Working Paper Series, N. 1, January 2007

Testing for Linearity in Markov Switching Models: A Bootstrap Approach

Silvestro Di Sanzo
Department of Statistical Sciences
University of Padua
Italy

Abstract: Testing for linearity in the context of Markov switching models is complicated because standard regularity conditions for likelihood based inference are violated. This is due to the fact that, under the null hypothesis of linearity, some parameters are not identified and scores are identically zero. Thus the asymptotic distribution of the test statistic of interest does not possess the standard χ^2-distribution. In this paper we propose a bootstrap resampling scheme to approximate the distribution of the test statistic of interest under the null of linearity. The procedure is relatively easy to program and the computation requirements are reasonable. We investigate the performance of the bootstrap-based test using Monte Carlo simulations. We find that the test works well and outperforms the Hansen test and the Carrasco *et al.* test. The use of the various methods is also illustrated by means of empirical examples.

Keywords: Markov switching autoregressive models; Nuisance parameters; Bootstrap procedures; Monte Carlo simulation.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Silvestro Di Sanzo
silvio@unipd.it

Testing for Linearity in Markov Switching Models: A Bootstrap Approach

Silvestro Di Sanzo
Department of Statistical Sciences
University of Padua
Italy

Abstract: Testing for linearity in the context of Markov switching models is complicated because standard regularity conditions for likelihood based inference are violated. This is due to the fact that, under the null hypothesis of linearity, some parameters are not identified and scores are identically zero. Thus the asymptotic distribution of the test statistic of interest does not possess the standard χ^2-distribution. In this paper we propose a bootstrap resampling scheme to approximate the distribution of the test statistic of interest under the null of linearity. The procedure is relatively easy to program and the computation requirements are reasonable. We investigate the performance of the bootstrap-based test using Monte Carlo simulations. We find that the test works well and outperforms the Hansen test and the Carrasco *et al.* test. The use of the various methods is also illustrated by means of empirical examples.

Keywords: Markov switching autoregressive models; Nuisance parameters; Bootstrap procedures; Monte Carlo simulation.

1 Introduction

The class of Markov switching (MS) models introduced by Hamilton (1989), where changes between states (or regimes) are governed by the outcome of an unobservable Markov process, have been widely applied to many economic and financial time series[1] (see, among others, Hamilton (1989), Cecchetti *et al.* (1990) and Lam (1990) for aggregate output, Engel and Hamilton (1990) for exchange rates and Hamilton (1988) and Garcia and Perron (1996) for interest rates). A key problem which arises in empirical applications is how to determine the number of states (or regimes) required for an MS model to be an adequate characterization of the data. Hamilton (1989) offers suggestive evidence that a two-state MS model outperforms linear models in terms of forecasts, but no statistical tests. A more formal statistical procedure for determining the number of regimes is to test the null hypothesis of a linear model against the alternative hypothesis of an MS model using likelihood ratio (LR) tests. Such tests are complicated because usual regularity conditions

[1] Closely related parametric models have been proposed by Goldfield and Quandt (1973), Sclove (1983) and Cosslet and Lee (1985), among others.

required to apply the asymptotic theory are no longer respected. In fact, under the null of linearity, the likelihood function is nonquadratic and *flat* with respect to the nuisance parameters at the optimum and the scores are identically zero. Hence, in this case, conventional statistics do not have an asymptotic standard χ^2-distribution under the null hypothesis. Hansen (1992, 1996) proposes a theory of hypothesis testing which avoids these problems, but his procedure only gives bounds for the LR statistic and requires extensive computation, which involves large scale simulation and optimization over a three-dimensional grid. A less computationally demanding test procedure is discussed in Garcia (1998), but his method is theoretically little attractive since it overlooks the problem of the singular information matrix. Recently, Carrasco *et al.* (2004) propose a new test for the stability of parameters in MS models which only requires the estimation of the model under the null. This is a great advantage over the tests of Hansen (1992, 1996) and Garcia (1998), since it is well-known that the estimation of MS models is particularly burdensome. In this paper we propose an alternative way to perform a test of linearity in this framework. To calculate an appropriate p-value for the test of linearity, we propose a bootstrap procedure which is an extension of the parametric bootstrap proposed by McLachan (1987) in finite mixture models.[2]. We analyze the finite sample properties of the bootstrap test by means of Monte Carlo simulations. More precisely, we compare the size and power obtained with the bootstrap procedure with those obtained using Hansen (1992, 1996) and Carrasco *et al.* (2004) tests. The aim is to provide to applied researchers a guide about the best method to test for linearity.

The paper proceeds as follows. Section 2 gives a description of the Markov switching autoregressive model and outlines the maximum likelihood estimation procedure. Section 3 reviews the test procedures of interest. In section 4, we propose a bootstrap resampling scheme to compute the p-value for a linearity test under the framework of interest. Section 5 discusses the design of Monte Carlo experiments that are used to investigate the small-sample performance of various test procedures and presents the results of the experiments. Section 6 provides two illustrative examples involving real-world data. Section 7 concludes. Tables are relegated to the Appendix.

2 Autoregressive Models with Markov Switching Regime

We focus on the class of models of the form:

$$x_t = \mu(s_t) + \sum_{\tau=1}^{m} \phi_\tau(s_t)\left\{x_{t-\tau} - \mu(s_{t-\tau}\right\} + \sigma(s_t)u_t, \qquad (1)$$

$$\mu(s_t) = \sum_{i=1}^{r} \mu^{(i)} I(s_t = i),$$

$$\sigma(s_t) = \sum_{i=1}^{r} \sigma^{(i)} I(s_t = i),$$

$$\phi_\tau(s_t) = \textstyle\sum_{i=1}^{r} \phi_\tau^{(i)} I(s_t = i), \qquad \tau = 1, ..., m,$$

[2] Feng and McCulloch (1996) provide the mathematical background to justify the use of bootstrap likelihood ratios in this framework.

where m is a positive integer, $\{u_t\}$ is a sequence of independent and identically distributed (i.i.d.) real-valued random variables with mean zero and unit variance, s_t is a random variable which takes values in the finite set $\Upsilon = (1, ..., r)$ and which indicates the unobservable state of the system at time t, $\phi_\tau^{(i)}$, $\sigma^{(i)}$, $\mu^{(i)}$ $(i = 1.2, ..r)$ are real constants and $I(A)$ is an indicator function of event A. The process $\{s_t\}$ is assumed to form a strictly stationary, time-homogenous, first-order Markov chain on $\Upsilon = (1, ..., r)$ with transition probability matrix $P = (p_{ij})'_{i,j\in\Upsilon}$, where

$$p_{ij} = \Pr(s_t = j | s_{t-1} = i), \qquad i, j \in \Upsilon$$

and $\sum_{j=1}^{r} p_{ij} = 1$ for $i \in \Upsilon$. It is also assumed that $\{s_t\}$ is independent of $\{u_t\}$ and that P is ergodic.

Model (1) represents a Markov mixture of r autoregressive models and hereafter will be called a r-state m-order Markov switching autoregressive ($MSAR(r, m)$) model.

We can obtain the maximum likelihood estimator (MLE) of the vector of parameters of interest $\theta = \{p_{ij}, \phi_\tau^{(i)}, \sigma^{(i)}, \mu^{(i)}\}$ using the filter proposed by Hamilton (1989). Suppose for simplicity the case of an $MSAR(2, 1)$. In this case, $\theta = \{p_{11}, p_{22}, \phi_1^{(1)}, \phi_1^{(2)}, \sigma^{(1)}, \sigma^{(2)}, \mu^{(1)}, \mu^{(2)}\}$.

The inference is performed iteratively for $t = 1, 2, ..., T$, with step t accepting as input the values

$$\xi_{it-1} = \Pr(s_{t-1} = i | \Omega_{t-1}; \theta) \qquad i \in \{1, 2\},$$

and producing as output

$$\xi_{jt} = \Pr(s_t = j | \Omega_t; \theta) \qquad j \in \{1, 2\},$$

where $\Omega_{t-1} = \{x_{t-1}, x_{t-2,...,}x_0\}$ denotes the information available up to time $t - 1$, and θ for now is assumed to be known.

The key magnitudes we need to perform this iteration are the densities under the two regimes,

$$f(x_t | s_t = j, s_{t-1} = i, \Omega_{t-1}; \theta) = \frac{1}{\sigma(j)\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma(j)^2}(x_t - \mu(j) - \phi_1(j)[x_{t-1} - \mu(i)])^2\},$$

for $i, j \in \{1, 2\}$ .Specifically, given the input ξ_{it-1}, we can derive the conditional density of the tth observation from

$$f(x_t | \Omega_{t-1}; \theta) = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{ij}\xi_{it-1} f(x_t | s_t = j, s_{t-1} = i, \Omega_{t-1}; \theta),$$

and the desiderated output is

$$\xi_{jt} = \frac{\sum_{i=1}^{2} p_{ij}\xi_{it-1} f(x_t|s_t = j, s_{t-1} = i, \Omega_{t-1}; \theta)}{f(x_t|\Omega_{t-1}; \theta)}.$$

A byproduct of the filter is the evaluation of the sample conditional log-likelihood of the observed data

$$L(\theta|\Omega_T) = \sum_{t=1}^{T} \log f(x_t|\Omega_{t-1}; \theta)$$

for the specified value of θ. Therefore, the MLE of θ can be obtained by maximizing the log-likelihood by numerical optimization.

To start the filter, we need of the value ξ_{i0}, $i \in \{1, 2\}$. We can set ξ_{i0} equal to its limiting unconditional probabilities

$$\xi_{i0} = \Pr(s_0 = i) = \frac{1 - p_{jj}}{2 - p_{ii} - p_{jj}}.$$

Other alternatives are simply to set $\xi_{i0} = 1/2$ or estimate ξ_{i0} by maximum likelihood.

In this paper we are interested in testing the null hypothesis of a linear model against the alternative hypothesis of an MS model. For example, for the model described above, the test of interest takes the form:

$$\begin{aligned} H_0 &: \mu^{(1)} = \mu^{(2)},\ \sigma^{(1)} = \sigma^{(2)},\ \phi_1^{(1)} = \phi_1^{(2)}, \\ H_1 &: \mu^{(1)} \neq \mu^{(2)},\ \sigma^{(1)} \neq \sigma^{(2)},\ \phi_1^{(1)} \neq \phi_1^{(2)}. \end{aligned}$$

Note that transition probabilities, p_{11} and p_{22}, are unidentified under the null since any value between 0 and 1 leaves the likelihood function unchanged. Moreover, the scores are identically zero when evaluated at the null hypothesis. Under these conditions the asymptotic null distribution of the likelihood ratio test statistic is not a χ^2 with one degree of freedom.

3 Tests for Linearity in MS Models

In this section, we describe the most important hypothesis testing theories proposed in the literature to test for linearity in MS models. The linearity tests considered include the Hansen test and the Carrasco *et al.* test. In our subsequent Monte Carlo analysis, we consider both procedures and compare their size and power with those obtained with the bootstrap method. Since these tests have been discussed extensively in the literature, our description here will be relatively brief.

3.1 The Hansen Test

Hansen (1992, 1996) proposes a theory of hypothesis testing which allows to perform a test of linearity in the presence of nuisance parameters and scores identically zero under the null hypothesis, which is the case of MS models. This author considers the likelihood function as a function of unknown parameters and employes the empirical process theory to bound the asymptotic distribution of a standardized LR statistic. Now, we discuss formally the procedure.

The parameter vector θ is first split into two subvectors: a subvector of parameters of interest α, and a subvector of nuisance parameters λ, where the subvector α includes the transition probabilities and the regime-switching parameters.

The subvector α is further partitioned into β and γ. The null hypothesis takes the form

$$H_0 : \beta = 0, \qquad H_1 : \beta \neq 0,$$

and γ is not identified under H_0.

For example, for the case of an $MSAR(2,1)$ with switches only in the mean of the process, we obtain:

$$\begin{aligned} \theta &= \left\{\mu^{(1)},\ \mu^{(2)},\ p_{11},\ p_{22},\ \phi_1,\ \sigma^2\right\}, \quad \alpha = \left\{\mu^{(1)} - \mu^{(2)},\ p_{11},\ p_{22}\right\}, \\ \lambda &= \left\{\phi_1,\ \sigma^2, \mu^{(1)}\right\}, \quad \beta = \left\{\mu^{(1)} - \mu^{(2)}\right\}, \quad \gamma = \{p_{11},\ p_{22}\}. \end{aligned}$$

Let $f_t(\alpha, \lambda)$ be the conditional log-likelihood of the tth observation evaluated at α and λ:

$$f_t(\alpha, \lambda) = \log f(x_t|\Omega_{t-1}; \alpha, \lambda).$$

For any α, let $\widehat{\lambda}(\alpha)$ denote the value of λ that maximizes the log-likelihood with respect to λ taking α as given. Define

$$q_t(\alpha) = f_t[\alpha, \widehat{\lambda}(\alpha)] - f_t[\alpha_0, \widehat{\lambda}(\alpha_0)],$$

where α_0 is the value of α under the null. The sample mean of this variable is $\overline{q}(\alpha) = T^{-1}\sum_{t=1}^{T} q_t(\alpha)$. The LR test of the null hypothesis that $\alpha = \alpha_0$ against the specific alternative represented by α could be represented as $T\overline{q}(\alpha)$. Hansen (1992, 1996) suggests calculating the following *standardized* LR test statistic:

$$\widehat{H} = \max_{\alpha \in \Gamma} \{T\overline{q}(\alpha)(\sum_{t=1}^{T}[q_t(\alpha) - \overline{q}(\alpha)]^2)^{-1/2}\},$$

where Γ is a grid containing the possible values of α.

He shows that under the null $\alpha = \alpha_0$, for large samples, the probability that $\widehat{H}$ would exceed a critical value z is less than the probability that the statistic below would exceed the same value z:

$$\max_{\alpha\in\Gamma}\{(1+M)^{-1/2}(\sum_{k=0}^{M}\sum_{t=1}^{T}[q_t(\alpha)-\overline{q}(\alpha)]u_{t-k})^{-1}(\sum_{t=1}^{T}[q_t(\alpha)-\overline{q}(\alpha)]^2)^{-1/2}\}$$

where u_t is an i.i.d. $N(0,1)$ sequence generated by Monte Carlo and M is a parameter corresponding to the maximum order of autocorrelation allowed at $q_t(\alpha)$.

As we point out in the introduction, this procedure has two main shortcomings. First, it requires to set a grid for each element of the vector α, that is for each switching parameter plus the transition probabilities. For each value of this grid, we need to optimize the likelihood function with respect to the nuisance parameters of the model. Clearly, this becomes computationally burdensome as the grid search becomes more extensive and the model becomes more complex. Second, as we have seen above, this procedure provides a bound for the LR statistic and not a critical value, which means that this test may be conservative, i.e. under-rejection of the null hypothesis when it is true.

3.2 The Carrasco et al. Test

Carrasco *et al.* (2004) propose an optimal test for the stability of parameters in the framework of MS models. They derive a class of information matrix-type tests, strongly related as that proposed by White (1982), and show that it is equivalent to the LR test. Hence, their test is asymptotically optimal. To describe the procedure, let $f_t(\theta)$ denote the conditional log-likelihood of the tth observation under the null hypothesis of no Markov switching and let $f_t^1(\theta)$ and $f_t^2(\theta)$ denote its gradient and Hessian, respectively.

Define:

$$\gamma_t(\rho,\widehat{\theta}) = h'[f^2(\widehat{\theta}) + [f_t^1(\widehat{\theta})f_t^1(\widehat{\theta})'] + 2\sum_{s<t}\rho^{t-s}f_t^1(\widehat{\theta})f_s^1(\widehat{\theta})']h$$

where h and ρ are nuisance parameters. Specifically, parameter h measures the difference between the states and ρ is a parameter characterizing the serial correlation of the Markov chain for s_t under the alternative hypothesis of Markov-switching. $\widehat{\theta}$ is the MLE under the null hypothesis of constant parameters.

Finally, let $\widehat{\varepsilon}_t(\rho,h,\widehat{\theta})$ denote the tth residual from the OLS regression of $(1/2)\gamma_t(\rho,\widehat{\theta})$ on $f_t^1(\widehat{\theta})$. Carrasco *et al.* (2004) propose calculating

$$\sup TS = \sup_{\{h,\rho:\|h\|=1,\underline{\rho}<\rho<\overline{\rho}\}}[\max\{0,\frac{\sum_{t=1}^{T}\gamma_t(\rho,h)}{2\sqrt{\sum\widehat{\varepsilon}_t(\rho,h,\widehat{\theta})^2}}\}]^2$$

To calculate this test we need to use a grid search over the two-dimensional space (h,ρ). To execute this grid search, we have to pick a region over which to search. Carrasco *et al.* (2004) suggest to generate h uniformly over the unit sphere and select ρ from an equispaced grid. The asymptotic distribution of this test is not free of nuisance parameters, then we have to rely on Monte Carlo simulations to compute critical values.

4 The Bootstrap Algorithm

In this section we discuss a bootstrap approximation of the distribution of the LR test statistic under the null of linearity. We follow a similar approach to that used in McLachlan (1987) for bootstrapping the LR test statistic for the number of components in a normal mixture, though now we assess the number of components in a MS model. This bootstrap approximation can be used to calculate the p-values of the linearity test. The procedure is not very difficult to program and computation requirements are quite reasonable.

The bootstrap runs as follows:

STEP1: Construct some estimation of the coefficients of the model under the null hypothesis of linearity, $\widehat{\theta}_0 = \{\widehat{\mu}, \widehat{\phi}_\tau, \widehat{\sigma}^2\}$ with $\tau = 1, 2, ..., m$. For example, using the MLE.

STEP2: Compute the estimated residuals under H_0 as follows:

$$\widehat{u}_t = x_t - \widehat{\mu} - \sum_{\tau=1}^{m} \widehat{\phi}_\tau \left\{x_{t-\tau} - \widehat{\mu}\right\}, \qquad t = m+1, ..., T.$$

STEP3: Estimate the model under H_1 and compute the LR statistic,

$$LR = 2\left[L(\widehat{\theta}|\Omega_T) - L(\widehat{\theta}_0|\Omega_T)\right],$$

where $\widehat{\theta}$ denotes the unrestricted MLE of θ, and $\widehat{\theta}_0$ denotes the MLE estimate under the null hypothesis.

STEP4: Generate the bootstrap errors u_t^*, $t = m+1, ..., T$, by sampling with replacement from the residual series $\widehat{u}_t$. Construct the bootstrap sample as follows:

$$x_t^* = \widehat{\mu} + \sum_{\tau=1}^{m} \widehat{\phi}_\tau \left\{x_{t-\tau}^* - \widehat{\mu}\right\} + u_t^*$$

We need to establish a set of initial values for $(x_0, x_{-1}, x_{-2}, ..., x_{-m+1})$. We take the simple approach of conditioning on the observed values, so hold these values fixed in repeated samples. Naturally, some other ways to model the initial conditions can be used. The distribution of x_t^* is the bootstrap distribution of the data.

STEP5: Use the bootstrap sample x_t^* to calculate the LR statistic. Call its value LR^*. The distribution of LR^* is the bootstrap distribution of LR.

The experiment consist of repeating the above steps B times. We compute the bootstrap p-value as $p_B = card(LR^* \geq LR)/B$, that is the fraction of LR^* values that are greater than the observed value LR.

5 Monte Carlo Study

In this section we report a series of Monte Carlo experiments designed to compare the finite sample performance of the bootstrap procedure with the performance of the tests of Hansen and Carrasco *et al.*. We begin describing the data-generating processes and the experimental design used in our simulations. A discussion of the results of the experiments follows.

5.1 Experimental Design

To assess the size (rejection frequency under the null) of the tests, we have to generate the data under the null hypothesis of linearity. We use as data generating process (DGP) a first-order autoregressive linear model:

DGP_0 :

$$\begin{aligned} x_t &= \mu + \phi(X_{t-1} - \mu) + u_t, \quad \text{with} \quad u_t \sim N(0,1), \\ \mu &= 1.5,\ \phi = 0.3 \end{aligned}$$

On the other hand, to assess the power (rejection frequency under the alternative) of the tests, we need to generate data under the alternative hypothesis of Markov switching. In our experiments, to keep the calculations manageable, we use as DGP an $MSAR(2,1)$ model with $u_t \sim N(0,1)$. We consider the following parameterizations:

DGP_1 :

$$\begin{aligned} \mu^{(1)} &= 1, \quad \mu^{(2)} \in \{2,3\}, \\ \sigma^{(1)} &= 1, \quad \sigma^{(2)} \in \{1, \sqrt{1.5}\}, \\ \phi_1^{(1)} &= 0.3, \quad \phi_1^{(2)} \in \{0.3, 0.0.6, 0.9\}, \\ (p_{11}, p_{22}) &\in \{(0.6,0.4),(0.9,0.9),(0.9,0.98)\}. \end{aligned}$$

Transition probabilities $(p_{11}, p_{22}) = (0.6, 0.4)$ imply that the regime indicator variables $\{s_t\}$ are uncorrelated. This means that the state of the system at time t does not depend on the state at time $t-1$. On the other hand, the pairs $(p_{11}, p_{22}) = (0.9, 0.9)$ and $(p_{11}, p_{22}) = (0.9, 0.98)$ allow the regimes to be highly persistent with the regime corresponding to $s_t = 2$ being almost absorbing in the latter case (the stationary distribution of $\{s_t\}$ is $(0.5, 0.5)$ and $(0.1667, 0.8333)$, respectively).

Note that in the class of $MSAR(2,1)$ models presented under DGP_1, the conditional distribution of a realization depends upon the previous value of the Markov process and the model parameters varies between states. These are the primary sources of intensive computational requirements. In this framework, as pointed out by Hansen (1992, 1996), to perform a Monte Carlo study is not feasible because the computational requirement is enormous. Therefore, to compare the bootstrap and Carrasco *et al.* tests with the Hansen test we use a simpler class of DGPs, which are employed by Hansen (1992, 1996) in his Monte Carlo study.

To generate data under the null, we consider a simple model with no autoregressive parameters:

DGP_1^H :

$$\begin{aligned} X_t &= \mu + u_t, \\ \mu &= 0.057,\ \sigma = 0.983. \end{aligned}$$

To study the power, we use the following nonlinear model:

DGP_2^H :

$$\begin{aligned} X_t &= \mu(s_t) + u_t, \\ \mu^{(1)} &= -0.359,\ \mu^{(2)} = 1.522,\ \sigma = 0.769,\ p_{11} = 0.904,\ p_{22} = 0.755. \end{aligned}$$

To see if the autoregressive parameters affect the size and the power of the tests we employ DGP_3^H and DGP_4^H, respectively:

DGP_3^H :

$$\begin{aligned} X_t &= \mu + \sum_{\tau=1}^{4} \phi_\tau x_{t-\tau} + u_t, \\ \mu &= 0.557,\ \sigma = 0.983,\ \phi_1 = 0.310,\ \phi_2 = 0.127,\ \phi_3 = -0.121,\ \phi_4 = -0.089. \end{aligned}$$

DGP_4^H :

$$\begin{aligned} X_t &= \mu(s_t) + \sum_{\tau=1}^{4} \phi_\tau x_{t-\tau} + u_t, \\ \mu^{(1)} &= -0.447, \mu^{(2)} = 1.560,\ \sigma = 0.789,\ p_{11} = 0.912,\ p_{22} = 0.669, \\ \phi_1 &= 0.112,\ \phi_2 = 0.065,\ \phi_3 = -0.126,\ \phi_4 = -0.136. \end{aligned}$$

From DGP_1^H to DGP_4^H, the parameters are the maximum likelihood estimates obtained by fitting the models to the U.S. real GNP series ranging from 1952:2 to 1984:4. Note also that in DGP_2^H and DGP_4^H, the conditional likelihood only depends upon the current state. As a result, the computational burdens are much less demanding.

The experiments proceed by generating an artificial time series of length $T + 50$ according to each one of the DGPs under study, with $T \in \{100, 250\}$ for DGP_0 and DGP_1 and $T = 131$ for $DGP_1^H - DGP_4^H$. Initial values are set to zero in all specifications. The initial value of the Markov chain $\{s_t\}$ is drawn randomly from its stationary distribution, that is, $p(s_t = 1) = (1 - p_{22})/(2 - p_{11} - p_{22})$ and $p(s_t = 2) = (1 - p_{11})/(2 - p_{11} - p_{22})$. The first 50 pseudo-data points are then discarded to minimize the effect of initial conditions and the remaining T points are used to compute the test statistics.

For each simulated sample we test the null hypothesis of linearity using the bootstrap procedure and the Carrasco *et al.* test described in the previous section. We employ $B = 500$ bootstrap replications, and for the Carrasco *et al.* test we generate

h over the unit sphere and ρ is selected from an equispaced grid of $(-0.7, 0.7)$. The number of Monte Carlo replication experiments for each design point is $R = 1000$. The processing time becomes excessive when greater values of B or R are used.

The relevant test is deemed to reject at 10%, 5% and 1% level if its associated p-value is smaller than 0.10, 0.05 and 0.01, respectively. Since no substantial differences are observed across significance levels, results are only reported for the 5% significance level. Moreover, for DGP_1, we only report the results obtained with state-dependent variances because when the variance is kept constant across regimes we obtain similar results.

We start analyzing the size and the power properties of the bootstrap and Carrasco *et al.* tests obtained with DGP_0 and DGP_1, respectively. Table 1 records the Monte Carlo estimates of the empirical size, and Tables 2 shows the estimates of the power. The empirical size of the tests is remarkably close to its nominal value even for a small sample size $T = 100$. If the sample size is increased to $T = 250$, the empirical size is more accurate.

Regarding the power, it is evident that the performance of the tests improves as the difference between the values of the parameters in the two regimes increases. This is true for both changes in autoregressive parameters, $\phi_1(s_t)$, and changes in the mean parameters, $\mu(s_t)$. For example, when we consider the largest changes included in our experiments ($\mu^{(2)} - \mu^{(1)} = 2$, $\phi^{(2)} - \phi^{(1)} = 0.6$), both tests have an excellent power even for $T = 100$. Using the bootstrap test, the rejection rate ranges from 95.1 to 98.7. If the Carrasco *et al.* test is used, the rejection rate ranges from 93.8 to 96.3. In the context of normal mixture models, McLachlan (1987) also finds that the power of the bootstrap is very good if the components of the mixture are widely separated.

The magnitude of the transition probabilities also affects the power. More precisely, when the sample size and the changes between parameters are small, both tests suffer a significant loss of power in the case of uncorrelared Markov chains, especiallly for the Carrasco *et al.* test . This may be due to the frequent regime transitions that take place when $p_{11} = 1 - p_{22}$, combined with small changes in parameters, tend to make the $MSAR(2, 1)$ series look very much like a heteroskedastic white noise[3]. However, the performance of the tests improves considerably when the magnitude of the parameter changes and the persistence of the Markov regimes increase, especially when we use the bootstrap test.

Turning to the properties of individual tests, it is clear that the bootstrap procedure outperforms the Carrasco *et al.* test. When the sample size is small and parameter changes are moderate, the bootstrap is far superior to the Carrasco *et al.* test. For example, for $T = 100$, $\mu^{(2)} - \mu^{(1)} = 1$ and $\phi^{(2)} - \phi^{(1)} = 0$, the power of the bootstrap ranges from 52.8 to 71.2, while in the Carrasco *et al.* test it ranges from 21.3 to 47.6. When T increases, the power of both procedures improves substantially.

Now, we consider the second set of DGPs to compare the performance of the bootstrap and Carrasco *et al.* tests with the Hansen test. We report the empirical size and power obtained with DGP_1^H and DGP_3^H in Table 3. The size and power obtained by Hansen (1992, 1996) for the corresponding DGPs are reported in Table

[3]Psaradakis and Spagnolo (2003) obtain similar results in the analysis of the properties of complexity-penalized criteria to select the correct state dimension of an $MSAR(r, m)$ model.

4. As we pointed out in the previous section, the Hansen test may be conservative, in fact, we see that this test under-rejects with an empirical size ranging from 0 to 2. On the other hand, for the Carrasco *et al.* test and the bootstrap test, the empirical size is remarkably close to its nominal value. All tests have a good power. In Table 5, we report the results for the bootstrap test and the Carrasco *et al.* test obtained with DGP_2^H and DGP_4^H. The corresponding results for the test of Hansen are reported in Table 6. In this case, we obtain an interesting result. The Hansen and the Carrasco *et al.* tests suffer a significant loss of power. The Hansen test has a deterioration of power from 74 to 20, and the Carrasco *et al.* test from 69.0 to 40.7, while the power of the bootstrap erodes from 83.3 to 70.0. Therefore, the effect of the autoregressive parameters is quite strong, indicating that for the Hansen and Carrasco *et al.* tests the cost of over-fitting is higher.

To summarize, the following general conclusions can be drawn from the discussion above:

1) In general, the bootstrap method is the best procedure overall. In fact, it works well when the size of the sample is small, parameters changes are moderate, or the Markov chain is not persistent.

2) For simple models with no autoregressive components, the three testing procedures have a similar performance. In this case, the Carrasco *et al.* test and the bootstrap procedure are preferable because they are much easier to implement.

3) For models with more complicated behaviour, the Carrasco *et al.* and the bootstrap tests outperform the Hansen test. But the bootstrap is the method with the highest power. It has a lower over-fitting cost.

6 Empirical Examples

To illustrate the practical use of the test procedures descibed before, we provide two examples involving real-world data sets. More precisely, we analyze

the following time series:

1. The quarterly three-month US treasury bill rates for the period 1962:1-1987:3.

2. The quarterly precentage changes in real US GNP for the period 1952:2-1984:4.

For the nominal interest rate data, Hamilton (1988) found support for an MSAR (2,4) model with state-independent autoregressive coefficients and switching in the mean and the variance of the process. For the real GNP series, Hamilton (1989) fitted a MSAR(2,4) model with no switching in the innovation variance or the autoregressive coefficients. On the other hand, Albert and Chib (1993) found evidence for a specification with mean switching but no autoregressive dynamics. However, we wish investigate here whether the more parsimonious MSAR (2,0) model proposed by Albert and Chib (1993) is statistically significant. In Table 7 we report the P-values obtained with the test procedures described before. For the real GNP data, all three tests reject the null of linearity. Therefore, the model proposed by Albert and Chib (1993) is statistically significant. This result is in according with that obtained by Psaradakis (1998). For the nominal interest rate , the Carrasco *et al* and the bootstrap test reject the null of linearity, while the Hansen test provides

evidence in favour of a linear specification. In this case no definite answer is possible. However, our simulation evidence suggested that the power of the Hansen test is low in presence of autoregressive dyamics. This is certainly not conclusive evidence, but it is fair to conclude that model is statistically significant.

7 Conclusion

In this paper we propose a bootstrap algorithm to calculate an appropriate p-value for the test of linearity in the framework of Markov switching models inspired by the bootstrap likelihood ratio proposed by McLachlan (1987) in the context of finite mixture models . We analyze the finite sample properties of the test procedure by means of Monte Carlo simulations. We find that the bootstrap-based test works well and that it outperforms the tests of Hansen and Carrasco *et al.*. We conclude that, in applied research, the bootstrap procedure may be a valid alternative to Carrasco *et al.* and Hansen tests.

8 Appendix

TABLE 1: Monte Carlo size, $DGP_{0,}$ 5% nominal level

	$T = 100$	$T = 250$
Bootstrap	4.7	5.0
Carrasco *et al.*	4.3	5.2

TABLE 2: Monte Carlo power, $DGP_{1,}$ 5% nominal level

	$T = 100$	$T = 250$
$\mu^{(1)} = 1$, $\mu^{(2)} = 2$, $\sigma^{(1)} = 1$, $\sigma^{(2)} = \sqrt{1.5}$		
$\phi_1^{(1)} - \phi_1^{(2)} = 0$		
$(p_{11}, p_{11}) = (0.6, 0.4)$		
Bootstrap	58.8	63.4
Carrasco *et al.*	21.3	41.2
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	71.2	82.0
Carrasco *et al.*	47.6	56.8
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	70.1	77.5
Carrasco *et al.*	44.5	49.2
$\phi_1^{(1)} - \phi_1^{(2)} = 0.3$		
$(p_{11}, p_{11}) = (0.6, 0.4)$		
Bootstrap	73.3	75.1
Carrasco *et al.*	53.0	71.4
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	79.5	85.3
Carrasco *et al.*	67.0	75.7
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	76.2	83.1
Carrasco *et al.*	63.4	71.7
$\phi_1^{(1)} - \phi_1^{(2)} = 0.6$		
$(p_{11}, p_{11}) = (0.6, 0.4)$		
Bootstrap	80.6	85.3
Carrasco *et al.*	65.1	84.2
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	86.0	92.4
Carrasco *et al.*	75.9	85.0
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	83.9	86.3
Carrasco *et al.*	73.7	84.5

TABLE 2 (Continued)

	$T = 100$	$T = 250$
$\mu^{(1)} = 1, \mu^{(2)} = 3, \sigma^{(1)} = 1, \sigma^{(2)} = \sqrt{1.5}$		
$\phi_1^{(1)} - \phi_1^{(2)} = 0$		
$(p_{11}, p_{11}) = (0.6, 0.4)$		
Bootstrap	68.0	76.1
Carrasco *et al.*	43.6	72.4
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	87.0	90.3
Carrasco *et al.*	81.5	89.4
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	78.6	81.6
Carrasco *et al.*	75.2	76.7
$\phi_1^{(1)} - \phi_1^{(2)} = 0.3$		
$(p_{11}, p_{11}) = (0.6, 0.4)$		
Bootstrap	82.5	88.6
Carrasco *et al.*	80.1	86.9
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	91.9	93.4
Carrasco *et al.*	88.6	91.2
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	84.3	87.5
Carrasco *et al.*	83.0	85.3
$\phi_1^{(1)} - \phi_1^{(2)} = 0.6$		
$(p_{11}, p_{11}) = (0.5, 0.5)$		
Bootstrap	95.1	97.7
Carrasco *et al.*	93.8	96.6
$(p_{11}, p_{11}) = (0.9, 0.9)$		
Bootstrap	98.7	99.3
Carrasco *et al.*	96.3	97.4
$(p_{11}, p_{11}) = (0.9, 0.98)$		
Bootstrap	98.0	98.8
Carrasco *et al.*	95.2	97.3

TABLE 3: Monte Carlo with no autoregressive components 5% nominal level

	Size	Power
Bootstrap	5.2	83.3
Carrasco *et al.*	4.4	69.0

TABLE 4: Monte Carlo with autoregressive components 5% nominal level

	Size	Power
Bootstrap	5.4	70.0
Carrasco *et al.*	4.0	40.7

TABLE 5: Monte Carlo with no autoregressive components, 5% nominal level, (Results reported in Hansen (1996, p. 197))

M=0	size	0
	power	74
M=1	size	0
	power	74
M=2	size	2
	power	74
M=3	size	2
	power	74
M=4	size	2
	power	74

TABLE 6: Monte Carlo with autoregressive components, 5% nominal level, (Results reported in Hansen (1996, p. 197))

M=0	size	14
	power	24
M=1	size	16
	power	24
M=2	size	16
	power	22
M=3	size	16
	power	24
M=4	size	16
	power	20

TABLE 7: Test P-values

	Hansen test	Bootstrap test	Carrasco *et al.* test
Interest rate	0.21	0.003	0.004
Real output	0.034	0.000	0.000

9 References

Albert, J.H., and S. Chib (1993). *Bayes inference via Gibbs Sampling of autoregressive time series subject to Markov mean and variance shifts.* Journal of Business and Economic Statistics, 11, 1-15.

Carrasco, M, Liang H., and Ploberger W. (2004). *Optimal test for Markov Switching.* Working paper, University of Rochester,

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

Cosslett, S.R., Lee, L.F., (1985). *Serial correlation in latent discrete variable models.* Journal of Econometrics 27, 79-87.

Engel, C., Hamilton, J.D., (1990). *Long swings in the dollar: are they in the data and do the markets know it?* American Economic Review 80, 689-713.

Feng, Z.D and McCulloch, C.E. (1996). *Using bootstrap likelihood ratios in finite mixture models.* Journal of the Royal Statistical Society series B 58, 609-617.

Garcia, R., Perron, P., (1996). *An analysis of the real interest rate under regime shifts.* Review of Economics and Statistics 78, 111-125.

Garcia, R. (1998) *Asymptotic null distribution of the likelihood ratio test in Markov switching models*, International Economic Reviews, 39,763-788.

Goldfield, S.M., Quandt, R.E., (1973). *A Markov model for switching regression.* Journal of Econometrics 1, 3-16.

Hamilton, J.D, (1988). *Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rate.* Journal of Economic

Dynamics and Control 12, 385-423.
Hamilton, J.D. (1989). *A new approach to the economic analysis of nonstationary time series and the business cycle.* Econometrica 57(2), 357-384.
Hansen, B.E. (1992). *The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP.* Journal of Applied Econometrics 7(5), S61-82.
Hansen, B.E. (1996). Erratum: *The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP.* Journal of Applied Econometrics, 11(2), 195-98.
Lam, P., (1990). T*he Hamilton model with a general autoregressive component: estimation and comparison with other models of economic time series.* Journal of Monetary Economics 26, 409-432
McLachlan, G.J. (1987).*On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture.* Applied Statistics, Vol.36, No.3,318-324.
Psaradakis, Z.,(1998). *Bootstrap based evaluation of Markov-switching time series models.* Econometric Reviews, 17, 275-88.
Psaradakis, Z., and Spagnolo N., (2003). *On the Determination of the Number of Regimes in Markov-Switching Autoregressive Models.* Journal of Time Series, vol.24, n.2, 237-252.
Sclove, S. L., (1983). *Time-series segmentation: a model and a method.* information Sciences 29, 7-25.
White, H., (1982). *Maximum likelihood estimation of misspecified model*s. Econometrica, 50, 1-25.

Acknowledgements

We would like to thank Máximo Camacho, Juan Mora and Gabriel Pérez-Quirós for their valuable comments, which resulted in a great improvement of an earlier version of this paper. Useful comments were also received from seminar participants at the University of Alicante. All errors are my own responsibility.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it
Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy