Monte Carlo Likelihood Ratio Tests for Markov Switching Models

Gabriel Rodriguez Rondon*† Jean-Marie Dufour‡

February 22, 2023

Abstract

Markov switching models have wide applications in economics, finance, and other fields. Most studies focusing on identifying the number of regimes in a Markov switching model have been limited to testing the null hypothesis of only one regime (i.e., a linear model with no switching) against an alternative hypothesis with two regimes. Even in such simple cases, this type of problem raises issues of nonstandard asymptotic distributions, identification failure, and nuisance parameters. In this paper, we propose Monte Carlo test methods [Dufour (2006)] which deal transparently with these distributional issues, even allowing for finite-sample inference. The procedure is applied to likelihood ratio statistics. The tests circumvent the issues plaguing conventional hypothesis testing. This also allows one to deal with non-stationary processes, models with non-Gaussian errors and multivariate settings, which have received little attention in the literature. An important contribution of this paper is the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT), which is an identifications-robust valid test procedure both in finite samples and asymptotically. Further, the methods proposed are applicable to more general settings where a null hypothesis with M_0 regimes is tested against an alternative with $M_0 + m$ regimes where both $M_0 \geq 1$ and $m \geq 1$. This allows one to compare different Markov switching models and Hidden Markov Models. Simulation results are provided for both univariate and multivariate settings and suggest the proposed tests are able to control the level of the test and have good power. Finally, an empirical application using U.S. GNP growth data suggests three regimes are appropriate for modeling the series from 1951Q2 - 2022Q3. Our results confirm evidence about the Great Moderation in the sense that two regimes have positive growth but experience a change (reduction) in variance in the mid-80s. The third regime is consistent with recessionary periods as it matches NBER suggested recession dates. Like Gadea, Gómez-Loscos, and Pérez-Quirós (2018) we find that the low volatility regime continues after the Great Recession but the high volatility period returns following the recent recession induced by the COVID-19 pandemic.

Key words: Hypothesis testing, Monte Carlo tests, Likelihood ratio, Markov switching, Hidden Markov Model, Nonlinearity, Regimes

JEL codes:

*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. 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 (MSM) were first introduced by Goldfeld and Quandt (1973) and later popularized by Hamilton (1989) as an alternative approach to modelling U.S. GNP growth. These models allow one to treat a series as a nonlinear process where the nonlinearity arises from discrete shifts. The process before and after a shift can be described as two separate regimes, and Hamilton (1989) describes these regimes as episodes where the behavior of the series is significantly different. Using U.S. GNP growth as an example, one regime can characterize a period of positive growth, while the other represents a period of negative growth due to recessions. Due to this flexibility, they have since been widely used in macroeconomics and finance. For example, MSMs have been applied to the identification of business cycles [Chauvet 1998; Chauvet and Hamilton 2006; Chauvet, Juhn, and Potter 2002; Diebold and Rudebusch 1996; Hamilton 1989; Kim and Nelson 1999; Qin and Qu 2021], interest rate dynamics (Garcia and Perron 1996), financial markets (Marcucci 2005), conditional heteroskedasticity models [Augustyniak 2014; Gray 1996; Haas, Mittnik, and Paolella 2004; Hamilton and Susmel 1994; Klaassen 2002], conditional correlations (Pelletier 2006) and identification of structural VAR models [Herwartz and Lütkepohl 2014; Lanne, Lütkepohl, and Maciejowska 2010; Lütkepohl et al. 2021] to name a few. More complete surveys of this literature include Hamilton (2010), Hamilton (2016) and Ang and Timmermann (2012). Applications of MSMs outside of the macroeconomic and financial literature include: environmental and energy economics [Cevik, Yıldırım, and Dibooglu 2021; Charfeddine 2017; Chevallier 2011], industrial organization (Resende 2008), health economics (Anser et al. 2021) and many others. An alternative but related model is the Hidden Markov Model (HMM). Like MSMs, HMMs are used to describe a process Y_t which depends on a latent Markov process S_t. However, HMMs depend only on S_t, which takes discrete values $\{1, \dots, M\}$ where M is the number of regimes. In contrast, as described by An et al. (2013), when the process Y_t also depends on lags of Y_t (*e.g.*, $\{Y_{t-1}, \dots, Y_{t-p}\}$), it is called a Hidden Markov-switching model, or simply a Markov-switching model. The dependence on past observations allows for more general interactions between Y_t and S_t, which can be used to model more complicated causal links between economic or financial variables of interest, so that MSMs are a generalization of the basic HMM. However, it is worth noting that HMMs have many applications including computational molecular biology [Baldi et al. 1994; Krogh, Mian, and Haussler 1994], handwriting and speech recognition [Jelinek 1997; Nag, Wong, and Fallside 1986; Rabiner and Juang 1986; Rabiner and Juang 1993], computer vision and pattern recognition (Bunke and Caelli 2001), and other machine learning applications.

An important issue with MSMs and HMMs is that the number of states or regimes must be determined *a priori*. Since the number of regimes is not always known, it is of interest to test the fit of a given model with a certain number of regimes (*e.g.*, M_0 regimes), against an alternative model with a different number of regimes (*e.g.*, $M_0 + m$ regimes). This highlights the importance of valid test procedures to determine the number of regimes in these types of models. However, standard hypothesis testing techniques are not easily applicable in this setting, because certain parameters of the model are unidentified under the null hypothesis, and usual regularity conditions needed to derive the asymptotic distribution of test statistics are not satisfied. The study of the asymptotic distribution of the likelihood ratio test for MSMs is a problem that has received a lot of attention [see Carter and Steigerwald 2012; Cho and White 2007; Garcia 1998; Hansen 1992; Kasahara and Shimotsu 2018; Qu and Zhuo 2021]. Most procedures focusing on the likelihood ratio test approach currently available are only able to deal with settings where the null hypothesis is that of a linear model (*i.e.*, $H_0 : M_0 = 1$) and the alternative hypothesis is a MSM with two regimes (*i.e.*, $H_1 : M_0+m = 2$, where $M_0 = m = 1$). The exception is Kasahara and Shimotsu (2018) which study the asymptotic distribution of the likelihood ratio test statistic when the null hypothesis is of a model with M_0 regimes and the alternative hypothesis is that of a model with $M_0 + 1$ regimes where $M_0 \geq 1$ (and $m = 1$). Interestingly, in this setting, the authors establish the asymptotic validity of the parametric bootstrap procedure (see Proposition 21 of Kasahara and Shimotsu 2018). Qu and Zhuo (2021) also show the asymptotic validity of the parametric bootstrap for specific data generating processes in the more simple setting where we would like to compare a linear model to a MSM with two regimes. At the same time, others have proposed alternative test procedures based on moments of least-squares residuals (see Dufour and Luger 2017), parameter stability (see Carrasco, Hu, and Ploberger 2014), or other moment-matching conditions (see Antoine et al. 2022).

In Carrasco, Hu, and Ploberger (2014), the authors are interested in the case of testing a linear model against a MSM with only two regimes. This is because, as a test of parameter stability, the null hypothesis must always be that of a linear model and in this case the test only has good power against local alternatives. As a result, this test cannot be used to compare different general MSMs (with $M_0 > 1$). Other limitations of the test procedures mentioned so far (with the exception of those proposed in Dufour and Luger (2017)) is that they are aimed at establishing an asymptotic distribution of the test statistic, so they depend on assumptions needed to obtain asymptotic results which may be restrictive in many cases. For example, a common assumption is that the process studied is stationary with Gaussian errors. Within the likelihood ratio test literature, it is also common to assume a concentrated parameter space, in order to avoid the parameter boundary problem. On the other hand, Dufour and Luger (2017) propose a valid test for the null hypothesis of a linear model against an alternative of a MSM with two regimes, which relies on Monte Carlo test techniques. Specifically, the authors propose four test statistics based on the moments of the least-squares residuals, which are meant to capture different characteristics of a two-component mixture distribution. Approximate marginal p-values are computed for each moment specific test statistic and are combined using either the minimum or the product. This leads to a single test statistic that is free of nuisance parameters when there are no autoregressive lags in the model (*i.e.*, HMMs). When autoregressive lags are included in the model, the test procedure is no longer free of nuisance parameters. In this case, the authors use the Local Monte Carlo (LMC) and Maximized Monte Carlo (MMC) test procedure described in Dufour (2006).

In this paper, we also use the Monte Carlo procedures described in Dufour (2006) to deal with the issues plaguing conventional testing procedures discussed so far, but in a likelihood ratio test setting. This also allows us to deal with nuisance parameters in the distribution of the likelihood ratio test statistic. Specifically, we propose the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT), which can be used to identify the number of regimes in both MSMs and HMMs in the more general case where we would like to compare models with M_0 regimes under the null hypothesis against models with $M_0 + m$ regimes under the alternative, where here both $M_0 \geq 1$ and $m \geq 1$. Since MSMs are more general than HMMs in the sense that we can recover a HMM by simply setting the number of autoregressive lags to zero, we focus on MSMs throughout this study but the results of the tests proposed here are also applicable to HMMs. An important contribution of the MMC-LRT proposed here is that it is an exact test for determining the number of regimes in a MSM and is valid both in finite samples and asymptotically. Further, since we are not working with the asymptotic distribution of the test statistic, we can also relax some of the assumptions typically required to obtain asymptotic results. There are four main advantages of using such a framework. The first is that the violation of the regularity conditions needed to drive an asymptotic distribution are no longer problematic. This means that we can consider the full nuisance parameter space rather than a concentrated parameter space as in Qu and Zhuo (2021) and Kasahara and Shimotsu (2018).

The second advantage is that this allows us to determine the appropriate number of regimes even when dealing with a non-stationary process Y_t. The third advantage is that we can deal with cases where the asymptotic distribution is more complicated to obtain or even infeasible, such as specific cases where the errors are non-Gaussian. Finally, the fourth advantage is that LMC-LRT and MMC-LRT can be applied to multivariate settings (*e.g.*, Markov-switching VAR models or multivariate HMM). It is also worth noting that non-stationary processes, non-Gaussian errors and multivariate settings have not received a lot of attention in the literature on hypothesis testing for the number of regimes in MSMs. Simulation results indicate that both the LMC-LRT and MMC-LRT procedures presented here are able to control the probability of a type I error as suggested by the theory proposed in Dufour (2006), and have better power than other tests proposed in the literature and considered here for comparison in the univariate setting. We are also the first to tabulate such simulation results in the multivariate setting. Another noteworthy contributions of this paper includes tabulating results of the test proposed by Dufour and Luger (2017) when the process Y_t is non-stationary. All tests results presented in this paper are obtained using the R package **MSTest** described in a companion paper Rodriguez Rondon and Dufour (2022).

The next sections are structured as follows. Section 2 reviews the Markov-switching autoregressive model we are interested in and briefly discusses estimation procedures. Section 3 introduces our testing methodology and reviews the testing procedures we use for comparison purposes. Section 4 shows and discusses simulation results for the size and power of the proposed testing procedures. Section 5 presents an empirical example where we use the testing procedures proposed here to identify the number of regimes when modelling U.S. GNP growth as originally considered in Hamilton (1989) and Hansen (1992). Specifically, we consider the sample from 1951Q2-2010Q4 used by Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) and also consider an extended sample covering the period of 1951Q2-2022Q3. When using this extended sample, we find that three regimes are needed for this sample. This result confirms evidence about the Great Moderation in the sense that two regimes have positive growth but experience a change in variance in the mid-80s. Specifically, we find that there is a lower variance regime starting after the mid-80s. The third regime that is consistent with recessionary periods as it matches NBER suggested recession dates. Finally, section 6 provides concluding remarks.

2 Markov-switching Model

In general, a MSM can be expressed as

$$y_t = x_t\beta + z_t\delta_{s_t} + \sigma_{s_t}\epsilon_t \tag{1}$$

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). For our testing procedures, other distributions on the error processes may also be assumed but for simplicity, here we assume normally distributed errors. Similarly, if interested in a multivariate setting, we could allow y_t to be a $(1\times q)$ vector (i.e., $y_t = (y_{1,t},\dots,y_{q,t})$) and ϵ_t also a $(1\times q)$ vector that is for example distributed as $\mathcal{N}(\mathbf{0},\Sigma_{s_t})$ or $\mathcal{N}(\mathbf{0},\Sigma)$ if the variance-covariance matrix does not depend on the latent Markov process S_t. In order to have a MSM, as described above, lags of y_t must be included in either x_t or z_t depending on whether we want to allow the autoregressive coefficients to depend on the regimes.

This general setting also allows us to consider a trend function within x_t or z_t. A HHM can also be recovered by considering only a constant term in z_t and excluding x_t. For the sake of exposition, in the following we consider a MSM where only the mean and the variance may be subject to change and autoregressive coefficient remain constant. That is, $x_t = (y_{t-1}, \dots, y_{t-p})$ and $z_t = 1$. We also reformulate the model to make the dependence on the mean, μ_{s_t}, more explicit. This results in the following model

$$y_t = \mu_{s_t} + \sum_{k=1}^{p} \phi_k (y_{t-k} - \mu_{s_{t-k}}) + \sigma_{s_t} \epsilon_t \tag{2}$$

where we can see that the mean and variance of the observed process y_t are governed by the latent Markov chain process S_t. As described in Hamilton (1994), for a model with M regimes, the one-step transition probabilities can be gathered into a transition matrix such as

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

where for example $p_{ij} = P(S_t = j|S_{t-1} = i)$ is the probability of state i being followed by state j. If for example, if we consider a model with only two regimes, we only need a (2×2) transition matrix to summarize the transition probabilities $\mathbf{P}$:

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

In either case, the columns of the transition matrix must sum to one in order to have a well defined transition matrix (i.e, $\sum_{j=1}^{M} = p_{ij} = 1$). We can also obtain the ergodic probabilities, $\pi = (\pi_1, \pi_2)'$, which are given by

$$\pi_1 = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \qquad\qquad \pi_2 = 1 - \pi_1 \tag{3}$$

in a setting with two regimes or 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 $\mathbf{I}_{M+1}$.

Continuing with the example of a MSM such as the one given by equation (2) and where $S_t = \{1, 2\}$ (i.e., $M = 2$ regimes), the sample log likelihood conditional on the first p observations of y_t is given by

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

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

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

where we let

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

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

Typically, MSMs are estimated using the Expectation Maximization (EM) algorithm (see Dempster, Laird, and Rubin 1977), Bayesian methods or through the use of the Kalman filter if using the state-space representation of the model. In very simple cases, MSMs 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. In this study, we use the EM algorithm when estimating MSMs. 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. We omit a detailed explanation of the EM algorithm as our focus is on the hypothesis testing procedures proposed here. For the interested reader, the estimation of a Markov switching model via the EM algorithm is describe in detail in Hamilton (1990) and Hamilton (1994) and the in Krolzig (1997) for the a Markov-switching VAR model.

3 Monte Carlo likelihood ratio tests

In this section, we introduce the Maximized Monte Carlo likelihood ratio test (MMC-LRT) and the Local Monte Carlo likelihood ratio test (LMC-LRT) obtained assuming normally distributed errors and a model of the form (2). For simplicity, 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. However, it is easy to see this methodology can be extended to more general cases with $M_0 \geq 1$ and $m \geq 1$. 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 LRT approach requires that we estimate the model both under the null and alternative hypothesis, so that we can obtain the log-likelihoods for each model. The log-likelihood for the model under the alternative (and under the null hypothesis if $M_0 > 1$) is given by (4) - (6):

$$L_T(\theta_1) = \log f(y_1^T \,|\, y_{-p+1}^0;\, \theta_1) = \sum_{t=1}^{T} \log f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_1) \tag{7}$$

where

$$\theta_1 = (\mu_1,\, \mu_2,\, \sigma_1,\, \sigma_2,\, \phi_1,\, \ldots\,,\, \phi_p,\, p_{11},\, p_{22})' \in \Omega\,. \tag{8}$$

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

$$L_T^0(\theta_0) = \log f(y_1^T \,|\, y_{-p+1}^0;\, \theta_0) = \sum_{t=1}^{T} \log f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_0) \tag{9}$$

where

$$f(y_t \,|\, y_{-p+1}^{t-1};\, \theta_0) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ \frac{-[y_t - \mu - \sum_{k=1}^{p} \phi_k (y_{t-k} - \mu)]^2}{2\sigma^2} \right\}, \tag{10}$$

$$\theta_0 = (\mu,\, \sigma^2, \phi_1,\, \ldots\, ,\, \phi_p)' \in \bar{\Omega}_0. \tag{11}$$

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

$$H_0 : \delta_1 = \delta_2 = \delta \quad \text{for some unknown } \delta = (\mu,\, \sigma)\,, \tag{12}$$

$$H_1 : (\delta_1,\, \delta_2) = (\delta_1^*,\, \delta_2^*) \;\; \text{for some unknown } \delta_1^* \neq \delta_2^*\;, \tag{13}$$

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

$$L_T^0(\theta_0) = L_T(\theta_1)\,, \quad \theta_1 \in \Omega_0, \tag{14}$$

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 is in fact completely determined by θ_0.

The likelihood ratio statistic for testing H_0 against H_1 can then written as

$$LR_T = 2[\bar{L}_T(H_1) - \bar{L}_T(H_0)] \tag{15}$$

where

$$\bar{L}_T(H_1) = \sup\{L_T(\theta_1) : \theta_1 \in \Omega\}\,, \tag{16}$$

$$\bar{L}_T(H_0) = \sup\{L_T^0(\theta_0) : \theta_0 \in \bar{\Omega}_0\} = \sup\{L_T(\theta_1) : \theta_1 \in \Omega_0\}\,. \tag{17}$$

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

$$LR(N,\, \theta_0) := [LR_T^{(1)}(\theta_0),\, \ldots\, ,LR_T^{(N)}(\theta_0)]', \qquad \theta_0 \in \bar{\Omega}_0\,. \tag{18}$$

Let us denote $LR_T^{(0)} := LR_T$ the test statistic based in the observed data. Given the model considered, we can assume [as in (4.10) of Dufour (2006)] that:

$$\begin{gathered}\text{the random variables } \; LR_T^{(0)},\, LR_T^{(1)}(\theta_0),\, \ldots\, ,LR_T^{(N)}(\theta_0) \; \text{ are exchangeable for some} \\ \theta_0 \in \bar{\Omega}_0, \;\; \text{each with distribution function } \; F[x \,|\, \theta_0]\,. \end{gathered} \tag{19}$$

Set

$$\hat{F}_N[x \,|\, \theta_0] := \hat{F}_N[x;\, LR(N,\, \theta_0)] = \frac{1}{N} \sum_{i=1}^{N} I[LR_T^{(i)}(\theta_0) \leq x] \tag{20}$$

$$\hat{G}_N[x \,|\, \theta_0] := \hat{G}_N[x;\, LR(N,\, \theta_0)] = 1 - \hat{F}_N[x;\, LR(N,\, \theta_0)] \tag{21}$$

where $I(C) := 1$ if condition C holds, and $I(C) = 0$ otherwise. $\hat{F}_N[x \,|\, \theta_0]$ is the sample distribution of the simulated statistics, and $\hat{G}_N[x \,|\, \theta_0]$ is the corresponding survival function. Then, the Monte Carlo p-value is given by

$$\hat{p}_N[x \,|\, \theta_0] = \frac{N\hat{G}_N[x \,|\, \theta_0] + 1}{N+1}\,. \tag{22}$$

Alternatively, using the relationship

$$\begin{aligned} R_{LR}[LR_T^{(0)};\, N] &= N\hat{F}_N[x;\, LR(N,\, \theta_0)] \\ &= \sum_{i=1}^{N} I[LR_T^{0} \geq LR_T^{i}(\theta_0)] \end{aligned} \tag{23}$$

we can define a Monte Carlo p-value as

$$\hat{p}_N[x \,|\, \theta_0] = \frac{N + 1 - R_{LR}[LR_T^{(0)};\, N]}{N+1} \tag{24}$$

where, as can be seen from (23), $R_{LR}[LR_T^{(0)}; N]$ simply computes the rank of the test statistic using the observed data within the generated series $LR(N, \theta_0)$. As discussed in Dufour (2006), a critical region with level α is then given by

$$\sup_{\theta_0 \in \bar{\Omega}_0} \hat{p}_N[LR_T^{(0)} \,|\, \theta_0] \leq \alpha \tag{25}$$

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

$$\mathbb{P}\left[\sup\{\hat{p}_N[LR_T^{(0)} \,|\, \theta_0] : \theta_0 \in \bar{\Omega}_0\} \leq \alpha\right] \leq \alpha \tag{26}$$

under the null hypothesis: we get a valid test with level α for H_0; see Proposition 4.1 in Dufour (2006). In the present case, we call this procedure the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT).

The parameter space, however, can be very large. Specifically, it grows as the number of autoregressive components increases and as the number of regimes increases. Additionally, the solution 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. Such algorithms include: Generalized Simulated Annealing, Genetic Algorithms, and Particle Swarm [see Dufour 2006; Dufour and Neves 2019].

In order to facilitate optimization [as described in Dufour (2006)], it is also possible to search within a smaller consistent set of the parameter space C_T. 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

$$C_T = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < c\} \tag{27}$$

where c is a fixed positive constant that does not depend on T and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$. An empirically interesting consistent set that we consider in this study is $C_T^{*} = C_T^{CI} \cup C_T^{\epsilon}$ where

$$C_T^{CI} = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < 2 \times S.E.(\hat{\theta}_0)\} \tag{28}$$

$$C_T^{\epsilon} = \{\theta_0 \in \bar{\Omega}_0 : \parallel \hat{\theta}_0 - \theta_0 \parallel < \epsilon\} \tag{29}$$

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 of the autoregressive parameters and the transition probabilities, depending on the choice of ϵ, while also constraining the values considered for the mean and the variances to a reasonable region.

Finally, 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 (LMC-LRT). Here, the consistent set includes only the consistent point estimate $\hat{\theta}_0$. Generic conditions for the asymptotic validity of such a test are discussed in section 5 of Dufour (2006), but these are more restrictive than those for the MMC-LRT procedure. The LMC test can be interpreted as the finite-sample analogue of the parametric bootstrap. To reflect this, we replace $\hat{F}_N[x \mid \theta_0]$ with $\hat{F}_{TN}[x \mid \theta_0] = \hat{F}_N[x; LR_T(N, \theta_0)]$ and $\hat{G}_N[x \mid \theta_0]$ with $\hat{G}_{TN}[x \mid \theta_0] = \hat{G}_N[x; LR_T(N, \theta_0)]$ where the subscript T is meant to allow the test statistics and functions to change based on increasing sample sizes. As a result, the Monte Carlo p-value is given by

$$\hat{p}_{TN}[x \mid \theta_0] = \frac{N\hat{G}_{TN}[x \mid \theta_0] + 1}{N + 1} \tag{30}$$

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 $\hat{F}[x \mid \theta_0]$. 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. In fact, as can be seen from equations (24) and (30), the number of replications N is taken into account in the calculation of the p-value both in the numerator and the denominator so that it essentially remains fixed as N increases. As discussed in Dufour (2006), building a test with level $\alpha = 0.05$ requires as few as 19 replications but using more replications can increase the power of the test. For this reason, in our simulations results we use $N = 99$ for our Monte Carlo procedure as in Dufour and Khalaf (2001) and Dufour and Luger (2017) though it is also possible to use the procedure described in Davidson and MacKinnon (2000) to determine the optimal number of simulations to minimize experimental randomness and loss of power.

4 Simulation Evidence

We begin by reviewing the moment-based test proposed by Dufour and Luger (2017) and the parameter stability test proposed by Carrasco, Hu, and Ploberger (2014), which we use for comparison in the univariate setting where we consider the null hypothesis of one regime (linear model) against the alternative hypothesis of two regimes because this is the only setting where these tests are applicable. Simulation results for the two Monte Carlo likelihood ratio tests proposed here are also shown below and cover more complicated DGPs with two regimes under the null hypothesis and multivariate settings.

4.1 Moment-based Tests for Markov-switching

Here, we review the test proposed by Dufour and Luger (2017). This test avoids some of the statistical issues described above for likelihood ratio type tests by using moments of the residuals aimed at capturing characteristics of a mixture normal distribution. Additionally, their test is less costly computationally in comparison to the test proposed by Carrasco, Hu, and Ploberger (2014) and the likelihood ratio-based tests mentioned previously including the one proposed here. This is because it only requires estimating the model under the null hypothesis. It also allows the econometrician to perfectly control the level of the test through the use of Monte Carlo (MC) test methods discussed in Dufour (2006).

As previously mentioned, this procedure is aimed at detecting the presence of two regimes and so, in their setup the parameters μ_{s_t} and σ_{s_t} shown in (2) can be given as

$$\mu_{s_t} = \mu_1\mathbb{I}\{S_t = 1\} + \mu_2\mathbb{I}\{S_t = 2\} \qquad \sigma_{s_t} = \sigma_1\mathbb{I}\{S_t = 1\} + \sigma_2\mathbb{I}\{S_t = 2\} \tag{31}$$

Hence the hypothesis to be tested can be formulated as

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \ \ \& \ \ \sigma_1 = \sigma_2 \\ H_1 &: \mu_1 \neq \mu_2 \ \ \& \ \ \sigma_1 \neq \sigma_2 \end{aligned} \tag{32}$$

An important distinction from other testing procedures however, is that, like Cho and White (2007) they assume that under the alternative hypothesis the observed data y_t is governed by a mixture of normal distributions weighted by the ergodic proprieties of the markov-chain. That is, under the alternative y_t is given by

$$y_t = \pi_1 N(\mu_1, \sigma_1^2) + \pi_2 N(\mu_2, \sigma_2^2) \tag{33}$$

where π_1 and π_2 are given by (3).

To test the the hypothesis of a linear model against that of a AR-MS model, they develop test statistics that are based on the moments of the least-square residuals of autoregressive models. More Specifically, they focus on the mean, variance, skewness and excess kurtosis of the least-square residuals. To do this, they first transform the series to correct for the autoregressive parameters which do not change according to the regime and get

$$z_t = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} \tag{34}$$

and then obtain the residuals $\hat{\epsilon}_t = z_t - \bar{z}_t$. From here, the residual moments are calculate as

$$M(\hat{\epsilon}) = \frac{|m_2 - m_1|}{\sqrt{s_1^2 + s_2^2}} \tag{35}$$

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

$$V(\hat{\epsilon}) = \frac{\vartheta_2(\hat{\epsilon})}{\vartheta_1(\hat{\epsilon})} \tag{36}$$

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 $\hat{\sigma}^2 = T^{-1}\Sigma_{t=1}^{T}\hat{\epsilon}_t{}^2$

$$S(\hat{\epsilon}) = \left|\frac{\Sigma_{t=1}^{T}\hat{\epsilon}_t{}^3}{T(\hat{\sigma}^2)^{3/2}}\right| \tag{37}$$

and

$$K(\hat{\epsilon}) = \left| \frac{\Sigma_{t=1}^{T} \hat{\epsilon_t}^4}{T(\hat{\sigma}^2)^2} - 3 \right| \tag{38}$$

It is important to note that if we consider (34) as our series of interest, then these test statistics do not depend on any nuisance parameters since they can be obtained strictly from the vector of standardized residuals $\hat{\epsilon}/\hat{\sigma}$ and as such they are pivotal statistics. The testing procedure involves calculating these test statistics from the observed data, obtaining the sample distribution $\hat{F}(x)$ by simulating the test statistic under the null hypothesis and then computing the empirical p-values for each moment.

From here, we can obtain individual MC p-values for each test statistic but must find a way to combine them. One approach would be to set the level α_i for each test such that the the overall level $\alpha = \sum_i \alpha$ is exactly equal to the desired level of 0.05. Alternatively, there are two different methods proposed in Dufour and Luger (2017) for combining independent test statistics that are used to combine the p-values. The idea is to essentially treat the combined p-value as a test statistic which requires a second layer of simulations to obtain a combined p-value. The first step is to use the approximate distribution functions taking the simple logistic form:

$$\hat{F}[x] = \frac{exp(\hat{\gamma}_0 + \hat{\gamma}_1 x)}{1 + exp(\hat{\gamma}_0 + \hat{\gamma}_1 x)} \tag{39}$$

where $\hat{\gamma}_0$ and $\hat{\gamma}_1$ are estimated by nonlinear least-squares (NLS). Then the the approximate p-value of, for example $M(\hat{\epsilon})$, is computed as $\hat{G}_M[M(\hat{\epsilon})] = 1 - \hat{F}_M[M(\hat{\epsilon})]$. Once we have done this for each test statistic, they can be combined in one of two ways. The first method is based on the min of the p-values and was initially suggested by Tippett et al. (1931) and Wilkinson (1951). Here, the test statistic becomes,

$$F_{min}(\hat{\epsilon}) = 1 - min\{\hat{G_M}[M(\hat{\epsilon})], \hat{G_V}[V(\hat{\epsilon})], \hat{G_S}[S(\hat{\epsilon})], \hat{G_K}[K(\hat{\epsilon})]\} \tag{40}$$

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

$$F_{\times}(\hat{\epsilon}) = 1 - \{\hat{G_M}[M(\hat{\epsilon})] \times \hat{G_V}[V(\hat{\epsilon})] \times \hat{G_S}[S(\hat{\epsilon})] \times \hat{G_K}[K(\hat{\epsilon})]\} \tag{41}$$

Dufour et al. (2004) and Dufour, Khalaf, and Voia (2014) provide further discussion of these methods of combining test statistics for the interested reader. Finally, the Monte Carlo p-value of the combined statistics is given by

$$G_{F_{min}}[F_{min}(\hat{\epsilon}); N] = \frac{N + 1 - R_{F_{min}}[F_{min}(\hat{\epsilon}); N]}{N} \tag{42}$$

and

$$G_{F_{\times}}[F_{\times}(\hat{\epsilon}); N] = \frac{N + 1 - R_{F_{\times}}[F_{\times}(\hat{\epsilon}); N]}{N} \tag{43}$$

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 where $\hat{\eta} = \eta - \bar{\eta}$ and $\eta \sim N(0, I_T)$.

The result is four test procedure. Specifically: LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$. An advantage of these procedure is that they deal with less nuisance parameters than the LMC-LRT and MMC-LRT procedures proposed here. However, unlike LMC-LRT and MMC-LRT, they can only be used to compare a linear model under the null hypothesis to a MSM with two regimes under the alternative.

4.2 Optimal test for Regime Switching:

Carrasco, Hu, and Ploberger (2014) (CHP) proposes a test that they describe as an optimal test for the consistency of parameters in random coefficient and markov-switching models. Specifically, their testing procedure is more generally suited to detect parameter heterogeneity but includes the AR-MS model as a special case. As previously mentioned, an advantage of this method is that it only requires estimating the model under the null hypothesis. 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.

The authors formulate the hypothesis in the following way:

$$\begin{aligned} H_0 &: \theta_t = \theta_0 \\ H_1 &: \theta_t = \theta_0 + \eta_t \end{aligned} \tag{44}$$

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.

In order to deal with the presence of nuisance parameters, the authors propose two alternatives. The first is a Sup-type test as in Davies (1987). 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

$$\mu_{2,t}(\beta, \theta) = \frac{1}{2}c^2 h'[(\frac{\partial l_t}{\partial\theta\partial\theta'} + (\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta})') + 2\sum_{s<t}\rho^{(t-s)}(\frac{\partial l_t}{\partial\theta})(\frac{\partial l_t}{\partial\theta})']h \tag{45}$$

which allows us to get the expression

$$supTS = \sup_{\{h,p:||h||=1,\underline{\rho}<\rho<\bar{\rho}\}} = \frac{1}{2}(max(0, \frac{\Gamma_T^*}{\sqrt{\hat{\epsilon^*}'\hat{\epsilon^*}}}))^2 \tag{46}$$

where $\Gamma_T^* = \Gamma_T^*(\beta, \theta) = \sum_t \mu_{2,t}^*(\beta, \theta)/\sqrt{T}$ and $\mu_{2,t}^*(\beta, \theta) = \mu_{2,t}(\beta, \theta)/c^2$ 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. Here, the notation $l_t^{(k)}$ denotes the kth derivative of the log-likelihood with respect to the parameters θ.

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, CHP also suggest using an Exponential-type test as in Andrews and Ploberger (1994). This statistic is given by:

$$expTS = \int_{\{\underline{\rho}\leq\rho\leq\bar{\rho},||h||<1\}} \Psi(h, \rho) d\rho dh \tag{47}$$

where

$$\Psi(h,\rho) = \begin{cases} \sqrt{2\pi} exp[\frac{1}{2}(\frac{\Gamma_T^*}{\sqrt{\hat{\epsilon^*}'\hat{\epsilon^*}}} - 1)^2]\Phi(\frac{\Gamma_T^*}{\sqrt{\hat{\epsilon^*}'\hat{\epsilon^*}}} - 1) & \text{if } \hat{\epsilon^*}'\hat{\epsilon^*} \neq 0. \\ 1 & \text{otherwise.} \end{cases} \tag{48}$$

The authors also derive the asymptotic distribution of both test statistics they propose. Their test has since been used in Hamilton (2005), Warne and Vredin (2006), Kahn and Rich (2007), Hu and Shin (2008), Morley and Piger (2012) and Dufrénot, Mignon, and Péguin-Feissolle (2011).

4.3 Simulation Results

Now, we present tables summarizing the empirical size and power (in percentage) of the two tests proposed in this paper. We also provide the same empirical results for the test procedures described previously in this section for comparison. In what follows, the nominal level is set to be $\alpha = 0.05$ and results are based on 1000 replications of the data generating process (DGP). Throughout, we will consider a simple AR(1) model given by

$$y_t = \mu_{s_t} + \phi_1(y_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t}\epsilon_t \tag{49}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$, such that only the mean and variance are governed by the Markov process S_t. It is understood that the LMC-LRT, supTS, and expTS procedures should perform better in large sample sizes since they are asymptotic tests. However, many economic applications using quarterly observations are limited to as few as 100 to 200 observations so in the following we consider these sample sizes to get an idea of the finite sample performance of these tests. sizes to get an idea of the finite sample performance of these tests. Also, whenever considering a Maximized Monte Carlo test, we use the set $C_T^* = C_T^{CI} \cup C_T^{0.1}$. All results presented here were obtained using the R-package **MSTest** described in the companion paper Rodriguez Rondon and Dufour (2022).

We begin by considering a null hypothesis of a linear AR(1) model (i.e, $H_0 : M_0 = 1$) and an alternative hypothesis of a MSM with one lag and two regimes (i.e, $H_0 : M_0 + m = 2$).

Table 1: Empirical size of test when $H_0 : M_0 = 1$ and $H_1 : M_0 + m = 2$

Test	$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$	
	T=100	T=200	T=100	T=200	T=100	T=200
LMC-LRT	5.8	5.3	4.9	4.8	4.9	4.9
MMC-LRT	0.4	0.5	0.9	0.9	0.9	0.6
LMC$_{min}$	4.4	5.9	4.7	4.7	4.2	4.8
LMC$_{prod}$	5.3	5.2	4.9	5.1	4.9	4.3
MMC$_{min}$	0.2	0.2	0.2	0.7	0.2	0.0
MMC$_{prod}$	0.1	0.2	0.4	0.8	0.2	0.5
supTS	4.8	5.1	6.0	4.5	-	-
expTS	6.8	6.2	5.4	6.9	-	-

Table 1 reports the empirical size of the tests (in percentage) when $\phi = 0.1$ in the first two columns, when $\phi = 0.9$ in the next two columns, and when $\phi = 1.0$ in the last two columns. For each value of ϕ we have a sample size of $T = 100$ or $T = 200$. As suggested by the theory laid out in Dufour (2006), the maximized Monte Carlo tests has empirical rejection frequencies $\leq 5\%$ under the null hypothesis. The local Monte Carlo based tests, LMC-LRT, LMC$_{min}$, and LMC$_{prod}$, also appear perform very well with a rejection frequency of approximately 5% even in finite samples. The simulation results for the supTS and expTS show that these tests also appear to control the empirical size well but less so than the Monte Carlo based tests as they have a small degree of over-rejection in some cases. This however, should be expected from an asymptotic test given that the sample sizes are fairly small.

Table 2: Empirical power of test when $H_0 : M_0 = 1$ and $H_1 : M_0 + m = 2$

Test	$(p_{11}, p_{22}) = (0.9, 0.9)$						$(p_{11}, p_{22}) = (0.9, 0.5)$					
	$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$		$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$	
	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$\Delta\mu = 2$, $\Delta\sigma = 0$												
LMC-LRT	48.7	81.2	19.5	33.8	12.5	15.8	34.0	72.3	27.4	48.3	20.8	24.5
MMC-LRT	30.6	51.0	4.6	8.4	4.6	2.4	20.3	41.0	7.1	16.9	4.4	8.1
LMC_{min}	3.8	5.8	14.6	21.3	18.8	28.5	14.8	29.9	13.5	21.8	16.3	28.4
LMC_{prod}	4.1	5.9	15.2	24.2	19.6	30.8	11.9	22.4	15.0	23.4	16.9	29.0
MMC_{min}	0.0	0.1	2.8	4.1	1.9	5.0	1.0	3.7	1.4	2.6	1.6	2.1
MMC_{prod}	0.1	0.1	1.9	4.0	2.3	5.2	1.8	3.3	1.2	3.8	1.6	3.6
supTS	24.3	49.9	8.4	12.5	-	-	23.8	47.0	11.9	18.2	-	-
expTS	15.6	25.4	21.7	32.6	-	-	24.6	47.1	22.1	33.5	-	-
$\Delta\mu = 0$, $\Delta\sigma = 1$												
LMC-LRT	66.6	93.4	67.0	94.5	67.2	94.6	55.7	88.7	61.6	88.4	58.4	89.9
MMC-LRT	39.4	69.6	35.6	73.0	31.7	64.7	36.4	69.4	30.7	60.8	29.2	54.5
LMC_{min}	36.5	64.7	42.5	64.9	39.6	64.4	48.7	69.4	49.4	72.0	45.6	74.4
LMC_{prod}	40.6	66.8	43.3	69.1	42.5	65.3	48.8	70.7	49.5	71.6	47.8	73.2
MMC_{min}	9.0	30.2	11.2	27.8	10.4	16.3	20.5	44.4	19.7	49.2	16.6	34.3
MMC_{prod}	10.7	31.4	10.8	31.0	7.9	18.0	20.3	42.6	18.8	44.5	16.2	31.2
supTS	32.4	58.0	32.2	67.4	-	-	29.9	46.4	30.0	50.3	-	-
expTS	40.1	62.6	54.1	84.7	-	-	43.9	68.3	52.8	78.6	-	-
$\Delta\mu = 2$, $\Delta\sigma = 1$												
LMC-LRT	83.7	99.4	45.3	77.2	29.5	43.9	83.4	99.4	60.2	88.1	53.1	67.4
MMC-LRT	60.3	90.1	24.0	52.0	14.0	29.5	66.2	91.6	41.5	74.5	29.7	53.7
LMC_{min}	51.9	81.6	39.9	62.3	35.4	57.7	84.1	99.0	65.9	89.4	63.7	88.2
LMC_{prod}	45.9	74.2	42.5	65.1	38.1	60.9	83.7	99.2	68.5	91.6	63.9	89.9
MMC_{min}	10.5	39.0	10.2	24.0	8.8	13.7	47.3	50.3	28.1	49.8	23.4	46.7
MMC_{prod}	14.3	37.9	11.8	29.6	9.0	18.1	48.3	44.6	34.9	43.6	28.3	42.1
supTS	72.7	96.2	34.6	62.9	-	-	80.8	96.9	53.2	79.8	-	-
expTS	75.6	97.0	53.9	77.9	-	-	86.6	99.4	75.1	94.7	-	-

Table 2 reports the empirical power of the tests (in percentage) when the underlying DGP is a MSM with two regimes. We consider the same values for ϕ and T as in Table 1 but this time, in the first six columns the transition probabilities are $(p_{11}, p_{22}) = (0.9, 0.9)$ resulting in $\boldsymbol{\pi} = (0.5, 0.5)$ and in the last six columns the transition probabilities are $(p_{11}, p_{22}) = (0.90, 0.50)$ so that $\boldsymbol{\pi} = (0.83, 0.17)$. The first case corresponds to having spending, on average in the long run, the same amount of time in both regimes whereas the second case suggests the, on average in the long run, more time is spent in the first regime over the entire sample. The first panel corresponds to a DGP where only the mean changes, the second panel a DGP where only the variance changes and the third bottom panel a DGP where both the mean and the variance are different across regimes. As can be seen from this tables, the power of the test is lowest when only the mean is subject to change for all tests. The LMC_{min}, LMC_{prod}, MMC_{min}, and MMC_{prod} procedures have the lowest power when only the mean is subject to change. The LMC-LRT procedure proposed here on the other hand has comparable power to the supTS and expTS and in some cases even does better. When the variance is subject to change, all tests have higher power. The LMC-LRT test proposed here appears to have the highest power in most cases when when the variance is subject to change with the supTS and expTS having comparable results in some cases even in finite samples. Given that the MMC-LRT procedure considers a wider set of nuisance parameter values in comparison to the LMC-LRT procedure, the power of the MMC-LRT is lower than that of the LMC-LRT in all cases. The same is true when comparing the MMC_{min}, and MMC_{prod} power results to those of LMC_{min} and LMC_{prod} and this should be expected. However, we find that for the set C_T^*, the power of the MMC-LRT procedure proposed here is quite good and in some cases even outperforms the LMC_{min} and LMC_{prod} procedures when only the mean is subject to change.

Now we consider a null hypothesis of a linear AR(1) model (i.e, $H_0 : M_0 = 1$) and an alternative hypothesis of a MSM with one lag and three regimes (i.e, $H_0 : M_0 + m = 3$).

Table 3: Empirical size of test when $H_0 : M_0 = 1$ and $H_1 : M_0 + m = 3$

Test	$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$	
	T=100	T=200	T=100	T=200	T=100	T=200
LMC-LRT	4.1	5.2	3.3	4.3	4.4	2.9
MMC-LRT	0.5	1.1	0.5	0.4	0.3	0.0

Table 3 reports the empirical size of the tests (in percentage) in this setting. The DGPs are the same as those described for table 1 and the results are very comparable as the MMC-LRT procedure has an empirical rejection frequency $\leq 5\%$ and the LMC-LRT also appear perform very well with a rejection frequency of approximately 5%.

Table 4: Empirical power of test when $H_0 : M_0 = 1$ and $H_1 : M_0 + m = 3$

Test	$(p_{11}, p_{22}, p_{33}) = (0.9, 0.9, 0.9)$						$(p_{11}, p_{22}, p_{33}) = (0.9, 0.5, 0.5)$					
	$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$		$\phi = 0.1$		$\phi = 0.9$		$\phi = 1.0$	
	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200	T=100	T=200
$\mu = (-2, 0, 2)$, $\sigma = (1, 1, 1)$												
LMC-LRT	41.8	57.4	43.6	70.0	31.7	43.1	82.8	97.1	57.1	85.7	39.9	59.8
$\mu = (0, 0, 0)$, $\sigma = (1, 2, 3)$												
LMC-LRT	69.8	95.9	70.0	94.3	72.2	95.2	85.0	99.1	83.3	99.0	83.6	99.4
$\mu = (-2, 0, 2)$, $\sigma = (1, 2, 3)$												
LMC-LRT	78.9	97.8	50.0	85.1	36.1	46.5	98.3	100	87.6	99.5	75.5	83.3

Table 4 reports the empirical power of the LMC-LRT procedure (in percentage) when the underlying DGP is a MSM with three transition probabilities are either $(p_{11}, p_{22}, p_{33}) = (0.9, 0.9, 0.9)$ or $(p_{11}, p_{22}, p_{33}) = (0.9, 0.5, 0.5)$. The other transition probabilities are given by $(1 - p_{ii})/2$ so that all columns of the transition matrix sum to 1 as required. As before, the power of LMC-LRT is lowest when only the mean is subject to change and is much higher when the variances or both the mean and the variance are subject to change.

The LMC_{min}, LMC_{prod}, MMC_{min}, and MMC_{prod} procedures have been omitted in this setting because those tests are not equipped to consider an alternative MSM with three regimes. Similarly, the supTS and expTS procedures have also been omitted because, although the null in still a linear model (i.e, parameter stability), these tests only have power against local alternatives and as a result do not perform well in this setting.

Now we consider a null hypothesis of a MSM with one lag and two regimes (i.e, $H_0 : M_0 = 2$) and an alternative hypothesis of a MSM with one lag and three regimes (i.e, $H_0 : M_0 + m = 3$). For this comparison, we use the DGP used in Kasahara and Shimotsu (2018). Specifically, for the Boot-LRT we borrow the results obtained by Kasahara and Shimotsu (2018) which perform 3000 replications and only use $N = 199$ simulations. We provide results for LMC-LRT to make sure our results are comparable to those obtained by Kasahara and Shimotsu (2018). Additionally, we only consider the case where the variance is constant and the mean changes according to the regime.

Table 5: Empirical size of test when $H_0 : M_0 = 2$ and $H_1 : M_0 + m = 3$

Test	$(p_{11}, p_{22}) = (0.5, 0.5)$			$(p_{11}, p_{22}) = (0.7, 0.7)$		
	T=100	T=200	T=500	T=100	T=200	T=500
$(\phi, \mu_1, \mu_2, \sigma) = (0.5, -1, 1, 1)$						
LMC-LRT	6.80	6.30	4.60	6.00	6.00	4.80
MMC-LRT	0.10	0.20	0.10	0.00	0.10	0.00
Boot-LRT	-	7.16	4.43	-	6.07	4.20

Table 5 reports the empirical size of the LMC-LRT, MMC-LRT and Boot-LRT (in percentage) in this setting for a sample size of $T = 100$, $T = 200$, and $T = 500$. Kasahara and Shimotsu (2018) do not consider a sample sixe of $T = 100$ so results for Boot-LRT are not provided for this case. We find that all test have fairly good finite sample results but that the Boot-LRT and

LMC-LRT procedures display some over-rejection when the sample size is $T = 200$ or $T = 100$ for the LMC-LRT procedure. As mentioned above, these are asymptotic tests in the sense that we need the sample size to be large so that the consistent estimators of the nuisance parameters converge in probability to the true parameters. To this point, when the sample size is $T = 500$, we see that both tests are better able to control the level of the test with a rejection frequency closer to 5%. On the other hand, we find that the MMC-LRT procedure still performs as suggested by the theory and maintains a rejection frequency $\leq 5\%$ when with very small sample sizes. This highlights the contribution of the MMC-LRT procedure as a valid test procedure both in finite samples and asymptotically.

Finally, we consider a multivariate setting. Such a setting has, to the best of our knowledge, not been considered in the literature on hypothesis testing for the number of regimes in MSMs. Specifically, we consider a null hypothesis with a linear VAR(1) model (i.e, $H_0 : M_0 = 1$) and an alternative hypothesis of a MS-VAR model with one lag and two regimes (i.e, $H_0 : M_0 + m = 2$) so that we may show that the test procedures proposed here are applicable in multivariate settings. The DGP is a simple bivariate VAR model given by

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & 0.25 \\ -0.25 & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} \tag{50}$$

As before, simulation results are obtained using 1000 replications and $N = 99$.

Table 6: Empirical Size of Test for Bivariate Markov-Switching VAR

Test	$\phi_{11} = \phi_{22} = 0.1$		$\phi_{11} = \phi_{22} = 0.9$	
	T=100	T=200	T=100	T=200
LMC-LRT	5.5	5.3	4.1	4.0

Table 5 reports the empirical size of the LMC-LRT in this setting. We only consider $\phi = 1.0$ and $\phi = 0.9$ but it would also be of interest to consider a setting where we have a cointegrated system. The current results suggest that the LMC-LRT procedure performs well and is able to maintain a rejection frequency of approximately 5% as we would like.

5 Empirical Results

In this section, we use the LMC-LRT procedure proposed in this paper to determine the appropriate number of regimes for modelling the log difference of U.S. real GNP. Hamilton (1989) first used U.S. GNP growth data spanning from 1951Q2 to 1984Q4 when proposing the MSM discussed here. This data was then considered by Hansen (1992), Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) to showcase their testing procedures. In all three studies, the authors failed to reject the null hypothesis of linear model when compared to a MSM with two regimes under the alternative. Carrasco, Hu, and Ploberger (2014) even consider hypothesis testing when only the mean changes, as in Hamilton 1989 and when both mean and variance are assumed to change according to the regime. In both cases, the null hypothesis could not be rejected. Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) also consider a larger sample spanning from 1951Q2 to 2010Q4. In this paper we consider this second sample and a third sample spanning from 1951Q2 to 2022Q3 which includes the recessionary period induced by the COVID-19 pandemic. Figures A1 and A2 in the appendix section show graphs of the time series of the two samples considered here.

Table 7:U.S. real GNP growth MSM estimates

	1951Q2 - 2010Q4			1951Q2 - 2022Q3			
	M=1	M=2	M=3	M=1	M=2	M=3	M=4
μ_1	0.763	0.713	0.832	1.517	1.516	1.874	2.008
μ_2	-	0.812	0.078	-	2.169	3.575	3.844
μ_3	-	-	0.826	-	-	1.248	1.239
μ_4	-	-	-	-	-	-	1.697
σ_1^2	0.783	1.221	1.148	1.48	0.447	1.092	0.470
σ_2^2	-	0.177	1.002	-	14.199	65.082	66.963
σ_3^2	-	-	0.142	-	-	0.229	0.231
σ_4^2	-	-	-	-	-	-	1.981
ϕ_1	0.335	0.305	0.252	0.124	0.364	0.292	0.282
ϕ_2	0.124	0.21	0.199	0.202	0.237	0.192	0.181
ϕ_3	-0.083	-0.12	-0.111	0.038	-0.024	-0.053	-0.042
ϕ_4	-0.074	-0.053	-0.067	0.003	-0.042	-0.004	-0.023
p_{11}	-	0.033	0.005	-	0.962	0.005	0.095
p_{12}	-	0.967	0	-	0.038	0	0.009
p_{13}	-	-	0.995	-	-	0.995	0.897
p_{14}	-	-	-	-	-	-	0.000
p_{21}	-	0.956	0.187	-	0.42	0.21	0.000
p_{22}	-	0.044	0.813	-	0.58	0.515	0.210
p_{23}	-	-	0	-	-	0.275	0.276
p_{24}	-	-	-	-	-	-	0.514
p_{31}	-	-	0.951	-	-	0.983	0.000
p_{32}	-	-	0.043	-	-	0.017	0.983
p_{33}	-	-	0.006	-	-	0	0.000
p_{34}	-	-	-	-	-	-	0.017
p_{41}	-	-	-	-	-	-	0.857
p_{42}	-	-	-	-	-	-	0.000
p_{43}	-	-	-	-	-	-	0.143
p_{44}	-	-	-	-	-	-	0.000

Table 7 reports the estimated coefficients for an AR(4) model with the number of regimes ranging from $M = 1$ to $M = 3$ when considering the data up to 2010Q4 and up to $M = 4$ when considering the larger sample ending in 2022Q3. In the case of the MSMs we assume both the mean and the variance are governed by the latent Markov process S_t. Estimates are obtained using the R-package **MSTest** described in Rodriguez Rondon and Dufour (2022). Specifically, we estimate each model using the Expected Maximization (EM) algorithm with various initial values and keep the estimates that give the highest log likelihood value. We then use these estimates as initial values in a second stage where to obtain maximum likelihood estimates. The second column of table 8 reports the log-likelihood of each model.

Table 8: Determining Number of Regimes

		m=1		m=2		m=3	
M_0	log-like	LMC-LRT	p-value	LMC-LRT	p-value	LMC-LRT	p-value
U.S. GNP Growth, 1951Q2 - 2010Q4							
1	-304.23	50.18	0.01	60.67	0.01	-	-
2	-279.14	**10.49**	**0.25**	-	-	-	-
3	-273.90	-	-	-	-	-	-
U.S. GNP Growth, 1951Q2 - 2022Q3							
1	-454.92	193.02	0.01	250.02	0.01	261.22	0.01
2	-358.41	57.01	0.01	68.20	0.01	-	-
3	-329.91	**11.19**	**0.31**	-	-	-	-
4	-324.31	-	-	-	-	-	-

The first column specifies the number of regimes under the null hypothesis. As we move down the columns towards the right we show the LMC-LRT statistic when the alternative has one additional regime and the corresponding p-value. The following two columns correspond to an alternative hypothesis with two additional regimes and the last two columns correspond to an alternative hypothesis with three additional regimes. For example, the fifth column of the first row gives the

LMC-LRT statistic when comparing a linear model (i.e., $M_0 = 1$) to a MSM with three regimes (i.e., $M_0 + m = 3$) and the sixth column reports the corresponding p-value. We find that for the sample spanning from 1951Q2 to 2010Q4, a MSM with two regimes is appropriate. In Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) the authors also find that a MSM with two regimes better explains the data than a model with only one regime as we do here. However, in our case, we confirm that a MSM with more than two regimes is not appropriate for this sample, which had not been considered before for U.S. GNP data. Figure A3 in the appendix shows the smoothed probabilities of both regimes for this sample. We find that the first regime (black) is mostly consistent with the period from 1951Q2 to the mid 1980s and the second regime (red) is consistent with the period following the mid 1980s. The exception is that the probability of being in the first regime increases during the 1990s, early 2000s and during the 2008 recessionary period (i.e., Great Recession). However, we do not necessarily identify all other recessionary periods as in Hamilton (1989). This is in part due to the fact that we allow the variance to change whereas, in Hamilton (1989), the author considers a model where only the mean changes and further their sample only covers 1951Q2 to 1984Q4, which is within our first regime where out results suggest the variance is constant. What we see here is that the first regime is consistent with a period of high volatility whereas the second is marked by low volatility, which explains the return of the first regime during some recessionary periods within this sample, but not all. Interestingly this result already provides evidence of the Great Moderation, which states that there was a structural decline in the variance of many macroeconomic variables occurring in the mid 1980s, as the preferred model suggests there is a change in variance that occurred in the mid-1980s. When we consider the extended sample spanning from 1951Q2 to 2022Q3 we find that a MSM with three regimes is appropriate. Figure A4 in the appendix shows the smoothed probabilities of all three regime for this sample. In this case the second (red) regime is consistent with severe recessionary periods but misses the recessions that are comparatively shallow and that occurred before the Great Recession of 2008. The first regime (black) is again mostly consistent with the period from 1951Q2 to the mid 1980s excluding recessionary periods and the third regime (blue) is consistent with the period from mid 1980s to 2019 excluding recessionary periods. From table 7 we see that for this model the the variance in the regime beginning in 1951Q2 and up to mid 1980s is $\sigma_1^2 = 1.092$ and the variance in the regime beginning in the mid 1980s and going up to the recession induced by the COVID-19 pandemic is $\sigma_3^2 = 0.229$ and so we have a decrease in variance in non-recessionary state after the mid 1980s which further confirms evidence about the Great Moderation. Like Gadea, Gómez-Loscos, and Pérez-Quirós (2018) and Gadea, Gómez-Loscos, and Pérez-Quirós (2019) we also find that the low volatility period continues after the Great recession of 2008 in both samples. However, our new extended sample provides some evidence that the high volatility period makes a return after this recent recession, which has not been documented for U.S. GNP data so far. The second panel of figure A3 and A4 show the smoothed probabilities of each regime when considering an additional regime over the number suggested by our testing procedure. In both cases, the additional regime seems to pick up some of the comparatively more shallow recessions.

6 Conclusion

We have shown how to use the Monte Carlo procedures described in Dufour (2006) for the setting of a likelihood ratio test for MSMs. In doing so, we propose the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT) and the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) that can be used to determine the number of regimes in MSMs and in HMMs. Specifically, the tests proposed here are general enough where they can deal with settings where we are interested in comparing models with M_0 regimes under the null hypothesis against models with $M_0 + m$ regimes under the alternative, where here both $M_0, m \geq 1$. Further, they can also be applied to settings where we have a non-stationary process, a process with non-Gaussian errors, and multivariate settings. To the best of our knowledge, we are the first to consider hypothesis testing of multivariate MSMs. Although we work with the sample distribution of the test statistic, asymptotic results have not been provided for LRT in a multivariate setting which brings forward an interesting direction for future research. The simulation results suggest that both versions of the Monte Carlo likelihood ratio test are able to control the level of the test very well. An important contribution is the MMC-LRT, which performs well and maintains a rejection frequencies $\leq \alpha$ in all cases as suggested by the theory proposed in Dufour (2006) and is an identification-robust procedure that is valid in finite samples and asymptotically. Further, simulation results also suggest both test have good power. Specifically, the LMC-LRT has comparable or better power than the supTS and expTS tests while both the LMC-LRT and MMC-LRT outperform their moment-based counterparts. Finally, we use the LMC-LRT to determine the appropriate number of regimes for modelling real U.S. GNP growth and confirm that a model with two regimes best described data spanning 1951Q2 to 2010Q4 and find that a model with three regimes best describes the extended sample we consider here which begins in 1951Q2 and ends in 2022Q3. In doing so, our results confirms evidence about the Great Moderation in the sense that two of the three regimes have positive growth but experience a reduction in variance following the mid-1980s as suggested by Gadea, Gómez-Loscos, and Pérez-Quirós (2018), Gadea, Gómez-Loscos, and Pérez-Quirós (2019) among others but, in addition, we find that the period of high volatility returns following the COVID-19 recession.

References

An, Yonghong, Yingyao Hu, Johns Hopkins, and Matt Shum. 2013. *Identifiability and inference of hidden Markov models.* Technical report. Technical report.

Andrews, Donald WK, and Werner Ploberger. 1994. “Optimal tests when a nuisance parameter is present only under the alternative.” *Econometrica: Journal of the Econometric Society:* 1383–1414.

Ang, Andrew, and Allan Timmermann. 2012. “Regime changes and financial markets.” *Annu. Rev. Financ. Econ.* 4 (1): 313–337.

Anser, Muhammad Khalid, Danish Iqbal Godil, Muhammad Azhar Khan, Abdelmohsen A Nassani, Khalid Zaman, and Muhammad Moinuddin Qazi Abro. 2021. “The impact of coal combustion, nitrous oxide emissions, and traffic emissions on COVID-19 cases: a Markov-switching approach.” *Environmental Science and Pollution Research* 28 (45): 64882–64891.

Antoine, Bertille, Lynda Khalaf, Maral Kichian, and Zhenjiang Lin. 2022. “Identification-Robust Inference With Simulation-Based Pseudo-Matching.” *Journal of Business & Economic Statistics:* 1–18.

Augustyniak, Maciej. 2014. “Maximum likelihood estimation of the Markov-switching GARCH model.” *Computational Statistics & Data Analysis* 76:61–75.

Baldi, Pierre, Yves Chauvin, Tim Hunkapiller, and Marcella A McClure. 1994. “Hidden Markov models of biological primary sequence information.” *Proceedings of the National Academy of Sciences* 91 (3): 1059–1063.

Bunke, Horst, and Terry Michael Caelli. 2001. *Hidden Markov models: applications in computer vision.* Vol. 45. World Scientific.

Carrasco, Marine, Liang Hu, and Werner Ploberger. 2014. “Optimal test for Markov switching parameters.” *Econometrica* 82 (2): 765–784.

Carter, Andrew V, and Douglas G Steigerwald. 2012. “Testing for regime switching: a comment.” *Econometrica* 80 (4): 1809–1812.

Cevik, Emrah Ismail, Durmuş Çağrı Yıldırım, and Sel Dibooglu. 2021. “Renewable and non-renewable energy consumption and economic growth in the US: A Markov-Switching VAR analysis.” *Energy & Environment* 32 (3): 519–541.

Charfeddine, Lanouar. 2017. “The impact of energy consumption and economic development on ecological footprint and CO2 emissions: evidence from a Markov switching equilibrium correction model.” *Energy Economics* 65:355–374.

Chauvet, Marcelle. 1998. “An econometric characterization of business cycle dynamics with factor structure and regime switching.” *International economic review:* 969–996.

Chauvet, Marcelle, and James D Hamilton. 2006. “Dating business cycle turning points.” *Contributions to Economic Analysis* 276:1–54.

Chauvet, Marcelle, Chinhui Juhn, and Simon Potter. 2002. “Markov switching in disaggregate unemployment rates.” *Empirical Economics* 27 (2): 205–232.

Chevallier, Julien. 2011. "Evaluating the carbon-macroeconomy relationship: Evidence from threshold vector error-correction and Markov-switching VAR models." *Economic Modelling* 28 (6): 2634–2656.

Cho, Jin-Seo, and Halbert White. 2007. "Testing for regime switching." *Econometrica* 75 (6): 1671–1720.

Davidson, Russell, and James G MacKinnon. 2000. "Bootstrap tests: how many bootstraps?" *Econometric Reviews* 19 (1): 55–68.

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

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

Diebold, Francis X, and Glenn D Rudebusch. 1996. "Measuring business cycles: A modern perspective." *The Review of Economics and Statistics* 78.

Dufour, Jean-Marie. 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, Jean-Marie, and Lynda Khalaf. 2001. "Monte Carlo test methods in econometrics." *Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, UK:* 494–519.

Dufour, Jean-Marie, Lynda Khalaf, Jean-Thomas Bernard, and Ian Genest. 2004. "Simulation-based finite-sample tests for heteroskedasticity and ARCH effects." *Journal of Econometrics* 122 (2): 317–347.

Dufour, Jean-Marie, Lynda Khalaf, and Marcel Voia. 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, Jean-Marie, and Richard Luger. 2017. "Identification-robust moment-based tests for Markov switching in autoregressive models." *Econometric Reviews* 36 (6-9): 713–727.

Dufour, Jean-Marie, and Julien Neves. 2019. "Finite-sample inference and nonstandard asymptotics with Monte Carlo tests and R." In *Handbook of statistics,* 41:3–31. Elsevier.

Dufrénot, Gilles, Valérie Mignon, and Anne Péguin-Feissolle. 2011. "The effects of the subprime crisis on the Latin American financial markets: An empirical assessment." *Economic Modelling* 28 (5): 2342–2357.

Fisher, Ronald Aylmer. 1932. *Statistical methods for research workers.* Genesis Publishing Pvt Ltd.

Gadea, María Dolores, Ana Gómez-Loscos, and Gabriel Pérez-Quirós. 2018. "Great Moderation and Great Recession: From plain sailing to stormy seas?" *International Economic Review* 59 (4): 2297–2321.

———. 2019. "The decline in volatility in the US economy. A historical perspective." *Oxford Economic Papers* 72 (1): 101–123.

Garcia, Rene. 1998. "Asymptotic null distribution of the likelihood ratio test in Markov switching models." *International Economic Review:* 763–788.

Garcia, Rene, and Pierre Perron. 1996. “An analysis of the real interest rate under regime shifts.” *Review of Economics and Statistics* 78:111–125.

Goldfeld, Stephen M, and Richard E Quandt. 1973. “A Markov model for switching regressions.” *Journal of econometrics* 1 (1): 3–15.

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

Haas, Markus, Stefan Mittnik, and Marc S Paolella. 2004. “A new approach to Markov-switching GARCH models.” *Journal of financial econometrics* 2 (4): 493–530.

Hamilton, James D. 1989. “A new approach to the economic analysis of nonstationary time series and the business cycle.” *Econometrica* 57 (2): 357–384.

———. 1990. “Analysis of time series subject to changes in regime.” *Journal of econometrics* 45 (1-2): 39–70.

———. 2016. “Macroeconomic regimes and regime shifts.” *Handbook of macroeconomics* 2:163–201.

———. 2010. “Regime switching models.” In *Macroeconometrics and time series analysis,* 202–209. Palgrave Macmillan, London.

———. 2005. *What’s real about the business cycle?* Technical report. National Bureau of Economic Research.

Hamilton, James D, and Raul Susmel. 1994. “Autoregressive conditional heteroskedasticity and changes in regime.” *Journal of econometrics* 64 (1-2): 307–333.

Hamilton, James Douglas. 1994. *Time series analysis.* Princeton university press.

Hansen, Bruce E. 1992. “The likelihood ratio test under nonstandard conditions: testing the Markov switching model of GNP.” *Journal of applied Econometrics* 7 (S1): S61–S82.

Herwartz, Helmut, and Helmut Lütkepohl. 2014. “Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks.” *Journal of Econometrics* 183 (1): 104–116.

Hu, Liang, and Yongcheol Shin. 2008. “Optimal test for Markov switching GARCH models.” *Studies in Nonlinear Dynamics & Econometrics* 12 (3).

Jelinek, Frederick. 1997. *Statistical methods for speech recognition.* MIT press.

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

Kasahara, Hiroyuki, and Katsumi Shimotsu. 2018. “Testing the number of regimes in Markov regime switching models.” *arXiv preprint arXiv:1801.06862.*

Kim, Chang-Jin, and Charles R Nelson. 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.

Klaassen, Franc. 2002. “Improving GARCH volatility forecasts with regime-switching GARCH.” In *Advances in Markov-switching models,* 223–254. Springer.

Krogh, Anders, I Saira Mian, and David Haussler. 1994. "A hidden Markov model that finds genes in E. coli DNA." *Nucleic acids research* 22 (22): 4768–4778.

Krolzig, Hans-Martin. 1997. *Markov-Switching Vector Autoregressions.* Springer.

Lanne, Markku, Helmut Lütkepohl, and Katarzyna Maciejowska. 2010. "Structural vector autoregressions with Markov switching." *Journal of Economic Dynamics and Control* 34 (2): 121–131.

Lütkepohl, Helmut, Mika Meitz, Aleksei Netšunajev, and Pentti Saikkonen. 2021. "Testing identification via heteroskedasticity in structural vector autoregressive models." *The Econometrics Journal* 24 (1): 1–22.

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

Morley, James, and Jeremy Piger. 2012. "The asymmetric business cycle." *Review of Economics and Statistics* 94 (1): 208–221.

Nag, R, K Wong, and Frank Fallside. 1986. "Script recognition using hidden Markov models." In *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing,* 11:2071–2074. IEEE.

Pearson, Karl. 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.

Pelletier, Denis. 2006. "Regime switching for dynamic correlations." *Journal of econometrics* 131 (1-2): 445–473.

Qin, Anlong, and Zhongjun Qu. 2021. "Modeling Regime Switching in High-Dimensional Data with Applications to U.S. Business Cycles." *Working paper.*

Qu, Zhongjun, and Fan Zhuo. 2021. "Likelihood Ratio-Based Tests for Markov Regime Switching." *The Review of Economic Studies* 88 (2): 937–968.

Rabiner, Lawrence, and Biinghwang Juang. 1986. "An introduction to hidden Markov models." *ieee assp magazine* 3 (1): 4–16.

Rabiner, Lawrence R., and B. H. Juang. 1993. *Fundamentals of speech recognition.* Prentice Hall.

Resende, Marcelo. 2008. "Mergers and acquisitions waves in the UK: a Markov-switching approach." *Applied Financial Economics* 18 (13): 1067–1074.

Rodriguez Rondon, Gabriel, and Jean-Marie Dufour. 2022. "MSTest: An R-package for Testing Markov-Switching Models." *Working paper.*

Tippett, Leonard Henry Caleb, et al. 1931. "The methods of statistics." *The Methods of Statistics.*

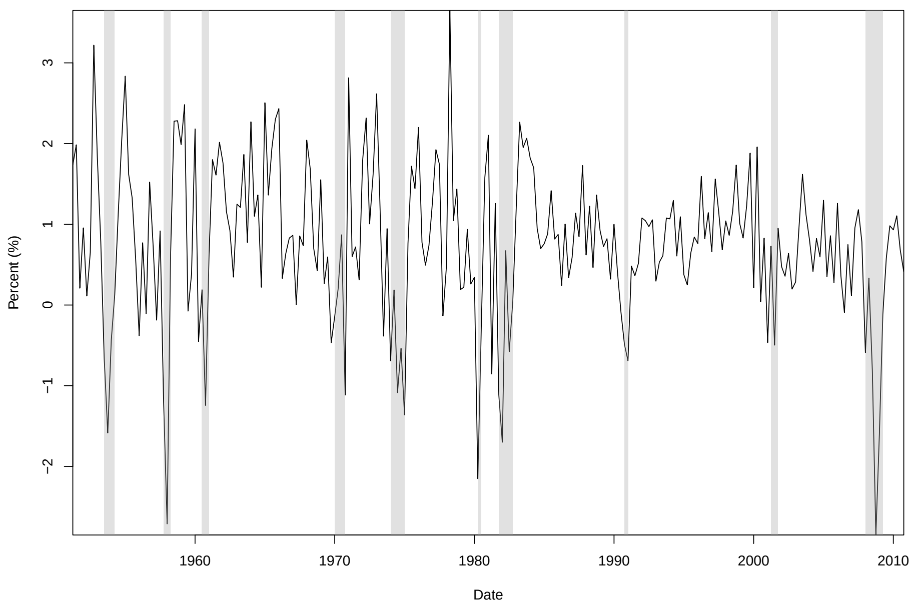
Warne, Anders, and Anders Vredin. 2006. "Unemployment and inflation regimes." *Studies in Nonlinear Dynamics & Econometrics* 10 (2).

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

Appendix

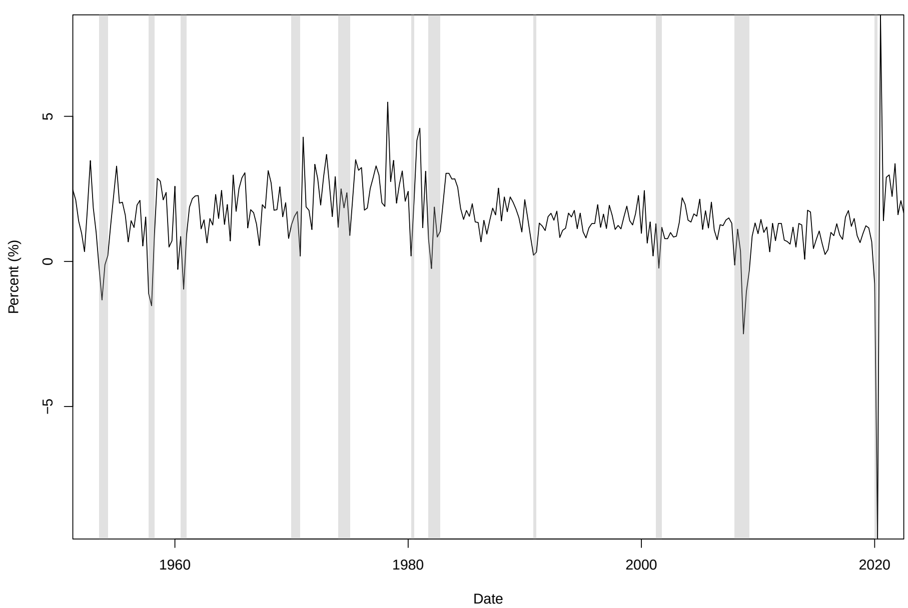
A: Figures

Figure A1: U.S. GNP % change from 1951Q2 - 2010Q4



Note: Highlighted region corresponds to NBER recession dates

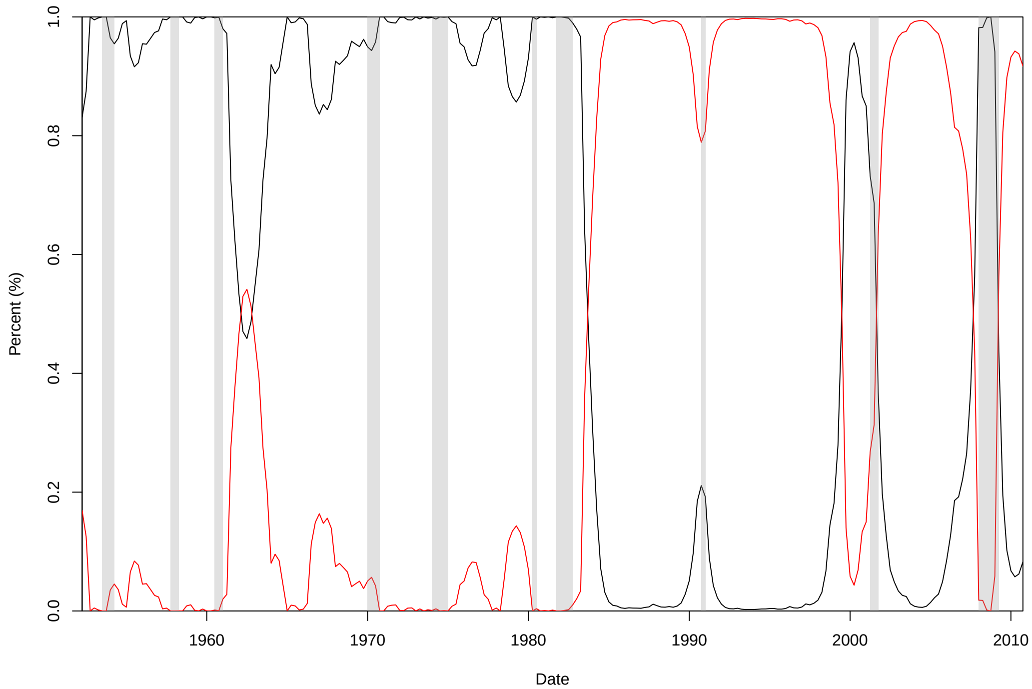
Figure A2: U.S. GNP % change from 1951Q2 - 2022Q3



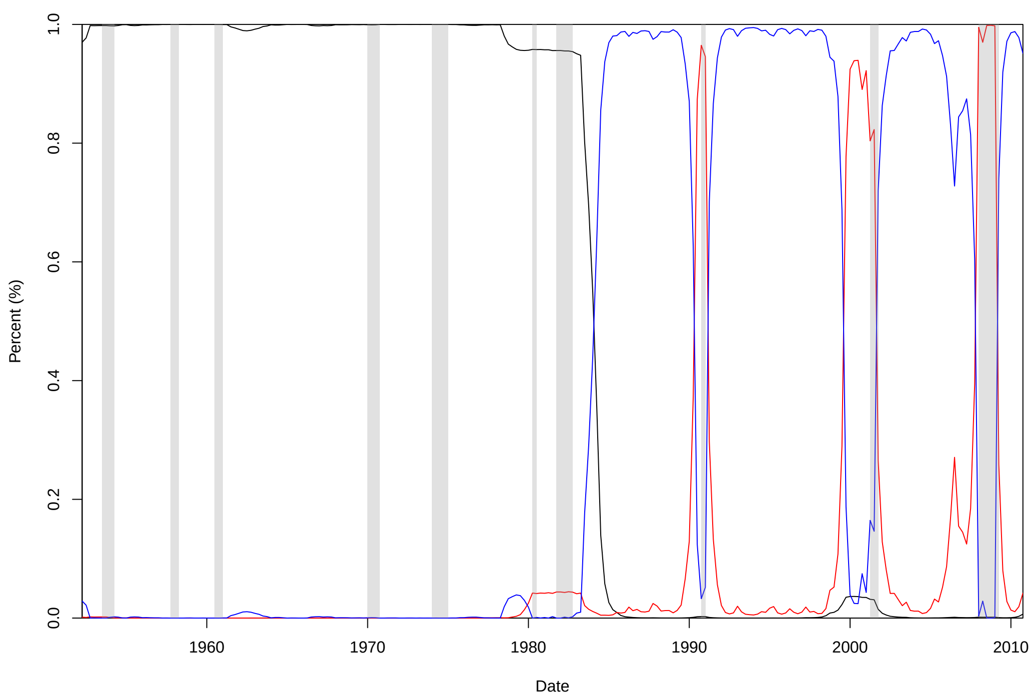
Note: Highlighted region corresponds to NBER recession dates

Figure A3: Smoothed Probabilities of Regimes for U.S. GNP % change from 1951Q2 - 2010Q4

Panel A: Markov switching model with $M = 2$ regimes



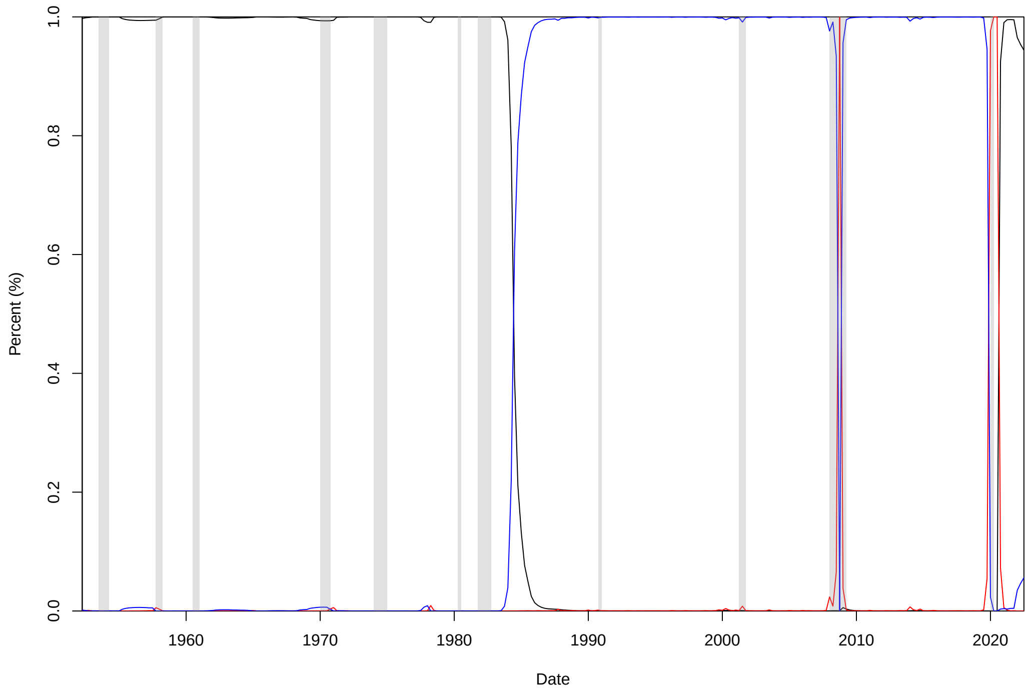
Panel B: Markov switching model with $M = 3$ regimes



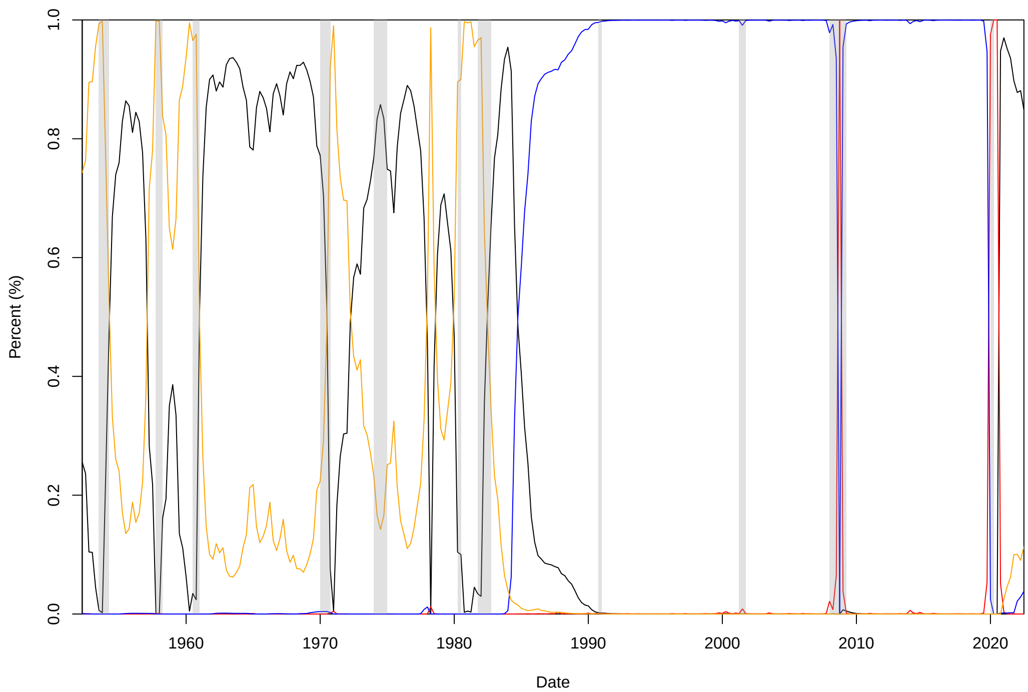
Note: Highlighted region corresponds to NBER recession dates

Figure A4: Smoothed Probabilities of Regimes for U.S. GNP % change from 1951Q2 - 2022Q3

Panel A: Markov switching model with $M = 3$ regimes



Panel B: Markov switching model with $M = 4$ regimes



Note: Highlighted region corresponds to NBER recession dates