MSTest: An R-Package for Testing Markov Switching Models

Gabriel Rodriguez-Rondon*† Jean-Marie Dufour‡

May 2, 2024

Abstract

We describe the R package **MSTest**, which implements hypothesis testing procedures that can be used to identify the number of regimes in Markov switching model. Markov switching models have wide applications in economics, finance, and many other fields. The package **MSTest** makes available hypothesis testing procedures that are valid (including exact Monte Carlo test procedures) and can be used to compare Markov switching models with different numbers of regimes. The package also allows users to simulate Markov regime switching processes, estimate Markov switching autoregressive and vector autoregressive models as well as hidden Markov models using the expectation maximization (EM) algorithm or maximum likelihood estimation (MLE). We illustrate the functionality of the **MSTest** package using both simulation experiments and data for U.S. GNP growth.

Key Words: Hypothesis testing, Monte Carlo tests, Likelihood ratio, Exact inference, Markov switching, Nonlinearity, Regimes, R software
JEL Classification:

*This work was supported by the Fonds de recherche sur la société et la culture Doctoral Research Scholarships (B2Z).

†Mailing address: Department of Economics, McGill University, 855 Sherbrooke St. West Montreal, QC H3A 2T7. e-mail: gabriel.rodriguezrondon@mail.mcgill.ca. Wed page: https://grodriguezrondon.com

‡William Dow Professor of Economics, McGill University, Centre interuniversitaire de recherche en analyse des organisations (CIRANO), and Centre interuniversitaire de recherche en economie quantitative (CIREQ). Mailing address: Department of Economics, McGill University, Leacock Building, Room 414, 855 Sherbrooke Street West, Montreal, Quebec H3A 2T7, Canada. e-mail: jean-marie.dufour@mcgill.ca. Web page: http://www.jeanmariedufour.com

Contents

1 Introduction

Markov switching models were first introduced by Baum and Petrie, 1966 but was later popularized and became an active area of research in economics since Hamilton, 1989 introduced the approach when proposing that we model the first difference of U.S. GNP as a nonlinear stationary process, rather than a linear stationary process as was typically done. The non-linearity here arises from discrete shifts in the process. Hamilton describes these regimes as episodes where the behavior of the series is significantly different. For example, one regime can characterize a period of positive growth while the other, a period of negative growth due to recessions. Many applications of regime switching models such as Garcia and Perron, 1996, Hamilton and Susmel, 1994, Gray, 1996, Kim and Nelson, 1999, Marcucci, 2005 and others have since considered Markov switching in their model specification. Hamilton, 2016 provides a detailed survey of regime switching in the macroeconomics. Markov switching models and normal-mixture models are commonly used to capture this non-linearity. Regimes in Markov switching models are governed by an unobservable two-state Markov-chain. In normal-mixture models, the process is a sum of M distinct normal distributions, where M is the number of regimes, also known as components, and weights equal to the proportion of data originating from each distribution. As discussed by Cho and White, 2007 and Carter and Steigerwald, 2012, using mixture models allows us to ignore temporal dependencies of the Markov-chain. On the other hand, the use of mixture models also has some limitations such as derivatives of the likelihood function being a linear combination of other derivatives, which creates problems of weak identification. Both models, however, provide an adequate framework to attempt to identify the number of regimes of a given process by testing the null hypothesis of a linear model (a single regime) against the alternative hypothesis of a non-linear model with, for example, two regimes. There are three common difficulties encountered when trying to identify the number of regimes a model should have. One is that some parameters are not identified under the null hypothesis. Specifically, transition probabilities and parameters which vary with regimes are not identified under the null hypothesis. Also, transition probabilities may also take the value of 0 or 1 introducing the parameter boundary problem, which is discussed in Andrews, 1999 and Andrews, 2001 and was first considered by Cho and White, 2007 in a regime testing framework. The third is that the score of the likelihood function may be identically zero under the null hypothesis. Andrews and Ploberger, 1994a, Andrews and Ploberger, 1995, Davies, 1977 and Davies, 1987 all investigate the problem where a nuisance parameter is identified only under the alternative hypothesis. The approaches taken in these studies with regards to non-identified nuisance parameters have been widely used in testing procedures including the ones we present in this paper.

Notable contributions to testing the null hypothesis of a linear model against a model with two-regimes include Hansen, 1992, Hansen, 1996a, Garcia, 1998, Cho and White, 2007, Marmer, 2008, Carrasco et al., 2014, Kasahara et al., 2014 and Dufour and Luger, 2017. Testing the null of M regime against the alternate of $M+1$ regimes for $M \geq 2$ has been a difficult task and an ongoing area of research. Kasahara and Shimotsu, 2018 recently develop a test for this hypothesis by developing the asymptotic limiting distribution of the Likelihood Ratio Test (LRT) statistic using the Difference in Quadratic Mean (DQM) first used by Liu and Shao, 2003 for the loss of identifiability. They also make use of a reparameterization and higher order expansion of the log likelihood function first introduced in Kasahara and Shimotsu, 2015. Testing the number of regimes a model should include is an important step when deciding on the specification of your model. However, performing these tests is not necessarily trivial and performing more than one test can quickly become very tedious. As a result, we introduce **MSTest**, an R-Package which can be used to test the null hypothesis of M regimes against the alternative hypothesis of $M+1$ regimes for

$M = 1$. Future versions of this package will include Kasahara and Shimotsu, 2018 test for the case where $M \geq 2$. The purpose of this R-package is to allow the econometrician to test the number of regimes a model should include for a given process. It aims to facilitate the comparison of different tests and the determination of number of regimes in a model for economic research. **MSTest** uses Rcpp (Eddelbuettel and François, 2011) and RccpArmadillo (Eddelbuettel and Sanderson, 2014) for computational efficiency. This is especially important given the computational burden of testing with the presence of nuisance parameters, as is the case here.

The **MSTest** package includes the methodologies presented in Hansen, 1992, Hansen, 1996a, Carrasco et al., 2014 and Dufour and Luger, 2017. It does not include the QLR test proposed by Cho and White, 2007 due to the limitations of their test with regards to autoregressive models as discussed in Carter and Steigerwald, 2012. The extension of the QLR test proposed by Qu and Zhuo, 2021 who consider the temporal dependence of the regimes but add some restrictions on the transition probabilities of latent regimes may be included instead in future versions of **MSTest**. Addtionally, the current version of the **MSTest** package doesn't include the methodology proposed by Garcia, 1998 but may in the future. The methodology introduced by Hansen, 1992 and Hansen, 1996a is a Likelihood Ratio test and can be used to test for switching mean, variances and autoregressive components. Carrasco et al., 2014 use barlett-identities to develop an information-type test which only requires the estimation of the model under the null, which significantly facilitates the estimating process and testing overall. The authors also show that their test is optimal in the Neyman-Pearson sense and is useful for random coefficient models as well. Since the authors also derive the asymptotic distribution of their test statistic, we contribute by adding a Monte Carlo version of their test. Other contributions of **MSTest** include extending the test proposed by Dufour and Luger, 2017 to a more general ARMA setting, models with no ARMA components and models with independent explanatory variables instead. The test proposed by Dufour and Luger, 2017 take the approach suggested by Cho and White, 2007 and Carter and Steigerwald, 2012 and uses a normal mixture model, which allows them to ignore the temporal dependencies of the Markov-chain. The authors use moments of the errors to test the null hypothesis of a linear model. Since this test uses errors of the proposed model, it was trivially extended to the more general models introduced here. The package also has available the Maximized Monte Carlo (MMC) version of the moment test as proposed in Dufour and Luger, 2017, which allows the user to treat ARMA coefficients or slope coefficients as nuisance parameters when testing, while setting bounds for the set of permissible parameter values.

In section 2 we introduce the hypothesis our package is designed to test. We discuss the identification difficulties and the methodologies developed in each framework to test our hypothesis. Section 3 described the package in more detail and how functions can be used to test the hypothesis of a linear model. This section will also serve as a manual for the **MSTest** package. Section 4 includes an empirical example where we apply our tests to the U.S. GNP data of Hamilton, 1989, the extended data used in Carrasco et al., 2014 and Dufour and Luger, 2017 and a further extension of the data ranging from 1951Q2 to 2018Q4. We compare the different models and provide tables with: test statistics, critical values and p-values. We also briefly provide some thoughts for future work. Finally we provide brief concluding remarks in section 5.

2 Markov switching models

In this paper, we consider Markov switching models where only the mean and the variance are governed by the Markov process S_t and hence are the only parameters that are subject to change according to the regime. We also consider a special case of the Markov switching model: the Hidden

Markov model, where no autoregressive coefficients are included as explanatory variables. For now, we do not consider other exogenous explanatory variables that can be included in more general Markov switching and do not consider changes in the autoregressive coefficients as the current version of the R package **MSTest** does not include this level of generality yet. However, this may be added in future versions.

2.1 First-order Markov process

We begin by describing the first-order Markov process S_t that governs the changes in the parameters of the Markov switching model. We assume that the process S_t is unobserved and evolves according to a first-order ergodic Markov chain with a $(M \times M)$ transition probability matrix given by

$$\mathbf{P} = \begin{bmatrix} p_{11} & \cdots & p_{M1} \\ \vdots & \ddots & \vdots \\ p_{1M} & \cdots & p_{MM} \end{bmatrix}$$

where for example $p_{ij} = P(S_t = j|S_{t-1} = i)$ is the probability of state i being followed by state j and M is the total number of regimes. Specifically, if we consider M regimes, the process takes integer values $S_t = \{1, \dots, M\}$. Additionally, the columns of the transition matrix $\mathbf{P}$ must sum to one in order to have a well defined transition matrix (i.e., $\sum_{j=1}^{M} p_{ij} = 1$).

Considering the example in Hamilton, 1994 where the Markov process has only two regimes, we only need a (2×2) transition matrix to summarize the transition probabilities $\mathbf{P}$ as follows:

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix}$$

We can also obtain the ergodic probabilities, $\pi = (\pi_1, \pi_2)'$, which are given by

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \qquad\qquad \pi_2 = 1 - \pi_1 \tag{1}$$

in a setting with two regimes. More generally, for any number of M regimes we could use

$$\boldsymbol{\pi} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{e}_{N+1} \;\;\&\;\; \mathbf{A} = \begin{bmatrix} \mathbf{I}_M - \mathbf{P} \\ \mathbf{1}' \end{bmatrix}$$

where $\mathbf{e}_{M+1}$ is the $(M+1)$th column of $\mathbf{I}_{M+1}$. These ergodic probabilities tell us on average, in the long-run, the proportion of time the process S_t spends in each regime.

2.2 Markov switching autoregressive models

In general, a Markov switching model can be expressed as

$$y_t = x_t\beta + z_t\delta_{s_t} + \sigma_{s_t}\epsilon_t \tag{2}$$

where, in a univariate setting, y_t is a scalar, x_t is a fixed (or predetermined) $(1 \times n)$ vector of variables who's coefficient do not depend on the latent Markov process S_t, z_t is a $(1 \times \nu)$ vector of variables who's coefficient do depend on the Markov process S_t and ϵ_t is the errors process, which for example may be disturbed as a $\mathcal{N}(0, 1)$ and is multiplied by the standard deviation σ_{s_t} which may also depend on the Markov process S_t or remain constant throughout (i.e., σ). Here, we consider the following version of the Markov switching model which only includes autoregressive lags as explanatory variables

$$y_t = \mu_{s_t} + \sum_{k=1}^{p} \phi_k (y_{t-k} - \mu_{s_{t-k}}) + \sigma_{s_t} \epsilon_t \tag{3}$$

This Markov switching autoregressive model is labelled as "`MSARmdl`" in the R package **MSTest** and the version that is most commonly used in various economic and financial applications, as well as other time series related applications. Other error distribution, such as a Student-t distribution, may also be considered. Currently the R package **MSTest** only considers Markov switching autoregressive models with normally distributed errors when simulating and estimating and so we focus on this setup in this paper. The use of other error distributions may also be added in future version of the package.

Continuing with the example where a Markov switching model given by equation (3) has $M = 2$ regimes, such that $S_t = \{1, 2\}$, the sample log likelihood conditional on the first p observations of y_t is given by

$$L_T(\theta) = log f(y_1^T | y_{-p+1}^0; \theta) = \sum_{t=1}^{T} log f(y_t | y_{-p+1}^{t-1}; \theta) \tag{4}$$

where $\theta = (\mu_1, \mu_2, \sigma_1, \sigma_2, \phi_1, \dots, \phi_p, p_{11}, p_{22})$ and

$$f(y_t | y_{-p+1}^{t-1}; \theta) = \sum_{s_t=1}^{2} \sum_{s_{t-1}=1}^{2} \cdots \sum_{s_{t-p}=1}^{2} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \dots, S_{t-p} = s_{t-p} | y_{-p+1}^{t-1}; \theta) \tag{5}$$

and more specifically

$$\begin{aligned} f(y_t, S_t = s_t, \dots, S_{t-p} = s_{t-p} | y_{-p+1}^{t-1}; \theta) = {} & \frac{P(S_t^* = s_t^* | y_{-p+1}^{t-1}; \theta)}{\sqrt{2\pi\sigma_{s_t}^2}} \\ & \times exp\left\{ \frac{-[y_t - \mu_{s_t} - \sum_{k=1}^{p} \phi_k (y_{t-k} - \mu_{s_{t-k}})]^2}{2\sigma_{s_t}^2} \right\} \end{aligned} \tag{6}$$

where we let

$$S_t^* = s_t^* \text{ if } S_t = s_t, S_{t-1} = s_{t-1}, \dots, S_{t-p} = s_{t-p}$$

and $P(S_t^* = s_t^* | y_{-p+1}^{t-1}; \theta)$ is the probability that this occurs.

2.3 Markov switching VAR model

Krolzig, 1997 generalized the the univariate Markov switching autoregressive model to the multivariate setting and hence introduced the Markov switching Vector Autoregressive (MS-VAR) model. A MS-VAR model can be expressed as

$$\boldsymbol{y}_t = \boldsymbol{\mu}_{S_t} + \boldsymbol{\Phi}_1 (\boldsymbol{y}_{t-1} - \boldsymbol{\mu}_{S_{t-1}}) + \dots + \boldsymbol{\Phi}_p (\boldsymbol{y}_{t-p} - \boldsymbol{\mu}_{S_{t-p}}) + \boldsymbol{\Sigma}_{S_t}^{1/2} \boldsymbol{\epsilon}_t \tag{7}$$

where $\boldsymbol{y}_t = [y_{1,t}, \dots, y_{q,t}]'$, $\boldsymbol{\mu}_{S_t} = [\mu_{1,S_t}, \dots, \mu_{q,S_t}]'$, $\boldsymbol{\epsilon}_t = [\epsilon_{1,t}, \dots, \epsilon_{q,t}]'$, $\boldsymbol{\Phi}_k$ is a $(q \times q)$ matrix containing the autoregressive parameters at lag k, and $\boldsymbol{\Sigma}_{S_t} = \boldsymbol{\Sigma}_{S_t}^{1/2} (\boldsymbol{\Sigma}_{S_t}^{1/2})'$ is the regime dependent covariance matrix. This is the version that is considered in the R package **MSTest** and are labelled as "`MSVARmdl`" within the package. More sophisticated versions of the MS-VAR model, as well as their likelihood functions, are described in Krolzig, 1997 and we direct the interested reader to consider this reference to learn more about these models and their components.

2.4 Hidden Markov model

Hidden Markov models (HMM) can be shown to be a special case of the more general Markov switching model we defined above. For example, we can recover the univariate version from equation (3) by simply excluding x_t and setting $z_t = 1$ to obtain the following HMM

$$y_t = \mu_{s_t} + \sigma_{s_t}\epsilon_t \tag{8}$$

This version and its multivariate counterpart are the HMMs considered in the R package **MSTest** and are labelled as "`HMmdl`". Hence, in this case, the process y_t is $i.i.d$ conditional on S_t and depends only on S_t.

As described by An et al., 2013, when the process Y_t also depends on lags $\{Y_{t-1}, \dots, Y_{t-p}\}$, it is referred to as a Hidden Markov-switching Model, or simply Markov-switching as it is commonly known in the economics and finance literature. The dependence on past observations allows for more general interactions between Y_t and S_t which be used to model more complicated causal links between economic or financial variables of interest. Including past observations is a very common practice in economic time series applications as a way to control for stochastic trends, which may explain why Markov-switching models are more popular than the basic HMM in this literature.

2.5 Model estimation

Typically, Markov switching models are estimated using the Expectation Maximization (EM) algorithm (see Dempster et al., 1977), Bayesian methods or through the use of the Kalman filter if using the state-space representation of the model. In very simple cases, Markov switching models can be estimated using Maximum Likelihood Estimation (MLE). However, since the Markov process S_t is latent and more importantly the likelihood function can have several modes of equal height in addition to other unusual features that can complicate estimation by MLE this is not often used. The R package **MSTest** allows one to estimate the models described above through the use of the EM algorithm by setting "`control = list(method = 'EM')`" or by Maximum likelihood estimation by setting "`control = list(method ='MLE')`" in the estimation functions which are described in more detail below in section four. It is worth noting that in practice, however, empirical estimates can sometimes be improved by using the results of the EM algorithm as initial values in a Newton-type of optimization algorithm. This two-step estimation procedure is used to obtain results presented in the empirical section of this paper and those of Rodriguez-Rondon and Dufour, 2023a as well as others. We omit a detailed explanation of the EM algorithm and MLE as our focus is on describing the R package **MSTest**. For the interested reader, the estimation of a Markov switching model via the EM algorithm and MLE is describe in detail in Hamilton, 1990 and Hamilton, 1994 and the in Krolzig, 1997 for the a Markov-switching VAR model.

3 Hypothesis testing for number of regimes

When estimating a Markov switching model, one typically needs to know the number of regimes they would like to estimate the model with as this is not determined endogenously during the estimation process. However, it is well understood in the literature that when considering testing for the number of regimes in a Markov switching model, conventional hypothesis testing procedures are no longer valid as a result of certain regularity conditions not being met. For this reason various studies have proposed different methods to circumvent these issues and yield a valid testing procedure. In this section we describe some important procedures with emphasis on the procedures included in the R package **MSTest**. As previously mentioned in the introduction the purpose of **MSTest** is to make available the most useful of these procedures to a general audience and here we describe where these procedures fit within the literature regarding hypothesis testing for Markov switching models.

In general, the hypothesis test of interest is as follow

$$H_0 : M = M_0 \tag{9}$$
$$H_1 : M = M_0 + m \tag{10}$$

where both $M_0, m \geq 1$. However, as we will see, most available test procedures can only address the case when $M_0 = m = 1$. Hence,

$$H_0 : M = 1 \tag{11}$$
$$H_1 : M = 2 \tag{12}$$

in which case, a linear model (one regime) is being considered under the null hypothesis and is being compared against a Markov switching model with two regimes under the alternative hypothesis. Currently the only procedure able to deal with the very general setting where $M_0, m \geq 1$ are the Monte Carlo likelihood ratio tests described in Rodriguez-Rondon and Dufour, 2023a but this flexibility comes at the cost of being relatively computationally intensive. For this reason **MSTest** also includes other procedures such as the moment-based tests of Dufour and Luger, 2017 which are useful when considering the more simple case of $M_0 = m = 1$. The procedures described in these two studies use Monte Carlo techniques first proposed in Dufour, 2006 to deal with the presence of nuisance parameters under the null hypothesis which is one of the issues plaguing conventional hypothesis testing procedures. Other procedures deal with this issue differently and we also describe these next.

3.1 Likelihood ratio tests

Hansen, 1992 was the first to propose a testing procedure for Markov switching models when $M_0 = m = 1$ and so we begin with a review of this procedure which is available in **MSTest**. In Hansen, 1992 the author provides a good description of the problems that plague the likelihood ratio approach to testing the number of regimes in a Markov switching model. First it is typically assumed that the likelihood function is locally quadratic in the region where the null hypothesis and the globally optimal estimated parameters are found. However, as the author describes, since some parameters are not identified under the null this region is likely flat with respect to those unidentified parameters and not quadratic. Unidentified nuisance parameter under the null hypothesis are issues that have been addressed by Davies, 1977, Davies, 1987, Andrews and Ploberger, 1994b, and Dufour, 2006 for example. Second it is also commonly assumed that the score is positive, as he describes it can actually be identically 0 under the restricted MLE (null hypothesis) of a linear model. Third some parameters, such as the transition probabilities, can take values of 0 and 1 and so we are met with the parameter boundary problem discussed in Andrews, 1999 and Andrews, 2001. Further, the likelihood surface can have multiple local optimums and so the null hypothesis may not lie in the same region as the global optimum.

Hansen, 1992 introduces a new approach that does not require these assumptions. Instead, they model the likelihood function as an empirical process of the unknown parameters. They use empirical process theory to obtain a bound for the asymptotic distribution of the standardized likelihood ratio test. Hansen, 1992 formulates the hypothesis as follows:

$$\begin{aligned} H_0 &: \alpha = \alpha_0 \\ H_1 &: \alpha \neq \alpha_0 \end{aligned} \tag{13}$$

where α_0 represents the parameter values under the null and α the parameter values under the alternative. They begin by decomposing the likelihood ratio as as such:

$$\begin{aligned} LR_n(\alpha) &= L_n(\alpha) - L_n(\alpha_0) \\ &= \Sigma_{i=0}^n [l_i(\alpha) - l_i(\alpha_0)] \\ &= R_n(\alpha) + Q_n(\alpha) \end{aligned} \tag{14}$$

where $R_n(\alpha) = E[LR_n(\alpha)]$ is the expectation of the likelihood ratio function and $Q_n(\alpha) = \Sigma_{i=1}^n q_i(\alpha) = [l_i(\alpha) - l_i(\alpha_0)] - E[l_i(\alpha) - l_i(\alpha_0)]$ is the deviations from the mean. Fluctuations in Q play an important role in identifying an optimum as:

$$\begin{aligned} \frac{1}{\sqrt{n}} LR_n(\alpha) &= \frac{1}{\sqrt{n}} R_n(\alpha) + \frac{1}{\sqrt{n}} Q_n(\alpha) \\ &= \frac{1}{\sqrt{n}} R_n(\alpha) + Q(\alpha) + o_p(1) \end{aligned} \tag{15}$$

and by using the fact that $R_n(\alpha) \leq 0$ for all α when the null hypothesis is true, we can see that $\frac{1}{\sqrt{n}} LR_n(\alpha) \leq \frac{1}{\sqrt{n}} Q_n(\alpha)$ and so it follows that,

$$P[\frac{1}{\sqrt{n}} LR_n \geq x] \leq P[\sup_{\alpha} \frac{1}{\sqrt{n}} Q_n(\alpha) \geq x] \rightarrow P[\sup_{\alpha} Q(\alpha) \geq x] \tag{16}$$

Thus, the distribution of the empirical process Q can provide a bound for the asymptotic distribution of the LR statistic. The test statistic is further standardized as such:

$$LR_n^* = \sup_{\alpha} LR_n^*(\alpha) \tag{17}$$

where

$$LR_n^*(\alpha) = \frac{LR_n(\alpha)}{V_n(\alpha)^{1/2}} \tag{18}$$

As suggested by equation (18), they resolve the issue of nuisance parameters by evaluating the standardized LR statistic for different parameter values α. Specifically, they evaluate the test statistic over a grid of different parameter values and optimize with respect to those nuisance parameters values. To be clear, Hansen, 1992 sets $\alpha = (\mu_2, \sigma_2, p_{11}, p_{22})$ as the vector of nuisance parameter, which includes the second state parameters and transition parameters, as nuisance parameters. The first regime parameters $\theta = (\mu_1, \sigma_1, \phi_1, ...\phi_p)$ are fully identified. He further splits α into $\beta = (\mu_2, \sigma_2)$ and $\gamma = (p_{11}, p_{22})$, where β is treated as a parameter of interest and set to be identical to μ_1 and σ_1 under the null. However, the process Q may also be serially correlated for some values of α and so Hansen, 1996a adds a correction that should be used to calculating the asymptotic distribution of the test statistic which is also included in the implementation of this test procedure in **MSTest**.

There are two main drawbacks to be aware of when considering using this likelihood ratio procedure for testing Markov switching models. The first is that this test only provides a bound for the standardized LR statistic and can therefore be conservative. As such, it is important to remember that the critical values provided by this test in the **MSTest** package are not necessarily those of the standardized LR statistic but rather of the process Q which provides a bound for the standardized LR statistic. The second is that it involves optimizing for the value of the nuisance parameters over a grid-search. Although this process is not too cumbersome when analyzing only a few variables as switching between regimes in addition to the transition parameters p_{11} and p_{22}, it can quickly become too computationally intensive for models that want to consider more parameters as switching between regimes. Despite some of these drawbacks, we include this test proposed in **MSTest** because it provides a good first step in testing the number of regimes and has often been used as a benchmark for comparison in the literature.

Garcia, 1998 also reviewed the problem of testing the number of regimes in a Markov switching model using a likelihood ratio approach. Garcia, 1998 builds on Hansen, 1992's approach but differs in that he only only treats γ as a nuisance parameter over which we have to optimize. This simplifies some of the computational burden. The β parameter is still a parameter of interest but is added to θ, the identified parameters. We do not include this test in this version of **MSTest** because the author assumes that the LR test can be written as the superemum of a chi-square functional asymptotically under the null, which Hansen, 1996b and Andrews and Ploberger, 1994b suggest cannot be done for MS models. It may however, still be included in future versions of **MSTest** for completeness.

Cho and White, 2007 also consider hypothesis testing when the null hypothesis is a linear model and the alternative is a Markov switching model with two regimes. Cho and White, 2007 is further concerned with a difficulty that was not addressed by Hansen, 1992 or Garcia, 1998 which is the case where parameters are on the boundary of the parameter space. The authors argue that the null hypothesis can be classified into two mutually exclusive subsets. In one subset, $p \in (0,1)$ and in the other $p = 0$ or $p = 1$. The second case is the one with the boundary parameter problem. The authors build on Andrews, 1999 and Andrews, 2001 to address this case and develop a QLR test statistic which considers this boundary problem. Additionally, Cho and White, 2007 use a normal mixture model framework and develop the QLR test under this setting. They argue that this approach allows them to ignore certain time series dependence properties implied by the regime-switching process. The authors describe the QLR test as being sensitive to the mixture aspect of the regime-switching process, which they say allows it to deliver a test with appealing power under the alternative. An important finding of their work is that the critical values of the asymptotic distribution of the test statistic are affected by the inclusion or exclusion of the boundary problem. In other words, they show that the boundary problem is important to consider when one is interested in the asymptotic distribution of the test statistic. However, Carter and Steigerwald, 2012 point out an important flaw in their proposed test. They argue that this test can ignore the time dependency of the Markov-chain but not the time dependencies that enter into an autoregressive model when parameters change with regimes, which the authors acknowledge in Cho and White, 2011 and for this reason the test proposed by Cho and White, 2007 is not included in **MSTest** since autoregressive models are especially of interest in many economic and financial applications.

Qu and Zhuo, 2021 present a novel characterization of the conditional regime probabilities for a family of likelihood ratio based tests and establishes the asymptotic distribution for the test statistics. Like Hansen, 1992, Garcia, 1998, Cho and White, 2007, and Carrasco et al., 2014 they derive an approximation of the likelihood ratio, as an empirical process where $\{(p_{11}, p_{22}) : \epsilon \leq p_{11}, p_{22} \leq 1 - \epsilon and p_{11} + p_{22} \geq 1 + \epsilon\}$ and ϵ is a small constant. They also provide a finite sample refinement that can correct some of the over-rejections that can occur in specific cases. As a result they are able to study the asymptotic null distribution and find that, even though the null hypothesis has only one regime and hence parameters do not switch, the nuisance parameters can affect the limiting distribution and this distribution will depend on which parameters are allowed to switch. Furthermore, Qu and Zhuo, 2021 describe how some of these results can explain why specific bootstrap procedures can be inconsistent (e.g., when including weakly exogenous regressors), and why standard information criteria such as the BIC can be sensitive to the hypothesis and the model structure. Although currently the R package **MSTest** does not include the test procedure proposed by Qu and Zhuo, 2021, future version will likely include this test procedure as it provides the best approximation to the asymptotic distribution of he likelihood ratio test statistic and can hence be useful to users interested in making use of these asymptotic results.

Another likelihood ratio based test is that of Kasahara and Shimotsu, 2018. In Kasahara and Shimotsu, 2015, the authors introduce a re-parameterization and higher order expansion of the likelihood ratio function. In Kasahara and Shimotsu, 2018 they use this technique along with the Difference in Quadratic Mean (DQM) approximation introduced by Liu and Shao, 2003 to avoid some of the issues discussed above including the parameter boundary problem to estimate and study the asymptotic distribution of the likelihood ratio test statistic for a Markov switching models. Additionally, they do this in a more general setting where $M_0 \geq 1$ and $m = 1$ which allows them to consider a null hypothesis with more than one regime and hence is more general than the likelihood ratio test procedures discussed so far in that regard. In doing, like Qu and Zhuo, 2021 they also show that the parametric bootstrap procedure is an asymptotically valid test procedure, at least when considering Markov switching autoregressive models. The parametric bootstrap procedure is not directly available in **MSTest** but it can easily be implemented by using specific options while using the Local Monte Carlo likelihood ratio test proposed by Rodriguez-Rondon and Dufour, 2023a which we discuss next.

In Rodriguez-Rondon and Dufour, 2023a, the authors propose the Maximized Monte Carlo likelihood ratio test (MMC-LRT) and the Local Monte Carlo likelihood ratio test (LMC-LRT) which can be used to compare very general Markov switching models such as those given by equation (2). These procedures are the most general type of testing procedures as they can be used to consider hypothesis testing in the most general form when $M_0, m \geq 1$. Further, as described in Rodriguez-Rondon and Dufour, 2023a these Monte Carlo likelihood ratio tests can also be used when the process y_t is non-stationary, has non-Gaussian errors, and when considering multivariate settings which include the Markov switching VAR and multivariate Hidden Markov models previously described. To be more precise, the MMC-LRT and LMC-LRT are the only test procedures available in the literature and in the R package **MSTest** that can be used to test multivariate Markov switching models.

Here, we summarize the MMC-LRT and LMC-LRT procedures but readers interested in further details are refereed to the more formal description provided in Rodriguez-Rondon and Dufour, 2023a and Dufour, 2006 for further details on the Monte Carlo techniques used in these procedures. For simplicity of exposition, we use an example where we are interested by a null hypothesis of linear model (*i.e.*, only $M_0 = 1$ regime) and an alternative hypothesis of $M_0 + m = 2$ regimes. Since this is a likelihood ratio-based approach, the log-likelihood values under the null and alternative hypothesis are required. The log-likelihood for the model under the alternative (and under the null hypothesis if $M_0 > 1$) is given by (4) - (6):

$$L_T(\theta_1) = \log f(y_1^T \,|\, y_{-p+1}^0;\, \theta_1) = \sum_{t=1}^{T} \log f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_1) \tag{19}$$

where

$$\theta_1 = (\mu_1,\, \mu_2,\, \sigma_1,\, \sigma_2,\, \phi_1,\, \dots\,,\, \phi_p,\, p_{11},\, p_{22})' \in \Omega\,. \tag{20}$$

The subscript of 1 underscores the fact that θ_1 is the parameter vector under the alternative hypothesis. The set Ω satisfies any theoretical restrictions we may wish to impose on θ_1 [such as $\sigma_1 > 0$ and $\sigma_2 > 0$]. On the other hand, the log-likelihood under the null hypothesis ($M_0 = 1$) is given by

$$L_T^0(\theta_0) = \log f(y_1^T \,|\, y_{-p+1}^0;\, \theta_0) = \sum_{t=1}^{T} \log f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_0) \tag{21}$$

where

$$f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-[y_t - \mu - \sum_{k=1}^{p} \phi_k (y_{t-k} - \mu)]^2}{2\sigma^2} \right\}, \tag{22}$$

$$\theta_0 = (\mu,\, \sigma^2, \phi_1,\, \ldots\,,\, \phi_p)' \in \bar{\Omega}_0. \tag{23}$$

Note that $\bar{\Omega}_0$ has lower dimension than Ω. The null and alternative hypotheses can be written as:

$$H_0 : \delta_1 = \delta_2 = \delta \quad \text{for some unknown } \delta = (\mu,\, \sigma)\,, \tag{24}$$

$$H_1 : (\delta_1,\, \delta_2) = (\delta_1^*,\, \delta_2^*) \;\; \text{for some unknown } \delta_1^* \neq \delta_2^*\,, \tag{25}$$

where $\delta_1 = (\mu_1,\, \sigma_1)$ and $\delta_2 = (\mu_2,\, \sigma_2)$. Clearly, H_0 is a restricted version of H_1: for each $\theta_0 \in \bar{\Omega}_0$, we can find θ_1 such that

$$L_T^0(\theta_0) = L_T(\theta_1)\,, \quad \theta_1 \in \Omega_0, \tag{26}$$

where Ω_0 is the subset of vectors $\theta_1 \in \Omega$ such that θ_1 satisfies H_0. Under H_0, the vector $\theta_0 \in \bar{\Omega}_0$ is a nuisance parameter: the null distribution of any test statistic for H_0 depends on $\theta_0 \in \bar{\Omega}_0$. In this problem, the null distribution of the test statistic, LR_T, is in fact completely determined by $\theta_0 \in \bar{\Omega}_0$. As in Garcia, 1998 and the parametric bootstrap procedure describe in Qu and Zhuo, 2021 and Kasahara and Shimotsu, 2018, we assume that the null hypothesis depends only on the mean, variance and autoregressive coefficients. The likelihood ratio statistic for testing H_0 against H_1 can then written as

$$LR_T = 2[\bar{L}_T(H_1) - \bar{L}_T(H_0)] \tag{27}$$

where

$$\bar{L}_T(H_1) = \sup\{L_T(\theta_1) : \theta_1 \in \Omega\}\,, \tag{28}$$

$$\bar{L}_T(H_0) = \sup\{L_T^0(\theta_0) : \theta_0 \in \bar{\Omega}_0\} = \sup\{L_T(\theta_1) : \theta_1 \in \Omega_0\}\,. \tag{29}$$

Since the model is parametric, we can generate a vector N i.i.d replications of LR_T for any given value of $\theta_0 \in \bar{\Omega}_0$:

$$LR(N,\, \theta_0) := [LR_T^{(1)}(\theta_0),\, \ldots\,,\, LR_T^{(N)}(\theta_0)]', \qquad \theta_0 \in \bar{\Omega}_0\,. \tag{30}$$

As discussed in Rodriguez-Rondon and Dufour, 2023a, the main assumptions required are that the random variables $LR_T^{(0)}$, $LR_T^{(1)}(\theta_0)$, $\ldots$, $LR_T^{(N)}(\theta_0)$ are exchangeable for some $\theta_0 \in \bar{\Omega}_0$ each with distribution function $F[x \,|\, \theta_0]$ (i.e., they are *i.i.d*). From here, we can compute the Monte Carlo p-value which is given by

$$\hat{p}_N[x \,|\, \theta_0] = \frac{N + 1 - R_{LR}[LR_T^{(0)};\, N]}{N + 1} \tag{31}$$

where

$$R_{LR}[LR_T^{(0)};\, N] = \sum_{i=1}^{N} I[LR_T^0 \geq LR_T^i(\theta_0)] \tag{32}$$

and $I(C) := 1$ if condition C holds, and $I(C) = 0$ otherwise. As can be seen from (32), $R_{LR}[LR_T^{(0)};$ $N]$ simply computes the rank of the test statistic from the observed data within the generated series $LR(N, \theta_0)$. As discussed in Dufour, 2006 and Rodriguez-Rondon and Dufour, 2023a, a critical region with level α is then given by

$$\sup_{\theta_0 \in \bar{\Omega}_0} \hat{p}_N[LR_T^{(0)} \,|\, \theta_0] \leq \alpha \tag{33}$$

More precisely, if $(N+1)\alpha$ is an integer, then

$$\mathbb{P}\left[\sup\{\hat{p}_N[LR_T^{(0)}\,|\,\theta_0] : \theta_0 \in \bar{\Omega}_0\} \leq \alpha\right] \leq \alpha \tag{34}$$

under the null hypothesis and so we get a valid test with level α for H_0.

The Maximized Monte Carlo likelihood ratio test requires searching for the maximum Monte Carlo p-value over the nuisance parameter space $\bar{\Omega}_0$. Since this space can be very large and grows as the number of autoregressive components and the number of regimes increases, the authors propose a more efficient alternative which involves searching over a consistent set C_T [as described in Dufour, 2006]. A consistent set can be defined using the consistent point estimate. For example, let $\hat{\theta}_0$ be the consistent point estimate of θ_0. Then, we can define

$$C_T = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < c\} \tag{35}$$

where c is a fixed positive constant that does not depend on T and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$. The consistent set considered in this Rodriguez-Rondon and Dufour, 2023a is $C_T^* = C_T^{CI} \cup C_T^{\epsilon}$ where

$$C_T^{CI} = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < 2 \times S.E.(\hat{\theta}_0)\} \tag{36}$$

$$C_T^{\epsilon} = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < \epsilon\} \tag{37}$$

Hence C_T^{CI} is defined by a 95% confidence interval of the consistent point estimates, while C_T^{ϵ} is defined using a fixed constant ϵ that does not depend on T. The union of these two sets allows us to consider values that may be outside the confidence interval for some parameters and within the confidence interval for other depending on the choice of ϵ. The **MSTest** R package users to use the define the consistent set C_T using only a fixed positive constant ϵ, only the confidence interval, or the union of both. As discussed in Dufour, 2006 and Rodriguez-Rondon and Dufour, 2023a the solution to this optimization problem may not be unique in the sense that the maximum p-value may be obtained by more than one parameter vector. For this reason, numerical optimization methods that do not depend on the use of derivatives are recommended to find the maximum Monte Carlo p-value within the nuisance parameter space. The R package **MSTest** allows users to use the Generalized Simulated Annealing algorithm, Genetic Algorithm, and Particle Swarm algorithm [see Yang Xiang et al., 2013, Zambrano-Bigiarini et al., 2013, Scrucca et al., 2013, Dufour, 2006, and Dufour and Neves, 2019].

Finally, as described in Rodriguez-Rondon and Dufour, 2023a, we can also define C_T to be the singleton set $C_T = \{\hat{\theta}_0\}$, which gives us the Local Monte Carlo Likelihood Ratio Test. Here, the consistent set includes only the consistent point estimate $\hat{\theta}_0$. As a result, the Monte Carlo p-value depends only on $\hat{\theta}_0$. The LMC version of test can be interpreted as the finite-sample analogue of the parametric bootstrap. The asymptotic validity in this case refers to the estimate $\hat{\theta}_0$ converging asymptotically to the true parameters in θ_0 as the sample size increases. This is not related to the asymptotic validity of the critical values as desired in Hansen, 1992, Garcia, 1998, Cho and White, 2007, Qu and Zhuo, 2021 and Kasahara and Shimotsu, 2018. Specifically, like the parametric bootstrap, the LMC procedure is only valid asymptotically as $T \to \infty$ but, unlike the parametric bootstrap, we do not need a large number of simulations ($i.e.$, $N \to \infty$), since we do not try to approximate the asymptotic critical values nor assume that the distribution of the test statistic converges asymptotically but rather work with the critical values from the sample distribution. This allows the procedure to be computationally efficient in the sense that we will not need to perform a large number of simulations with the aim of obtaining asymptotically valid critical values. It is now worth noting that the parametric bootstrap procedure discussed in

Qu and Zhuo, 2021 and Kasahara and Shimotsu, 2018 can be implemented by constraining the parameter space of the transition probabilities according to the assumptions used in those studies and using a larger number of simulations to attempt to approximate the asymptotic critical values in the cases and settings where authors have shown that the parametric bootstrap procedure is valid. Nevertheless, as discussed in Rodriguez-Rondon and Dufour, 2023a and as has now become apparent, the LMC-LRT and MMC-LRT procedures are the most general likelihood ratio test procedure available.

3.2 Moment-based tests

Dufour and Luger, 2017 propose a different way to test Markov switching models that also avoids the statistical issues described above for likelihood ratio type tests. Additionally, their test is less costly computationally in comparison to all test mentioned above, the parameter stability test discussed next, and allows the econometrician to perfectly control the level of the test through the use of the Monte Carlo test methods described in Dufour, 2006 and used in Rodriguez-Rondon and Dufour, 2023a. They are also interested in a hypothesis test of the form given by (24) - (25) but their proposed method can only deal with this case (i.e., when $M_0 = m = 1$). The moment-based test of Dufour and Luger, 2017 involves computing moments of the least-square residuals of autoregressive models under the null hypothesis. More Specifically, they focus on the mean, variance, skewness and excess kurtosis of the least-square residuals. These moments are calculate as

$$M(\hat{\epsilon}) = \frac{|m_2 - m_1|}{\sqrt{s_1^2 + s_2^2}} \tag{38}$$

where, $m_1 = \frac{\Sigma_{t=1}^T \hat{\epsilon}_t \mathbb{1}[\hat{\epsilon}_t<0]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon}_t<0]}$, $m_2 = \frac{\Sigma_{t=1}^T \hat{\epsilon}_t \mathbb{1}[\hat{\epsilon}_t>0]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon}_t>0]}$, $s_1^2 = \frac{\Sigma_{t=1}^T (\hat{\epsilon}_t - m_1)^2 \mathbb{1}[\hat{\epsilon}_t<0]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon}_t<0]}$ and $s_2^2 = \frac{\Sigma_{t=1}^T (\hat{\epsilon}_t - m_2)^2 \mathbb{1}[\hat{\epsilon}_t>0]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon}_t>0]}$

$$V(\hat{\epsilon}) = \frac{\vartheta_2(\hat{\epsilon})}{\vartheta_1(\hat{\epsilon})} \tag{39}$$

where, $\vartheta_1 = \frac{\Sigma_{t=1}^T \hat{\epsilon_t}^2 \mathbb{1}[\hat{\epsilon_t}^2<\hat{\sigma}^2]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon_t}^2<\hat{\sigma}^2]}$, $\vartheta_2 = \frac{\Sigma_{t=1}^T \hat{\epsilon_t}^2 \mathbb{1}[\hat{\epsilon_t}^2>\hat{\sigma}^2]}{\Sigma_{t=1}^T \mathbb{1}[\hat{\epsilon_t}^2>\hat{\sigma}^2]}$ and $\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^T \hat{\epsilon_t}^2$

$$S(\hat{\epsilon}) = |\frac{\Sigma_{t=1}^T \hat{\epsilon_t}^3}{T(\hat{\sigma}^2)^{3/2}}| \tag{40}$$

and

$$K(\hat{\epsilon}) = |\frac{\Sigma_{t=1}^T \hat{\epsilon_t}^4}{T(\hat{\sigma}^2)^2} - 3| \tag{41}$$

The testing procedure involves calculating the test statistic for each moment, obtaining the individual p-values and using two different methods of combining independent test statistics. The first method is based on the min of the p-values and was suggested by Tippett et al., 1931 and Wilkinson, 1951. Here, the test statistic becomes,

$$F_{min}(\hat{\epsilon}) = 1 - min\{\hat{G_M}[M(\hat{\epsilon})], \hat{G_V}[V(\hat{\epsilon})], \hat{G_S}[S(\hat{\epsilon})], \hat{G_K}[K(\hat{\epsilon})]\} \tag{42}$$

where for example, $\hat{G_M}[M(\hat{\epsilon})] = 1 - \hat{F_M}[M(\hat{\epsilon})]$ is the Monte Carlo p-value of $M(\hat{\epsilon})$. The second method of combining the test statistics involves taking the product of them. This method of combining test statistics was suggested by Fisher, 1932 and Pearson, 1933. In this case the the test statistic becomes,

$$F_{\times}(\hat{\epsilon}) = 1 - \{\hat{G_M}[M(\hat{\epsilon})] \times \hat{G_V}[V(\hat{\epsilon})] \times \hat{G_S}[S(\hat{\epsilon})] \times \hat{G_K}[K(\hat{\epsilon})]\} \tag{43}$$

Interested readers should see Dufour et al., 2004 and Dufour et al., 2014 which provide further discussion of these methods of combining test statistics. Finally, the Monte Carlo p-value of the combined statistics is given by

$$G_{F_{min}}[F_{min}(\hat{\epsilon}); N] = \frac{N + 1 - R_{F_{min}}[F_{min}(\hat{\epsilon}); N]}{N} \tag{44}$$

and

$$G_{F_{\times}}[F_{\times}(\hat{\epsilon}); N] = \frac{N + 1 - R_{F_{\times}}[F_{\times}(\hat{\epsilon}); N]}{N} \tag{45}$$

where $R_{F_{min}}$ and $R_{F_{\times}}$ are the ranks of $F_{min}(\hat{\epsilon})$ and $F_{\times}(\hat{\epsilon})$ in $F_{min}(\hat{\eta}_1)$, ..., $F_{min}(\hat{\eta}_{N-1})$ and $F_{\times}(\hat{\eta}_1)$, ...,$F_{\times}(\hat{\eta}_{N-1})$ respectively, when ordered. Also, $\hat{\eta} = \eta - \bar{\eta}$ and $\eta \sim N(0, I_T)$.

The computational efficiency of this test makes it easily extendable to the use of Maximized Monte Carlo when nuisance parameters are present. Furthermore, this test is not subject to the same level of statistical difficulties such as unidentified parameters under the null as in Hansen, 1992, Garcia, 1998 and Carrasco et al., 2014. This is because, transition probability parameters, the mean and the variance do not need to be treated as nuisance parameters. Parameters of explanatory variables are the only ones which may be unidentified under the null and so only these are treated as nuisance parameters. Garcia, 1998 also reduced the nuisance parameter space by treating only the transition probabilities p_{11} and q_{22} as nuisance parameters, however, this is an even further reduction of the nuisance parameter space and makes the treatment of autoregressive models with more lags more tractable. Although the moment-based test can only be used to compare linear models against Markov switching models with two regimes, it is the least computationally intensive procedure available and takes only seconds to compute, even when considering the MMC version of the test, and for this reason it is included in the R package **MSTest**.

3.3 Optimal test for regime switching

Carrasco et al., 2014 proposes a different type of test. The authors describe their test as an optimal test for the consistency of parameters in random coefficient and Markov switching models. They add that their test can be understood as an extension of White, 1982's information matrix test and argue that it shares some similar advantages, such as only needing to estimate the model under the null hypothesis which as we saw is also the case for the moment-based test of Dufour and Luger, 2017. On the other hand, the likelihood ratio tests proposed by Hansen, 1992, Garcia, 1998, Cho and White, 2007, Qu and Zhuo, 2021, Kasahara and Shimotsu, 2018, and Rodriguez-Rondon and Dufour, 2023a all require estimating the model both under the null and alternative hypothesis. This is a favourable feature because the non-linearity in estimating Markov switching models introduces multiple local optimums making the likelihood ratio test procedures relatively more computationally intensive. Another noteworthy advantage of this test is that it can also be applied to test Markov switching GARCH models. Furthermore, the authors appeal to the Neyman-Pearson lemma to prove the optimality of their test and show that it is asymptotically locally equivalent to the likelihood ratio test. However, this method also involves a bootstrap procedure while performing a search over the nuisance parameter space which can make obtaining asymptotic critical values, especially when the variance is allowed to switch, very computationally intensive.

The authors formulate the hypothesis in the following way:

$$\begin{aligned} H_0 &: \theta_t = \theta_0 \\ H_1 &: \theta_t = \theta_0 + \eta_t \end{aligned} \tag{46}$$

where the switching variable η_t is unobservable, stationary and may depend on nuisance parameters β. Their test makes use of the second derivatives of the log-likelihood and the outer products of the scores as in the information matrix test with the addition of an extra term, which captures the serial dependence of the time-varying coefficients. This means that the form of the test depends on the latent process η_t only through its second-order properties. Additionally, the distribution of η_t is assumed to exist even under the null, but does not play a role with regards to the distribution of the data $(y_T, y_{T-1}, y_{T-2}, ...y_1)$ under the null. That is, under the null, they are mutually exclusive.

The authors first propose a Sup-type test as in Davies, 1987 to combat the presence of nuisance parameters. They set $\eta_t = chS_t$, where c is a scalar specifying the amplitude of the change, h a vector specifying the direction of the alternative and S_t is a Markov-chain, which follows an autoregressive process such as $S_t = \rho S_{t-1} + e_t$, where e_t is i.i.d. U[-1,1] and $-1 < \rho < 1$ so that S_t is bounded by support $(-1/(1-|\rho|), 1/(1-|\rho|)$ and has zero mean. Letting $\beta = (c^2, h', \rho')$ be our vector of nuisance parameters, we can write

$$\mu_{2,t}(\beta,\theta) = \frac{1}{2}c^2 h'[(\frac{\partial l_t}{\partial\theta\partial\theta'} + (\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta})') + 2\sum_{s<t}\rho^{(t-s)}(\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta})']h \tag{47}$$

which allows us to get the expression

$$supTS = \sup_{\{h,p:||h||=1,\underline{\rho}<\rho<\bar{\rho}\}} = \frac{1}{2}(max(0, \frac{\Gamma_T^*}{\sqrt{\hat{\epsilon^{*}}'\hat{\epsilon^*}}}))^2 \tag{48}$$

where $\mu^*_{2,t}(\beta,\theta) = \mu_{2,t}(\beta,\theta)/c^2$, $\Gamma^*_T = \Gamma^*_T(\beta,\theta) = \sum_t \mu^*_{2,t}(\beta,\theta)/\sqrt{T}$ and $\hat{\epsilon^*}$ are the residuals from regressing $\mu^*_{2,t}(\beta,\theta)$ on $l_t^{(1)}(\hat{\theta})$ so that Γ^*_T and $\hat{\epsilon^*}$ are not dependent on c^2. As previously mentioned, this methodology involves bootstrapping over the distributions of the nuisance parameters. As a result, one must choose a prior distribution for the nuisance parameters. The most commonly used distribution in this case is the uniform distribution, but since the parameter c^2 is not necessarily bounded from above, a uniform distribution may not always be appropriate. As a result, Carrasco et al., 2014 also suggest using an Exponential-type test as in Andrews and Ploberger, 1994b. They propose the following statistic:

$$expTS = \int_{\{\underline{\rho}\leq\rho\leq\bar{\rho},||h||<1\}} \Psi(h,\rho)d\rho dh \tag{49}$$

where

$$\Psi(h,\rho) = \begin{cases} \sqrt{2\pi}exp[\frac{1}{2}(\frac{\Gamma^*_T}{\sqrt{\hat{\epsilon^{*}}'\hat{\epsilon^*}}}-1)^2]\Phi(\frac{\Gamma^*_T}{\sqrt{\hat{\epsilon^{*}}'\hat{\epsilon^*}}}-1) & \text{if } \hat{\epsilon^{*}}'\hat{\epsilon^*} \neq 0. \\ 1 & \text{otherwise.} \end{cases} \tag{50}$$

These tests proposed by Carrasco et al., 2014 have been used in the empirical applications of Hamilton, 2005, Warne and Vredin, 2006, Kahn and Rich, 2007, Morley and Piger, 2012, and Dufrénot et al., 2011, in testing MS-GARCH models by Hu and Shin, 2008, and in Dufour and Luger, 2017, Qu and Zhuo, 2021, and Rodriguez-Rondon and Dufour, 2023a as a benchmark to compare their test procedures. Due to their wide use and optimality results, the **MSTest** package also includes these test procedures.

Additionally, the **MSTest** package includes a Monte Carlo Test version of this test procedure which is performed much quicker. This version of the test is novel and unique to this package and to the best of knowledge, has not been considered in other works.

4 The R package MSTest

4.1 Data sets

The R package **MSTest** includes three samples of U.S. real GNP that can be easily called once the package is loaded. Specifically, it makes available the original sample used in Hamilton, 1989 which was later used by Hansen, 1992, Carrasco et al., 2014, and Dufour and Luger, 2017 among others to test the null hypothesis of a linear model (one regime) against an alternative hypothesis of a Markov switching autoregressive model with two regimes and showcase the performance of their test procedures.

Label	Descriptions
`hamilton84GNP`	Sample originally considered in Hamilton, 1989. This sample ranges from 1951Q2 to 1984Q4.
`chp10GNP`	This sample is the second sample of the U.S. real GNP used in Carrasco et al., 2014 and Dufour and Luger, 2017. This sample ranges from 1951Q2 to 2010Q4.
`USGNP`	This sample is used in Rodriguez-Rondon and Dufour, 2023a. The sample ranges from 1947Q2 to 2022Q3.

Table 1: U.S. real GNP samples in **MSTest**

table 1 provides the label used to identify each sample in the R package **MSTest** and describes the span of each sample.

These data sets can be loaded using the following commands in R once **MSTest** has been loaded:

```
R> data("hamilton84GNP", package = "MSTest")
R> data("chp10GNP", package = "MSTest")
R> data("USGNP", package = "MSTest")
```

all three data sets include three columns: (1) `DATE`, (2) `GNP_logdiff`, and (3) `GNP`. The first is a `Date` type variable defined using the `as.Date()` function, the second is a growth rate of U.S. real GNP and the third column is the levels of U.S. real GNP.

4.2 Simulation

Here we describe a set of functions available in **MSTest**, which allow the user to simulate different types of processes. These functions are used by some hypothesis testing procedures, specifically those that involve using a simulations procedure to obtain the (sample or asymptotic) null distribution of the test statistic such as `LMCLRTest`, `MMCLRTest`, `DLMCTest`, `DLMMCTest`, and `CHPTest`. Users interested in developing new estimation or testing procedures for Markov switching models may find these functions useful to test the performance of such procedures and compare them to the ones that are available here in **MSTest**. table 4 provides the label used to identify each simulation function in the R package **MSTest** and describes what process they can be used to simulate. Each function requires a `list` object as input and next wee describe what this list must contain for each function and how they can be used in R.

As described in 4, simulating a normally distributed process can be done using command the command `simuNorm` in R. Specifically, the function require that the user input a `list` with the sample size (`n`), the number of series to be simulated `q` (i.e., if `q=1` a univariate process is simulated and if `q>1` a multivariate process is simulated), a $(q \times 1)$ vector of means for each series, and a

Function	Description
`simuNorm`	Simulates a normally distributed process.
`simuAR`	Simulate an autoregressive process with p lags (AR(p)).
`simuVAR`	Simulate a vector autoregressive process with p lags (VAR(p)).
`simuMSAR`	Simulate a Markov switching autoregressive process with p lags (MSAR(p)).
`simuMSVAR`	Simulate a Markov switching vector autoregressive process with p lags (MSVAR(p)).
`simuHMM`	Simulate a hidden Markov process (HMM).

Table 2: Simulation functions available in the R package **MSTest**

`simuNorm` ($q \times q$) covariance matrix. The user may also specify the number of additional observations to simulate and discard from the beginning using the `burnin` input. By default, this input for `simuNorm` is negative. As we will see with autoregressive processes, this value is typically higher to avoid any dependence on the initialization used in simulating the process. The user also have the option to provide a ($(n + burnin) \times q$) matrix of errors if they do not which to use normally distributed errors. This is done by defining the element `eps` in the input `list` using that matrix.

```
R> # Define DGP of multivariate normal process
R> mdl_norm <- list(n     = 500,
+                   q     = 2,
+                   mu    = c(5, -2),
+                   sigma = rbind(c(5.0, 1.5),
+                                 c(1.5, 1.0)))
R> # Simulate process
R> simu_norm <- simuNorm(mdl_norm)
```

Simulating an autoregressive process can be done using

```
R> # Define DGP of AR(2) process
R> mdl_ar <- list(n     = 500,
+                 mu    = 5,
+                 sigma = 1,
+                 phi   = c(0.75))
R> # Simulate process
R> simu_ar <- simuAR(mdl_ar)
```

Simulating a vector autoregressive process can be done using

```
R> # Define DGP of VAR(2) process
R> mdl_var <- list(n     = 500,
+                  p     = 1,
+                  q     = 2,
+                  mu    = c(5, -2),
+                  sigma = rbind(c(5.0, 1.5),
+                                c(1.5, 1.0)),
+                  phi   = rbind(c(0.50, 0.30),
+                                c(0.20, 0.70)))
```

```
R> # Simulate process
R> simu_var <- simuVAR(mdl_var)

R> # Define DGP of MS AR process
R> mdl_ms <- list(n     = 500,
+                 mu    = c(5,10),
+                 sigma = c(1,2),
+                 phi   = c(0.75),
+                 k     = 2,
+                 P     = rbind(c(0.95, 0.10),
+                               c(0.05, 0.90)))
R> # Simulate process
R> simu_msar <- simuMSAR(mdl_ms)

R> # Define DGP of MS VAR process
R> mdl_msvar <- list(n     = 500,
+                    p     = 1,
+                    q     = 2,
+                    mu    = rbind(c(5, -2),
+                                  c(10, 2)),
+                    sigma = list(rbind(c(5.0, 1.5),
+                                       c(1.5, 1.0)),
+                                 rbind(c(7.0, 3.0),
+                                       c(3.0, 2.0))),
+                    phi   = rbind(c(0.50, 0.30),
+                                  c(0.20, 0.70)),
+                    k     = 2,
+                    P     = rbind(c(0.95, 0.10),
+                                  c(0.05, 0.90)))
R> # Simulate process
R> simu_msvar <- simuMSVAR(mdl_msvar)

R> # Define DGP of HMM
R> mdl_hmm <- list(n     = 500,
+                  q     = 2,
+                  mu    = rbind(c(5, -2),
+                                c(10, 2)),
+                  sigma = list(rbind(c(5.0, 1.5),
+                                     c(1.5, 1.0)),
+                               rbind(c(7.0, 3.0),
+                                     c(3.0, 2.0))),
+                  k     = 2,
+                  P     = rbind(c(0.95, 0.10),
+                                c(0.05, 0.90)))
R> # Simulate process
R> simu_hmm <- simuHMM(mdl_hmm)
```

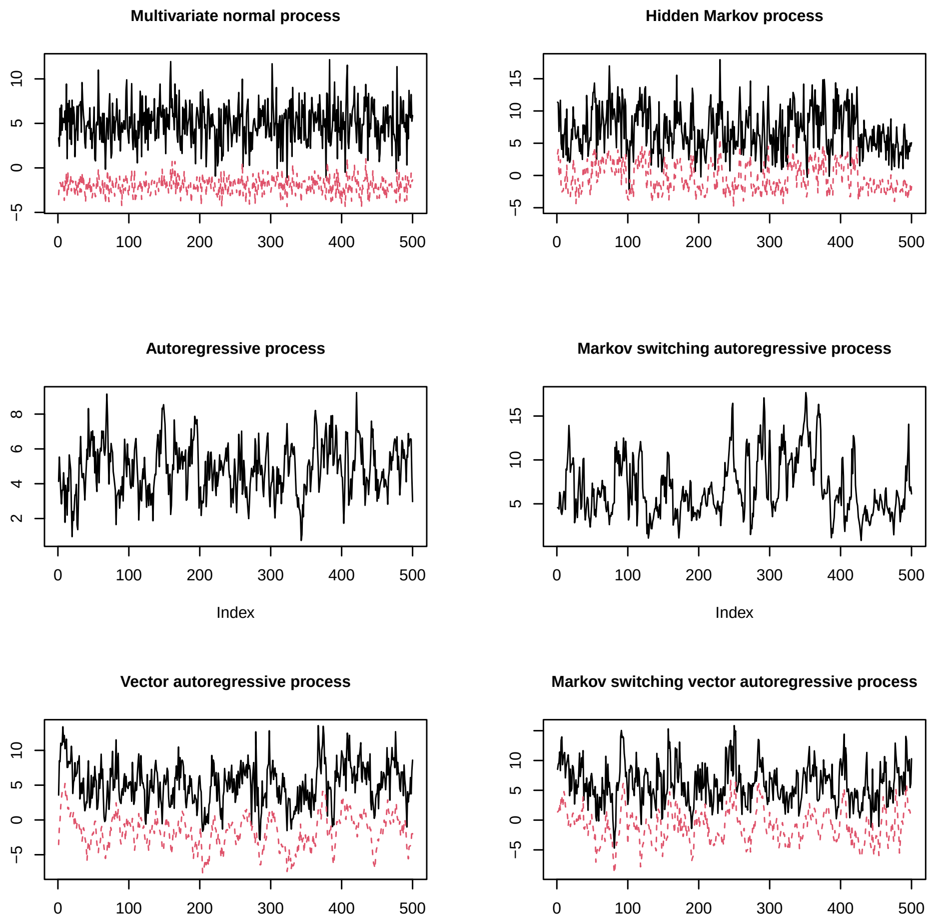

Figure 1: Simulated processes. Linear process on left and Markov switching process on right.

Model	Label	Equation
$\mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	`Nmdl`	$\boldsymbol{y}_t = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$
$\mathrm{AR}(p)$	`ARmdl`	$y_t = \mu + \sum_{k=1}^{p} \phi_k(y_{t-k} - \mu) + \sigma\epsilon_t$
$\mathrm{VAR}(p)$	`VARmdl`	$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k(\boldsymbol{y}_{t-k} - \boldsymbol{\mu}) + \boldsymbol{\Sigma}^{1/2}\boldsymbol{\epsilon}_t$
$\mathrm{MS\text{-}AR}(p)$	`MSARmdl`	$y_t = \mu_{s_t} + \sum_{k=1}^{p} \phi_k(y_{t-k} - \mu_{s_t}) + \sigma_{s_t}\epsilon_t$
$\mathrm{MS\text{-}VAR}(p)$	`MSVARmdl`	$\boldsymbol{y}_t = \boldsymbol{\mu}_{s_t} + \sum_{k=1}^{p} \boldsymbol{\Phi}_k(\boldsymbol{y}_{t-k} - \boldsymbol{\mu}_{s_{t-k}}) + \boldsymbol{\Sigma}_{s_t}^{1/2}\boldsymbol{\epsilon}_t$
HMM	`HMmdl`	$\boldsymbol{y}_t = \boldsymbol{\mu}_{s_t} + \boldsymbol{\Sigma}_{s_t}^{1/2}\boldsymbol{\epsilon}_t$

Table 3: Models and their specifications available in the R package **MSTest**

4.3 Model estimation

```
R> # Estimate model
R> mdl_est_norm <- Nmdl(simu_norm[["y"]])
R> mdl_est_norm

Normally Distributed Model
```

```
          coef     s.e.
mu_1    5.00020 0.100940
mu_2   -1.98240 0.044126
sig_11  5.09410 0.322830
sig_12  1.52100 0.120850
sig_22  0.97354 0.061695

log-likelihood =  -1661.175
AIC =  3332.35
BIC =  3291.277

R> # Set options for model estimation
R> control <- list(const  = TRUE,
+                  getSE  = TRUE)
R> # Estimate model
R> mdl_est_ar <- ARmdl(simu_ar[["y"]], p = 1, control)
R> mdl_est_ar

Autoregressive Model
         coef     s.e.
mu    4.86110 0.162830
sig   1.04320 0.066175
phi_1 0.71919 0.031227

log-likelihood =  -718.0947
AIC =  1442.189
BIC =  1417.552

R> # Set options for model estimation
R> control <- list(const  = TRUE,
+                  getSE  = TRUE)
R> # Estimate model
R> mdl_est_var <- VARmdl(simu_var[["y"]], p = 1, control)
R> mdl_est_var

Vector Autoregressive Model
             coef     s.e.
mu_1      5.24650 0.418380
mu_2     -1.63710 0.404700
sig_11    4.70470 0.298450
sig_12    1.34530 0.113490
sig_22    0.97581 0.061902
phi_1,11  0.44985 0.054058
phi_1,12  0.34717 0.068028
phi_1,21  0.19076 0.024619
phi_1,22  0.70559 0.030982

log-likelihood =  -1670.295
```

```
AIC =  3358.591
BIC =  3284.677

R> # Set options for model estimation
R> control <- list(msmu   = TRUE,
+                  msvar  = TRUE,
+                  method = "EM",
+                  use_diff_init = 10)
R> # Estimate model
R> mdl_est_msar <- MSARmdl(simu_msar[["y"]], p = 1, k = 2, control)
R> mdl_est_msar

Markov Switching Autoregressive Model
           coef     s.e.
mu_1  10.174000 0.316780
mu_2   4.786000 0.210500
sig_1  3.075900 0.362110
sig_2  0.875460 0.078859
phi_1  0.762070 0.027317
p_11   0.923750 0.078638
p_12   0.076248 0.019789
p_21   0.044319 0.011521
p_22   0.955680 0.057576

log-likelihood =  -902.9518
AIC =  1823.904
BIC =  1749.99

R> # Change estimation method
R> control\$method <- "MLE"
R> control\$init_theta <- mdl_est_msar\$theta
R> control\$mle_theta_low <- c(0,0,0.1,0.1,-0.90,0.02,0.02,0.02,0.02)
R> control\$mle_theta_upp <- c(20,20,20,20,0.90,0.98,0.98,0.98,0.98)
R> # Estimate model
R> mdl_est_msar_mle <- MSARmdl(simu_msar[["y"]], p = 1, k = 2, control)
R> mdl_est_msar_mle

Markov Switching Autoregressive Model
          coef     s.e.
mu_1  9.871100 0.318360
mu_2  4.814500 0.211880
sig_1 3.024200 0.349120
sig_2 0.865760 0.078514
phi_1 0.764450 0.027046
p_11  0.911690 0.078361
p_12  0.088313 0.022572
p_21  0.051667 0.013216
p_22  0.948330 0.058185
```

```
log-likelihood =  -901.7898
AIC =  1821.58
BIC =  1747.666

R> # Set options for model estimation
R> control <- list(msmu   = TRUE,
+                  msvar  = TRUE,
+                  method = "EM",
+                  use_diff_init = 10)
R> # Estimate model
R> mdl_est_msvar <- MSVARmdl(simu_msvar[["y"]], p = 1, k = 2, control)
R> mdl_est_msvar

Markov Switching Vector Autoregressive Model
              coef     s.e.
mu_1,1    4.890200 0.315090
mu_2,1   -2.023500 0.347620
mu_1,2    8.892100 0.368430
mu_2,2    1.413000 0.382430
sig_11,1  4.643900 0.370050
sig_12,1  1.401100 0.149540
sig_22,1  1.059800 0.085850
sig_11,2  6.445000 0.786710
sig_12,2  2.387700 0.359030
sig_22,2  1.573300 0.221670
phi_1,11  0.375830 0.056838
phi_1,12  0.306690 0.068385
phi_1,21  0.152740 0.027689
phi_1,22  0.735630 0.033098
p_11      0.940940 0.049999
p_12      0.059055 0.013463
p_21      0.138560 0.031708
p_22      0.861440 0.047686

log-likelihood =  -1843.327
AIC =  3722.654
BIC =  3574.828

R> # Set options for model estimation
R> control <- list(msmu   = TRUE,
+                  msvar  = TRUE,
+                  method = "EM",
+                  use_diff_init = 10)
R> # Estimate model
R> mdl_est_hmm <- HMmdl(simu_hmm[["y"]], k = 2, control)
R> mdl_est_hmm
```

```
Hidden Markov Model
               coef     s.e.
mu_1,1     9.801200 0.188920
mu_2,1     1.994300 0.094119
mu_1,2     4.972500 0.138300
mu_2,2    -1.977700 0.060805
sig_11,1  6.808900 0.717760
sig_12,1  2.469300 0.321760
sig_22,1  1.603100 0.178630
sig_11,2  5.387200 0.459070
sig_12,2  1.480400 0.170530
sig_22,2  1.006400 0.088702
p_11      0.888340 0.066222
p_12      0.111660 0.023834
p_21      0.076838 0.016780
p_22      0.923160 0.055813

log-likelihood =  -1892.784
AIC =  3813.569
BIC =  3698.564
```

4.4 Hypothesis testing

In this section, we describe how to use each function available in the **MSTest** package. It serves as a sort of manual for the package and should facilitate the use of it. Throughout the different testing procedures, we are mainly interested in understanding whether the distribution of the process of interest $y_t = (y_1, ...y_T)$ can be characterized by two different regimes with statistically different means, variances. In some cases we are interested in the conditional density of y_t conditional on autoregressive components or explanatory variables in which case their coefficients may also vary by state. As a result, all tests in the **MSTest** package only requires the user to provide the variable y_t (and X when relevant) and can choose further specifications through function inputs.

Table 1 shows all the functions currently available in the MSTest package, briefly describes them and shows each test's possible inputs and default values. We provide further detail on all inputs for each testing procedure next. All functions return the test-statistic, critical-values of the test-statistic distribution, the p-value and parameter estimates under the null when appropriate. It is important to note that, even though they are called critical-values, the values returned by **HLRtest()** are the critical-values of the process Q discussed in Hansen, 1992 which provide a bound for the LR but is not the critical-values of the LR test for Markov switching. Likewise, values returned by **DLMCtest()** and **DLMMCtest()** described as critical-values are in fact the percentiles of the p-value cumulative distribution after the respective method of combining the individual moment p-values.

4.4.1 Monte Carlo likelihood ratio test

Local Monte Carlo Likelihood ratio test

```
R> # Set options for testing procedure
R> lmc_control = list(N = 99,
```

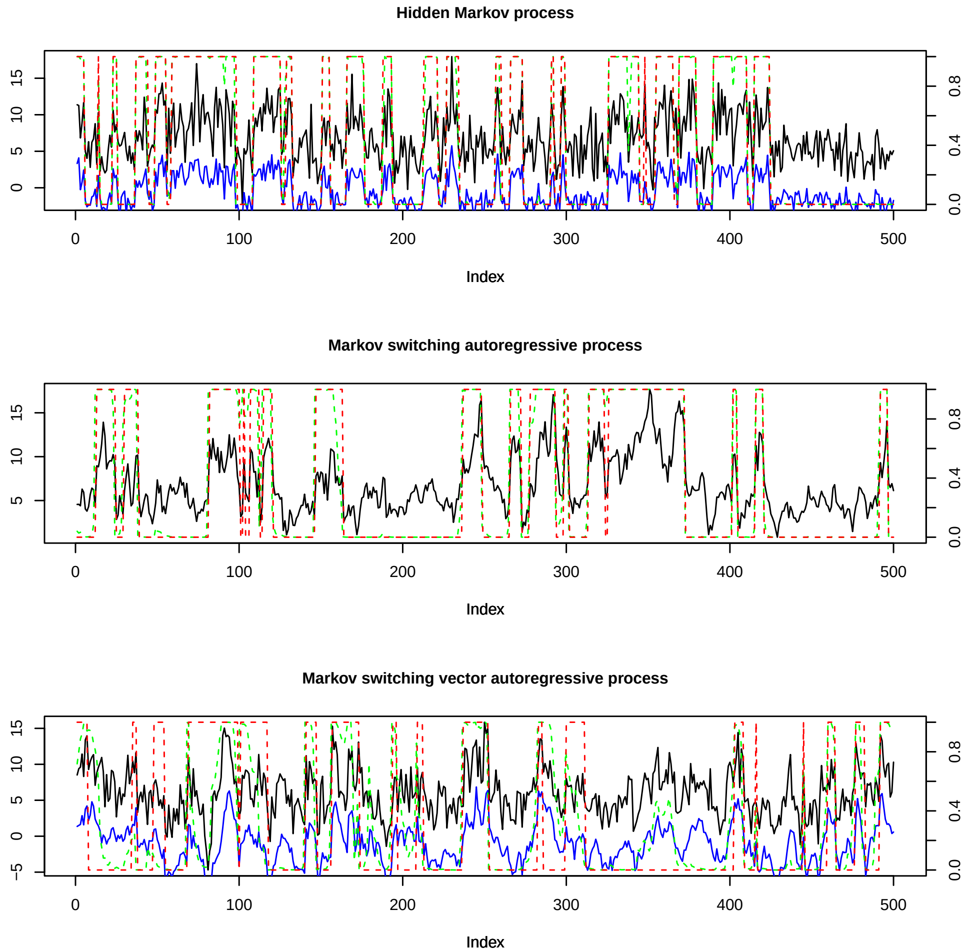

Figure 2: Simulated processes (black and blue), true regime states (red - dashed), and model estimated smoothed probabilities (green - dashed).

```
+                       mdl_h0_control = list(const  = TRUE,
+                                             getSE  = TRUE),
+                       mdl_h1_control = list(msmu   = TRUE,
+                                             msvar  = TRUE,
+                                             getSE  = TRUE,
+                                             method = "EM",
+                                             use_diff_init = 10),
+                       use_diff_init_sim = 5)
R> # Perform Rodriguez-Rondon & Dufour (2023) LMC-LRT
R> lmclrt <- LMCLRTest(simu_msar[["y"]], p = 1, k0 = 1 , k1 = 2, lmc_control)
R> lmclrt

Restricted Model
         coef     s.e.
mu    6.84040 0.520710
sig   3.14660 0.199610
```

Function	Description
`LMC-LRT`	Local Monte Carlo Likelihood Ratio test proposed in Rodriguez-Rondon and Dufour, 2023a
`MMC-LRT`	Maximized Monte Carlo Likelihood Ratio test proposed in Rodriguez-Rondon and Dufour, 2023a
`DLMCtest`	Dufour and Luger, 2017 moment-based test for Markov switching autoregressive models. Extended to more general ARMA models and models with explanatory variables.
`DLMMCtest`	Dufour and Luger, 2017 Maximized Monte Carlo (MMC) version of the moment-based test. This function allows the user to set some explanatory variables as nuisance parameters.
`CHPtest`	Carrasco et al., 2014 optimal test for Markov switching parameters.
`CHPMCtest`	Monte Carlo Test version of Carrasco et al., 2014 optimal test for Markov switching parameters.
`HLRtest`	Hansen, 1992 likelihood ratio test. Uses empirical process theory to estimate asymtptotic distribution of bound for LR test.

Table 4: Available tests in the R package **MSTest**

```
phi_1 0.84749 0.023751

log-likelihood =  -993.5581
AIC =  1993.116
BIC =  1968.478

Unrestricted Model
           coef      s.e.
mu_1  10.174000 0.316780
mu_2   4.786000 0.210500
sig_1  3.075900 0.362110
sig_2  0.875460 0.078859
phi_1  0.762070 0.027317
p_11   0.923750 0.078638
p_12   0.076248 0.019789
p_21   0.044319 0.011521
p_22   0.955680 0.057576

log-likelihood =  -902.9518
AIC =  1823.904
BIC =  1749.99

Rodriguez-Rondon & Dufour (2023) Local Monte Carlo Likelihood Ratio Test
         LRT_0  0.90%  0.95%  0.99% p-value
LMC_LRT 181.21 5.7014 6.5639 8.4872    0.01
```

Maximized Monte Carlo Likelihood ratio test

```
R> # Set options for testing procedure
R> mmc_control = list(N = 99,
```

```
+                      eps = 0.3,
+                      threshold_stop = 0.05 + 1e-6,
+                      type = "pso",
+                      workers  = 11,
+                      CI_union = FALSE,
+                      phi_low = -0.99,
+                      phi_upp = 0.99,
+                      mdl_h0_control = list(const  = TRUE,
+                                            getSE  = TRUE),
+                      mdl_h1_control = list(msmu   = TRUE,
+                                            msvar  = TRUE,
+                                            getSE  = TRUE,
+                                            method = "EM"),
+                      type_control   = list(maxit  = 100,
+                                            maxf   = 1000))
R> # start cluster
R> doParallel::registerDoParallel(mmc_control[["workers"]])
R> # Perform Rodriguez-Rondon & Dufour (2023) MMC-LRT
R> mmclrt <- MMCLRTest(simu_ar[["y"]], p = 1, k0 = 1 , k1 = 2, mmc_control)
S=13, K=3, p=0.2135, w0=0.7213, w1=0.7213, c.p=1.193, c.g=1.193
v.max=NA, d=1.023, vectorize=FALSE, hybrid=off
Converged
R> mmclrt

Restricted Model
         coef     s.e.
mu    4.86110 0.162830
sig   1.04320 0.066175
phi_1 0.71919 0.031227

log-likelihood =  -718.0947
AIC =  1442.189
BIC =  1417.552

Unrestricted Model
          coef     s.e.
mu_1  4.985300 1.294900
mu_2  4.837700 0.320620
sig_1 0.473810 0.606380
sig_2 1.173400 0.101650
phi_1 0.713050 0.057719
p_11  0.033205 2.296700
p_12  0.966800 2.288700
p_21  0.240330 0.571260
p_22  0.759670 0.569630

log-likelihood =  -717.4822
AIC =  1452.964
```

```
BIC =  1379.051

Rodriguez-Rondon & Dufour (2023) Maximized Monte Carlo Likelihood Ratio Test
        LRT_0 p-value
MMC_LRT 1.2251    0.73
R> # stop cluster
R> doParallel::stopImplicitCluster()
```

We could continue searching for the parameter values under the null that give the maximum p-value but for the purpose of exposition we set the `threshold` parameter to be $0.05 + 1e - 6$ so that we stop searching once the test fails to reject the null hypothesis.

4.4.2 Moment-based tests

The Monte Carlo version of the moment-based test proposed by Dufour and Luger, 2017 is called by using **DLMCtest()**. Like all other functions, it requires Y to be specified by the user. However, unlike the other tests, this test is based on the errors of the process. As a result, the number of autoregressive components is not required and is set to NULL by default. Users may even choose a number of moving average components q which is not possible in the other tests. This is another extension provided by **MSTest** as in Dufour and Luger, 2017 the authors are mainly concerned with autoregressive models only. Additionally, we extend this model to accept independent explanatory variables x. This variable can be provided by the user and must be a matrix or vector with each variable in a different column. This variable is set to NULL as default and so is not required of the user unless needed. The variable N can be used to set the number of Monte Carlo replications the user wishes to use. The default value for N is 100 as Dufour et al., 2004 show that having more than 100 replications has a small affect on power. This test also involves approximating the distribution of the p-value for each statistic through a separate round of simulation. The number of replications used in this approximation is determined by $N2$ and it is set to $10,000$ by default. For example, if Y is the U.S. GNP % change data, we could test Hamilton, 1989 AR(4) model by simply running the following command:

Moment-based local Monte Carlo test

```
R> # Set options for testing procedure
R> lmc_control = list(N = 99,
+                     simdist_N = 10000,
+                     getSE = TRUE)
R> # Perform Dufour & Luger (2017) LMC test
R> lmcmoment <- DLMCTest(simu_msar[["y"]], p = 1, control = lmc_control)
R> lmcmoment

Restricted Model
        coef     s.e.
mu    6.84040 0.520710
sig   3.14660 0.199610
phi_1 0.84749 0.023751

log-likelihood =  -993.5581
AIC =  1993.116
BIC =  1968.478
```

```
Dufour & Luger (2017) Moment-Based Local Monte Carlo Test
           phi_1   M(e)   V(e)     S(e)   K(e) F(e)    0.90%   0.95%   0.99% p-value
LMC_min  0.84749 1.4557 14.757 0.054697 1.9073    1 0.98073 0.99012 0.99566    0.01
LMC_prod 0.84749 1.4557 14.757 0.054697 1.9073    1 0.99873 0.99956 0.99993    0.01
```

Moment-based maximized Monte Carlo test As previously discussed, the test proposed by Dufour and Luger, 2017 has the advantage of being computationally efficient, which makes treating explanatory parameters as nuisance parameters and optimizing with respect to these nuisance parameter using the MMC approach discussed in Dufour, 2006 much more tractable. The **DLMMCtest()** command in **MSTest** allows the user to use the MMC version if the moment-based test proposed in Dufour and Luger, 2017. The $N3$ variable which was not available in the regular MC approach is used to determine the number of points to consider in the MMC procedure. The **searchType** variable can be used to choose a gridsearch when set to $searchType = "gridSearch"$. By default, a gridsearch will search over the interval $[-1, 1]$ for autoregressive parameters and explanatory variables. Users are require to change these limits using **optimOptions** if desired. This can be done for the interval $[-5, 5]$ by using the following command:

```
R> # Set options for testing procedure
R> mmc_control <- list(N = 99,
+                     getSE = TRUE,
+                     eps = 1e-9,
+                     CI_union = TRUE,
+                     phi_low = -0.99,
+                     phi_upp = 0.99,
+                     optim_type = "GenSA",
+                     threshold_stop = 0.05 + 1e-6,
+                     type_control = list(maxit = 300))
R> # Perform Dufour & Luger (2017) MMC test
R> mmcmoment <- DLMMCTest(simu_msar[["y"]], p = 1, control = mmc_control)
R> mmcmoment

Restricted Model
         coef     s.e.
mu    6.84040 0.520710
sig   3.14660 0.199610
phi_1 0.84749 0.023751

log-likelihood =  -993.5581
AIC =  1993.116
BIC =  1968.478

Dufour & Luger (2017) Moment-Based Maximized Monte Carlo Test
           M(e)   V(e)     S(e)   K(e) F(e) p-value
MMC_min  1.4557 14.757 0.054697 1.9073    1    0.01
MMC_prod 1.4557 14.757 0.054697 1.9073    1    0.01
```

We again set the `threshold` parameter to be $0.05 + 1e - 6$ so that we stop searching once the test fails to reject the null hypothesis and this time also set `eps` to be a very small number so that it is not theoretically used and the search is performed over the C.I. as in Dufour and Luger, 2017.

The user can also choose to do a stochastic search within 2 standard deviations of the parameter estimates as in the empirical example in Dufour and Luger, 2017, by setting $searchType$ = "$randSearch\_paramCI''$. The two optimization methods described so far, only consider parameter values which allow the process to still be stationary. Other available optimization algorithms include: Simulated Annealing using the "GenSA" package introduced by Yang Xiang et al., 2013, Particle Swarm algorithm using the "PSO" package introduced byZambrano-Bigiarini et al., 2013 and Genetic algorithm using the "GA" introduced by Scrucca et al., 2013. These algorithms do not discriminate between stationary and non-stationary models but future versions of **MSTest** may include the option to constrain the optimization to stationary models in future versions as part of the **optimOptions** function. These optimization algorithms can be used by setting $searchType$ = "$GenSA$", "PSO", or "GA".

4.4.3 Parameter stability test

The test proposed by CHP is called by using **CHPtest()** where again Y must be provided by the user. The user may also again specify the number of autoregressive components they would like to consider by setting p equal to that number. By default, it is set to one. N, the number of bootstraps is set to 3000 and ρ, the bound for one of the nuisance parameters, is set to 0.7 as in CHP. By setting $\rho = 0.7$ the gridsearch occurs over the parameter space $\rho \in [-0.7, 0.7]$. The user can set the number of bootstraps to 1000 and the bound for $\rho = 0.9$ by running:

```
R> # Set options for testing procedure
R> chp_control = list(N = 1000,
+                     rho_b = 0.7,
+                     msvar = TRUE)
R> # Perform Carrasco, Hu, & Ploberger (2014) test
R> pstabilitytest <- CHPTest(simu_ar[["y"]], p = 1, control = chp_control)
R> pstabilitytest

Restricted Model
         coef     s.e.
mu    4.86110 0.162830
sig   1.04320 0.066175
phi_1 0.71919 0.031227

log-likelihood =  -718.0947
AIC =  1442.189
BIC =  1417.552

Carrasco, Hu, & Ploberger (2014) Parameter Stability Test
- Switch in Mean and Variance
      test-stat  0.90%  0.95%  0.99% p-value
supTS   0.68188 2.5745 3.2278 4.7506   0.678
expTS   0.62505 1.2230 1.5283 2.1055   0.560
```

MSTest also includes a Monte Carlo Test version of the test propose by CHP. This MC version of the CHP test is meant provide a computationally efficient complement to their original test and has the same default settings as **CHPtest()** described above. This version of the test is called by using **CHPMCtest()**. The inputs and output are the same as before and only the methodology for obtaining the null distribution change internally and so we may expect that the values of the outputs will change accordingly.

4.4.4 Stochastic likelihood ratio test

In **MSTest**, the test proposed by Hansen, 1992 is called by using the command **HLRtest()**. This function requires the user to provide Y, the variable of interest. This test was developed to test the null hypothesis of linearity in autoregressive models. As such, a value for the number of autoregressive components to include must be provided. By default, this value is set to one. The following three inputs are indicator variables taking the value of 1 when we want to allow the mean or variance to switch between regimes and bound q by $q = 1 - p$ and 0 otherwise. The **meanS** variable is by default set to one. If the user would like to test the alternative hypothesis with regime switching mean and variance he must set **varS** = 1 by using the command

```
R> # Set options for testing procedure
R> hlrt_control  <- list(ix          = 1,
+                        msvar       = TRUE,
+                        gridsize    = 5,
+                        p_gridsize  = 9,
+                        p_stepsize  = 0.1,
+                        mugrid_from = 0,
+                        mugrid_by   = 1)
R> # Perform Hansen (1992) likelihood ratio test
R> hlrt <- HLRTest(simu_msar[["y"]], p = 1, control = hlrt_control)
R> hlrt

Restricted Model
         coef     s.e.
mu    6.84040 0.520710
sig   3.14660 0.199610
phi_1 0.84749 0.023751

log-likelihood =  -993.5581
AIC =  1993.116
BIC =  1968.478

Hansen (1992) Likelihood Ratio Bound Test -  Switch in Mean and Variance
      test-stat 0.90 % 0.95 % 0.99 % p-value
M = 0    7.4246 2.9242 3.1342 3.5934       0
M = 1    7.4246 2.9303 3.2393 3.7130       0
M = 2    7.4246 2.9806 3.3019 3.7690       0
M = 3    7.4246 3.0196 3.2675 3.7443       0
M = 4    7.4246 3.0530 3.3661 4.0110       0
```

Since in Hansen, 1992 these parameters are treated as nuisance parameters the user should note that setting $varS = 1$ will require more time to compute as the function must optimize over a larger parameter space. **gridSize** determines the parameter space over which we search. The grid begins from 0.1 and increases by 0.1 up to the number of points specified by **gridSize**. It is set to 20

Test procedure	1951Q2-1984Q4		1951Q2-2010Q4		1951Q2-2022Q3	
	test-stat	p-value	test-stat	p-value	test-stat	p-value
Panel A: Change in mean						
LMC-LRT	4.73	0.33	15.67	0.01	105.38	0.01
MMC-LRT	0.26	0.82	0.05	0.79	52.45	0.01
supTS	0.03	0.61	0.81	0.18	0.90	0.13
expTS	0.44	0.77	0.61	0.31	0.76	0.27
H-LRT	2.13	0.34	1.85	0.56	2.37	0.19
Panel B: Change in mean & variance						
LMC-LRT	9.04	0.23	50.14	0.01	196.93	0.01
MMC-LRT	7.41	0.45	50.14	0.01	196.93	0.01
LMC_{min}	0.82	0.63	1.00	0.01	1.00	0.01
LMC_{prod}	0.96	0.68	1.00	0.01	1.00	0.01
MMC_{min}	0.35	1.00	1.00	0.01	1.00	0.01
MMC_{prod}	0.77	0.99	1.00	0.02	1.00	0.01
supTS	1.86	0.26	15.27	0.00	37.41	0.00
expTS	0.95	0.25	355.07	0.00	$5.71\text{x}10^{10}$	0.00
H-LRT	1.36	1.00	5.19	0.00	10.95	0.00

Table 5: Hypothesis test results with data sets available in **MSTest**.

by default so that the we search for $\mu_2, \sigma_2 \in [0.1, 2]$ and $\phi_{i,2} \in [-1, 1]$ divided into **gridSize** points. The last input N determines the number of bootstrap replications the test should use to compute the p-values. As in Hansen, 1992, this value is set to 1000 by default.

5 Empirical example

For the empirical section of this paper, we apply all the tests currently available in **MSTest** to the three data sets also available in **MSTest** described above. Specifically, we will apply the test procedures to the growth rates of U.S. GNP. According to Kim and Nelson, 1999, we should find that a model with two regimes should be able to capture the structural decline in the volatility of the business cycle fluctuations that began in the mid 1980s. This structural change has also been named the *Great Moderation*. This makes the growth rate of U.S. GNP a good data set to use as an empirical example for the tests provided by **MSTest**. If the statement above is true and we allow the variance to change according to the regime, we should find that our tests fail to reject the null hypothesis of one regime (i.e., a linear model) when using data from the first sample since this entire sample belongs to the period marked with higher volatility. The second sample however, includes more data from both the high volatility period and the subsequent period with lower levels of volatility. As a result, if we allow the variance to change according the the regime, we should find that our test would reject the null hypothesis of a linear model in favour of the non-linear Markov switching model with 2 regimes. We also consider the third data set which extends further and ranges from 1951Q2 to 2022Q3. Here, we should also reject the null hypothesis of a linear model. As discussed in Rodriguez-Rondon and Dufour, 2023a we should also find that the high volatility period returns following the recession induced by the COVID-19 pandemic. Additionally, we also consider the same test procedures but only allowing the mean to change according to the regime. Table 5 provides the results of the test procedures for all three samples. The first panel shows the results when only the mean is allowed to change. Here we see that all tests fail to reject the null hypothesis of a linear model when considering the first two samples. On the other hand, they all reject the null hypothesis when considering the third, larger sample. The second panel shows results when both the mean and the variance are allowed to change.

Although it would be interesting to further test for the presence of a third regime in the second and third sample, this is done in Rodriguez-Rondon and Dufour, 2023a and since the authors used the **MSTest** package to generate their results, we direct the interested reader to their paper for such results.

6 Conclusion

The importance of testing the number of regimes in Markov switching models inspired various contributions, most of which propose different methods for dealing with the statistical and computational difficulties that plague this problem. Notable contributions include Hansen, 1992, Carrasco et al., 2014 and Dufour and Luger, 2017 for testing the null hypothesis of one regime (i.e., a linear model) against the alternative of two regimes. More recently Rodriguez-Rondon and Dufour, 2023a proposed a set of Monte Carlo test procedures that can be used to test a null hypothesis of M_0 regimes against an alternative of $M_0 + m$ regimes where both $M_0 \geq 1$ and $m \geq 1$. The R package **MSTest** makes available the test procedures in these four studies. This paper provided a review of these test procedures and described how the **MSTest** package can be used to implement them as well as the other features of the package such as simulation and model estimation. Hence, the purpose of **MSTest** is to provide researchers with a ready to use R package that allows them to implement some of these tests procedures to facilitate making inferences and deciding an appropriate model specification for their data.

Computational details

The results in this paper were obtained using R 4.2.2 R Core Team, 2022 with the packages **MSTest** 0.1.2 Rodriguez-Rondon and Dufour, 2023b, **Rcpp** 1.0.10 Eddelbuettel and François, 2011, **RcppArmadillo** 0.11.4.3.1 Eddelbuettel and Sanderson, 2014, **GenSA** 1.1.7 Yang Xiang et al., 2013, **pso** 1.0.4 Bendtsen., 2022, **foreach** 1.5.2 Microsoft and Weston, 2022, and **doParallel** 1.0.17 Corporation and Weston, 2022. Computations were performed on Apple macOS Ventura Version 13.2.1 10 x86_64-apple-darwin17.0/x64 (64-bit) with Intel(R) Core i9 2.30 GHz. Code for the computations is available in the R script article.R, available in the GitHub repository at https://github.com/roga11/MSTest. R itself and all packages used are available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/.

Acknowledgments

This work was supported by the Fonds de recherche sur la société et la culture Doctoral Research Scholarships (B2Z).

References

An, Y., Hu, Y., Hopkins, J., & Shum, M. (2013). *Identifiability and inference of hidden markov models* (tech. rep.). Technical report.

Andrews, D. W. (1999). Estimation when a parameter is on a boundary. *Econometrica*, *67*(6), 1341–1383.

Andrews, D. W. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, *69*(3), 683–734.

Andrews, D. W., & Ploberger, W. (1994a). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383–1414.

Andrews, D. W., & Ploberger, W. (1994b). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica: Journal of the Econometric Society*, 1383–1414.

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

Baum, L. E., & Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The annals of mathematical statistics*, *37*(6), 1554–1563.

Bendtsen., C. (2022). *Pso: Particle swarm optimization* [R package version 1.0.4]. https://CRAN.R-project.org/package=pso

Carrasco, M., Hu, L., & Ploberger, W. (2014). Optimal test for markov switching parameters. *Econometrica*, *82*(2), 765–784.

Carter, A. V., & Steigerwald, D. G. (2012). Testing for regime switching: A comment. *Econometrica*, *80*(4), 1809–1812.

Cho, J.-S., & White, H. (2007). Testing for regime switching. *Econometrica*, *75*(6), 1671–1720.

Cho, J.-S., & White, H. (2011). Testing for regime switching: Rejoinder. *Unpublished Manuscrpt, University of California, San Diego*.

Corporation, M., & Weston, S. (2022). *Doparallel: Foreach parallel adaptor for the 'parallel' package* [R package version 1.0.17]. https://CRAN.R-project.org/package=doParallel

Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, *64*(2), 247–254.

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

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B*, *39*(1), 1–38.

Dufour, J.-M. (2006). Monte carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics. *Journal of Econometrics*, *133*(2), 443–477.

Dufour, J.-M., Khalaf, L., Bernard, J.-T., & Genest, I. (2004). Simulation-based finite-sample tests for heteroskedasticity and arch effects. *Journal of Econometrics*, *122*(2), 317–347.

Dufour, J.-M., Khalaf, L., & Voia, M. (2014). Finite-sample resampling-based combined hypothesis tests, with applications to serial correlation and predictability. *Communications in Statistics-Simulation and Computation*, *44*(9), 2329–2347.

Dufour, J.-M., & Luger, R. (2017). Identification-robust moment-based tests for markov switching in autoregressive models. *Econometric Reviews*, *36*(6-9), 713–727.

Dufour, J.-M., & Neves, J. (2019). Finite-sample inference and nonstandard asymptotics with monte carlo tests and r. In *Handbook of statistics* (pp. 3–31, Vol. 41). Elsevier.

Dufrénot, G., Mignon, V., & Péguin-Feissolle, A. (2011). The effects of the subprime crisis on the latin american financial markets: An empirical assessment. *Economic Modelling*, *28*(5), 2342–2357.

Eddelbuettel, D., & François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, *40*(8), 1–18. https://doi.org/10.18637/jss.v040.i08

Eddelbuettel, D., & Sanderson, C. (2014). Rcpparmadillo: Accelerating r with high-performance c++ linear algebra. *Computational Statistics and Data Analysis*, *71*, 1054–1063. http://dx.doi.org/10.1016/j.csda.2013.02.005

Fisher, R. A. (1932). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, 763–788.

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

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27–62.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics*, *45*(1-2), 39–70.

Hamilton, J. D. (2005). *What's real about the business cycle?* (Tech. rep.). National Bureau of Economic Research.

Hamilton, J. D. (2016). Macroeconomic regimes and regime shifts. In *Handbook of macroeconomics* (pp. 163–201, Vol. 2). Elsevier.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, *64*(1-2), 307–333.

Hamilton, J. D. (1994). *Time series analysis*. Princeton university press.

Hansen, B. E. (1992). The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp. *Journal of applied Econometrics*, *7*(S1), S61–S82.

Hansen, B. E. (1996a). Erratum: The likelihood ratio test under nonstandard conditions: Testing the markov switching model of gnp. *Journal of Applied econometrics*, *11*(2), 195–198.

Hansen, B. E. (1996b). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, 413–430.

Hu, L., & Shin, Y. (2008). Optimal test for markov switching garch models. *Studies in Nonlinear Dynamics & Econometrics*, *12*(3).

Kahn, J. A., & Rich, R. W. (2007). Tracking the new economy: Using growth theory to detect changes in trend productivity. *Journal of Monetary Economics*, *54*(6), 1670–1701.

Kasahara, H., Okimoto, T., & Shimotsu, K. (2014). Modified quasi-likelihood ratio test for regime switching. *The Japanese Economic Review*, *65*(1), 25–41.

Kasahara, H., & Shimotsu, K. (2015). Testing the number of components in normal mixture regression models. *Journal of the American Statistical Association*, *110*(512), 1632–1645.

Kasahara, H., & Shimotsu, K. (2018). Testing the number of regimes in markov regime switching models. *arXiv preprint arXiv:1801.06862*.

Kim, C.-J., & Nelson, C. R. (1999). Has the us economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *Review of Economics and Statistics*, *81*(4), 608–616.

Krolzig, H.-M. (1997). The markov-switching vector autoregressive model. In *Markov-switching vector autoregressions* (pp. 6–28). Springer.

Liu, X., & Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, *31*(3), 807–832.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching garch models. *Studies in Nonlinear Dynamics & Econometrics*, *9*(4).

Marmer, V. (2008). Testing the null hypothesis of no regime switching with an application to gdp growth rates. *Empirical Economics*, *35*(1), 101–122.

Microsoft & Weston, S. (2022). *Foreach: Provides foreach looping construct* [R package version 1.5.2]. https://CRAN.R-project.org/package=foreach

Morley, J., & Piger, J. (2012). The asymmetric business cycle. *Review of Economics and Statistics*, *94*(1), 208–221.

Pearson, K. (1933). On a method of determining whether a sample of size n supposed to have been drawn from a parent population having a known probability integral has probably been drawn at random. *Biometrika*, *25*(3/4), 379–410.

Qu, Z., & Zhuo, F. (2021). Likelihood ratio-based tests for markov regime switching. *The Review of Economic Studies*, *88*(2), 937–968.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Rodriguez-Rondon, G., & Dufour, J.-M. (2023a). Monte carlo likelihood ratio tests for markov switching models. *Working paper*.

Rodriguez-Rondon, G., & Dufour, J.-M. (2023b). *Mstest: Hypothesis testing for markov switching models* [R package version 0.1.2.9000]. https://github.com/roga11/MSTest

Scrucca, L., et al. (2013). Ga: A package for genetic algorithms in r. *Journal of Statistical Software*, *53*(4), 1–37.

Tippett, L. H. C., et al. (1931). The methods of statistics. *The Methods of Statistics.*

Warne, A., & Vredin, A. (2006). Unemployment and inflation regimes. *Studies in Nonlinear Dynamics & Econometrics*, *10*(2).

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 1–25.

Wilkinson, B. (1951). A statistical consideration in psychological research. *Psychological bulletin*, *48*(2), 156.

Yang Xiang, Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized simulated annealing for efficient global optimization: The GenSA package for R. *The R Journal Volume 5/1, June 2013*. https://journal.r-project.org/archive/2013/RJ-2013-002/index.html

Zambrano-Bigiarini, M., Clerc, M., & Rojas, R. (2013). Standard particle swarm optimisation 2011 at cec-2013: A baseline for future pso improvements. *2013 IEEE Congress on Evolutionary Computation*, 2337–2344.