear error correction mechanism in the interest rates could be successfully replaced by a regime-switching mechanism (Teräsvirta, Tjøstheim, and Granger, 2010). Therefore, the nonlinear models presented in this section are applied to log-differenced time series, supposing that the conditional mean follows a regime-switching mechanism. Here, we are mainly focusing on understanding whether the number of detected regimes is somehow coherent with the number of regimes supposed in the paper of Tsay (1998). Still, we compare the in-sample mean squared error of the VECM applied on $\ln(Y_{1,t})$ and $\ln(Y_{2,t})$ (to guarantee that the prediction is on growth rates) with the one obtained from both a VLSTAR and VTAR model with the number of regimes determined by the sequential procedure presented in this article, to understand if the correct identification of the number of regimes has provided a practical benefit.

As aforementioned, to obtain weakly stationary time series, the data have been considered as growth rates, *i.e.*, $\mathbf{y}_t = \left(y_{1,t}, y_{2,t}\right)'$, with $y_{i,t} = \ln(Y_{i,t}) - \ln(Y_{i,t-1})$ for $i = 1, 2$. The plots of the time series are shown in Fig. 1.

On the basis of what stated before, as a candidate transition variable, we select the maturity spread computed as $x_t = \ln(Y_{1,t}) - \ln(Y_{2,t})$ since, according to the inverted yield curve theory (Harvey, 1988), this would reflect the business cycle of the U.S. economy. Following Tsay (1998), to avoid random fluctuations in the interest rates term structure, we use the 3-month moving average of x_t as a transition variable, therefore $s_t = (x_t + x_{t-1} + x_{t-2})/3$ (see the green line in Fig. 1). To select the VLSTAR lag length, we use AIC and BIC criteria which suggest a single lag specification, *i.e.*, $p = 1$.

The results of the tests are reported in Table 11. It can be noticed in the first column that all the tests strongly reject the null hypothesis of linearity at any significance level. This means that, in line with what was supposed and proved by Tsay (1998), a nonlinearity is present in the dynamics of the interest rates and the top-down procedure for the selection of the number of regimes presented in this article can be applied. The procedure foresees testing the null hypothesis of $m = 2$ regimes against the alternative of $m = 3$. Once again, the null hypothesis is rejected in all the test statistics (second and third columns of the table) introduced in this article. According to the results from the test statistics, a 2-regime model is not enough to consider all the nonlinearity in the model. Consequently, it can be deduced that the optimal number of regimes for these time series is greater than two, which is also what was originally supposed in Tsay (1998). We also report the values of the AIC in Table 12. When computed on the estimated VLSTAR, the AIC suggests a 2-regime model, while the estimated VTAR suggests a 3-regime model.

To assess whether our nonlinear approach offers an improvement over standard linear methods for cointegrated series, we compare its in-sample fit to that of a benchmark linear VECM. The linear VECM yields an MSE of 0.0483. In contrast, the 3-regime VLSTAR model selected by our procedure achieves a lower MSE of 0.0471, while the 3-regime VTAR model exhibits an MSE of 0.0434. This demonstrates that our approach, using a statistically appropriate testing procedure on stationary variables, with a transition variable motivated by the system's long-run properties, not only identifies significant nonlinearity but also provides a better fit of the data than a standard linear model.

6.2. River flows data

As a second empirical example, the linearity and no remaining nonlinearity tests have been applied to the daily Icelandic river flow data for the period from 1972 to 1974. The time series in this dataset include river flows in cubic meters per second for two rivers, the Jökulsá and the Vatndalsá, as well as the temperature and the precipitation, see Fig. 2. It may be noticed from the figure that the flow is stronger in the spring, when the snow melts, and severely slows down in summer. Spring seems to hit

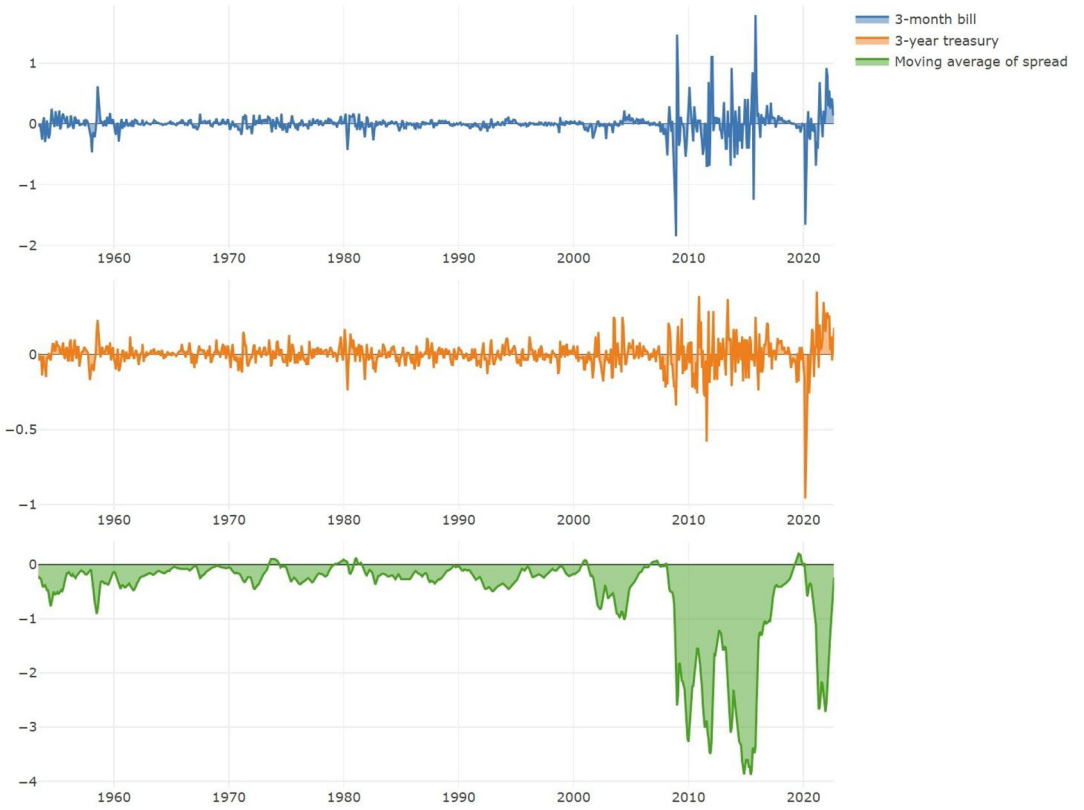

Figure 1. Time series plots of U.S. interest rates.
Note: The time series of the monthly 3-bill rate growth (blue line) and the 3-year Treasury rate growth (orange line) highlight the possible existence of several regimes. The 3-month moving average of the maturity spread (green line) is used as a transition variable.

Table 11. Tests for linearity and additive nonlinearity in the empirical application with interest rates

	$H_0: m = 1$	$H_0: m = 2$	
	Linearity test	Smooth transition	ST approach
LM_3	105.307 * ($p<0.001$)	67.146 * ($p<0.001$)	78.228 * ($p<0.001$)
LM_3^{resc}	5.801 * ($p<0.001$)	3.667 * ($p<0.001$)	4.309 * ($p<0.001$)
Wilks	107.982 * ($p<0.001$)	67.900 * ($p<0.001$)	79.600 * ($p<0.001$)

Note: In the first column the linearity test is reported, in column two we test $H_0: m = 2$ versus $H_1: m = 3$ regimes using the no-additive nonlinearity test for a VLSTAR, while column three reports the multivariate version of the sequential procedure proposed by Strikholm and Teräsvirta (2006), denoted by "ST". The asterisk identifies a significant result at a significance level $\alpha = 0.01$, while p-values are reported in brackets.

Table 12. AIC in the empirical application with different rates.

	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
AIC_{VLSTAR}	-2120.226	-5118.166	-5051.915
AIC_{VTAR}	-2120.226	-6902.252	-6955.521

Vatndalsá's flow more than Jökulsá, probably due to the fact that the drainage area of the latter contains a glacier which smooths the river's flow. Accordingly, temperature may somehow drive the regime changes of the conditional mean of a multivariate model. For these reasons, several former applications have

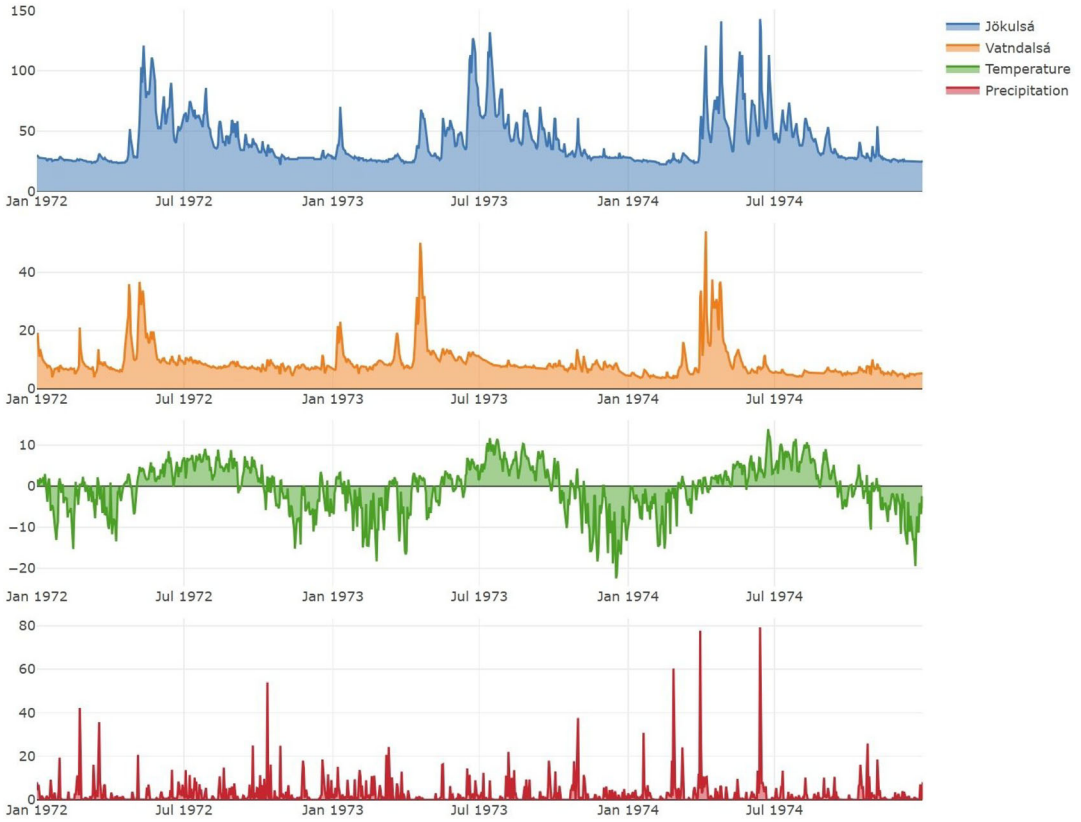

Figure 2. Time series plots of Icelandic river flow data.
Note: The time series of the Jökulsá (blue line) and the Vatndalsá (orange line) river flows highlight a nonlinear path for both rivers. The temperature (green line) and the precipitation level (red line) can be used as a transition variable.

used these time series to test regime-switching nonlinear models. For instance, Tong, Thanoon, and Gudmundsson (1985) use a univariate threshold model to estimate their relationship with temperature and precipitation, Chen and Tsay (1993) test their nonlinear additive AR model on the same data, while Tsay (1998), Teräsvirta and Yang (2014b), and Livingston Jr and Nur (2020) apply a multivariate nonlinear model.

Following Tsay (1998) and Teräsvirta and Yang (2014b), we first select the lagged temperature as the transition variable for both flow equations (Panel A of Table 13) and we compute the sequential procedure to identify the number of regimes in these series. According to the linearity test statistics, the null hypothesis of linearity is strongly rejected. The procedure introduced in this article foresees to use the additive nonlinearity test on an increasing number of regimes when the null of linearity is rejected. Therefore, we perform the additive nonlinearity test when the null hypothesis is $H_0: m = 2$. The null of $m = 2$ is rejected at any significance level in all the tests using the ST approach (third column), while the test based on the VLSTAR model is not capable of rejecting the null hypothesis. Based on these results, the top-down sequential procedure using the no-additive nonlinearity test for a VLSTAR model leads to 2 regimes, while the multivariate version of the ST method points to $m \geq 3$ regimes as the optimal number of regimes for the Icelandic rivers' time series, which is in line with what found in Teräsvirta and Yang (2014b). This discrepancy between the two approaches may be due to the fact that in the VLSTAR-based sequential procedure, we let γ and c be different for each equation, which may lead to potentially different results in terms of number of identified regimes. Moreover, when the transition

Table 13. Tests for linearity and additive nonlinearity in the empirical application with river flows

	Panel A: Lagged temperature as the transition variable		
	$H_0: m = 1$	$H_0: m = 2$	
	Linearity test	Smooth transition	ST approach
LM_3	274.926 * ($p<0.001$)	13.643 (0.752)	122.018 * ($p<0.001$)
LM_3^{resc}	15.176 * ($p<0.001$)	0.748 (0.763)	6.735 * ($p<0.001$)
Wilks	307.672 * ($p<0.001$)	13.592 (0.755)	127.185 * ($p<0.001$)
	Panel B: Lagged precipitation as the transition variable		
	$H_0: m = 1$	$H_0: m = 2$	
	Linearity test	Smooth transition	ST approach
LM_3	84.763 * ($p<0.001$)	41.218* (0.001)	36.153* (0.007)
LM_3^{resc}	4.679 * ($p<0.001$)	2.261 * (0.002)	1.996 * (0.008)
Wilks	85.794 * ($p<0.001$)	41.273* (0.001)	36.158* (0.007)

Note: In the first column the linearity test is reported, while in column two we test $H_0: m = 2$ versus $H_1: m = 3$ regimes. The test statistics with the subscript "ST" denotes the multivariate version of the sequential procedure proposed by Strikholm and Teräsvirta (2006). The asterisk identifies a significant result at a significance level $\alpha = 0.01$, while p-values are reported in brackets.

Table 14. AIC in the empirical application with river flows

	Panel A: Lagged temperature as the transition variable		
	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
AIC_{VLSTAR}	12096.007	7733.896	7670.739
AIC_{VTAR}	12096.007	5584.780	5796.052
	Panel B: Lagged precipitation as the transition variable		
	$\hat{m} = 1$	$\hat{m} = 2$	$\hat{m} = 3$
AIC_{VLSTAR}	12096.007	8163.618	8106.787
AIC_{VTAR}	12096.007	5791.642	5703.663

variable smoothly moves between possible regimes, like the temperature (see the green line in Fig. 2), distinguishing between regimes using a smooth transition function could be harder.

We also apply the tests using the lagged precipitation as a transition variable (Panel B of Table 13). As for the case of the lagged temperature as a transition variable, the null hypothesis of linearity is rejected for all the tests. Once rejected the hypothesis of a linear model, we conduct the top-down procedure for the selection of the number of regimes. In this case, both the no-additive nonlinearity tests and linearity tests in the ST approach reject the null hypothesis of $m = 2$ regimes at 1% significance level. This means that, with the lagged precipitation as a transition variable, the optimal model may be at least a 3-regime model. The selected model through AIC slightly differs from what was selected through our sequential procedure, see Table 14. In fact, when using the lagged temperature as a transition variable, the AIC indicates a VLSTAR with three regimes and a VTAR with two regimes. Instead, in accordance with what found in Table 13, a three-regime model is at least necessary when using lagged precipitation as a transition variable.

7. Conclusions

In this article, we developed a simple method for selecting the number of regimes in multivariate nonlinear models with no restrictions on the number of dependent variables and transitions, also giving a more formal context for the linearity and no additional nonlinearity tests introduced in Teräsvirta and Yang (2014a).

The finite-sample properties of the tests highlight that the empirical sizes of the linearity test well approximate the nominal sizes, no matter what the persistence level and sample size. The analysis of the selection frequencies when simulating from multiple-regime models highlights that the sequential

procedure is capable of correctly identifying the number of regimes for any sample size. Furthermore, our simulation study underlines that the sequential procedure introduced in this article can be applied either to detect the number of regimes in smoothly changing time series and abrupt regime-changing time series. Nevertheless, when the DGP has two regime switches, the smooth transition approximation of the indicator function in a VTAR model somehow affects the ability of the sequential procedure to identify the correct number of regimes, which is instead properly detected by the ST approach. Finally, when we apply the sequential test procedure to real data, we can observe that the procedure identifies a number of regimes consistent with findings from previous studies.

Despite the promising results, certain challenges remain. One key issue is the lack of general stationarity and ergodicity conditions for the VLSTAR model, particularly in the presence of multiple regimes. While conditions exist for specific cases, as noted by Saikkonen (2008) and Kheifets and Saikkonen (2020), a comprehensive theoretical framework is still needed. This limitation implies that the asymptotic properties of the test are guaranteed only under specific circumstances, such as a single-transition null hypothesis. Moreover, our approach employs a sequential test that begins with testing the null hypothesis of linearity (*i.e.*, a single–regime model) and then proceeds to test for additional regimes only if the null is rejected. In this framework, the overall error rate is affected by the fact that multiple tests are conducted sequentially. In principle, if the individual tests were independent, one could relate the overall significance level to the individual levels through a Bonferroni-like correction. However, because the tests in our sequential procedure are not independent (errors at an early stage propagate into subsequent tests), the overall type I error rate does not simply equal the nominal level chosen for each individual test. While it is clear that the sequential nature of the procedure can lead to an accumulation of testing errors, our simulation evidence, alongside previous work (*e.g.*, Strikholm and Teräsvirta, 2006 and Leeb and Pötscher, 2005), suggests that, for realistic sample sizes and under common data-generating scenarios, the overall performance remains acceptable. We acknowledge that the optimal way to choose the size in a sequential setting remains an open question, and further work could explore adjustments to better control the overall type I error rate. Finally, our framework focuses on the conditional mean, but a joint test for nonlinearity in the conditional variance as well could be considered for future work. However, this is a complex task, as it would introduce unidentified nuisance parameters under the null hypothesis for both the mean and variance components, complicating inference (Teräsvirta, Tjøstheim, and Granger, 2010). It should be recalled that, while our framework does not explicitly model shifts in volatility, the LM test for the conditional mean remains consistent even in the presence of GARCH-type errors (Hansen, 1996).

A possible implementation for future research could also foresee the use of the procedure to detect the number of structural breaks in a multivariate linear model. In fact, if the transition variable is a temporal trend, the VLSTAR model becomes a time-varying parameter model, and the changes in regimes coincide with smooth structural breaks. Moreover, exploring the applicability of the sequential testing procedure in high-dimensional problems or using exogenous variables as a transition variable could enhance its utility. Another promising extension is the adaptation of our procedure for the real-time monitoring of the number of regimes. Sequentially applying the test as new observations arrive would allow for the dynamic identification of structural changes, which is crucial for forecasting and interpretation.

Acknowledgments

I offer my sincerest gratitude to Timo Tërasvirta who has contributed to the early version of this article, and whose comments have helped me improve the theoretical background behind the sequential procedure. I also thank the participants of the International Association for Applied Econometrics (IAAE 2025) for their valuable suggestions. Moreover, I thank the Editor-in-Chief, the Associate Editor, and the referees for their suggestions and comments which, I believe, have led to an improved and clearer version of the manuscript. I retain the responsibility for any errors and shortcomings in this work.

Funding

This article is the result of the project "Monitoring risks in financial markets", code 2022NEL482, CUP D53D23008350001, funded by the PRIN 2022 program, financed by the European Union-Next Generation EU-PNRR Mission 4 Component 2 Investment 1.1.

Data availability

The author confirms that the data and the R code supporting the findings of this study are available at https://github.com/andbucci/Sequentialregimes.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. 3rd ed. New York: Wiley.

Bai, J., Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* 66(1):47. doi: 10.2307/2998540

Basawa, I., Feigin, P., Heyde, C. (1976). Asymptotic properties of maximum likelihood estimators for stochastic processes. *Sankhyā, A* 38(3):259–270.

Bera, A. K., Byron, R. P., Jarque, C. M. (1981). Further evidence on asymptotic tests for homogeneity and symmetry in large demand systems. *Economics Letters* 8(2):101–105. doi: 10.1016/0165-1765(81)90001-X

Breusch, T. S., Pagan, A. R. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *The Review of Economic Studies* 47(1):239 doi: 10.2307/2297111

Caggiano, G., Castelnuovo, E., Figueres, J. M. (2017). Economic policy uncertainty and unemployment in the united states: A nonlinear approach. *Economics Letters* 151:31–34. doi: 10.1016/j.econlet.2016.12.002

Camacho, M. (2004). Vector smooth transition regression models for US GDP and The composite index of leading indicators. *Journal of Forecasting* 23(3):173–196. doi: 10.1002/for.912

Campbell, J. Y., Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *The Review of Economic Studies* 58(3):495 doi: 10.2307/2298008

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics* 21:520–533.

Chen, R., Tsay, R. S. (1993). Nonlinear additive ARX models. *Journal of the American Statistical Association* 88(423):955–967. doi: 10.1080/01621459.1993.10476363

Cho, J. S., Ishida, I., White, H. (2011). Revisiting tests for neglected nonlinearity using artificial neural networks. *Neural Computation* 23(5):1133–1186. doi: 10.1162/NECO_a_00117

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 74(1):33–43. doi: 10.1093/biomet/74.1.33

Diebold, F. X., Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics* 130(2):337–364. doi: 10.1016/j.jeconom.2005.03.005

Eitrheim, Ø., Teräsvirta, T. (1996). Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics* 74(1):59–75. doi: 10.1016/0304-4076(95)01751-8

Estrella, A., Mishkin, F. S. (1998). Predicting U.S. Recessions: financial variables as leading indicators. *Review of Economics and Statistics* 80(1):45–61. doi: 10.1162/003465398557320

Giacomini, R., White, H. (2006). Tests of conditional predictive ability. *Econometrica* 74(6):1545–1578. doi: 10.1111/j.1468-0262.2006.00718.x

Giannerini, S., Goracci, G., Rahbek, A. (2024). The validity of bootstrap testing for threshold autoregression. *Journal of Econometrics* 239(1):105379. doi: 10.1016/j.jeconom.2023.01.004

Gonzalo, J., Pitarakis, J.-Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *Journal of Econometrics* 110(2):319–352. doi: 10.1016/S0304-4076(02)00098-2

Hamilton, J. D. (2016). *Chapter 3- Macroeconomic Regimes and Regime Shifts*, Vol. 2. Amsterdam, Netherlands: Elsevier, pp. 163–201.

Hansen, B. (1999). Testing for linearity. *Journal of Economic Surveys* 13(5):551–576. doi: 10.1111/1467-6419.00098

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64(2):413. doi: 10.2307/2171789

Harvey, C. R. (1988). The real term structure and consumption growth. *Journal of Financial Economics* 22(2):305–333. doi: 10.1016/0304-405X(88)90073-6

He, C., Teräsvirta, T., González, A. (2008). Testing parameter constancy in stationary vector autoregressive models against continuous change. *Econometric Reviews* 28(1-3):225–245. doi: 10.1080/07474930802388041

Hendry, D. F., Mizon, G. E. (1998). Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK. *Empirical Economics* 23(3):267–294. doi: 10.1007/s001810050021
Honda, Y. (1988). A size correction to the lagrange multiplier test for heteroskedasticity. *Journal of Econometrics* 38(3):375–386. doi: 10.1016/0304-4076(88)90052-8
Hubrich, K., Teräsvirta, T. (2013). *Thresholds and Smooth Transitions in Vector Autoregressive Models*, Advances in Econometrics, in: VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims. Emerald Group Publishing Limited. Vol. 32. doi: 10.1108/S0731-9053(2013)0000031008
Jiang, Y., Wang, G.-J., Ma, C., Yang, X. (2021). Do credit conditions matter for the impact of oil price shocks on stock returns? evidence from a structural threshold VAR model. *International Review of Economics & Finance* 72:1–15. doi: 10.1016/j.iref.2020.10.019
Kheifets, I. L., Saikkonen, P. J. (2020). Stationarity and ergodicity of vector STAR models. *Econometric Reviews* 39(4):407–414. doi: 10.1080/07474938.2019.1651489
Laitinen, K. (1978). Why is demand homogeneity so often rejected?. *Economics Letters* 1(3):187–191. doi: 10.1016/0165-1765(78)90022-8
Lanne, M., Virolainen, S. (2024). A Gaussian smooth transition vector autoregressive model: An application to the macroeconomic effects of severe weather shocks. *Journal of Economic Dynamics and Control* 178:105162. doi: https://dx.doi.org/10.1016/j.jedc.2025.105162
Leeb, H., Pötscher, B. M. (2005). Model selection and inference: facts and fiction. *Econometric Theory* 21(1):21–59. doi: 10.1017/S0266466605050036
Livingston, G., Nur, D. (2020). Bayesian estimation and model selection of a multivariate smooth transition autoregressive model. *Environmetrics* 31(6):2615.
Lütkepohl, H. (1996). *Handbook of Matrices*. New York: Wiley.
Luukkonen, R., Saikkonen, P., Teräsvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* 75(3):491–499. doi: 10.1093/biomet/75.3.491
Meisner, J. F. (1979). The sad fate of the asymptotic slutsky symmetry test for large systems. *Economics Letters* 2(3):231–233. doi: 10.1016/0165-1765(79)90027-2
Saikkonen, P. (2008). Stability of regime switching error correction models under linear cointegration. *Econometric Theory* 24(1):294–318.
Seo, M. H. (2011). Estimation of nonlinear error correction models. *Econometric Theory* 27(2):201–234.
Seong, D., Cho, J. S., Teräsvirta, T. (2022). Comprehensively testing linearity hypothesis using the smooth transition autoregressive model. *Econometric Reviews* 41(8):966–984. doi: 10.1080/07474938.2022.2091713
Sims, C. A. (1980). Macroeconomics and reality. *Econometrica* 48(1):1. doi: 10.2307/1912017
Strikholm, B., Teräsvirta, T. (2006). A sequential procedure for determining the number of regimes in a threshold autoregressive model. *The Econometrics Journal* 9(3):472–491. doi: 10.1111/j.1368-423X.2006.00194.x
Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89(425):208–218. doi: 10.1080/01621459.1994.10476462
Teräsvirta, T., Tjøstheim, D., Granger, C. W. (2010). *Modelling Nonlinear Economic Time Series*. Oxford: Oxford University Press.
Teräsvirta, T., Yang, Y. (2014a). Linearity and misspecification tests for vector smooth transition regression models, CREATES Research paper, Aarhus University, Aarhus, Denmark.
Teräsvirta, T., Yang, Y. (2014b). Specification, estimation and evaluation of vector smooth transition autoregressive models with applications, Technical report, Department of Economics and Business Economics, Aarhus University. CREATES Research Paper, Aarhus University, Aarhus, Denmark.
Tong, H., Thanoon, B., Gudmundsson, G. (1985). Threshold time series modeling of two icelandic riverflow systems. *JAWRA Journal of the American Water Resources Association* 21(4):651–662. doi: 10.1111/j.1752-1688.1985.tb05380.x
Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* 93(443):1188–1202. doi: 10.1080/01621459.1998.10473779
Wang, G., Zhu, K. E., Shao, X. (2022). Testing for the martingale difference hypothesis in multivariate time series models. *Journal of Business & Economic Statistics* 40(3):980–994. doi: 10.1080/07350015.2021.1889568
Watson, M. W., Engle, R. F. (1985). Testing for regression coefficient stability with a stationary AR(1) alternative. *The Review of Economics and Statistics* 67(2):341. doi: 10.2307/1924737
White, H., Cho, J. S. (2012). Higher-order approximations for testing neglected nonlinearity. *Neural Computation* 24(1):273–287. doi: 10.1162/NECO_a_00225