ELSEVIER

Journal of Econometrics 109 (2002) 239–273

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Misspecified Structural Change, Threshold, and Markov-switching models

Marine Carrasco *

Department of Economics, University of Rochester, Harkness Hall 226, Rochester, NY 14627, USA

Abstract

Sudden perturbations of a large amplitude occur frequently in macroeconomic and financial time series. A usual practice is to test linearity against a permanent structural change. However, changes can also be captured by nonlinear stationary models such that Threshold and Markov-switching models. In this paper, we show that tests designed for a threshold alternative have also power against parameter instability originating from Structural Change or Markov-switching models. On the other hand, it is shown that tests for structural change have no power if the data are generated by a Markov-switching or Threshold model. Therefore, it appears that testing the null of parameter stability against a threshold alternative is a robust way to detect parameter instability in economic and financial time series. A Monte Carlo analysis based on several models studied in the literature illustrates how the tests perform in practice.

JEL classification: C12; C22

Keywords: Asymptotic theory; Hypothesis testing; Nonlinear models; Nuisance parameter

1. Introduction

Economic time series often exhibit sudden changes of a large amplitude, as a result of a technological change, a change of policy, a supply shock, a financial crash or extreme events such as a war. Linear models such as Gaussian autoregressive moving-average (ARMA) models fail to exhibit such dramatic changes. Therefore, several nonlinear dynamic models have been recently proposed. One can oppose the Structural Change model where the change is permanent to models subject to cycles (Markov-switching

* Tel.: +1-585-275-7307; fax: +1-585-256-2309.
E-mail address: csco@troi.cc.rochester.edu (M. Carrasco).

PII: S0304-4076(02)00112-4

and Threshold models). We consider autoregressive versions of these three models:

$$y_t = (\mu_0 + \mu_1 y_{t-1} + \cdots + \mu_l y_{t-l}) + (\mu_0^* + \mu_1^* y_{t-1} + \cdots + \mu_l^* y_{t-l})\Delta_{tT} + \varepsilon_t,$$
$$\varepsilon_t \sim \text{i.i.d.}(0, \sigma^2), \tag{1}$$

where Δ_{tT} is an indicator variable that takes two values, 0 (low regime) and 1 (high regime) whose specification varies across models. The Structural Change model denoted SCA (Andrews, 1993) exhibits a jump at an exogenous date. $\Delta_{tT} = 0$ before a change point and $\Delta_{tT} = 1$ afterward. In the Markov-switching model denoted MSA (Hamilton, 1989, 1994), changes are driven by on unobservable exogenous Markov chain, S_t. For this model, $\Delta_{tT} = S_t$. The Threshold model denoted TAR (Tong, 1990) induces asymmetric, periodic behavior including an amplitude–frequency dependence. In this case, $\Delta_{tT} = 1$ if y_{t-d} was less than a threshold value and $\Delta_{tT} = 0$ if y_{t-d} was greater than the threshold. Contrary to Gaussian ARMA models, these three models are irreversible (Tong, 1990, pp. 12, 197). This is an attractive feature since most financial and macroeconomic time series are characterized by an asymmetry between the upward and downward movements. These nonlinear models have been widely applied in economics. [1]

Policy debates depend on whether a macroeconomic or financial time series is best characterized by a linear or a nonlinear model. Using a linear model with fixed parameters to describe a macroeconomic time series may lead to a wrong quantitative assessment of policy effects if structural changes occur during the period of study. Therefore, many papers concerned with macroeconomic or financial modelling have focused on testing for linearity. Often, econometricians test linearity by testing structural change in the mean or slope parameters. However, we know since Andrews (1993, p. 826) that such tests will have no power if the true alternative is a stationary model. Our paper illustrates that the supWald, supLM and supLR tests do not have power against two particularly relevant alternatives which are the Threshold and Markov switching models. Moreover, we show that when the DGP is either SCA or MSA, the test designed for a threshold alternative has asymptotic power equal to one. It has an important implication, namely that in large samples, the TAR model detects the presence of a shift whatever its nature is. Therefore, this test could be used as pre-test to detect parameter instability. We note, however, that this test has power against local alternative in $T^{1/4}$ but not $T^{1/2}$; therefore, it will not be able to detect small changes.

In the three models under consideration, testing for parameter stability is problematic because of the presence of nuisance parameters which are not identified under the null hypothesis. These nuisance parameters are the following: (i) the timing of the change in the Structural Change model, (ii) the transition probabilities in the Markov-switching model, (iii) the value of the threshold in the Threshold model. Testing procedures of these models have been provided by Davies (1987), Andrews (1993), Andrews and Ploberger (1994), and Hansen (1996) among others. We focus our attention on sup tests

[1] See Hansen (1996), Potter (1995), Koop and Potter (1999, 2001) for the Threshold model, Cecchetti et al. (1990), Garcia and Perron (1996), Hamilton (1989), Kaminsky (1993), and Raymond and Rich (1997) for the Markov-switching model, Banerjee et al. (1992), Perron (1990), Stock and Watson (1996) for the Structural Change model, among others.

initially proposed by Davies. We provide an analytical expression of the asymptotic distributions of the tests under misspecification and show how it is affected by the value of the parameters of the DGP. The focus of our paper is closely related to that of Koop and Potter (1999, 2001). They investigate by Monte Carlo how their linearity tests behave against another alternative than that for which it was designed; they also point out the difficulty to discriminate between TAR and MSA. However, their tests are Bayesian and differ from ours that are classical. Bayesian tests allow them to treat the MSA case that we are not able to address here. [2]

The paper is organized as follows: Section 2 presents the models and describes the testing method. Section 3 derives analytically the asymptotic distribution of the tests when the models are misspecified. Section 4 presents Monte Carlo experiments. Section 5 concludes. All the technical proofs are in the appendix.

Throughout the paper, the symbol $\Rightarrow$ denotes weak convergence of probability measures with respect to the uniform metric, $\xrightarrow{\text{d}}$ denotes convergence in distribution and $\xrightarrow{\text{P}}$ denotes convergence in probability.

2. Estimation and testing method

In this section, we present the null hypothesis of interest and the test statistics considered in this paper.

2.1. Null hypothesis

Data are generated according to Model (1) defined in the Introduction. The number of lags l is fixed a priori. The observations are given by $\{y_{t-m+1}, y_{t-m}, \dots y_T\}$ where $m = \max(d, l)$ and T is the sampling period. Δ_{tT} depends on some unknown vector of parameters $v \in \Gamma$. The unknown parameters of interest are the components of $\theta = (\mu^{*'}, \mu', \sigma)' \in \Theta \subset \mathbb{R}^{l+1} \times \mathbb{R}^{l+1} \times \mathbb{R}_+^*$ where $\mu = (\mu_0, \dots, \mu_l)'$ and $\mu^* = (\mu_0^*, \dots, \mu_l^*)'$. The null hypothesis of parameter stability is H_0: $\mu^* = 0$. Under H_0, y_t does not depend on Δ_{tT}, therefore v is a nuisance parameter that is not identified under H_0. The alternative hypothesis of interest is H_1: $\mu^* \neq 0$, $v \in \Gamma$. An important point is that under H_0, the three models are the same. Eq. (1) becomes an autoregressive AR(l) model. The basic assumption on the distribution of the error is the following:

Assumption 1. $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with mean zero and finite fourth moment, and with absolutely continuous distribution. $\{\varepsilon_t\}$ is independent of $\{y_{t-1}, y_{t-2} \dots\}$.

Note that if ε_t are i.i.d normally distributed $\mathcal{N}(0, \sigma^2)$, Assumption 1 is satisfied. The following assumption guarantees stationarity of the process under H_0:

[2] Some of the usual assumptions are not satisfied and we are not able to derive the asymptotic distribution of the stability test against a Markov-switching alternative.

Assumption 2. The roots of the polynomial $1 - \mu_1 z - \cdots - \mu_l z^l$ are outside the unit circle.

Under Assumptions 1 and 2 and if H_0 is satisfied, the series $\{y_t\}$ is geometrically β-mixing (Doukhan, 1994, p. 99). This means that the law of large numbers (LLN) and functional central limit theorem (CLT) apply. For a definition of β-mixing (also called absolute regularity) and an example of application of this property, see Hansen (1996).

2.2. Testing

The null hypothesis of interest can be written as H_0: $A\theta = 0$, where A is a suitable matrix. Denote $\theta_0 = (\mu_0^{*\prime}, \mu_0', \sigma_0^2)'$ the true value of θ. Let $\mathrm{LR}_T(v)$ denote the standard likelihood ratio test statistic to test H_0 against H_1 given the parameter v. When the value of the nuisance parameter is known a priori, $\mathrm{LR}_T(v)$ converges to a chi-square distribution. However, when the nuisance parameter is unknown, this statistic fails to be a chi-square. To handle this problem, Davies (1987) proposed a supLR test, the maximum of the LR test with respect to the nuisance parameters over a restricted interval Γ. Later, Andrews and Ploberger (1994) proposed a class of tests denoted ExpLR. These tests are constructed to be optimal if the amplitude of the change follows a specific normal distribution. On the other hand, Andrews and Ploberger (1995) show that the supLR test is optimal for a uniform prior. Since the sup tests are the most widely applied, we shall focus on the supLR test and sup versions of specification robust Wald and Lagrange Multiplier tests proposed by White (1982).

Let us define $Q_T(\theta, v)$ as the mean log-likelihood function, $\hat{\theta}(v)$ as the maximum likelihood estimator of θ under $H_1(v)$, and $\tilde{\theta}$ the restricted estimate of θ under H_0. Our statistics are based on the following matrices:

$$S_T(\theta, v) = \frac{\partial}{\partial\theta} Q_T(\theta, v),$$

$$M_T(\theta, v) = -\frac{\partial^2}{\partial\theta\partial\theta'} Q_T(\theta, v),$$

$$\Omega_T(\theta, v) = M_T(\theta, v)^{-1} V_T(\theta, v) M_T(\theta, v)^{-1},$$

where $V_T(\theta, v)$ is an estimate of $\mathrm{Var}(\sqrt{T} S_T(\theta, v))$.

The pointwise likelihood ratio test statistics is given by

$$\mathrm{LR}_T(v) = 2T[Q_T(\hat{\theta}(v), v) - Q_T(\tilde{\theta})].$$

The pointwise Wald statistics is

$$W_T(v) = T\hat{\theta}(v)' A' [A\Omega_T(\hat{\theta}(v), v) A']^{-1} A\hat{\theta}(v).$$

The pointwise Lagrange multiplier statistics is

$$\mathrm{LM}_T(v) = \mathrm{TS}_T(\tilde{\theta}, v)' M_T(\tilde{\theta}, v)^{-1} A' [A\Omega_T(\tilde{\theta}, v) A']^{-1} \mathrm{AM}_T(\tilde{\theta}, v)^{-1} S_T(\tilde{\theta}, v).$$

In the following, we shall consider the following sup-tests:

$$\mathrm{Sup\,W_T} = \sup_{v \in \Gamma} W_T(v), \quad \mathrm{Sup\,LM_T} = \sup_{v \in \Gamma} \mathrm{LM}_T(v), \quad \mathrm{Sup\,LR_T} = \sup_{v \in \Gamma} \mathrm{LR}_T(v).$$

The choice of Γ is delicate. It must be bounded away from the maximum values of v, otherwise, the sup statistics will diverge. Note that Γ is a subset of $\mathbb{R}$ in the case of a SCA and TAR and is a subset of $\mathbb{R}^2$ in the case of a MSA. The choice of Γ will be discussed below for each specific model.

Assumption 3. θ_0 is interior point of Θ_0, where Θ_0 is a compact subset of Θ that contains neighborhoods of θ_0.

Assumption 3 is a standard assumption in the literature, see for instance Andrews and Ploberger (1994). Other assumptions, that are typically required to derive properties of the maximum likelihood estimators, are not stated here because they are satisfied by construction of the models, for instance differentiability of the likelihood with respect to θ.

2.3. Structural Change Autoregressive model (SCA)

The SCA is an autoregressive model where all the coefficients jump at some unknown timing, $T\pi$, with $\pi \in (0,1)$.

$$\begin{aligned} & y_t = \mathbf{y}_t'\alpha + \mathbf{y}_t'\alpha^* I\{t > [T\pi]\} + \varepsilon_t, \\ & \varepsilon_t \sim \text{i.i.d.}(0, \sigma_1^2) \end{aligned} \tag{2}$$

where $\mathbf{y}_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-l})'$, $\alpha = (\alpha_0, \ldots, \alpha_l)'$ and $\alpha^* = (\alpha_0^*, \ldots, \alpha_l^*)'$. I denotes the indicator function and $[a]$ denotes the greatest integer value smaller than a. The proportion, π, of observations before the break-point is supposed to be constant. This can be valid only if the interval between observations approaches zero, this corresponds to a finer and finer discretization. In that sense, the asymptotic in this model differs from the two others.

The null hypothesis of interest is H_0: $\alpha^* = 0$ and the alternative hypothesis is H_1: $\alpha^* \neq 0$. π is a nuisance parameter that is not identified under H_0. The following assumption is supposed to hold:

Assumption 4. The roots of the polynomial $1 - (\alpha_1 + \alpha_1^*)z - \cdots - (\alpha_l + \alpha_l^*)z^l$ are outside the unit circle.

Under Assumptions 2 and 4, the SCA is piecewise stationary because $\{y_t,\ t \leqslant [T\pi]\}$ follows a stationary AR(l) process and $\{y_t,\ t > [T\pi]\}$ follows another stationary AR(l). This property is used to obtain the LLN and central limit theorem (CLT) for triangular arrays of temporally dependent random variables. However, it is clear that the process as a whole is not even mean stationary.

Without any extra assumptions, the SCA can be estimated by OLS. The test statistics will be based on a Gaussian likelihood. The following result was proven independently by Andrews (1993) and Davis et al. (1995).

Proposition 1. *Assume Assumptions* 1–4 *hold and* ε_t *is normally distributed. Then under* H_0: $\mu^* = 0$:

$$\mathrm{SupW_T}, \mathrm{SupLM_T}, \mathrm{SupLR_T} \xrightarrow{\mathrm{d}} \sup_{\pi\in\Pi}(B_{l+1}(\pi) - \pi B_{l+1}(1))'(B_{l+1}(\pi) - \pi B_{l+1}(1))/[\pi(1-\pi)] \tag{3}$$

where B_{l+1} *is a* $l+1$*-vector of independent Brownian motions on* $[0,1]$ *restricted to* Π.

Remark that for π fixed, the Brownian bridge $B(\pi) - \pi B(1)$ is simply a normal random variable with mean zero and variance $\pi(1-\pi)$. As a process, $\mathrm{LR}_T(\pi)$ converges to the sum of independent squared Brownian bridges. Therefore, for π fixed, we obtain the usual result that the LR_T test statistics converges to a chi-square distribution with $l+1$ degrees of freedom. The set Π must exclude values of π such as $\pi=0$ and $\pi=1$ where the information matrix is singular. Andrews (1993) has tabulated the critical values for various intervals Π, for $l=0,1,\dots,19$, and test level $\alpha=1\%$, 5%, 10%. The normality is needed to have asymptotic equivalence between the three tests because the test $\sup_{\pi\in\Pi}\mathrm{LR}_T(\pi)$ is not robust to misspecification. Its asymptotic distribution under more general assumptions can be found in Hansen (1991).

Assume $l=0$. Under H_0: $\alpha^*=0$, Eq. (3) simplifies to:

$$\sup_{\pi\in\Pi} W_T(\pi), \sup_{\pi\in\Pi}\mathrm{LM}_T(\pi), \sup_{\pi\in\Pi}\mathrm{LR}_T(\pi) \xrightarrow{\mathrm{d}} \sup_{\pi\in\Pi}\frac{(B(\pi)-\pi B(1))^2}{\pi(1-\pi)}.$$

2.4. Markov-Switching Autoregressive model (MSA)

The MSA is the counterpart of the SCA where the change of regime depends on an exogenous unobservable Markov chain.

$$y_t = \mathbf{y}_t'\beta + \mathbf{y}_t'\beta^* S_t + \omega_t,$$

$$\omega_t \sim \text{i.i.d.}\,\mathcal{N}(0,\sigma_2^2) \tag{4}$$

where $\beta=(\beta_0\cdots\beta_l)'$, and $\beta^*=(\beta_0^*\cdots\beta_l^*)'$. S_t is a two-state Markov chain with unknown transition probabilities $p=P(S_t=1|S_{t-1}=1)$ and $q=P(S_t=0|S_{t-1}=0)$, $p\in(0,1)$ and $q\in(0,1)$. Hamilton (1993, p. 235) points out that the specification (4) allows the possibility of a permanent change as a special case if $p=1$, $0<q<1$ or $q=1$, $0<p<1$. So that the SCA is a particular case of the MSA. For identifiability purpose, when we refer to an MSA, it is implicit that p and q are away from the boundaries. Moreover, the labels of states and submodels are interchangeable. To guarantee identifiability, it is assumed that $\beta_0^*>0$. However, this constraint is not imposed in the estimation.

The parameters $\theta = (\beta^{*\prime}, \beta', \sigma_2)'$ are estimated by maximum likelihood estimation (MLE). Since S_t is not observable, inference about the state is done applying an EM-algorithm developed by Hamilton (1989). The assumption of normality is crucial. Since the model is nonlinear, the ML estimators may fail to be consistent if the normality is not satisfied.

The null hypothesis of interest is H_0: $\beta^* = 0$. The alternative is H_1: $\beta^* \neq 0$. Under the null hypothesis, the transition probabilities p and q are not identified. However, a more serious problem occurs since the score functions are identically equal to zero under the null hypothesis, and thus the information matrix is singular. Under H_0, $P(S_t|y_1,\ldots,y_T)$ is, indeed, equal to $P(S_t = 1)$ that is constant. Hence, the general theory breaks down (see Garcia, 1998; Lee and Chesher, 1986). Nevertheless, Garcia (1998) derives the distribution of the supLR test and assesses its validity by a Monte Carlo analysis. To make sure that the process behaves properly under the alternative, the following assumption is needed:

Assumption 5. The MSA is stationary and geometrically β-mixing.

In a recent paper, Yao and Attali (2000) give a sufficient condition for geometric ergodicity of MSA. Geometric ergodicity implies that the stationary distribution exists and that if y_0 is drawn from this distribution, then $\{y_t\}$ is also stationary and geometrically β-mixing. If $l=1$, a sufficient condition is simply $\lambda \ln|\beta_1| + (1-\lambda)\ln|\beta_1 + \beta_1^*| < 1$, where $\lambda \equiv \mathrm{P}(S_t = 1) = (1-q)/(2-p-q)$. Note that an explosive root in one regime does not preclude global strict stationarity. Franck and Zakoian (2001) give sufficient conditions for second-order stationarity of multivariate MSA. They show in particular that stationarity within regimes is neither necessary nor sufficient for global second-order stationarity. For moments of MSA, see also Timmermann (2000).

When $l = 0$, Garcia (1998) shows that under H_0: $\beta^* = 0$:

$$\sup_{(p,q)\in\Psi^2} \mathrm{LR}_T(p,q) \xrightarrow{\mathrm{d}} \sup_{\lambda\in\Lambda} \frac{(B(\lambda) - \lambda B(1))^2}{\lambda(1-\lambda)} \tag{5}$$

with Ψ being the closed interval where p and q are supposed to lie. Λ is the closed interval where $\lambda(p,q) = (1-q)/(2-p-q)$ belongs. The supLR test is likely to converge to the distribution given in (3) in the case of a general MSA but no formal proof is available.

2.5. Threshold Autoregressive model (TAR)

We consider a Self-Exciting Threshold Autoregressive model, which has been studied extensively by Tong (1990).

$$\begin{aligned} y_t &= \mathbf{y}_t'\gamma + \mathbf{y}_t'\gamma^* I\{y_{t-d} \leqslant r\} + u_t, \\ u_t &\sim \text{ i.i.d. } (0, \sigma_3^2), \end{aligned} \tag{6}$$

where $\gamma = (\gamma_0,\ldots,\gamma_l)'$, $\gamma^* = (\gamma_0^*,\ldots,\gamma_l^*)'$. The null hypothesis of interest is H_0: $\gamma^* = 0$. The alternative is H_1: $\gamma^* \neq 0$. Both r and d are nuisance parameters that are not identified under the null. Chan (1990) gives a test of this hypothesis for a fixed d and an unknown

r, Hansen (1996) considers both r and d unknown. Since the issue of interest is the behavior of the test statistic when the model is misspecified, we can assume d known a priori and r unknown. By assumption, r is such that $\delta = P(y_t \leqslant r)$ belongs to $(0,1)$. As for the SCA, the model is estimated by OLS and tests are constructed using Gaussian likelihood. The following assumption ensures the geometric ergodicity of the process and the existence of a (strict sense) stationary distribution:

Assumption 6. $\max(\sum_{i=1}^{l} |\gamma_i|, \sum_{i=1}^{l} |\gamma_i + \gamma_i^*|) < 1$.

This result has been proved first by Chan and Tong (1985) for $d \leqslant l$ and by Bhattacharya and Lee (1995) for $d > l$.

Now, we study in detail the Threshold nonautoregressive model (case $l = 0$ and u_t normal). Its stationary distribution is a mixture of normal distributions:

$$f(y_t) = \delta \exp - \frac{(y_t - \gamma - \gamma^*)^2}{2\sigma_3^2} + (1 - \delta) \exp - \frac{(y_t - \gamma)^2}{2\sigma_3^2},$$

where $\delta = P(y_t \leqslant r) = \Phi(r - \gamma/\sigma_3)/[1 - \Phi(r - \gamma - \gamma^*/\sigma_3) + \Phi(r - \gamma/\sigma_3)]$ and Φ is the c.d.f. of $\mathcal{N}(0,1)$. No restriction on d, γ and γ^* is necessary to obtain the stationarity of $\{y_t\}$. We also have calculated the mean and the variance of y_t

$$E(y_t) \equiv m = \gamma + \delta\gamma^*,$$

$$V(y_t) \equiv \tau^2 = \sigma_3^2 + \delta(1 - \delta)\gamma^{*2}.$$

Lemma 2. *Assume that y_t is generated by* (6) *with $l = 0$. Let X_t denote $I\{y_{td} \leqslant r\}$. Then $\{X_t\}$ is a Markov chain.*

Lemma 2 is a straightforward generalization of a result by Gourieroux (1997, p. 13). It points out the close relationship between a TAR with $d = 1$ and an MSA. Note that the Threshold model with $d = 1$ is not really a Markov-switching model because the Markov chain $\{X_t\}$ is not exogenous. Moreover, Threshold model (6) with $l = 0$ and $d = 1$ is a Markov process (conditionally on y_{t-1}, y_t does not depend on $y_{t-2}, y_{t-3}, \ldots$) while a Markov-switching model (4) with $l = 0$ is not Markov (conditionally on y_{t-1}, y_t still depends on all past values of the series through S_t).

We wish to test the null hypothesis H_0: $\gamma^* = 0$ against H_1: $\gamma^* \neq 0$. The threshold r is assumed to lie inside an interval R. Chan (1990) and Chan and Tong (1990) show that

$$\sup_{r \in R} W_T(r), \sup_{r \in R} \mathrm{LM}_T(r), \sup_{r \in R} \mathrm{LR}_T(r) \xrightarrow{\mathrm{d}} \sup_{\delta \in \Delta} \frac{(B(\delta) - \delta B(1))^2}{\delta(1 - \delta)},$$

where the expression of $\delta(r) = \mathrm{E}(I\{y_{t-d} \leqslant r\})$ is given above and $\Delta \in (0,1)$ is the image of R under the map δ.

Note that under H_0, the distributions of the test statistics are asymptotically equivalent for the three nonautoregressive models as mentioned by Garcia (1998, Appendix 3).

In the case of a general TAR model, the asymptotic distribution of the test statistic depends on the unknown parameters of the model (see Chan, 1990, Theorem 2.3) which makes the computation of the critical values more difficult. To handle such cases, Garcia (1998) proposes an algorithm to simulate chi-square processes and compute the critical values, on the other hand, Hansen (1996) develops techniques to determine the p-values via simulations.

3. Asymptotic distribution of statistics under model misspecification

Most often, econometricians test linearity against a specific alternative they believe in, but ultimately they have to determine which model fits the data best via tests for parameter stability. Hence to make the right decision, it is essential to know the asymptotic behavior of these test statistics if the model is misspecified. In the following, we shall assume that the data are generated by a model, say M_1, but one estimates another model, say M_2, which is therefore misspecified. A test at level α is applied as if the model were correctly specified, that is, the critical region is $[l_\alpha, \infty)$ given by Andrews' tables. In order to calculate the probability of the critical region when the model is misspecified, we have to determine first the asymptotic distribution under the true model of the estimator $\hat{\theta}$ and then of the statistics defined in the previous section. The criterion, which has been maximized, is a quasi-likelihood function $Q_T(\theta, v)$. The pseudo-maximum likelihood estimate:

$$\hat{\theta}(v) = \arg\max_{\theta \in \Theta} Q_T(\theta, v)$$

converges asymptotically to a limit denoted θ_a that depends not only on v but also on the true value of the parameters associated with the DGP M_1 (including the nuisance parameters in that model). From now on, Θ_0 is a subset of Θ that contains neighborhoods of θ_a, Θ_0 will be assumed to satisfy the compactness of Assumption 3.

3.1. Test for an SCA when the data are generated by a TAR or MSA

Here a test of parameter stability is performed as if the model were SCA. However, the data are generated either by a TAR or MSA. When testing for the absence of structural change, the null hypothesis can be rewritten H_0: $A\theta = 0$, where $\theta = [\alpha^{*\prime}, \alpha', \sigma]'$, and $A = [I_{l+1} | \mathrm{O}_{l+1} | \mathrm{o}_{l+1}]$ with I_{l+1} is a $(l+1) \times (l+1)$ matrix identity, O_{l+1} is a $(l+1) \times (l+1)$ matrix of zeros, and o_{l+1} is a $(l+1) -$ vector of zeros. Define

$$M = \frac{1}{\sigma_a^2} \mathrm{E}[\mathbf{y}_t \mathbf{y}_t'] \quad \text{and} \quad S = \lim_{T \to \infty} V\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{1}{\sigma_a^2} \mathbf{y}_t \{y_t - \mathbf{y}_t' \alpha_a\}\right],$$

where α_a and σ_a^2 are the pseudo-true values defined below.

Assumption 7. $\mathrm{E}\|\mathbf{y}_t \mathbf{y}_t'\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$.

Assumption 8. M and S are positive definite.

Assumption 7 imposes the existence of higher order moments.

Proposition 3. *Suppose that Assumptions* 1, 2, 4–8 *hold. When the data are generated either by a stationary TAR or MSA, the pseudo-maximum likelihood estimators of the parameters of the SCA satisfy*

$$\hat{\alpha}(\pi) \xrightarrow{\mathrm{P}} \alpha_a = \mathrm{E}[\mathbf{y}_t \mathbf{y}_t']^{-1} \mathrm{E}[\mathbf{y}_t y_t],$$

$$\hat{\alpha}^*(\pi) \xrightarrow{\mathrm{P}} \alpha_a^* = 0,$$

$$\hat{\sigma}^2(\pi) \xrightarrow{\mathrm{P}} \sigma_a^2 = \mathrm{E}[\{y_t - \mathbf{y}_t'\alpha_a\}^2].$$

The sup LM *and* W *statistics satisfy*

$$\sup_{\pi \in \Pi} \mathrm{LM}_T(\pi), \sup_{\pi \in \Pi} W_T(\pi)$$

$$\Rightarrow \sup_{\pi \in \Pi} (B_{l+1}(\pi) - \pi B_{l+1}(1))'(B_{l+1}(\pi) - \pi B_{l+1}(1))/[\pi(1-\pi)].$$

The sup LR *statistic satisfies*

$$\sup_{\pi \in \Pi} \mathrm{LR}_T(\pi) \Rightarrow \sup_{\pi \in \Pi} \sum_{j=1}^{l+1} c_j(\theta_a) \frac{(B^j(\pi) - \pi B^j(1))^2}{\pi(1-\pi)},$$

where B^j *are independent scalar Brownian motions and* $c_1(\theta_a) \geqslant c_2(\theta_a) \geqslant \cdots \geqslant c_l(\theta_a) \geqslant 0$ *are the eigenvalues of the matrix* $M^{-1}S$.

For π known a priori, the LR test converges to a sum of independent chi-square processes weighted by the eigenvalues c_j (see Foutz and Srivastava, 1977). Again for π known, our specification robust LM and W tests converge to a chi-square of $(l+1)$ degrees of freedom, see White (1982). For π unknown, we get the counterpart of these results where independent chi-square random variables are replaced by independent squared Brownian bridges. The power of the test based on LM_T and W_T is exactly equal to the level of the test. In the case of LR_T, the power depends on the value of the parameters.

Let $M(\theta_a, \pi) = \lim_{T\to\infty} M_T(\theta_a, \pi)$ and $V(\theta_a, \pi) = \lim_{T\to\infty} V_T(\theta_a, \pi)$ where M_T and V_T are defined in Section 2. $M(\theta_a, \pi)$ and $V(\theta_a, \pi)$ should not be confused with M and S. Simple calculations show the following relationship:

$$[M(\theta_a, \pi)]_{11} = -\begin{bmatrix} M & (1-\pi)M \\ (1-\pi)M & (1-\pi)M \end{bmatrix} \quad \text{and}$$

$$[V(\theta_a, \pi)]_{11} = \begin{bmatrix} S & (1-\pi)S \\ (1-\pi)S & (1-\pi)S \end{bmatrix}, \tag{7}$$

where $[M(\theta_a, \pi)]_{11}$ and $[V(\theta_a, \pi)]_{11}$ are the upper-left blocks of $M(\theta_a, \pi)$ and $V(\theta_a, \pi)$ and correspond to the derivatives with respect to $(\alpha', \alpha^{*\prime})'$. Below, we specialize our results to simple nonautoregressive models.

3.1.1. The DGP is the Markov-switching model

One tests for the absence of structural change H_0: $\alpha^* = 0$ in Model (2) with $l = 0$ when in reality, the data generating process is a two-state Markov-switching one (4) with $l = 0$. Recall that $\lambda = P(S_t = 1) = (1-q)/(2-p-q)$.

Proposition 4. *Suppose that the Markov-switching model holds, then*

$$\sqrt{T}\begin{pmatrix} \hat{\alpha} - \beta - \lambda\beta^* \\ \hat{\alpha}^* \end{pmatrix} \xrightarrow{\mathrm{d}} \mathcal{N}$$

$$\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \left(\sigma_2^2 + \beta^{*2}\frac{\lambda(1-\lambda)(p+q)}{(2-p-q)}\right)\begin{pmatrix} \frac{1}{\pi} & -\frac{1}{\pi} \\ -\frac{1}{\pi} & \frac{1}{\pi(1-\pi)} \end{pmatrix}\right),$$

$$\hat{\sigma}_1^2 \xrightarrow{\mathrm{P}} \sigma_{1a}^2 = \sigma_2^2 + \beta^{*2}\lambda(1-\lambda).$$

Proposition 5. *Suppose that the Markov-switching model holds, then*

$$\sup_{\pi\in\Pi} W_T(\pi), \sup_{\pi\in\Pi} \mathrm{LM}_T(\pi) \xrightarrow{\mathrm{d}} \sup_{\pi\in\Pi} \frac{(B(\pi) - \pi B(1))^2}{\pi(1-\pi)},$$

$$\sup_{\pi\in\Pi} \mathrm{LR}_T(\pi) \xrightarrow{\mathrm{d}} \sup_{\pi\in\Pi} \frac{(B(\pi) - \pi B(1))^2}{\pi(1-\pi)}\left(\frac{\sigma_2^2}{\sigma_{1a}^2} + \frac{\beta^{*2}\lambda(1-\lambda)(p+q)}{\sigma_{1a}^2(2-p-q)}\right).$$

The sup LR statistic converges to the same function of a Brownian bridge as in the correctly specified case, multiplied by a coefficient

$$K_1 = \left(\frac{\sigma_2^2}{\sigma_{1a}^2} + \frac{\beta^{*2}\lambda(1-\lambda)(p+q)}{\sigma_{1a}^2(2-p-q)}\right)$$

which is greater or less than 1 according to the values of p and q. We have:

1. $p + q > 1 \Rightarrow K_1 > 1$,
2. $p + q < 1 \Rightarrow K_1 < 1$.

When $p + q < 1$, then the probability of a switch in regime is higher than the probability to stay in the same regime. The process switches a lot between the two regimes and therefore appears as a white noise with large variance. In that case, the probability of rejection is lower than the level of the test. That is to say that the test does not have any power. When $p + q > 1$, the probability of staying in the same regime is larger than the probability to switch. Therefore, the probability of rejection is greater than the level of the test and depends on the parameters β^*, p and q but, as the pseudo-true value of $\hat{\alpha}^*$ is equal to zero, the power of the test should be low.

3.1.2. The DGP is the threshold model

Now, the absence of structural change is tested when in fact the model is a Threshold one (6) with $l = 0$ and normal error.

Proposition 6. *Suppose that the data are generated by a Threshold model, then*

$$\sqrt{T}\begin{pmatrix} \hat{\alpha}-\gamma-\delta\gamma^* \\ \hat{\alpha}^* \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, K_2\begin{pmatrix} \frac{1}{\pi} & -\frac{1}{\pi} \\ -\frac{1}{\pi} & \frac{1}{\pi(1-\pi)} \end{pmatrix}\right),$$

$$\hat{\sigma}_1^2 \xrightarrow{P} \sigma_{1a}^2 = \sigma_3^2 + \delta(1-\delta)\gamma^{*2}$$

with

$$K_2 = \left(\sigma_3^2 + \gamma^{*2}\frac{\delta(1-\delta)}{1-\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right)+\Phi\left(\frac{r-\gamma}{\sigma_3}\right)}\frac{2\gamma^*}{1-\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right)+\Phi\left(\frac{r-\gamma}{\sigma_3}\right)}\right.$$
$$\left. - \left[\delta\varphi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right)+(1-\delta)\varphi\left(\frac{r-\gamma}{\sigma_3}\right)\right]\right),$$

where φ is the probability density function (p.d.f.) of the standard normal distribution.

Note that the expression of K_2 is more complicated than K_1 because the Markov chain $X_t = I\{y_t \leqslant r\}$ is correlated with u_t whereas S_t is exogenous and therefore uncorrelated with ω_t.

Proposition 7. *Under the Threshold model, we have*

$$\sup_{\pi\in\Pi} W_T(\pi), \sup_{\pi\in\Pi} \mathrm{LM}_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi} \frac{(B(\pi)-\pi B(1))^2}{\pi(1-\pi)},$$

$$\sup_{\pi\in\Pi} \mathrm{LR}_T(\pi) \xrightarrow{d} \sup_{\pi\in\Pi} \frac{K_2}{\sigma_{1a}^2}\frac{(B(\pi)-\pi B(1))^2}{\pi(1-\pi)}.$$

The coefficient K_2/σ_{1a}^2 is greater or less than 1 according to the values of γ^*:

1. $\gamma^* > 0 \Rightarrow K_2/\sigma_{1a}^2 < 1$,
2. $\gamma^* < 0 \Rightarrow K_2/\sigma_{1a}^2 > 1$.

When $\gamma^* > 0$, the process is much more likely to switch back and forth because when the process is in Regime 1, its mean is greater than in Regime 0 and then y_t is more likely to exceed the threshold value r. As in the Markov-switching model, a process that switches a lot is likely to be mistaken for a white noise with large variance. Therefore, the probability to reject is less than the level of the test. When $\gamma^* < 0$ the process switches less often. Hence, the probability to reject is greater than the level of the test but the power will remain low.

Note that the estimator of the expectation of y_t is

$$\hat{\mathrm{E}}(y_t) = \hat{\alpha} + \hat{\alpha}^*(1-\hat{\pi}).$$

This coincides (see the score with respect to α) with the empirical mean:

$$\hat{\mathrm{E}}(y_t) = \bar{y}.$$

Hence the misspecified Structural Change provides a consistent estimation of the expectation whatever the correct model is.

When the misspecified model is a Structural Change and the true model is a stationary one, the only hypothesis making the tested model stationary, that is $\alpha^* = 0$, will be the one having the highest probability of being chosen. In other words, the test is not consistent. This lack of power can be very misleading in applications if one does not investigate other forms of parameter instability. However, one can use these results to get information about the nature of the switch. Assume one tests the null hypothesis of absence of structural change. If the statistic is in the critical region, the switch is certainly a structural change but if the statistic is low, then the model could be either Threshold or Markov-switching.

3.2. Test for a TAR when the data are generated by an SCA or MSA

The null hypothesis H_0: $\gamma^* = 0$ is tested as if the data were generated by the TAR model (6).

Proposition 8. *Suppose that Assumptions* 1–5 *hold. If the DGP is either an SCA with* $\pi \in (0,1)$ *or a stationary MSA with* $p \in (0,1)$, $q \in (0,1)$, *and such that* $p+q \neq 1$, *then*

$$\sup_{r\in R} W_T(r), \sup_{r\in R} \mathrm{LM}_T(r), \sup_{r\in R} \mathrm{LR}_T(r) \xrightarrow{\mathrm{P}} +\infty.$$

When $p+q=1$, $P(S_t = 1|S_{t-1}=0) = 1-q = p = P(S_t=1|S_{t-1}=1)$, the $\{S_t\}$ are independent. In that case, the statistics converge to zero and you would accept H_0. In all other cases, you would reject H_0 asymptotically with probability 1. To get some insights into the behavior of the test statistics under a local alternative, we restrict our attention to nonautoregressive models.

3.2.1. The DGP is the Structural Change model

We test the hypothesis of parameter stability H_0: $\gamma^* = 0$ in Threshold model (6) with $l = 0$ when, in fact, the true model is Structural Change (2) with $l = 0$ and normal error.

Proposition 9. *Consider* $\hat{\gamma}$ *and* $\hat{\gamma}^*$ *the quasi-maximum likelihood estimators obtained in the Threshold model. When the Structural Change model holds, we have*

$$\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_a = \alpha + \alpha^*(1-\pi)\frac{1-\Phi\left(\dfrac{r-\alpha-\alpha^*}{\sigma_1}\right)}{1-\pi\Phi\left(\dfrac{r-\alpha}{\sigma_1}\right)-(1-\pi)\Phi\left(\dfrac{r-\alpha-\alpha^*}{\sigma_1}\right)},$$

$$\hat{\gamma}^* \underset{M_1}{\overset{T\to\infty}{\longrightarrow}} \gamma_a^* = \frac{\alpha^*(1-\pi)\pi\left[\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)-\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\right]}{\left[\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)+(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]\left[1-\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)-(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]},$$

where Φ is the cumulative distribution function of the standard normal.

Note that the pseudo-true value $\gamma_a^* = 0$ only if α^* is also equal to zero.

Proposition 10. *Suppose that the Structural Change model holds. Then*
(i)

$$\sqrt{T}(\hat{\gamma}^* - \gamma_a^*) \xrightarrow{d} \mathcal{N}(0, K_3),$$

where K_3 is given in Lemma A.3 of the appendix.
(ii) *Under a local alternative* $\mathrm{H}_{1T}(\pi)$: $\alpha^* = a/T^{1/4}$, $\pi \in (0,1)$, *we have*

$$W_T(r), \mathrm{LM}_T(r), \mathrm{LR}_T(r) \xrightarrow{d} \chi^2(1, \rho^2),$$

where $\chi^2(1,\rho^2)$ is a noncentral chi-square distribution with 1 degree of freedom and noncentrality parameter:

$$\rho^2 = a^4 \frac{(1-\pi)^2\pi^2\varphi^2((r-\alpha)/\sigma_1)}{\sigma_1^4\Phi((r-\alpha)/\sigma_1)(1-\Phi((r-\alpha)/\sigma_1))}.$$

This result is to be compared with what is obtained when testing for a structural change instead of a threshold. The tests designed for a structural change alternative have optimal power, that is, they detect alternatives of the type $a/\sqrt{T}$ while the tests designed for the threshold alternative have no power against alternatives $a/\sqrt{T}$. It is due to the pseudo-true value that is not linear in α^*. Indeed, an equivalent of γ_a^* when α^* approaches zero is given by

$$\gamma_a^*(\alpha^*) \sim -\alpha^{*2} \frac{(1-\pi)\pi\varphi(r-\alpha/\sigma_1)}{\sigma_1\Phi((r-\alpha)/\sigma_1)(1-\Phi((r-\alpha)/\sigma_1))}.$$

Therefore, γ_a^* converges to zero faster than α^*. It means that these tests will not have power against small jumps. However, under a fixed alternative, these statistics diverge to infinity in probability and the hypothesis H_0 will be rejected with probability going to 1. Note that from the point of view of the applied econometrician who is using a misspecified TAR model, the speed of the sequence of local alternatives is the standard one: $\gamma^* = g/\sqrt{T}$. Moreover, under the local alternatives the LR statistic has same limit as the Wald and LM tests. It is due to the fact that the limiting distribution is affected by the variance under the null only and under the null the model is correctly specified. However, the equivalence of the three tests does not hold under a fixed alternative.

3.2.2. The DGP is the Markov-switching model

As above, we test the parameter stability in Threshold model (6) with $l = 0$ when in fact the data are generated by a two-state Markov-switching model (4) with $l = 0$.

Proposition 11. *Suppose that the Markov-switching model holds, the pseudo-true values are given by*

$$\hat{\gamma} \xrightarrow{\mathrm{P}} \gamma_a = \beta + \beta^*\lambda - (p+q-1)^d \frac{\beta^*(1-\lambda)\lambda\left[\Phi\left(\dfrac{r-\beta-\beta^*}{\sigma_2}\right) - \Phi\left(\dfrac{r-\beta}{\sigma_2}\right)\right]}{1-(1-\lambda)\Phi\left(\dfrac{r-\beta}{\sigma_2}\right) - \lambda\Phi\left(\dfrac{r-\beta-\beta^*}{\sigma_2}\right)},$$

$$\hat{\gamma}^* \xrightarrow{\mathrm{P}} \gamma_a^* = \frac{(p+q-1)^d\beta^*(1-\lambda)\lambda\left[\Phi\left(\dfrac{r-\beta-\beta^*}{\sigma_2}\right) - \Phi\left(\dfrac{r-\beta}{\sigma_2}\right)\right]}{\left[(1-\lambda)\Phi\left(\dfrac{r-\beta}{\sigma_2}\right) + \lambda\Phi\left(\dfrac{r-\beta-\beta^*}{\sigma_2}\right)\right]\left[1-(1-\lambda)\Phi\left(\dfrac{r-\beta}{\sigma_2}\right) - \lambda\Phi\left(\dfrac{r-\beta-\beta^*}{\sigma_2}\right)\right]}.$$

The pseudo-true values in Proposition 11 are very close to those of Proposition 9; they depend in the same way of the probability to be in regime 1 (π and λ, respectively) and on the probability for an observation to be less than the threshold r. The only difference is that here the pseudo-true values depend on d while, in the previous case, it does not. It is because the Markov chain $\{S_t\}$ is autocorrelated while its counterpart in the Structural Change model, that is $I\{t \leqslant T\pi\}$, is not. Again, γ_a^* converges to zero faster than β^* because the equivalent of γ_a^* as β^* approaches zero is given by

$$\gamma_a^*(\beta^*) \sim -\beta^{*2}\frac{(p+q-1)^d(1-\lambda)\lambda\varphi((r-\beta)/\sigma_2)}{\sigma_2\Phi((r-\beta)/\sigma_2)(1-\Phi((r-\beta)/\sigma_2))}.$$

Proposition 12. *Under a local alternative* $\mathrm{H}_{1T}(\lambda)$: $\beta^* = b/T^{1/4}$, $\lambda \in (0,1)$, *we have*

$$W_T(r), \mathrm{LM}_T(r), \mathrm{LR}_T(r) \xrightarrow{\mathrm{d}} \chi^2(1,\rho^2)$$

with

$$\rho^2 = b^4(p+q-1)^{2d}\frac{(1-\lambda)^2\lambda^2\varphi^2((r-\beta)/\sigma_2)}{\sigma_2^4\Phi((r-\beta)/\sigma_2)(1-\Phi((r-\beta)/\sigma_2))}.$$

The variance of $\sqrt{T}\hat{\gamma}^*$ under the true model is complicated and not presented in this paper. Note that the estimator of the expectation of y_t is

$$\hat{\mathrm{E}}(y_t) = \hat{\gamma} + \hat{\gamma}^*\frac{1}{T}\sum_{t=1}^{T} I\{y_{t-d} \leqslant \hat{r}\},$$

which coincides with $\bar{y}$ (see the score with respect to γ). Hence as in the case of the Structural Change, the Threshold model delivers a consistent estimator of the expectation whatever the true model is.

We can summarize the results of this subsection as following. Tests designed for a threshold alternative have power against alternatives originating from a Structural Change or Markov-switching. Therefore, to test the Threshold model can be used to establish whether there is parameter stability. However, these tests fail to detect local alternatives as $b/\sqrt{T}$ and will have a low power against small jumps. This result has to hold for each r, hence it holds for the sup.

3.3. Test for an MSA

Assume one tests for parameter stability in Markov-switching model (4) when in fact the data are generated either by an SCA or TAR. Consider first the case of an SCA. This model is nested in the MSA with $p=1$ and/or $q=1$; therefore, the $\text{supLR}_{\text{MSA}}$ is expected to have power against a structural change. However, when $p=1$, p is on the boundary of its space $[0,1]$ and we do not know what the properties of the estimators are. Simulations show that the MSA captures correctly the single jump, indeed $\text{P}(S_t=1|y_{t-1},\ldots,y_1)$ goes from 0 to 1 at the right moment and the estimators obtained by fitting the MSA are basically the same as those obtained by fitting the SCA. Moreover, $\text{supLR}_{\text{MSA}}$ appears to have a very large power even in small samples.

Now assume that the DGP is a TAR model. The properties of the estimators depend crucially on how Hamilton's algorithm behaves. To illustrate our point, suppose that we fit an MSA with $l=0$. The pseudo-MLE $\hat{\beta}$ and $\hat{\beta}^*$ are solutions of the first-order conditions given by Hamilton (1994, p. 692):

$$\sum_{t=1}^{T}\sum_{s_t=0,1}\{y_t-\hat{\beta}-\hat{\beta}^*s_t\}P(S_t=s_t|y_{t-1},\ldots,y_1)=0,$$

$$\sum_{t=1}^{T}\{y_t-\hat{\beta}-\hat{\beta}^*\}P(S_t=1|y_{t-1},\ldots,y_1)=0.$$

We do not know what Hamilton's algorithm will capture as S_t. If it "identifies" $X_t=I(y_{t-d}\leqslant r)$ as the Markov chain that drives the changes of regimes even though X_t is not exogenous, then $\text{supLR}_{\text{MSA}}$ will have power. In the simulations, we see that $P(S_t=s_t|y_{t-1},\ldots,y_1)$ identifies the right regime most of the times and that the $\text{supLR}_{\text{MSA}}$ is powerful.

4. Empirical size and power

4.1. Methodology

The simulations displayed in Tables 1–4 intend to show how well the test performs in practice. We consider only simple models with no autoregressive terms. Simulations were performed using 2000 replications of small samples of size 100 and large samples of size 1000. The results are shown for tests at a 5% level. For each test, we report the empirical power, that is the percentage of rejections, denoted P, and the standard error of P calculated as $(P(1-P)/\text{rep})^{1/2}$, where rep is the number of replications. All computations were done using GAUSS software package and its random number generators.

Table 1
DGP: linear model

DGP Sample size	$\alpha = 1,$		$\alpha^* = 0,$		$\sigma^2 = 1$	
	$T = 100$		$T = 1000$		$T = 2000$	
$\sup W_{SCA}$	0.057	(0.005)	0.043	(0.004)	0.045	(0.006)
$\sup LM_{SCA}$	0.022	(0.003)	0.040	(0.004)	0.044	(0.006)
$\sup LR_{SCA}$	0.032	(0.004)	0.040	(0.004)	0.043	(0.006)
$\sup LR_{MSA}$	0.023	(0.003)	0.022	(0.003)	0.022	(0.005)
$\sup W_{TAR}$	0.066	(0.005)	0.053	(0.005)	0.061	(0.007)
$\sup LM_{TAR}$	0.032	(0.004)	0.047	(0.005)	0.058	(0.007)
$\sup LR_{TAR}$	0.040	(0.004)	0.047	(0.005)	0.060	(0.007)

Table 2
DGP: Structural Change model

DGP Sample size	Nile River flow				Exchange rate			
	$\alpha = 1151$ $\alpha^* = 800$		$\pi = 0.27$ $\sigma_1^2 = 12210$		$\alpha = 2.29$ $\alpha^* = 0.545$		$\pi = 0.8$ $\sigma_1^2 = 0.087$	
	$T = 100$		$T = 1000$		$T = 100$		$T = 1000$	
$\hat{\alpha}$	1151.03	(21.53)	1150.5	(6.58)	2.288	(0.032)	2.289	(0.011)
$\hat{\alpha}^*$	800.04	(25.34)	799.99	(7.69)	0.552	(0.073)	0.545	(0.023)
$\hat{\pi}$	0.27	(0)	0.27	(0)	0.798	(0.017)	0.800	(0.001)
$\hat{\sigma}_1^2$	12002	(1753)	12111	(541)	0.085	(0.012)	0.087	(0.004)
predict. SCA	1	(0)	1	(0)	0.992	(0.015)	0.999	(0.001)
$\sup W_{SCA}$	1	(0)	1	(0)	1	(0)	1	(0)
$\sup LM_{SCA}$	1	(0)	1	(0)	1	(0)	1	(0)
$\sup LR_{SCA}$	1	(0)	1	(0)	1	(0)	1	(0)
$\hat{\beta}$	1151.03	(21.53)	1150.5	(6.58)	2.280	(0.037)	2.283	(0.011)
$\hat{\beta}^*$	800.04	(25.34)	799.99	(7.69)	0.553	(0.080)	0.549	(0.029)
$\hat{p}$	0.95	(0.000)	0.95	(0.00)	0.940	(0.044)	0.949	(0.013)
$\hat{q}$	0.95	(0.000)	0.95	(0.00)	0.946	(0.037)	0.949	(0.013)
$\hat{\sigma}_2^2$	12002	(1753)	12112	(541)	0.082	(0.013)	0.084	(0.004)
predict. MSA	1	(0)	1	(0)	0.979	(0.054)	0.994	(0.019)
$\sup LR_{MSA}$	1	(0)	1	(0)	0.991	(0.002)	0.999	(0.000)
$\hat{\gamma}$	1950.90	(12.66)	1950.81	(4.25)	2.671	(0.109)	2.668	(0.046)
$\hat{\gamma}^*$	−772.81	(25.91)	−796.95	(7.81)	−0.359	(0.107)	−0.339	(0.042)
$\hat{r}$	1370.74	(55.35)	1463.48	(41.39)	2.622	(0.144)	2.693	(0.066)
$\hat{\sigma}_3^2$	17942	(2698)	12849	(601)	0.111	(0.016)	0.116	(0.005)
predict. TAR	0.990	(0.001)	0.999	(0.000)	0.842	(0.091)	0.868	(0.019)
$\sup W_{TAR}$	1	(0)	1	(0)	0.871	(0.007)	1	(0)
$\sup LM_{TAR}$	1	(0)	1	(0)	0.758	(0.009)	1	(0)
$\sup LR_{TAR}$	1	(0)	1	(0)	0.883	(0.007)	1	(0)

Table 3
DGP: Markov-switching model

DGP Sample size	Consumption				NYSE share volume			
	$\beta = 0.0228$ $\beta^* = -0.0926$		$p = 0.5279$ $q = 0.9761$ $\sigma_2^2 = 0.001$		$\beta = -7.3475$ $\beta^* = 21.1089$		$p = 0.05$ $q = 0.358$ $\sigma_2^2 = 150.3$	
	$T = 100$		$T = 1000$		$T = 100$		$T = 1000$	
$\hat{\alpha}$	0.017	(0.013)	0.018	(0.004)	1.214	(3.567)	1.172	(1.191)
$\hat{\alpha}^*$	0.000	(0.021)	0.000	(0.007)	−0.006	(5.933)	−0.013	(1.952)
$\hat{\pi}$	0.494	(0.238)	0.504	(0.240)	0.492	(0.236)	0.497	(0.237)
$\hat{\sigma}_1^2$	0.001	(0.000)	0.001	(0.000)	248.7	(32.84)	256.8	(10.37)
predict. SCA	0.565	(0.226)	0.537	(0.216)	0.518	(0.052)	0.507	(0.046)
$\sup W_{\text{SCA}}$	0.149	(0.008)	0.157	(0.008)	0.022	(0.003)	0.006	(0.001)
$\sup\text{LM}_{\text{SCA}}$	0.089	(0.006)	0.149	(0.008)	0.009	(0.002)	0.005	(0.001)
$\sup\text{LR}_{\text{SCA}}$	0.138	(0.008)	0.161	(0.008)	0.012	(0.002)	0.007	(0.002)
$\hat{\beta}$	0.028	(0.011)	0.024	(0.001)	−4.447	(4.395)	−5.615	(3.509)
$\hat{\beta}^*$	−0.064	(0.035)	−0.082	(0.016)	13.143	(10.12)	16.37	(8.633)
$\hat{p}$	0.493	(0.238)	0.454	(0.131)	0.215	(0.178)	0.132	(0.128)
$\hat{q}$	0.773	(0.263)	0.939	(0.070)	0.330	(0.170)	0.340	(0.079)
$\hat{\sigma}_2^2$	0.001	(0.000)	0.001	(0.000)	190.5	(64.57)	174.9	(48.27)
predict. MSA	0.807	(0.245)	0.969	(0.074)	0.824	(0.065)	0.856	(0.026)
$\sup\text{LR}_{\text{MSA}}$	0.434	(0.011)	0.967	(0.004)	0.342	(0.011)	0.765	(0.009)
$\hat{\gamma}$	0.021	(0.008)	0.021	(0.001)	−4.860	(3.072)	−3.775	(1.188)
$\hat{\gamma}^*$	−0.013	(0.020)	−0.014	(0.006)	11.041	(3.130)	8.401	(1.001)
$\hat{r}$	0.007	(0.027)	−0.011	(0.013)	2.590	(8.687)	4.384	(4.369)
$\hat{\sigma}_3^2$	0.001	(0.000)	0.001	(0.000)	227.3	(30.54)	240.8	(9.906)
predict. TAR	0.650	(0.230)	0.813	(0.10)	0.667	(0.064)	0.657	(0.028)
$\sup W_{\text{TAR}}$	0.149	(0.008)	0.748	(0.010)	0.726	(0.010)	1	(0)
$\sup\text{LM}_{\text{TAR}}$	0.082	(0.006)	0.738	(0.010)	0.611	(0.011)	1	(0)
$\sup\text{LR}_{\text{TAR}}$	0.192	(0.009)	0.832	(0.008)	0.638	(0.011)	1	(0)

4.1.1. Implementation of the sup tests

For the SCA, the statistics are maximized over all possible change-points t from $0.15T$ to $0.85T$, this corresponds to a choice of $\Pi = [0.15, 0.85]$ as advocated by Andrews (1993). For the TAR, the statistics are maximized over the values of r obtained by ranking the observations y_t and discarding 15% of the largest and smallest data. For the MSA, the likelihood is programmed using formulae [22.4.5]–[22.4.8] of Hamilton's book (1994). It is maximized by the DFP algorithm of the procedure Maxlik of GAUSS over all parameters, where p and q are restricted to a subset Ψ^2. Remark that for any symmetric subset of the form $\Psi = [\underline{p}, 1 - \underline{p}]$, $p \in \Psi$ and $q \in \Psi$ implies $\lambda \in \Psi$, hence $\Lambda = \Psi$. This result can be proved in the following way. Let $\underline{p} < p < 1 - \underline{p}$, λ satisfies the following inequality:

$$\frac{1-q}{2-\underline{p}-q} < \frac{1-q}{2-p-q} < \frac{1-q}{1+\underline{p}-q}.$$

Table 4
DGP: Threshold model

DGP Sample size	Car CPI				Unemployment			
	$\gamma=0.6133$ $\gamma^*=-0.3333$		$r=0.5828$ $\sigma_3^2=0.1682$		$\gamma=0.128$ $\gamma^*=-0.158$		$r=0.20$ $\sigma_3^2=0.03$	
	$T=100$		$T=1000$		$T=100$		$T=1000$	
$\hat{\alpha}$	0.393	(0.153)	0.389	(0.052)	0.0342	(0.081)	−0.033	(0.029)
$\hat{\alpha}^*$	0.003	(0.245)	0.001	(0.084)	0.002	(0.132)	−0.001	(0.050)
$\hat{\pi}$	0.490	(0.237)	0.497	(0.240)	0.487	(0.230)	0.505	(0.237)
$\hat{\sigma}_1^2$	0.179	(0.027)	0.191	(0.009)	0.031	(0.005)	0.035	(0.002)
predict. SCA	0.581	(0.100)	0.530	(0.084)	0.661	(0.105)	0.564	(0.050)
$\sup W_{\text{SCA}}$	0.245	(0.010)	0.260	(0.010)	0.552	(0.011)	0.624	(0.011)
$\sup\text{LM}_{\text{SCA}}$	0.152	(0.008)	0.248	(0.010)	0.417	(0.011)	0.613	(0.011)
$\sup\text{LR}_{\text{SCA}}$	0.197	(0.009)	0.250	(0.010)	0.465	(0.011)	0.613	(0.011)
$\hat{\beta}$	0.743	(0.140)	0.741	(0.038)	0.140	(0.072)	0.157	(0.020)
$\hat{\beta}^*$	−0.603	(0.110)	−0.564	(0.031)	−0.198	(0.090)	−0.211	(0.025)
$\hat{p}$	0.789	(0.151)	0.831	(0.029)	0.813	(0.188)	0.926	(0.048)
$\hat{q}$	0.723	(0.130)	0.721	(0.033)	0.800	(0.178)	0.893	(0.046)
$\hat{\sigma}_2^2$	0.112	(0.025)	0.118	(0.008)	0.025	(0.006)	0.025	(0.002)
predict. MSA	0.711	(0.099)	0.752	(0.019)	0.732	(0.116)	0.793	(0.027)
$\sup\text{LR}_{\text{MSA}}$	0.430	(0.011)	1	(0)	0.545	(0.011)	0.988	(0.002)
$\hat{\gamma}$	0.620	(0.086)	0.614	(0.024)	0.126	(0.037)	0.128	(0.009)
$\hat{\gamma}^*$	−0.350	(0.093)	−0.335	(0.028)	−0.157	(0.042)	−0.158	(0.011)
$\hat{r}$	0.557	(0.128)	0.582	(0.010)	0.182	(0.096)	0.199	(0.008)
$\hat{\sigma}_3^2$	0.162	(0.024)	0.168	(0.007)	0.029	(0.004)	0.03	(0.001)
predict. TAR	0.941	(0.094)	0.995	(0.006)	0.889	(0.064)	0.904	(0.010)
$\sup W_{\text{TAR}}$	0.858	(0.008)	1	(0)	0.914	(0.006)	1	(0)
$\sup\text{LM}_{\text{TAR}}$	0.793	(0.009)	1	(0)	0.856	(0.008)	1	(0)
$\sup\text{LR}_{\text{TAR}}$	0.839	(0.008)	1	(0)	0.902	(0.007)	1	(0)

The right-hand side of this inequality is a decreasing function of q and hence is maximized at $q=\underline{p}$ where it takes the value $1-\underline{p}$. The left-hand side is a decreasing function of q and reaches its minimum at $q=1-\underline{p}$ where it equals $\underline{p}$. We constrain p and q to lie in the interval $[0.05, 0.95]$ by using an arctan transformation. We use a larger interval than the usual $[0.15, 0.85]$ because the latter one turns out to be binding. Since the labels of the states 0 and 1 are arbitrary, we call state 0, the state with mean β and state 1 the state with mean $\beta+\beta^*$ where the sign of β^* is the same as that of α^*, β^*, or γ^* in the true model. For Tables 2–4, we adopt as starting values $p=0.5$, $q=0.5$, $\beta=2/3\bar{y}$, $\beta^*=2/3\bar{y}$, which correspond to the case $\beta=\beta^*$ and $P(S_t=1)=1/2$, and finally $\sigma_2^2=\sum_{t=1}^{T}(y_t-\bar{y})^2/T$. Note that these starting values do not correspond to a single state solution since $\beta^*\neq 0$ and is not a local maximum. To avoid local maxima under H_0 (Table 1), we investigate five sets of starting values and take the maximum over these sets as suggested by Garcia (1998). This way, the LR statistic is greater than −1e-4 with a 100% success rate. Since the three tests have the same asymptotic distribution under the null, we use the critical values given by Andrews (1993) for the appropriate interval: 8.85 for [0.15.0.85] and 9.84 for [0.05, 0.95].

4.1.2. Predictability

The three models under consideration have in common to exhibit two states. Let us call State 1 the state corresponding to the higher mean. It is interesting to investigate how well a model is able to predict State 1 conditionally on the observations $\{y_1, y_2, \ldots, y_T\}$. The state at date t is deterministic conditionally on $\hat{\pi}$ for the SCA and conditionally on $\hat{r}$ and the observation at date $t-1$ for the TAR. For the MSA, we calculate $P(S_t = 1|y_1, y_2, \ldots, y_T)$ using Kim's algorithm (see Hamilton, 1994, Appendix 22.A). When $P(S_t = 1|y_1, y_2, \ldots, y_T) > 0.5$, it is considered that the regime at date t is 1, otherwise it is 0. This way, we get a $T \times 1$ vector corresponding to the inferred states at date $t = 1, \ldots, T$. This list can be compared with the true states. For each model, we calculate the percentage of states that have been correctly predicted. In the lines "predict" of Tables 2–4, we report the mean and standard error of these percentages.

4.1.3. Empirical test size

To assess the properties of our tests under H_0, we performed simulations (reported in Table 1) on a linear model including just a Gaussian white noise plus an intercept. The number of repetitions for the sample size $T = 2000$ is 1000. $\text{SupLR}_{\text{MSA}}$ seems to underreject.[3] Other simulations (not reported here) show that the level does not seem to be sensitive to the innovation variance.

4.2. Results for the Structural Change model

We illustrate the methods on two models. The first one is based on Nile River flow data. This time series has been studied extensively in the statistical literature. We report below the parameter values found by Müller (1992). He fitted a Structural Change model on the annual volume of the Nile river from 1871 to 1970. In Table 2, we see that the sup tests against a Threshold alternative display a high power in accordance with the results of Section 3.2. The simulations corroborate the fact that SCA is a special case of MSA where one of the transition probability, p or q, is close to 1. Indeed, the MSA picks up most values of p and q at the boundary 0.95 and the associated test rejects almost all the time. SCA is also a special case of a TAR with a large r. It is striking how well these models perform in terms of predictability of the state (almost 100%) and in terms of estimation of the mean before and after the jump. Compare $\alpha = 1151$, $\alpha + \alpha^* = 1950.5$ with $\hat{\beta} = 1150.5$, $\hat{\beta} + \hat{\beta}^* = 1950.5$, and $\hat{\gamma} + \hat{\gamma}^* = 1153.4$, $\hat{\gamma} = 1950.5$.

The second model is obtained by testing linearity on monthly data on Deutsche marks per U.S. dollar from 1973-01 to 1986-01.[4] A break-point was found around June 1983.

[3] On the other hand, Garcia (1998) found that the $\text{supLR}_{\text{MSA}}$ tends to overreject. This difference may be due to the use of different optimization algorithms in presence of multiple local maxima.

[4] *Source*: International Financial Statistics.

Interestingly, Hamilton (1996) fitted a Markov-switching model on similar data (for a longer period) but could not exclude the possibility of a structural break. Here our results speak in favor of a Structural Change. Indeed $\text{supLR}_{\text{MSA}}$ and $\text{supLR}_{\text{TAR}}$ reject with a 100% success rate while data are generated by an SCA.

4.3. Results for the Markov-switching model

First, we consider as DGP the model found by Cecchetti et al. (1990) for consumption data. They estimated an equilibrium model of asset pricing where the growth rate of the endowment is assumed to follow a two-state Markov-switching model. They used annual consumption data from 1889 to 1985. Results are reported in Table 3. As expected, the sup tests with threshold alternative have a high power. And as expected also, the sup tests for structural change have a low power but somewhat larger than 5% because $p+q>1$. The lack of power of $\text{supLR}_{\text{MSA}}$ may have two explanations. First, its level might be lower than 5% (see Table 1). Second, the poor performance might stem from the fact that Regime 2 is very persistent, so that regime 1 might appear only rarely in the data. A really large sample would be necessary to exhibit enough changes and to permit the distinction between MSA and a white noise. Note that the TAR identifies relatively well the correct state (see line "predict. TAR").

Second, we examined monthly data on NYSE reported share volume [5] from 1980-01 to 1995-09. We fitted a Markov-switching model on the first difference of the log of the data in millions of dollars and used the resulting estimators as basis for our simulation experiment. This series has been investigated by Hamilton and Susmel (1994). They fit on weekly data an ARCH process where the coefficients vary according to an observed Markov chain. Since we do not take into account the presence of conditional heteroscedasticity we obtain a large variance. As a result, $\text{supLR}_{\text{MSA}}$ lacks of power in small sample. Because $p+q<1$, the power of $\text{supLR}_{\text{SCA}}$ is lower than its level.

4.4. Results for the Threshold model

The first set of values was obtained by fitting a Threshold model (with $d=1$) on the growth rate in percentage of the consumer price index [6] for new cars where we replaced three outliers by the sample mean. The data set included monthly data from 1975-01 to 1991-10. Koop and Potter (1999) have investigated the inflation on a longer period (1947–1998) and they found evidence in favor of a structural change. We consider a shorter series and our tests point toward a Threshold model. Simulations reported in Table 4 suggest that $\text{supLR}_{\text{MSA}}$ will have power when the alternative is TAR although this power might be low in small samples. It seems that Hamilton's algorithm does a relatively good job in identifying the changes originating from a TAR.

The second model is based on the growth rate of the U.S. Unemployment using quarterly data from 1970 I to 2001 III. [7] Unemployment is extensively studied by

[5] *Source*: Ramanathan (1998) and references therein.
[6] *Source*: Citibase Data Series.
[7] *Source*: www.economagic.com.

Hansen (1997). He uses monthly U.S. unemployment, denoted x_t, and finds that the differentiated series is better represented by a TAR with jumps driven by $x_{t-1} - x_{t-d}$ with $d = 12$. Using quarterly data, we find that fitting $y_t = x_t - x_{t-1}$ on the intercept and $I(x_{t-1} - x_{t-d} \leqslant r)$, the LR test is maximized for $d = 4$. We adopt this model and report the results in Table 4. This model does not exactly fit our setting. We do not actually regress y_t on $I(y_{t-d} \leqslant r)$. However, the results are interesting. Since we estimate the right Threshold model, we obtain an almost perfect fit. Because $\gamma^* < 0$, the power of the $\mathrm{LR}_{\mathrm{SCA}}$ is large.

5. Conclusion

This paper shows that testing only for a structural change might be very misleading and might result in adopting a linear model while the data are generated by another nonlinear model. On the other hand, the stability test based on a misspecified TAR model can detect parameter instability originating from SCA or MSA models.

These results suggest that the Structural Change model is easy to distinguish from the two other models. On the other hand, selecting between MSA and SCA seems to be much more challenging. In a companion paper (Carrasco, 2002), we propose to use a Wald Encompassing test (WET) developed by Gourieroux et al. (1983) and Mizon and Richard (1986) to discriminate between these two models.

Our approach has some limitations. We consider only autoregressive models with a given order. We allow only from a single break in the intercept and autoregressive coefficients but not in the innovation variance. Extensions to multiple breaks might raise problems of identifiability since an SCA with multiple breaks might look very much like an MSA or TAR. On the other hand, it should be relatively easy to accommodate a shift in the innovation variance.

In this paper, we restrict our attention to three specific models and assume that the DGP is one of those (or linear). It would be interesting to look at other nonlinear models like smooth threshold autoregressive models and bilinear models (Tong, 1990). However, for the focus of the paper, it is better to consider three relatively close alternatives. The more different the models, the easier to discriminate.

Acknowledgements

I would like to thank Jean-Pierre Florens, my former thesis advisor, for his insightful comments; the paper owes much to his detailed suggestions. I gratefully acknowledge helpful comments from Christophe Bontemps, Xiaohong Chen, Eric Ghysels, Lars Peter Hansen, Katerina Kyriazidou, Jean-François Richard, Michael Rockinger, Francisco Ruge-Murcia, and especially René Garcia, as well as from Ronald Gallant (the editor), an anonymous associate editor and two referees. I also benefited from comments made by participants of ESEM94, of ASSET Meeting, of ASU and of seminars at the University of Chicago, Crest (Paris), Core (Louvain-La-Neuve), Ohio State University, and Stanford. Of course, all errors are mine.

Appendix

Proof of Proposition 3. The Structural Change model can be rewritten as

$$y_t = \mathbf{y}_t'\alpha I\{t \leqslant [T\pi]\} + \mathbf{y}_t'\eta I\{t > [T\pi]\} + \varepsilon_t,$$

where $\eta = \alpha + \alpha^*$. Testing H_0: $\alpha^* = 0$ is equivalent to testing $\eta = \alpha$.

The mean log-likelihood is given by

$$Q_T = -\ln(\sqrt{2\pi}) - \ln(\sigma_1) - \frac{1}{2\sigma_1^2}\frac{1}{T}\sum_{t=1}^{T}\{y_t - \mathbf{y}_t'\alpha I\{t \leqslant [T\pi]\} + \mathbf{y}_t'\eta I\{t > [T\pi]\}\}^2.$$

The score functions are given by

$$\frac{\partial Q_T}{\partial \alpha} = -\frac{1}{\sigma_1^2}\frac{1}{T}\sum_{t \leqslant [T\pi]} \mathbf{y}_t\{y_t - \mathbf{y}_t'\alpha\},$$

$$\frac{\partial Q_T}{\partial \eta} = -\frac{1}{\sigma_1^2}\frac{1}{T}\sum_{t > [T\pi]} \mathbf{y}_t\{y_t - \mathbf{y}_t'\eta\},$$

$$\frac{\partial Q_T}{\partial \sigma_1^2} = -\frac{1}{2\sigma_1^2} + \frac{1}{2\sigma_1^4}\frac{1}{T}\sum_{t=1}^{T}\{y_t - \mathbf{y}_t'\alpha I\{t \leqslant [T\pi]\} + \mathbf{y}_t'\eta I\{t > [T\pi]\}\}^2.$$

From the first-order condition, $\partial/\partial\alpha\, Q_T = 0 \Leftrightarrow 1/T \sum_{t \leqslant [T\pi]} \mathbf{y}_t\{y_t - \mathbf{y}_t'\alpha\} = 0$, it follows that

$$\hat{\alpha}(\pi) = \left[\frac{1}{T}\sum_{t \leqslant [T\pi]} \mathbf{y}_t\mathbf{y}_t'\right]^{-1}\left[\frac{1}{T}\sum_{t \leqslant [T\pi]} \mathbf{y}_t y_t\right].$$

From the condition $\partial/\partial\eta\, Q_T = 0$, we get

$$\hat{\eta}(\pi) = \left[\frac{1}{T}\sum_{t > [T\pi]} \mathbf{y}_t\mathbf{y}_t'\right]^{-1}\left[\frac{1}{T}\sum_{t > [T\pi]} \mathbf{y}_t y_t\right].$$

By the LLN and the stationarity of $\{y_t\}$, we get

$$\hat{\eta}(\pi) - \hat{\alpha}(\pi) \xrightarrow{P} 0,$$

$$\hat{\alpha}(\pi) \xrightarrow{P} \alpha_a = E[\mathbf{y}_t\mathbf{y}_t']^{-1}E[\mathbf{y}_t y_t],$$

$$\hat{\sigma}^2(\pi) \xrightarrow{P} E[\{y_t - \mathbf{y}_t'\alpha_a\}^2],$$

where α_a denotes the pseudo-true value of α. The restricted estimator of α can be shown to converge also to α_a in probability.

The score functions can be considered as moment conditions. Our moment conditions are satisfied asymptotically for the values of parameters equal to the pseudo-true values. We are exactly in the same setting as in Andrews (1993), where the GMM conditions are exactly identified. To apply his results, we just need to check that Assumption 1 (pp. 830–831) are satisfied. This assumption guarantees the Functional Central Limit Theorem. We shall refer to m_t as the moment conditions:

$$m_t = \begin{cases} \dfrac{1}{\sigma_{1a}^2}\{\mathbf{y}_t(y_t - \mathbf{y}_t'\alpha_a)\}, \\ -\dfrac{1}{2\sigma_{1a}^2} + \dfrac{1}{2\sigma_{1a}^4}\{y_t - \mathbf{y}_t'\alpha_a\}^2. \end{cases}$$

Verification of Assumption 1 (Andrews, 1993):

Assumption 1(a) and 1(b): Near epoch dependence (NED) property.

By Assumption 5, the MSA process is geometrically β-mixing. On the other hand, the TAR process is β-mixing under Assumption 4. Indeed, Chan and Tong (1985) show that, under Assumption 4, a TAR is geometrically ergodic. This implies that the process is geometrically β-mixing using results by Doukhan (1994). Note that β-mixing implies strong mixing. Therefore, both processes TAR and MSA are stationary L^2-NED of arbitrarily large size on a strong mixing base (the process itself) of arbitrarily large size. Finally, $\{m_t\}$ has the same NED property as $\{y_t\}$ itself since the sum and product of NED processes are NED (see Gallant and White, 1988, Corollary 4.3). Assumption 1(b) imposes a moment condition $\mathrm{E}\|m_t\|^{2+\varepsilon} < \infty$ for some $\varepsilon > 0$ that is satisfied under Assumption 7. Therefore, conditions (a) and (b) of Andrews' Assumption 1 are satisfied.

Assumption 1(c) is satisfied as soon as S is definite positive.

Assumption 1(d) requires the verification of Assumption A by Andrews. It includes among others an identification assumption, a condition on the weighting matrix (trivially satisfied here since conditions are exactly identified) and a moment condition: $\mathrm{E}\sup_{\theta\in\Theta_0}\|m_t\|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$. This moment condition is satisfied as long as Θ_0 is compact and Assumption 7 is fulfilled.

Assumption 1(e) does not apply since we are in an exactly identified case. Assumption 1(f) includes a moment condition $\mathrm{E}\sup_{\theta\in\Theta_0}\|\partial/\partial\theta\, m_t\|^{1+\varepsilon} < \infty$ for some $\varepsilon > 0$ which is satisfied under Assumption 7. Assumption 1(g) is immediate to verify. Assumption 1(h) is satisfied for Π bounded away from zero and M nonsingular.

Andrews' Assumption 1 being satisfied, his results can be applied. The limiting distributions of $\mathrm{LM}_T(\pi)$ and $W_T(\pi)$ follow immediately from Andrews' Theorem 3. However, Andrews consider only a specification robust version of the LR test. Our LR test is not robust to misspecification. From Andrews' proof of Theorem 3 (p. 850), we get

$$\begin{aligned}\sqrt{T}\hat{\alpha}^*(\pi) &= \sqrt{T}(\hat{\eta} - \hat{\alpha})(\pi)\\ &\Rightarrow C\frac{B_{l+1}(\pi) - \pi B_{l+1}(1)}{\pi(1-\pi)},\end{aligned}$$

where $C = (M'S^{-1}M)^{-1}M'S^{-1/2}$. From Foutz and Srivastava (1977) and references therein, we know that the LR test is equivalent to

$$T\hat{\alpha}^*(\pi)'[AM(\theta_a,\pi)^{-1}A']^{-1}\hat{\alpha}^*(\pi),$$

where $M(\theta_a,\pi)$ is defined in (7) and $AM(\theta_a,\pi)^{-1}A' = M^{-1}/(\pi(1-\pi))$. We have

$$\mathrm{LR}_T(\pi) \Rightarrow (B_{l+1}(\pi) - \pi B_{l+1}(1))'C'M^{-1}C(B_{l+1}(\pi) - \pi B_{l+1}(1))/[\pi(1-\pi)],$$

where $M \equiv M(\theta_a)$. Using the properties of the quadratic form of normal variables (Johnson and Kotz, 1970), $\mathrm{LR}_T(\pi)$ converges to a sum of chi-square processes weighted by the eigenvalues of $CC'M = (M'S^{-1}M)^{-1}M = M^{-1}S$. Finally, the limiting distribution of supLR follows from the continuous mapping theorem. This concludes the proof of Proposition 3. □

The following lemma will be useful for the proofs of Propositions 4–12

Lemma A.1. *Consider $\{S_t\}$ a two-state Markov chain with transition probability $P(S_1 = 1|S_{t-1} = 1) = p$ and $P(S_1 = 0|S_{t-1} = 0) = q$. Then*

$$\mathrm{E}(S_t|S_{t-l}) = \lambda[1-(p+q-1)^l] + (p+q-1)^l S_{t-l},$$

$$\mathrm{cov}(S_t, S_{t-l}) = (p+q-1)^l \lambda(1-\lambda)$$

where $\lambda = (1-p)/(2-p-q)$.

This lemma follows from Cox and Miller (1965, p. 82).

Proof of Propositions 4 and 5. For π fixed the usual results apply, we have

$$\sqrt{T}(\hat{\theta} - \theta_a) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \Omega(\theta_a, \pi)).$$

where $\Omega(\theta_a, \pi) = M^{-1}(\theta_a, \pi)V(\theta_a, \pi)M^{-1}(\theta_a, \pi)$. By Lemma A.5 of Andrews (1993), the estimation of σ_1^2 does not affect the asymptotic distribution of $(\hat{\alpha}, \hat{\alpha}^*)'$. We can, therefore, focus of the upper-left block of the matrices M and V. Using Eq. (7), we obtain

$$\sqrt{T}\begin{pmatrix} \hat{\alpha} - \alpha_a \\ \hat{\alpha}^* \end{pmatrix} \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, M^{-1}SM^{-1}\begin{bmatrix} 1 & 1-\pi \\ 1-\pi & 1-\pi \end{bmatrix}^{-1}\right). \tag{8}$$

We apply Proposition 3. Note that $\mathbf{y}_t = 1$, therefore we have $M = 1/\sigma_{1a}^2$ and

$$\begin{aligned} S &= \frac{1}{\sigma_{1a}^4} \lim_{T\to\infty} \frac{1}{T} V\left(\sum_{t=1}^T y_t\right) \\ &= \frac{1}{\sigma_{1a}^4} \lim_{T\to\infty} \left\{ \frac{1}{T}\beta^{*2} V\left(\sum_{t=1}^T S_t\right) + \frac{1}{T} V\left(\sum_{t=1}^T \omega_t\right)\right\}. \end{aligned}$$

Using Lemma A.1, the variance of the numbers of periods spent in State 1 is asymptotically equal to $\lambda(1-\lambda)(p+q)/(2-p-q)$. We obtain

$$S = \frac{1}{\sigma_{1a}^4}\left\{\beta^{*2}\lambda(1-\lambda)\frac{p+q}{2-p-q} + \sigma_2^2\right\}. \qquad \square$$

The following lemma will be used in the proof of Propositions 6 and 7.

Lemma A.2. *If $\{y_t\}$ is generated by Model* (4) *with $l = 0$, we have*

$$\sum_{t=1}^{\infty} \mathrm{cov}(y_0, y_t) = \frac{\gamma^*[\mathrm{E}(yI\{y \leqslant r\}) - (\gamma + \delta\gamma^*)\delta]}{1 - \Phi((r-\gamma-\gamma^*)/\sigma_3) + \Phi((r-\gamma)/\sigma_3)}$$

with

$$\mathrm{E}(yI\{y \leqslant r\}) = \delta\gamma + \delta\gamma^* \Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \left[\delta\varphi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) + (1-\delta)\varphi\left(\frac{r-\gamma}{\sigma_3}\right)\right]$$

and

$$\delta = \frac{\Phi((r-\gamma)/\sigma_3)}{1-\Phi((r-\gamma-\gamma^*)/\sigma_3) + \Phi((r-\gamma)/\sigma_3)},$$

where φ and Φ are, respectively, the p.d.f. and c.d.f. of $\mathcal{N}(0,1)$.

Proof of Lemma A.2. We calculate the first term of the sum:

$$\begin{aligned}\mathrm{cov}(y_t, y_{t+d}) &= \mathrm{E}(y_t y_{t+d}) - m^2\\ &= \mathrm{E}[y_t \mathrm{E}(y_{t+d}|y_t)] - m^2\\ &= \gamma m + \gamma^* \mathrm{E}(yI\{y \leqslant r\}) - m^2.\end{aligned}$$

With a look on the p.d.f. of y_t, we see that $\mathrm{E}(yI\{y \leqslant r\})$ is the weighted sum of two truncated Gaussian density functions.

$$\begin{aligned}\mathrm{E}(yI\{y \leqslant r\}) &= \delta \int_{y\leqslant r} \frac{1}{\sqrt{2\pi}\,\sigma_3} y \exp - \frac{(y-\gamma-\gamma^*)^2}{2\sigma_3^2}\,\mathrm{d}y + (1-\delta)\\ &\quad \int_{y\leqslant r} \frac{1}{\sqrt{2\pi}\,\sigma_3} y \exp - \frac{(y-\gamma)^2}{2\sigma_3^2}\,\mathrm{d}y\\ &= \delta\left[(\gamma+\gamma^*)\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \sigma_3\varphi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right)\right]\\ &\quad + (1-\delta)\left[\gamma\Phi\left(\frac{r-\gamma}{\sigma_3}\right) - \varphi\left(\frac{r-\gamma}{\sigma_3}\right)\right]\\ &= \delta\gamma + \delta\gamma^*\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right)\\ &\quad - \left[\delta\varphi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) + (1-\delta)\varphi\left(\frac{r-\gamma}{\sigma_3}\right)\right].\end{aligned}$$

Before computing the covariance, note that from Lemma 1, $X_k = I\{y_{kd} \leqslant r\}$ is a two-state Markov chain with transition probabilities

$$P[X_k = 1|X_{k-1} = 1] = \Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right),$$

$$P[X_k = 0|X_{k-1} = 0] = 1 - \Phi\left(\frac{r-\gamma}{\sigma_3}\right).$$

Therefore, Lemma A.1 applies and we have

$$\mathrm{E}[X_k|X_0] = \delta\left\{1 - \left[\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \Phi\left(\frac{r-\gamma}{\sigma_3}\right)\right]^k\right\} + \left[\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \Phi\left(\frac{r-\gamma}{\sigma_3}\right)\right]^k X_0.$$

Now we turn to the covariance:

$$\begin{aligned}
&\mathrm{cov}(y_0, y_{(k+1)d})\\
&= \mathrm{E}[y_0 \mathrm{E}(y_{(k+1)d}|y_0)] - m^2\\
&= \mathrm{E}\{y_0[\gamma + \gamma^* \mathrm{E}(X_k|X_0)]\} - m^2\\
&= \gamma m + \gamma^*\delta m\left\{1 - \left[\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \Phi\left(\frac{r-\gamma}{\sigma_3}\right)\right]^k\right\}\\
&\quad + \gamma^* \mathrm{E}(yI\{y \leqslant r\})\left[\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \Phi\left(\frac{r-\gamma}{\sigma_3}\right)\right]^k - m^2.
\end{aligned}$$

Remark that $\gamma m - m^2 = -\gamma^* m\delta$. After simplification, we obtain

$$\mathrm{cov}(y_0, y_{(k+1)d}) = \gamma^*\left[\Phi\left(\frac{r-\gamma-\gamma^*}{\sigma_3}\right) - \Phi\left(\frac{r-\gamma}{\sigma_3}\right)\right]^k [\mathrm{E}(yI\{y \leqslant r\}) - m\delta] \quad (9)$$

We deduce the result:

$$\sum_{t=1}^{\infty} \mathrm{cov}(y_0, y_t) = \frac{\gamma^*[\mathrm{E}(yI\{y \leqslant r\}) - m\delta]}{1 - \Phi((r-\gamma-\gamma^*)/\sigma_3) + \Phi((r-\gamma)/\sigma_3)}. \qquad \square$$

Proof of Propositions 6 and 7. The asymptotic distribution of $(\hat{\alpha}, \hat{\alpha}^*)'$ is given by Eq. (8) given in the proof of Proposition 4. Again we apply Proposition 3. We have $M = 1/\sigma_{1a}^2$ and

$$S = \frac{1}{\sigma_{1a}^4} \lim_{T\to\infty} \frac{1}{T} V\left[\sum y_t\right] \to \frac{1}{\sigma_{1a}^4}\left\{V(y_0) + 2\sum_{j=1}^{\infty} \mathrm{cov}(y_0, y_j)\right\}.$$

By Lemma A.2, we obtain the desired result. $\square$

Proof of Proposition 8. The Threshold model can be rewritten as

$$y_t = \mathbf{y}_t'\gamma I\{y_{t-d} > r\} + \mathbf{y}_t'\phi I\{y_{t-d} \leqslant r\} + u_t,$$

where $\phi = \gamma + \gamma^*$ and $\mathbf{y}_t = (1,\ y_{t-1}, \ldots, y_{t-l})'$. The OLS estimates of γ and ϕ are given by

$$\begin{pmatrix} \hat{\gamma} \\ \hat{\phi} \end{pmatrix} = \begin{pmatrix} \left[\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} > r\}\right]^{-1} \sum_{t=1}^{T} \mathbf{y}_t y_t I\{y_{t-d} > r\} \\ \left[\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} \leqslant r\}\right]^{-1} \sum_{t=1}^{T} \mathbf{y}_t y_t I\{y_{t-d} \leqslant r\} \end{pmatrix}.$$

Then the estimator of γ^* is given by $\hat{\phi} - \hat{\gamma}$.

1. When the data are generated by a Structural Change model

$$\hat{\gamma}^* = \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} \leqslant r\}\right]^{-1} \frac{1}{T} \sum_{t=[T\pi]+1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} \leqslant r\}\alpha^*$$
$$- \left[\frac{1}{T}\sum_{t=1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} > r\}\right]^{-1} \frac{1}{T} \sum_{t=[T\pi]+1}^{T} \mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} > r\}\alpha^*.$$

   To obtain the limit, one should split the sum in two subsamples, before and after the change point. Except for the case $\alpha^* = 0$, $\hat{\gamma}^*$ does not converge to zero.
2. When the data are generated by a Markov-switching model:

$$\hat{\gamma}^* \xrightarrow{\mathrm{P}} [\mathrm{E}(\mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} \leqslant r\})]^{-1} \mathrm{E}(\mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} \leqslant r\} S_t)\beta^*$$
$$- [\mathrm{E}(\mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} > r\})]^{-1} \mathrm{E}(\mathbf{y}_t \mathbf{y}_t' I\{y_{t-d} > r\} S_t)\beta^*.$$

The convergence is guaranteed by the stationarity of the MSA. This limit is in general different from zero. It will be equal to zero in two cases. First, when $\beta^* = 0$, this is H_0. Second, when $p+q=1$, then $P(S_t=1|S_{t-1}=0)=1-q=p=P(S_t=1|S_{t-1}=1)$, that is to say $\{S_t\}$ are independent, therefore $\mathbf{y}_t$ and y_{t-d} are independent of S_t. The limit of γ^* is equal to zero. In all other cases, the pseudo-true value of γ^* should be different from zero. However, it will approach zero when d increases.

Since the pseudo-true value of γ^* is different from zero, the statistics diverge. □

Proof of Proposition 9. The objective function is given by

$$Q_T = \frac{1}{T}\sum_{t=1}^{T} -\frac{1}{2\sigma_3^2}(y_t - \gamma - \gamma^* I\{y_{t-d} \leqslant r\})^2.$$

The minimization of Q_T yield:

$$\hat{\gamma} = \frac{\bar{y} - 1/T \sum_{t=1}^{T} y_t I\{y_{t-d} \leqslant r\}}{1 - 1/T \sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}},$$

$$\hat{\gamma}^* = \frac{1/T \sum_{t=1}^{T} y_t I\{y_{t-d} \leqslant r\} - \bar{y} 1/T \sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}}{1/T \sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}(1 - 1/T \sum_{t=1}^{T} I\{y_{t-d} \leqslant r\})},$$

where $\bar{y} = 1/T \sum_{t=1}^{T} y_t$.

We have

$$\bar{y} \xrightarrow{P} \alpha + \alpha^*(1-\pi),$$

$$\frac{1}{T}\sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}$$

$$= \frac{T\pi + d}{T}\frac{1}{T\pi + d}\sum_{t \leqslant T\pi + d} I\{\alpha + \varepsilon_{t-d} \leqslant r\}$$

$$+ \frac{T - T\pi - d - 1}{T}\frac{1}{T - T\pi - d - 1}\sum_{t > T\pi + d} I\{\alpha + \alpha^* + \varepsilon_{t-d} \leqslant r\}$$

$$\xrightarrow{P} \pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right) + (1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right) \tag{10}$$

and

$$\frac{1}{T}\sum_{t=1}^{T} y_t I\{y_{t-d} \leqslant r\}$$

$$= \frac{1}{T}\sum_{t \leqslant T\pi} (\alpha + \varepsilon_t) I\{\alpha + \varepsilon_{t-d} \leqslant r\}$$

$$+ \frac{1}{T}\sum_{T\pi < t \leqslant T\pi + d} (\alpha + \alpha^* + \varepsilon_t) I\{\alpha + \varepsilon_{t-d} \leqslant r\}$$

$$+ \frac{1}{T}\sum_{t > T\pi + d} (\alpha + \alpha^* + \varepsilon_t) I\{\alpha + \alpha^* + \varepsilon_{t-d} \leqslant r\}$$

$$\xrightarrow{P} \pi\alpha\Phi\left(\frac{r-\alpha}{\sigma_1}\right) + 0 + (1-\pi)(\alpha+\alpha^*)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)$$

by the LLN. □

The following lemma is used in the proof of Proposition 10.

Lemma A.3. *Assume one estimates Model* (6), *while the DGP is* (2) *with* $l = 0$. *Denote* $g = (\gamma, \gamma^*)'$ *and* $M_{\mathrm{SC}} = -\partial^2 Q_T / \partial g \partial g'$. *We have*

$$M_{\mathrm{SC}}(\theta) = -\frac{1}{\sigma_3^2}\begin{bmatrix} 1 & \pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right) + (1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right) \\ \pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right) + (1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right) & \pi\Phi(\frac{r-\alpha}{\sigma_1}) + (1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right) \end{bmatrix}.$$

Denote $V_{\mathrm{SC}}(\theta) = \lim \mathrm{var}[\sqrt{T}\partial Q_T / \partial g]$. *It is equal to* $V^{\mathrm{SC}}/\sigma_{3a}^4$ *where the elements of the matrix* V^{SC} *are given by*

$$V_{11}^{\mathrm{SC}} = \sigma_1^2 + \gamma_a^{*2}\left[\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\left(1 - \Phi\left(\frac{r-\alpha}{\sigma_1}\right)\right)\right.$$

$$+(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\left(1-\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right)\Bigg]$$
$$+2\gamma_a^*\sigma\left[\pi\varphi\left(\frac{r-\alpha}{\sigma_1}\right)+(1-\pi)\varphi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right],$$

$$V_{22}^{\mathrm{SC}}=\sigma^2\left[\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)+(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]$$
$$+(\alpha-\gamma_a-\gamma_a^*)^2\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\left(1-\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\right)$$
$$+(\alpha+\alpha^*-\gamma_a-\gamma_a^*)^2(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\left(1-\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right)$$
$$-2\sigma\Bigg[(\alpha-\gamma_a-\gamma_a^*)\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\varphi\left(\frac{r-\alpha}{\sigma_1}\right)$$
$$+(\alpha+\alpha^*-\gamma_a-\gamma_a^*)(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\varphi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\Bigg],$$

$$V_{12}^{\mathrm{SC}}=\sigma^2\left[\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)+(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]$$
$$-\sigma\left[(\alpha-\gamma_a-\gamma_a^*)\pi\varphi\left(\frac{r-\alpha}{\sigma_1}\right)+(\alpha+\alpha^*-\gamma_a-\gamma_a^*)(1-\pi)\varphi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]$$
$$+\gamma^*\sigma\left[\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\varphi\left(\frac{r-\alpha}{\sigma_1}\right)+(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\varphi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right]$$
$$-\gamma^*\Bigg[(\alpha-\gamma_a-\gamma_a^*)\pi\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\left(1-\Phi\left(\frac{r-\alpha}{\sigma_1}\right)\right)$$
$$+(\alpha+\alpha^*-\gamma_a-\gamma_a^*)(1-\pi)\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\left(1-\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right)\right)\Bigg],$$

and we have $K_3=AM_{\mathrm{SC}}^{-1}(\theta_a)V_{\mathrm{SC}}(\theta_a)M_{\mathrm{SC}}^{-1}(\theta_a)A'$ *where* $A=[0,1]$.

Proof of Lemma A.3. We have

$$\frac{\partial^2 Q_T}{\partial\gamma^2}=-\frac{1}{\sigma_{3a}^2},$$

$$\frac{\partial^2 Q_T}{\partial\gamma^{*2}}=-\frac{1}{T\sigma_{3a}^2}\sum_{t=1}^{T}I\{y_{t-d}\leqslant r\},$$

$$\frac{\partial^2 Q_T}{\partial\gamma\partial\gamma^*}=-\frac{1}{T\sigma_{3a}^2}\sum_{t=1}^{T}I\{y_{t-d}\leqslant r\}.$$

The expression of M_{SC} follows from Eq. (10). For the terms of V_{SC}, we use

$$TV\left(\frac{\partial Q_T}{\partial \gamma}\right) = \frac{1}{\sigma_{3a}^4}\,TV\left(\frac{1}{T}\sum_{t=1}^{T}(y_t - \gamma_a^* I\{y_{t-d} \leqslant r\})\right)$$

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}(y_t - \gamma_a^* I\{y_{t-d} \leqslant r\})\\
&\quad = \frac{1}{T}\sum_{t=1}^{T\pi}(\alpha + \varepsilon_t - \gamma_a^* I\{\alpha + \varepsilon_{t-d} \leqslant r\})\\
&\qquad + \frac{1}{T}\sum_{t=T\pi+1}^{T\pi+d+1}(\alpha + \alpha^* + \varepsilon_t - \gamma_a^* I\{\alpha + \varepsilon_{t-d} \leqslant r\})\\
&\qquad + \frac{1}{T}\sum_{t=T\pi+d+1}^{T}(\alpha + \alpha^* + \varepsilon_t - \gamma_a^* I\{\alpha + \alpha^* + \varepsilon_{t-d} \leqslant r\}).
\end{aligned}$$

Note that the second summand converges to zero when T approaches to infinity. Moreover, the first and third terms are independent. The variance of the first term is given by:

$$\begin{aligned}
&TV\left(\frac{1}{T}\sum_{t=1}^{T\pi}(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\})\right)\\
&\quad = \pi V(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\})\\
&\qquad + 2\sum_{t=1}^{T\pi}\mathrm{cov}(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\}, \varepsilon_1 - \gamma_a^* I\{\varepsilon_{1-d} \leqslant r - \alpha\}).
\end{aligned}$$

Remark that the covariance is equal to zero for all $t \neq d+1$. We have

$$\begin{aligned}
&TV\left(\frac{1}{T}\sum_{t=1}^{T\pi}(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\})\right)\\
&\quad \overset{T\to\infty}{\to} \pi\left[\sigma^2 + \gamma_a^* \Phi\left(\frac{r-\alpha}{\sigma_1}\right)\left(1 - \Phi\left(\frac{r-\alpha}{\sigma_1}\right)\right) + 2\gamma_a^* \sigma_1 \pi \varphi\left(\frac{r-\alpha}{\sigma_1}\right)\right].
\end{aligned}$$

The variance of the second term may be calculated by the same way, the result follows. For the variance of the score with respect to γ^*,

$$TV\left(\frac{\partial Q_T}{\partial \gamma^*}\right) = \frac{T}{\sigma_{3a}^4}\,V\left(\frac{1}{T}\sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}(y_t - \gamma_a - \gamma_a^* I\{y_{t-d} \leqslant r\})\right)$$

the result is obtained using the same steps as above.

$$T\operatorname{cov}\left(\frac{\partial Q_T}{\partial \gamma}, \frac{\partial Q_T}{\partial \gamma^*}\right) = \frac{T}{\sigma_{3a}^4}\operatorname{cov}\left(\frac{1}{T}\sum_{t=1}^{T}(y_t - \gamma_a^* I\{y_{t-d} \leqslant r\}), \right.$$
$$\left. \times \frac{1}{T}\sum_{t=1}^{T} I\{y_{t-d} \leqslant r\}(y_t - \gamma_a - \gamma_a^* I\{y_{t-d} \leqslant r\})\right)$$

which is equal to

$$= \frac{T}{\sigma_{3a}^4}\operatorname{cov}\left(\frac{1}{T}\sum_{t=1}^{T\pi}(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\}), \right.$$
$$\left. \times \quad \frac{1}{T}\sum_{t=1}^{T\pi} I\{\varepsilon_{t-d} \leqslant r - \alpha\}(\varepsilon_t + \alpha - \gamma_a - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha\})\right)$$
$$+ \frac{T}{\sigma_{3a}^4}\operatorname{cov}\left(\frac{1}{T}\sum_{t=T\pi+d+1}^{T}(\varepsilon_t - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha - \alpha^*\}), \right.$$
$$\left. \times \frac{1}{T}\sum_{t=T\pi+d+1}^{T} I\{\varepsilon_{t-d} \leqslant r - \alpha - \alpha^*\}(\varepsilon_t + \alpha + \alpha^* - \gamma_a - \gamma_a^* I\{\varepsilon_{t-d} \leqslant r - \alpha - \alpha^*\})\right)$$

plus a term which converges to zero when T approaches infinity. □

Proof of Proposition 10. (i) follows directly from Lemma A.3. As in the estimation of an AR model, the limiting distribution of $(\hat{\gamma}, \hat{\gamma}^*)$ is not affected by that of $\hat{\sigma}_3^2$, we can proceed as if σ_3^2 were known a priori. Therefore, the expression of the variance of $\sqrt{T}\hat{\gamma}^*$ is given by K_3 that does not depend on σ_{3a}^2.

(ii) Again, we can proceed as if σ_3^2 were known. Here, it is important to distinguish between g_{Ta} and g_a. We denote $g_{Ta} = (\gamma_{Ta}, \gamma_{Ta}^*)$ the pseudo-true value given by Lemma 9 where α^* is replaced by $a/T^{1/4}$. Its limit as T approaches infinity is denoted g_a and equals $(\alpha, 0)'$. Note that, because the null hypothesis holds at the limit, $\sigma_{3a}^2 = \sigma_1^2$. We denote $\theta_a = (\alpha, 0, \sigma_1^2)'$. Using Taylor expansions (around the pseudo-true value θ_a) as in Gourieroux and Monfort (1989, p. 100), we obtain

$$\mathrm{LR}_T \sim T\hat{\gamma}^{*\prime}[AM_T^{-1}(\theta_a, r)A']^{-1}\hat{\gamma}^*, \tag{11}$$

where $A = [0, 1]$. Moreover, from Lemma A.3, we have

$$M_T(\theta_a, r) \xrightarrow{\mathrm{P}} \frac{1}{\sigma_1^2}\begin{bmatrix} 1 & \delta_0 \\ \delta_0 & \delta_0 \end{bmatrix},$$

where $\delta_0 = \Phi(r - \alpha/\sigma_1)$. Using Lemma A.3 and $\gamma_a = \alpha$, $\gamma_a^* = 0$, we have $V(\theta_a, r) = M(\theta_a, r)$. Moreover,

$$\sqrt{T}(\hat{\gamma}^* - \gamma_{Ta}^*) \xrightarrow{\mathrm{d}} \mathrm{N}(0, A\Omega(\theta_a, r)A') \tag{12}$$

with $\Omega(\theta_a, r) = M^{-1}(\theta_a, r)$ and hence $A\Omega(\theta_a)A' = \sigma_{3a}^2/(\delta_0(1-\delta_0))$. A Taylor expansion of $\Phi(r - \alpha - \alpha^*/\sigma_1)$ around $\alpha^* = 0$ yields the following result:

$$\Phi\left(\frac{r-\alpha-\alpha^*}{\sigma_1}\right) - \Phi\left(\frac{r-\alpha}{\sigma_1}\right) = -\frac{\alpha^*}{\sigma_1}\varphi\left(\frac{r-\alpha}{\sigma_1}\right).$$

Therefore,

$$\sqrt{T}\gamma_{Ta}^* \overset{T\to\infty}{\to} \tilde{\gamma} \equiv -a^2\frac{(1-\pi)\pi\varphi((r-\alpha)/\sigma_1)}{\sigma_1\delta_0(1-\delta_0)}.$$

From (11) and (12), we deduce that

$$\mathrm{LR}_T \overset{\mathrm{d}}{\to} \chi^2(1, \tilde{\gamma}^2(A\Omega(\theta_a, r)A')^{-1}).$$

The limiting distributions of W_T and LM_T can be derived in the same manner. □

Proof of Proposition 11.

$$Q_T = \frac{1}{T}\sum_{t=1}^{T} -\frac{1}{2\sigma_3^2}(y_t - \gamma - \gamma^* I\{y_{t-d} \leqslant r\})^2.$$

By stationarity and ergodicity of $\{y_t\}$, we have

$$Q_T \overset{T\to\infty}{\to} Q_\infty = -\frac{1}{2\sigma_{3a}^2}\,\mathrm{E}[(y_t - \gamma - \gamma^* I\{y_{t-d} \leqslant r\})^2].$$

γ_a and γ_a^* are the results of the maximization of Q_∞ with respect to γ and γ^*. We obtain

$$\gamma_a = \frac{\mathrm{E}(y_t) - \mathrm{E}(y_t I\{y_{t-d} \leqslant r\})}{1 - \mathrm{E}(I\{y_{t-d} \leqslant r\})},$$

$$\gamma_a^* = \frac{\mathrm{E}(y_t I\{y_{t-d} \leqslant r\}) - \mathrm{E}(y_t)\mathrm{E}(I\{y_{t-d} \leqslant r\})}{\mathrm{E}(I\{y_{t-d} \leqslant r\})(1 - \mathrm{E}(I\{y_{t-d} \leqslant r\}))}.$$

We will compute now the different terms of these equations. We obtain immediately

$$\mathrm{E}(y_t) = \beta + \beta^*\lambda,$$

$$\mathrm{E}(I\{y_{t-d} \leqslant r\}) = (1-\lambda)\Phi\left(\frac{r-\beta}{\sigma_2}\right) + \lambda\Phi\left(\frac{r-\beta-\beta^*}{\sigma_2}\right),$$

$$\mathrm{E}(y_t I\{y_{t-d} \leqslant r\}) = \mathrm{E}[I\{y_{t-d} \leqslant r\}\mathrm{E}(y_t|\, y_{t-d} \leqslant r)].$$

Firstly, we calculate $\mathrm{E}(y_t|\, y_{t-d} \leqslant r)$:

$$\begin{aligned}\mathrm{E}(y_t|y_{t-d} \leqslant r) &= \beta + \beta^*\mathrm{E}(S_t|\, y_{t-d} \leqslant r)\\ &= \beta + \beta^*\mathrm{E}[\mathrm{E}(S_t|\, S_{t-d})|\, y_{t-d} \leqslant r].\end{aligned}$$

Replacing $\mathrm{E}(S_t|\, S_{t-d})$ by its expression given by Lemma A.1, we obtain

$$\mathrm{E}(S_t|\, y_{t-d} \leqslant r) = \lambda[1 - (p+q-1)^d] + (p+q-1)^d\mathrm{E}(S_{t-d}|\, y_{t-d} \leqslant r).$$

Using the Bayes Formula, we establish that

$$P(S_t = 1 \mid y_t \leqslant r) = \frac{P(y_t \leqslant r|S_t = 1)P(S_t = 1)}{P(y_t \leqslant r)}$$

$$= \frac{\lambda\Phi((r - \beta - \beta^*)/\sigma_2)}{(1-\lambda)\Phi((r-\beta)/\sigma_2) + \lambda\Phi((r - \beta - \beta^*)/\sigma_2)}.$$

We deduce that

$$\mathrm{E}(y_t I\{y_{t-d} \leqslant r\}) = \mathrm{E}(I\{y_t \leqslant r\})[\beta + \beta^*\lambda(1 - (p+q-1)^d)]$$

$$+ \beta^*(p+q-1)^d\lambda\Phi\left(\frac{r - \beta - \beta_1}{\sigma}\right).$$

By replacing in the expression of γ_a and γ_a^*, we obtain the result. □

Proof of Proposition 12. It is enough to note that as T approaches infinity, the pseudo-true value converges to $\theta_a \equiv (\beta, 0, \sigma_2^2)'$ because $\sigma_{3a}^2 = \sigma_2^2$ under H_0. Since at the limit we are under H_0, $\Omega(\theta_a, r) = M^{-1}(\theta_a, r)$ with

$$M(\theta_a, r) = -\frac{1}{\sigma_2^2}\begin{bmatrix} 1 & \delta_1 \\ \delta_1 & \delta_1 \end{bmatrix}$$

and $\delta_1 = P(\beta + \varepsilon_{t-d} \leqslant r) = \Phi((r-\beta)/\sigma_{2a})$. The rest of the proof is similar to that of Proposition 10. □

References

Andrews, D.W.K., 1993. Tests for parameter instability and structural change point. Econometrica 61, 821–856.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383–1414.

Andrews, D.W.K., Ploberger, W., 1995. Admissibility of the likelihood ratio test when a nuisance parameter is present only under the alternative. The Annals of Statistics 23, 1609–1629.

Banerjee, J., Lumsdaine, R., Stock, J., 1992. Recursive and sequential tests for a unit root: theory and international evidence. Journal of Business & Economic Statistics 10, 271–287.

Bhattacharya, R., Lee, C., 1995. On geometric ergodicity of nonlinear autoregressive models. Statistics & Probability Letters 22, 311–315.

Carrasco, M., 2002. Model selection strategy for nonlinear specifications. Mimeo, University of Rochester.

Cecchetti, S.G., Lam, P., Mark, N.C., 1990. Mean reversion in equilibrium asset prices. The American Economic Review 80 (3), 398–418.

Chan, K.S., 1990. Testing for threshold autoregression. The Annals of Statistics 18, 1886–1894.

Chan, K.S., Tong, H., 1985. On the use of the deterministic Lyapunov functions for the ergodicity of stochastic difference equations. Advances in Applied Probability 17, 666–678.

Chan, K.S., Tong, H., 1990. On likelihood ratio tests for threshold autoregression. Journal of the Royal Statistical Society B 52, 469–476.

Cox, D.R., Miller, H.D., 1965. The Theory of Stochastic Processes. Methuen & Co. Ltd., London.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74 (1), 33–43.

Davis, R., Huang, D., Yao, Y.-C., 1995. Testing for a change in the parameter values and order of an autoregressive model. The Annals of Statistics 23, 282–304.

Doukhan, P., 1994. Mixing, Properties and Examples. Lectures Notes in Statistics. Springer, New York.

Foutz, R., Srivastava, R., 1977. The performance of the likelihood ratio test when the model is incorrect. The Annals of Statistics 5, 1183–1194.

Franck, C., Zakoian, J.-M., 2001. Stationarity of multivariate Markov-switching ARMA models. Journal of Econometrics 102, 339–364.

Gallant, R., White, H., 1988. A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models. Basil Blackwell, Oxford.

Garcia, R., 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. International Economic Review 39, 763–788.

Garcia, R., Perron, P., 1996. An analysis of real interest rate under regime shifts. Review of Economics and Statistics 78, 111–125.

Gourieroux, C., 1997. Arch Models and Financial Applications. Springer, New York.

Gourieroux, C., Monfort, A., 1989. Statistique et Modeles Econometriques, Vol. 2. Economica, Paris.

Gourieroux, C., Monfort, A., Trognon, A., 1983. Testing nested and non nested hypotheses. Journal of Econometrics 21, 83–115.

Hamilton, J.D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57, 357–384.

Hamilton, J.D., 1993. Estimation, inference and forecasting of time series subject to changes in regime. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics, Vol. 11. Elsevier, Amsterdam, pp. 231–260.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ.

Hamilton, J.D., 1996. Specification testing in Markov-switching time-series models. Journal of Econometrics 70, 127–157.

Hamilton, J.D., Susmel, R., 1994. Autoregressive conditional heteroskedasticity and changes in regime. Journal of Econometrics 64, 307–333.

Hansen, B.E., 1991. Inference when a nuisance parameter is not identified under the null hypothesis. Mimeo, University of Rochester.

Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.

Hansen, B.E., 1997. Inference in TAR models. Studies in Nonlinear Dynamics and Econometrics 2, 1–14.

Johnson, N., Kotz, S., 1970. Continuous Univariate Distributions II. Houghton Mifflin, Boston.

Kaminsky, G., 1993. Is there a Peso problem? Evidence from the Dollar/Pound exchange rate, 1976–1987. The American Economic Review 83, 450–472.

Koop, G., Potter, S., 1999. Bayes factors and nonlinearity: evidence from economic time series. Journal of Econometrics 88, 251–281.

Koop, G., Potter, S., 2001. Are apparent findings of nonlinearity due to structural instability in economic time series? Econometrics Journal 4, 37–55.

Lee, L.-F., Chesher, A., 1986. Specification testing when score test statistics are identically zero. Journal of Econometrics 31, 121–149.

Mizon, G.E., Richard, J.F., 1986. The encompassing principle and its application to testing non-nested hypotheses. Econometrica 54, 657–678.

Müller, H.-G., 1992. Change-points in nonparametric regression analysis. The Annals of Statistics 20, 737–761.

Perron, P., 1990. Testing for a unit root in a time series with changing mean. Journal of Business & Economic Statistics 8, 153–162.

Potter, S., 1995. A nonlinear approach to US GNP. Journal of Applied Econometrics 10, 109–125.

Ramanathan, R., 1998. Introductory Econometrics with Applications. The Dryden Press, Fort Worth.

Raymond, J., Rich, R., 1997. Oil and the macroeconomy: a Markov state-switching approach. Journal of Money Credit and Banking 29, 193–213.

Stock, J., Watson, M., 1996. Evidence on structural instability in macroeconomic time series relations. Journal of Business & Economic Statistics 14, 11–30.

Timmermann, A., 2000. Moments of Markov switching models. Journal of Econometrics 96, 75–111.

Tong, H., 1990. Non-linear Time Series. Oxford University Press, Oxford.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

Yao, J.-F., Attali, J.-G., 2000. On stability of nonlinear processes with Markov switching. Advances in Applied Probability 32, 394–407.