Monte Carlo Likelihood Ratio Tests for Markov Switching Models *

Gabriel Rodriguez-Rondon† Jean-Marie Dufour‡

October 16, 2024

Abstract

Markov switching models have wide applications in economics, finance, and other fields. Most studies focusing on testing the number of regimes often focus on the null hypothesis of a single regime (i.e., a linear model with no switching) versus two regimes. Even in such simple cases, this type of problem raises issues of nonstandard asymptotic distributions, identification failure, and nuisance parameters. This paper proposes Monte Carlo likelihood ratio tests for Markov switching models, which address these issues and are applicable to more general settings where a null hypothesis with M_0 regimes can be tested against an alternative with $M_0 + m$ regimes where both $M_0 \geq 1$ and $m \geq 1$. This allows one to compare a broad class of Markov switching and Hidden Markov Models. Applied to likelihood ratio statistics, our approach overcomes the limitations of conventional tests, allowing for broader applicability to non-stationary processes, non-Gaussian errors, and multivariate settings, which have seen little attention in the literature. An important contribution is the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT), an identification-robust procedure valid in finite samples and asymptotically. Simulation results show that the proposed tests effectively control the level of the test and can provide good power across different settings. An empirical application to U.S. GNP and GDP growth data suggests a three-regime model, confirm evidence of the Great Moderation and identifying a return to the low volatility regime post-Great Recession and post-COVID-19. In a second multivariate application, we use our test procedures with Markov switching VAR models to test business cycle synchronization. Preliminary evidence suggests that adding COVID data weakens the synchronization between the U.S. and Canada.

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

JEL codes: C12, C15, C22, C52

*The authors thank Zhongjun Qu, Pierre Perron, Lynda Khalaf, John Galbraith, Hafedh Bouakez, as well as several other Boston University Econometrics seminar participants, Carleton Brown Bag Lunch seminar participants, and CIREQ Lunch seminar participants for useful comments. 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.rodriguezzrondon@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://jeanmariedufour.com.

Contents

1 Introduction

Markov regime-switching models were first introduced by Goldfeld and Quandt (1973) and later popularized by Hamilton (1989). They have since been widely used in economics and finance due to their ability to treat a series as a non-linear process where the non-linearity arises from discrete shifts. The process before and after a shift can be described as two separate regimes. For example, using U.S. GNP growth, one regime could characterize a period of positive growth, while the other represents a period of negative growth during recessions. Due to this flexibility, Markov switching models have become widely used in macroeconomics and finance. For instance, Markov switching models 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), state-dependent impulse response functions [see Sims and Zha 2006; Caggiano, Castelnuovo, and Figueres 2017], and the identification of structural VAR models [Herwartz and Lütkepohl 2014; Lanne, Lütkepohl, and Maciejowska 2010; Lütkepohl et al. 2021] to name a few. More comprehensive surveys of this literature can be found in Hamilton (2010), Hamilton (2016), and Ang and Timmermann (2012). Markov switching models also have applications outside the macroeconomic and financial literature. Examples include climate change [see Golosov et al. 2014; Dietz and Stern 2015], environmental and energy economics [see Cevik, Yıldırım, and Dibooglu 2021; Charfeddine 2017], industrial organization [see Aguirregabiria and Mira 2007; Sweeting 2013], and health economics [see Hernández and Ochoa 2016; Anser et al. 2021], among others.

An important issue with Markov switching models is that the number of 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 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). 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 the 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 Markov switching models has received significant attention [see Carter and Steigerwald 2012; Cho and White 2007; Garcia 1998; Hansen 1992; Kasahara and Shimotsu

2018; Qu and Zhuo 2021]. A very important and noteworthy contribution within this setting is the SupLR(Λ_ϵ) test of Qu and Zhuo 2021. Still, most available procedures, including those focusing on the likelihood ratio test approach, can only address settings where the null hypothesis assumes a linear model (i.e., $H_0 : M_0 = 1$) and the alternative hypothesis is a Markov switching model with two regimes (i.e., $H_1 : M_0 + m = 2$, where $M_0 = m = 1$). The only exception is Kasahara and Shimotsu (2018), who establish the asymptotic validity of the parametric bootstrap procedure for the likelihood ratio test statistic when the null hypothesis involves a model with M_0 regimes and the alternative hypothesis is a model with $M_0 + 1$ regimes, where $M_0 \geq 1$ and $m = 1$ (see Proposition 21). However, this result applies to a limited class of models and relies on restrictive assumptions. In addition to proposing the SupLR(Λ_ϵ) test, Qu and Zhuo (2021) also demonstrate the asymptotic validity of the parametric bootstrap procedure for a wider class of models (see Proposition 1 and section 7), but still within the setting where $M_0 = m = 1$ and also requires restrictive assumptions. Meanwhile, other researchers 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 their seminal work, Carrasco, Hu, and Ploberger (2014) propose an optimal test for assessing the consistency of parameters in random coefficient and Markov switching models. However their test is best suited for cases where the null hypothesis is of linear model and the alternative hypothesis is of a model with two regimes. This is because the null hypothesis of parameter stability assumes a linear mode. As a result, this test cannot be used to compare general Markov switching models where both M_0 and $m > 1$. Furthermore, this parameter stability test and all other likelihood ratio test approaches mentioned thus far, are valid only asymptotically. These tests aim to establish an asymptotic distribution of the test statistic, which means they depend on assumptions required to obtain asymptotic results, which can be restrictive in many cases. For example, a common assumption is that the process under study is stationary with Gaussian errors. In the likelihood ratio test literature, it is also common to assume a constrained parameter space to avoid the parameter boundary problem. In contrast, Dufour and Luger (2017) propose a moment-based approach using the Monte Carlo techniques described in Dufour (2006) that allows one to relax many of these assumptions. Specifically, the authors introduce test statistics based on the moments of the least-squares residuals, designed to capture different characteristics of a two-component mixture distribution. As a result, their procedure still only applies to the basic case where $M_0 = m = 1$, similar to other methods, but by using Monte Carlo techniques, they propose a test procedure that is valid in finite samples and avoids many of the restrictive assumptions mentioned earlier.

In this paper, we build on the Monte Carlo procedures described in Dufour (2006) and apply them to a likelihood ratio test setting for Markov switching models. This approach allows us to address the issues that plague conventional hypothesis testing procedures, such as nonstandard asymptotic distributions and nuisance parameters, within the likelihood ratio framework. Specifically, we propose the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT) for Markov switching models, which can be used to determine the number of regimes in both Markov switching models and Hidden Markov Models. These tests allow us to consider very general settings which include comparing models with M_0 regimes under the null hypothesis against models with $M_0 + m$ regimes under the alternative, where both $M_0 \geq 1$ and $m \geq 1$. Further, The MMC-LRT is an exact test and is valid both in finite samples and asymptotically. It is also robust to identification issues, which are common when dealing with Markov switching models. Both the LMC-LRT and MMC-LRT eliminate the need for conditions typically required for asymptotic validity of likelihood ratio tests. For instance, we no longer need to assume stationarity, Gaussian errors, or constrained parameter spaces. Both tests can even be applied in multivariate settings, such as Markov-switching VAR models or multivariate Hidden Markov models, which are settings that have not received attention in previous literature. Specifically, these tests do not rely on the existence of an asymptotic distribution and, as a result, can be applied in settings where previous test procedures, including the parametric bootstrap procedure, are not asymptotically valid or settings where the asymptotic validity simply hasn't been established in the literature.

Notably, non-stationary processes, non-Gaussian errors, multivariate settings, and cases $m > 1$ have received limited attention in the literature on hypothesis testing for the number of regimes in Markov switching models, making this study a novel contribution. Simulation results demonstrate that both the LMC-LRT and MMC-LRT effectively control the size of the test and exhibit better power in many settings where alternative tests are available (i.e., when $M_0 = m = 1$). Another contribution of this paper is the application of the test proposed by Dufour and Luger (2017) to non-stationary processes y_t, offering new insights into this scenario. All test results in this paper are implemented using the R package **MSTest**, described in a companion paper Rodriguez-Rondon and Dufour (2024). Since Markov switching models are more general than Hidden Markov models, we focus on mainly on Markov switching models throughout this study. However, the proposed tests are also applicable to hidden Markov models.

The next sections are structured as follows. Section 2 reviews notation, the Markov switching models we are interested in, and briefly discusses estimation procedures. Section 3 introduces our testing methodology. Section 4 presents and discusses simulation results for the size and power of the proposed testing procedures, while comparing them the tests of Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017). Section 5 presents two empirical applications. In the first we use the testing procedures proposed here to identify the number of regimes when modelling U.S. GNP as previously considered in Hansen (1992)., Carrasco, Hu, and Ploberger (2014), and Dufour and Luger (2017) but also consider U.S. GDP growth as in Qu and Zhuo (2021) and Kasahara and Shimotsu (2018). In doing so, we find evidence of a three-regime model where both the mean and the variance are subject to change. Our results confirm evidence of the Great Moderation and identifying a return to the low volatility regime post-Great Recession and post-COVID-19. We also consider controlling for the Great Moderation and COVID period as known structural breaks to asses whether a simpler model, with fewer regimes, can be justified, but find that a three-regime model is still preferred. In a second multivariate application, we use our test procedures with Markov switching VAR models to test business cycle synchronization. Preliminary evidence suggests that adding COVID data weakens the synchronization between the U.S. and Canada. Finally, section 6 provides concluding remarks.

2 Markov-switching Model

A Markov switching model is described as follows. Let (y_t, w_t) be a sequence of random vectors. The vector w_t is a finite-dimensional vector, and in this work, we allow y_t to be either a scalar (univariate setting) or a finite-dimensional vector (multivariate setting). Further, let $S_t = \{1, \dots, M\}$ be a latent variable that determines the regimes at time t and let s_t denote the (observed) realization of S_t. We define the information set $\mathscr{Y}_{t-1} = \sigma\text{-field}\{\dots, w_{t-1}, y_{t-2}, w_t, y_{t-1}\}$. The Markov switching model can be expressed as

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

where, in a univariate setting, y_t is a scalar, x_t is a $(1 \times q_x)$ vector of variables whose coefficients do not depend on the latent Markov process S_t, z_t is a $(1 \times q_z)$ vector of variables whose coefficients do depend on the Markov process S_t, and ϵ_t is the error process. The number of regressors, q_x, that remain constant, and the number of regressors that change with S_t, q_z, must sum to $q = (q_y \times p) + q_w$, where $q_y = 1$ in the univariate setting or larger than 1 in the multivariate setting, p is the number of lags in the model, and q_w is the number of exogenous regressors. As can be seen from (1), the variance can also change according to the Markov process S_t. We can group all parameters in $\theta^{s_t} = (\beta, \delta_{s_t}, \sigma_{s_t}, vec(\mathbf{P}))$, where $vec(\cdot)$ is the vectorization operator that transforms a matrix to a vector, and $\mathbf{P}$ is the transition matrix, which we describe in more detail below. When considering the multivariate setting, we then have a covariance matrix Σ_{s_t} and make use of the $vech(\cdot)$ operator, which takes the values under and on the main diagonal of the matrix since, given the symmetry, these are the only parameters needed to summarize the covariance structure. In this case, β and δ_{s_t} are matrices and so we must use $vec(\beta)$ and $vec(\delta_{s_t})$ in θ^{s_t}.

We can assume, for example, that the error process is distributed as a $\mathcal{N}(0, I_{q_y})$. It is important to note, however, that for the testing procedure we propose below, the assumption of normality is not required, and other distributions can be considered instead by simply using the appropriate likelihood density function. Alternatively, even if the error process is not normally distributed, we can continue to use the normal density function. In this case, the test presented below becomes better described as a pseudo-Monte Carlo Likelihood Ratio Test for Markov switching models. However, as will be described in the next section, the test is still valid in this case and in other cases where the likelihood function may not be well-defined. For this reason, and for simplicity, we continue to present the model using this normality assumption in what follows.

A Markov switching model is typically described as having lags of y_t as explanatory variables. That is, lags must be included in either x_t or z_t depending on whether we want the autoregressive coefficients to change across regimes. This setting is very general and allows us to consider a trend function within x_t or z_t. On the other hand, Hidden Markov models typically do not include lags of the dependent variable. However, the dependence on past observations allows for more general interactions between the dependent variable and the Markov process S_t, which can be used to model more complex causal links between our variables of interest. Hence, a Hidden Markov model can be understood as a simplified version of a Markov switching model, and for this reason, we focus on the more general Markov switching case. Nonetheless, the results presented here still apply to Hidden Markov models, which, as previously discussed, have a wide range of interesting applications.

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, $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$ is the probability of state i being followed by state j. The columns of the transition matrix must sum to one to have a well-defined transition matrix (i.e., $\sum_{j=1}^{M} p_{ij} = 1$, $\forall i$). We can also obtain the ergodic probabilities, $\pi = (\pi_1, \dots, \pi_M)'$, which are given by

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

where $\mathbf{e}_{M+1}$ is the $(M+1)$th column of $\mathbf{I}_{M+1}$. These ergodic probabilities can be understood as representing, in the long-run on average, the proportion of time spent in each regime.

Let $f(y_t|\mathscr{Y}_{t-1};\theta)$ denote the conditional density of y_t given $\mathscr{Y}_{t-1}$, and assume it satisfies

$$y_t|(\mathscr{Y}_{t-1}, s_t) \sim \begin{cases} f(y_t|\mathscr{Y}_{t-1};\theta^1), & \text{if } \ s_t = 1 \\ \qquad\qquad \vdots \\ f(y_t|\mathscr{Y}_{t-1};\theta^M), & \text{if } \ s_t = M \end{cases} \tag{2}$$

for $t = 1, \dots, T$. The sample log likelihood conditional on the first p observations of y_t is given by

$$L_T(\theta) = \log f(y_1^T|y_{-p+1}^0;\theta) = \sum_{t=1}^{T} \log f(y_t|\mathscr{Y}_{t-1};\theta) \tag{3}$$

where $\theta = (\beta, \delta_1, \dots, \delta_M, \sigma_1, \dots, \sigma_M, vec(\mathbf{P}))$, and where the $vec(\cdot)$ operator should also be applied to β, δ_{s_t}, and Σ_{s_t} if working with a multivariate model. Here,

$$f(y_t|\mathscr{Y}_{t-1};\theta) = \sum_{s_t=1}^{M}\sum_{s_{t-1}=1}^{M}\cdots\sum_{s_{t-p}=1}^{M} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \dots, S_{t-p} = s_{t-p}|\mathscr{Y}_{t-1};\theta) \tag{4}$$

and more specifically

$$f(y_t, S_t = s_t, \ldots, S_{t-p} = s_{t-p} | \mathscr{Y}_{t-1}; \theta) = \frac{\Pr(S_t^* = s_t^* | \mathscr{Y}_{t-1}; \theta)}{\sqrt{2\pi\sigma_{s_t}^2}} \times \exp\left\{ \frac{-[y_t - x_t\beta - z_t\delta_{s_t^*}]^2}{2\sigma_{s_t}^2} \right\} \tag{5}$$

where we set

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

and $\Pr(S_t^* = s_t^* | \mathscr{Y}_{t-1}; \theta)$ is the probability that this occurs.

An alternative but related model is the Hidden Markov Model. Like Markov switching models, Hidden Markov models are used to describe a process y_t which depends on a latent Markov process S_t, but as discussed in An et al. (2013), these models are used in the case where the process y_t does not depend on its own lags. However, 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. As a result, Hidden Markov models are a special, more simple, case of Markov switching models and so the hypotheses testing procedure proposed in the next section will also apply to these models. Still, it is worth noting that Hidden Markov models 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.

Typically, Markov switching and Hidden Markov models 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, Markov switching models can also be estimated using Maximum Likelihood Estimation (MLE). However, since the Markov process S_t is unobservable, and more importantly, the likelihood function can have several modes of equal height, along with other unusual features that can complicate estimation by MLE, this approach is not often used, except for simple cases where M is small (e.g., $M = 2$). In this study, when necessary, we use the EM algorithm for estimating Markov switching models. It is worth noting that, in practice, empirical estimates can sometimes be improved by using the results of the EM algorithm as initial values in a Newton-type optimization algorithm. This two-step estimation procedure is used to obtain the 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 next. For the interested reader, the estimation of a Markov switching model via the EM algorithm is described in detail in Hamilton (1990) and Hamilton (1994), as well as in Krolzig (1997) for the 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) for Markov switching models, which we propose in this paper. Similar to Garcia (1998) and the parametric bootstrap procedures described in Qu and Zhuo (2021) and Kasahara and Shimotsu (2018), when parameters are not identified under the null hypothesis, we assume that the null distribution depends only on the remaining parameters. The LRT approach requires us to estimate the model under both the null and alternative hypotheses in order to obtain the log-likelihoods for each model. The log-likelihood for models with $M > 1$ regimes is given by equations (3) - (5):

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

where

$$\theta_i = (\beta, \delta_1, \ldots, \delta_M, \sigma_1, \ldots, \sigma_M, vec(\mathbf{P}))' \in \bar{\Omega}_i \,. \tag{6}$$

The subscript of i underscores the fact that θ_i represents the parameter vector under the null hypothesis when $i = 0$, or under the alternative hypothesis when $i = 1$. Note that in a multivariate setting, we simply treat β, δ_{s_t}, and Σ_{s_t} as matrices, and apply the $vec(\cdot)$ operator to vectorize them, as discussed in the previous section. The set $\bar{\Omega}_i$ satisfies any theoretical restrictions we wish to impose on θ_i (e.g., $\sigma_i > 0$). For example, as noted by Qu and Zhuo (2021) and Kasahara and Shimotsu (2018), for the asymptotic validity of the parametric bootstrap and the SupLR(Λ_ϵ), we would need to impose that $p_{i,j} \in (\epsilon, 1-\epsilon)$ on $\bar{\Omega}_i$. However, in our setting, this restriction is not necessary. When we consider the null hypothesis with $M = 1$, the log-likelihood is given by

$$L_T^0(\theta_0) = \log f(y_1^T \,|\, y_{-p+1}^0;\, \theta_0) = \sum_{t=1}^{T} \log f(y_t \,|\mathscr{Y}_{t-1};\, \theta_0) \tag{7}$$

where

$$f(y_t\,|\mathscr{Y}_{t-1};\,\theta_0) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{\frac{-[y_t - x_t\beta]^2}{2\sigma^2}\right\}\,, \tag{8}$$

$$\theta_0 = (\beta,\,\sigma)' \in \bar{\Omega}_0. \tag{9}$$

Here, δ_{s_t} and $vec(\mathbf{P})$ are excluded because there are no parameters that change under the null hypothesis of no Markov regime- switching. Also, note that in general $\bar{\Omega}_0$ has a lower dimension than $\bar{\Omega}_1$.

For simplicity of exposition, consider first the straightforward and common scenario where we want to compare a null hypothesis of $M_0 = 1$ regime (i.e., no Markov switching) against an alternative hypothesis of a Markov switching model with $M_0 + m = 2$ regimes. In this case, the null and alternative hypotheses can be expressed as:

$$H_0 : \delta_1 = \delta_2 = \delta \quad \text{for some unknown } \delta\,, \tag{10}$$

$$H_1 : (\delta_1,\,\delta_2) = (\delta_1^*,\,\delta_2^*) \;\; \text{for some unknown } \delta_1^* \neq \delta_2^*\;, \tag{11}$$

where δ_i includes any parameter we consider to be governed by the Markov process S_t. In general, when $M_0 \geq 1$ and $m \geq 1$, we need to consider different combinations of restrictions under the null hypothesis. For example, when considering $H_0 : M_0 = 2$ against $H_1 : M_0 + m = 3$, we must account for the following cases: i. $\delta_1 = \delta_2$ and $\delta_1 \neq \delta_3$, ii. $\delta_1 = \delta_3$ and $\delta_1 \neq \delta_2$, or iii. $\delta_2 = \delta_3$ and $\delta_2 \neq \delta_1$. Using the likelihood ratio test statistic allows us to consider these combinations directly by comparing the likelihoods of the null and alternative hypotheses. For this reason, and for convenience, we continue with the notion of comparing $H_0 : M_0$ against $H_1 : M_0 + m$, where both M_0 and $m \geq 1$.

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

$$L_T^0(\theta_0) = L_T(\theta_1)\,, \quad \theta_1 \in \Omega_0, \tag{12}$$

where Ω_0 is the subset of vectors $\theta_1 \in \bar{\Omega}_1$ such that θ_1 satisfies H_0. Under H_0, the vector $\theta_0 \in \bar{\Omega}_0$ consists of nuisance parameters: the null distribution of any test statistic for H_0 depends on $\theta_0 \in \bar{\Omega}_0$. In this context, 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 be expressed as

$$LR_T = 2[\bar{L}_T(H_1) - \bar{L}_T(H_0)] \tag{13}$$

where

$$\bar{L}_T(H_1) = \sup\{L_T(\theta_1) : \theta_1 \in \bar{\Omega}_1\}\,, \tag{14}$$

$$\bar{L}_T(H_0) = \sup\{L_T^0(\theta_0) : \theta_0 \in \bar{\Omega}_0\} = \sup\{L_T(\theta_1) : \theta_1 \in \Omega_0\}\,. \tag{15}$$

The null distribution of LR_T depends on the parameter $\theta_0 \in \bar{\Omega}_0$. Now, let $LR_T^{(0)}$ denote a real random variable, computed from observed data when the true parameter vector is θ_0. Since the model in (1) is parametric, we can use it to generate a vector of N i.i.d. replications of LR_T for any given value of $\theta_0 \in \bar{\Omega}_0$:

$$LR(N,\,\theta_0) := [LR_T^{(1)}(\theta_0),\,\ldots\,,LR_T^{(N)}(\theta_0)]', \qquad \theta_0 \in \bar{\Omega}_0\,. \tag{16}$$

That is, we will assume that

Assumption 3.1 $LR_T^{(0)}$ *is a real random variable and* $LR(N,\,\theta_0)$ *a real random vector, all defined on a common probability space* $(\mathcal{F}, \mathscr{Y}_{t-1}, P_{\theta_0})$ *such 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* $F[x\,|\,\theta_0]$*.*

Note that generating N i.i.d. replications of LR_T using (1) requires knowledge of the distribution of ϵ_t. The procedure proposed here is quite general, allowing us to consider any distribution for ϵ_t, including non-Gaussian distributions. In the case of non-Gaussian distributions, we simply need to use the appropriate likelihood function in (3) - (5) or (7) - (8). However, even when the distribution of ϵ_t is non-Gaussian or unknown, we can continue to work with the Gaussian density function. In such cases, we refer to this approach as Monte Carlo pseudo-likelihood ratio tests. Next, we define

$$\hat{F}_N[x\,|\,\theta_0] := \hat{F}_N[x;\, LR(N,\,\theta_0)] = \frac{1}{N}\sum_{i=1}^{N} I[LR_T^{(i)}(\theta_0) \leq x] \tag{17}$$

$$\hat{G}_N[x\,|\,\theta_0] := \hat{G}_N[x;\, LR(N,\,\theta_0)] = 1 - \hat{F}_N[x;\, LR(N,\,\theta_0)] \tag{18}$$

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

$$\hat{p}_N[x \mid \theta_0] = \frac{N\hat{G}_N[x \mid \theta_0] + 1}{N+1}\,. \tag{19}$$

Alternatively, using the relationship

$$\begin{aligned} R_{LR}[LR_T^{(0)};\, N] &= N\hat{F}_N[x;\, LR(N,\, \theta_0)] \\ &= \sum_{i=1}^{N} I[LR_T^{(0)} \geq LR_T^{i}(\theta_0)] \end{aligned} \tag{20}$$

we can define a Monte Carlo p-value as

$$\hat{p}_N[x \mid \theta_0] = \frac{N+1-R_{LR}[LR_T^{(0)};\, N]}{N+1} \tag{21}$$

where, as can be seen from (20), $R_{LR}[LR_T^{(0)}; N]$ simply computes the rank of the test statistic using the observed data within the generated series $LR(N, \theta_0)$. We also make the following assumption,

Assumption 3.2 *Let* $\sup\{\hat{G}_N[LR_T^{(0)}|\theta_0] \,:\, \theta_0 \in \bar{\Omega}_0\}$ *and* $\inf\{\hat{F}_N[LR_T^{(0)}|\theta_0] \,:\, \theta_0 \in \bar{\Omega}_0\}$ *be* $\mathscr{Y}_{t-1}$*-measurable and where* $\bar{\Omega}_0$ *is a nonempty subset of* Ω*.*

Now, we can make the following proposition

Proposition 3.1 *(Validity of MMC-LRT for Markov switching models). Let* $LR_T^{(0)}(\theta_0) = LR_T^{(0)}$*,* $\alpha(N+1)$ *be and integer, and suppose*

$$Pr[LR_T^{(i)} = LR_T^{(j)}] = 0 \;\; \textit{for} \;\; i \neq j, \;\;\; i,j = 1,\ldots,N. \tag{22}$$

Using assumptions 3.1 and 3.2, if $\theta_0 \in \bar{\Omega}_0$*, then for* $0 \leq \alpha_1 \leq 1$*,*

$$\Pr[\sup\{\hat{G}_N[LR_T^{(0)}|\theta_0] : \theta_0 \in \bar{\Omega}_0\} \leq \alpha_1] \leq \Pr[\inf\{\hat{F}_N[LR_T^{(0)}|\theta_0] : \theta_0 \in \bar{\Omega}_0\} \geq 1-\alpha_1] \tag{23}$$

$$\leq \frac{I[\alpha_1 N]+1}{N+1} \tag{24}$$

where $\Pr[\inf\{\hat{F}_N[LR_T^{(0)}|\theta_0] : \theta_0 \in \bar{\Omega}_0\} \geq 1-\alpha_1] = \Pr[LR_T^{(0)} \geq \sup\{\hat{F}_N^{-1}[1-\alpha_1|\theta_0] : \theta_0 \in \bar{\Omega}_0\}]$ *for* $0 < \alpha_1 < 1$ *and so*

$$\Pr[\sup\{\hat{p}_N[LR_T^{(0)}|\theta_0] : \theta_0 \in \bar{\Omega}_0\} \leq \alpha] \leq \alpha \;\; \textit{for} \;\; 0 \leq \alpha \leq 1. \tag{25}$$

where the last line follows from using (19), setting $\alpha_1 = \alpha - \frac{(1-\alpha)}{N}$, and noting that $\alpha = \frac{I[\alpha(N+1)]}{N+1}$

whenever α and N are chosen such that $\alpha(N+1)$ is an integer, as assumed. Additionally, here, $\hat{F}^{-1}$ denotes the quantile function of $\hat{F}$. In this context, we refer to this procedure as the Maximized Monte Carlo Likelihood Ratio Test for Markov switching models, and this proposition establishes the validity of the test. This follows from Proposition 4.1 in Dufour (2006), so the proof directly relies on the proof of Proposition 4.1.

This procedure is referred to as the Maximized Monte Carlo likelihood ratio test because (25) is maximized with respect to $\theta_0 \in \bar{\Omega}0$. However, this parameter space can be very large, specifically growing with the number of regressors considered and the number of regimes. Additionally, the solution may not be unique, as the maximum p-value could be obtained by more than one parameter vector. For this reason, numerical optimization methods that do not rely on 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]. As described in Dufour (2006), to facilitate optimization, it is also possible to search within a smaller consistent subset of the parameter space, denoted as 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

$$C_T = \{\theta_0 \in \bar{\Omega}_0 : \| \hat{\theta}_0 - \theta_0 \| < c\} \tag{26}$$

where c is a fixed positive constant that does not depend on T and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$.

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) for Markov switching models. 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. 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 Local Monte Carlo p-value is given by

$$\hat{p}_{TN}[x \mid \theta_0] = \frac{N\hat{G}_{TN}[x \mid \theta_0] + 1}{N + 1} \tag{27}$$

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, the LMC test can be interpreted as the finite-sample analogue of the parametric bootstrap. This is because, 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. Instead, we work with the critical values from the sample distribution $\hat{F}[x \mid \theta_0]$.

To be more specific, the MMC-LRT procedure will be valid even when an asymptotic distribution does not exist and the LMC-LRT procedure will also be valid as $T \to \infty$ if this is the case. This means the tests proposed here are much more general than the parametric bootstrap procedure as validity does not require stationarity or working with constrained parameter spaces, which are needed to obtain its asymptotic validity in the likelihood ratio setting (see Qu and Zhuo 2021 and Kasahara and Shimotsu 2018 for example). In most cases, these assumptions are needed because otherwise the likelihood function may not be well-defined. These are cases where our procedure may again be better described as Monte Carlo pseudo-likelihood ratio test procedures. Further, we are directly able to deal with cases where $m > 1$, non-Gaussian settings, and multivariate settings where the asymptotic validity of the parametric bootstrap procedure has simply not yet been established in the literature. Finally, this also 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 (21) and (27), 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.

At this point we have introduced the MMC-LRT and LMC-LRT for Markov switching models. We have also described how these tests are more general than the parametric bootstrap procedure and how they are useful even in settings where y_t is a vector (multivariate setting), y_t is non-stationary, and ϵ_t is non-Gaussian. For hypothesis testing, the generality of our procedure even extends to settings where $m > 1$, ensuring finite-sample validity for the MMC-LRT procedure, and does not require working with a constrained parameter space.

We believe this third feature is especially important because there may be cases where $L_T^0(\theta 0) = LT(\theta_1)$ for values $\theta_1 \in \Omega_0$ that lie on the boundary. Consider, for example, a scenario where $M = 2$ and $p_{1,1}, p_{2,1} \to 1$. In this case, the Markov switching model with $M = 2$ may be statistically equivalent to a one-regime (no Markov switching) model. Generally, similar arguments can be made for cases where $M > 2$. As a result, we believe allowing parameters, specifically transition probabilities, to take values on the boundary is an important feature for comparing M_0 with M_0+m regimes.

Another important aspect to consider is the case where regressors are weakly exogenous. So far, we have discussed simulating the test statistic by using the parametric model in (1) and i.i.d. replication of ϵ_t. In many applications of Markov switching models, where only lags of the observed data y_t are included as explanatory variables, this works perfectly fine. In fact, even in cases where other regressors are included, as long as they are fixed or strictly exogenous so that we can treat them as fixed in this context, we can proceed as previously discussed. However, as discussed in Qu and Zhuo (2021), for the parametric bootstrap procedure, weakly exogenous regressors can lead to size distortions. The same can be true for the LMC-LRT procedure proposed here. In such settings, if the joint distribution of the dependent variable and regressors is unknown, we propose assuming some functional form (e.g., an AR(p) model), use this relationship to jointly simulate them, and then proceed as previously discussed.

4 Simulation Evidence

This section presents simulation evidence on the performance of the Local Monte Carlo (LMC-LRT) and Maximized Monte Carlo Likelihood Ratio Tests (MMC-LRT) for Markov switching models that are proposed here. Throughout, we will consider DGPs with the following form

$$y_t = \mu_{s_t} + \phi_1(y_{t-1} - \mu_{s_{t-1}}) + \sigma_{s_t}\epsilon_t \tag{28}$$

where $\epsilon_t \sim \mathcal{N}(0,1)$, the mean and variance are allowed to switch according to the Markov process S_t. Similar data generating processes (DGPs) have been considered in Carrasco, Hu, and Ploberger

(2014), Dufour and Luger (2017), and Qu and Zhuo (2021). We specifically use some of the same DGPs as Dufour and Luger (2017) to present simulation evidence for a wide range of scenarios, including low persistence, high persistence, symmetric regimes, asymmetric regimes, changes in the mean only, changes in the variance only, and cases where both the mean and variance change simultaneously. Additionally, given the generality of our test procedure, we also consider cases where $M_0 > 1$ and $m = 1$, $M_0 = 1$ and $m > 1$, both $M_0 > 1$ and $m > 1$, when $\phi_1 = 1.00$, and when transition probabilities take values at the boundary of the parameter space (e.g., $p_{22} = 1$). We also consider three different sample sizes: $T = 100$, $T = 200$, and $T = 500$. We believe that these DGPs cover many empirically relevant settings that a researcher might encounter. For example, smaller sample sizes and asymmetric regimes may be of particular interest for macroeconomic applications, where quarterly observations are used, and some regimes are relatively short-lived. For cases where we consider a linear model under the null hypothesis (i.e., $H_0 : M_0 = 1$) against a Markov switching model with two regimes under the alternative hypothesis (i.e., $H_1 : M_0 + m = 2$), we compare the performance of our proposed test procedures with those of Dufour and Luger (2017) and Carrasco, Hu, and Ploberger (2014).

The tests proposed by Dufour and Luger (2017) are also based on the Monte Carlo procedure described in Dufour (2006), but they avoid some of the statistical issues associated with likelihood ratio tests by using the moments of the residuals from the restricted model. These moments aim to capture characteristics of a mixture normal distribution. The test relies on four moments of the residuals, resulting in four Monte Carlo (MC) p-values. To combine these p-values, two methods are proposed, which can broadly be understood as being based on either the minimum or the product of the four p-values. Dufour et al. (2004) and Dufour, Khalaf, and Voia (2014) provide further discussion on these methods of combining test statistics for interested readers. As a result, four test procedures are proposed in Dufour and Luger (2017): LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$. An advantage of these procedures is that they only require estimating the linear model without Markov switching under the null hypothesis. However, unlike the LMC-LRT and MMC-LRT, these tests can only be used to compare such linear models under the null with a Markov switching model with two regimes under the alternative.

Carrasco, Hu, and Ploberger (2014) propose a test that is optimal for assessing the consistency of parameters in random coefficient and Markov switching models. Their testing procedure is generally suited to detect parameter heterogeneity, with the Markov switching model included as a special case. Similar to the moment-based approach of Dufour and Luger (2017), a key advantage of this method is that it only requires estimating the model under the null hypothesis, but again only applies to the simple case where there is no Markov switching under the null hypothesis. To address the presence of nuisance parameters, the authors propose two alternatives: the first is a Sup-type test, referred to as supTS, as in Davies (1987), and the second is an Exponential-type test, referred to as expTS, as in Andrews and Ploberger (1994). As with Dufour and Luger (2017), when applying the supTS and expTS tests, we consider values of ρ in the interval $[\underline{\rho}, \overline{\rho}] = [-0.7, 0.7]$.

As previously mentioned, the consistency of the parametric bootstrap procedure when $m = 1$ is shown by Qu and Zhuo (2021) for $M_0 = 1$, and by Kasahara and Shimotsu (2018) for $M_0 > 1$ under more restrictive assumptions than those required for the test procedures proposed here. Specifically, using the parametric bootstrap test requires constraining the parameter space away from the boundary when simulating the null distribution. Furthermore, its consistency has only been demonstrated for univariate, stationary, and Gaussian settings, though Kasahara and Shimotsu (2018) consider some non-Gaussian settings also. In Kasahara and Shimotsu (2018), the authors additional make use of other constraints on the variance parameters when estimating the models. Given the similarity between the LMC-LRT and parametric bootstrap procedures for many of the DGPs considered here—particularly when the process is stationary and parameters are sufficiently far from the boundary—we do not include results from a parametric bootstrap procedure where such constraints would be enforced. Moreover, we believe the results of the LMC-LRT procedure provided below will shed light on the performance of the parametric bootstrap procedure when its required assumptions are met, as well as in cases where its consistency has not yet been established in the literature. It is important to emphasize, however, that the primary distinction between these two procedures in these specific scenarios lies in the estimation of the null distribution and, more fundamentally, in the differing assumptions about the existence and approximation of an asymptotic distribution.

All the test procedures discussed so far, including those proposed in this work, can be easily implemented using the R package **MSTest** (see Rodriguez-Rondon and Dufour 2024), which is available online through the Comprehensive R Archive Network (CRAN) and described in a companion paper by Rodriguez-Rondon and Dufour (2024). All simulation results provided below were computed using this R package. For these simulations, the nominal significance level is set at $\alpha = 0.05$, and the results are based on $1,000$ replications of the DGP.

The results under the null hypothesis of no Markov switching (i.e., $H_0 : M_0 = 1$) are reported in

Table 1: Empirical size of test when $H_0 : M_0 = 1$

Test	$\phi = 0.10$			$\phi = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500
			$H_1 : M_0 + m = 2$			
LMC-LRT	4.9	4.7	4.9	5.3	5.0	4.9
MMC-LRT	1.9	1.5	1.3	0.8	0.7	0.8
LMC$_{min}$	5.0	3.8	5.5	5.1	4.2	5.5
LMC$_{prod}$	4.0	4.1	4.6	4.7	4.3	4.8
MMC$_{min}$	1.7	1.3	4.3	1.3	1.7	4.1
MMC$_{prod}$	1.6	1.8	3.6	1.4	2.5	3.8
supTS	4.8	5.1	4.8	6.0	4.5	4.7
expTS	6.8	6.2	5.2	5.4	6.9	5.5
			$H_1 : M_0 + m = 3$			
LMC-LRT	5.2	5.4	4.8	4.6	4.1	5.3
MMC-LRT	2.5	2.3	1.5	1.2	0.8	1.0

Notes: The nominal level is 5%. LMC-LRT and MMC-LRT are the Local Monte Carlo and Maximized Monte Carlo Likelihood Ratio Tests proposed here, respectively. Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Table 1. The table includes two panels: the first considers the alternative hypothesis of a Markov switching model with two regimes, while the second panel explores the alternative hypothesis of a Markov switching model with three regimes. The rejection frequencies of the LMC-LRT proposed here are remarkably close to the nominal level of the test. As suggested by theory, the MMC-LRT proposed here has empirical rejection frequencies $\leq 5\%$ under the null hypothesis. The results of the moment-based test by Dufour and Luger (2017), namely LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$, are consistent with the results of our Monte Carlo likelihood ratio tests. However, the expTS test shows mild over-rejection in some cases with smaller sample sizes but performs very well when $T = 500$. This is expected, as it is an asymptotic test procedure. In contrast, the supTS test demonstrates excellent size properties.

To study the power properties, we consider DGPs with transition probabilities $(p_{11}, p_{22}) = (0.90, 0.90)$ and $(p_{11}, p_{22}) = (0.90, 0.50)$. In both cases, the other transition probabilities are given by $p_{ij} = (1 - p_{ii})$ for $j \neq i$. In the first case, both regimes are symmetric and relatively persistent. Given their symmetry, the vector $\boldsymbol{\pi} = (\pi_1, \pi_2) = (0.50, 0.50)$, meaning that, on average, equal time is spent in each regime in the long run. In contrast, the second case is asymmetric, where one regime is more persistent than the other, resulting in $\boldsymbol{\pi} = (0.83, 0.17)$ and more time spent in one regime on average. Table 2 reports the empirical power of the tests. Since the MMC-LRT procedure considers a wider set of nuisance parameter values compared to the LMC-LRT procedure, its power is lower in all cases. This also applies to the moment-based approach. The LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$ procedures exhibit the lowest power when only the mean changes and persistence is

Table 2: Empirical Power of Test when $M_0 = 1$, $m = 1$

Test	$(p_{11}, p_{22}) = (0.90, 0.90)$						$(p_{11}, p_{22}) = (0.90, 0.50)$					
	$\phi = 0.10$			$\phi = 0.90$			$\phi = 0.10$			$\phi = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500	T=100	T=200	T=500	T=100	T=200	T=500
						$\Delta\mu$						
LMC-LRT	60.2	88.6	98.3	14.7	20.5	43.9	24.9	51.3	92.8	21.4	39.3	74.6
MMC-LRT	58.0	81.7	90.0	7.5	14.7	31.3	21.6	42.3	84.5	14.0	30.0	62.0
LMC_{min}	5.3	5.4	3.7	14.5	20.9	42.1	14.8	30.2	70.6	13.7	18.8	40.3
LMC_{prod}	4.8	4.3	4.3	16.2	22.3	43.0	12.3	24.0	56.4	14.3	20.5	42.9
MMC_{min}	1.1	2.3	1.9	6.7	13.2	33.8	6.7	20.5	61.5	5.7	11.0	31.9
MMC_{prod}	0.9	2.4	2.0	6.9	14.5	34.2	7.0	16.5	49.2	6.6	12.9	35.7
supTS	36.4	64.0	96.5	5.5	3.9	6.1	7.6	7.1	11.3	5.7	8.4	24.0
expTS	35.6	60.9	95.4	5.4	3.9	6.4	7.3	8.6	11.7	8.0	9.2	22.6
						$\Delta\sigma$						
LMC-LRT	52.4	84.1	99.8	46.0	80.9	99.8	42.1	69.0	96.2	38.7	65.5	95.1
MMC-LRT	41.8	79.7	92.6	38.0	76.8	94.3	39.1	61.3	93.2	32.9	58.0	91.3
LMC_{min}	38.1	63.6	95.5	39.5	63.3	95.2	47.8	72.7	95.5	47.4	72.2	95.6
LMC_{prod}	40.5	66.3	96.3	39.7	66.5	96.5	48.9	72.9	95.4	48.8	72.8	95.1
MMC_{min}	25.8	51.8	92.9	24.8	52.4	92.6	35.0	65.2	94.1	33.1	65.3	94.2
MMC_{prod}	28.9	57.7	95.1	27.3	57.5	94.3	35.8	64.8	94.1	34.8	65.6	94.3
supTS	32.4	58.0	98.9	32.2	67.4	91.6	29.9	46.4	94.7	30.0	50.3	92.1
expTS	40.1	62.6	99.3	54.1	84.7	92.2	43.9	68.3	95.2	52.8	78.6	93.6
						$\Delta\mu$ & $\Delta\sigma$						
LMC-LRT	81.2	98.7	100.0	39.5	70.0	98.7	77.5	97.2	100.0	58.0	87.3	99.3
MMC-LRT	78.0	94.5	100.0	25.6	66.0	96.0	74.3	96.0	100.0	48.7	79.2	96.0
LMC_{min}	53.1	80.9	99.4	35.3	60.7	92.6	84.7	97.8	100.0	66.9	89.9	99.5
LMC_{prod}	46.1	74.1	98.7	38.7	63.9	95.3	84.6	98.3	100.0	69.2	91.9	99.7
MMC_{min}	37.2	69.6	99.0	22.9	49.3	89.4	74.6	96.0	100.0	52.2	85.4	99.3
MMC_{prod}	34.2	66.0	98.1	26.3	55.5	92.7	74.9	97.0	100.0	56.0	88.1	99.7
supTS	74.0	96.0	100.0	34.0	62.9	95.4	78.0	98.0	100.0	54.0	83.3	99.4
expTS	73.3	92.0	100.0	45.6	76.0	97.0	80.0	98.3	100.0	56.2	83.4	99.7

Notes: The nominal level is 5%. LMC-LRT and MMC-LRT are the Local Monte Carlo and Maximized Monte Carlo Likelihood Ratio Tests proposed here, respectively. Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

low. The supTS and expTS tests show very low power when only the mean changes and persistence is high. Qu and Zhuo (2021) provides a more detailed discussion on why the supTS test has lower power when persistence is higher. In contrast, the LMC-LRT and MMC-LRT proposed here have higher power in both of these cases when only the mean changes. Once the variance changes, all the tests show improved power performance, with the LMC-LRT and MMC-LRT still maintaining higher power in most cases. This remains true when both the mean and variance change, despite the increase in power for all the test procedures. Overall, for the case where $H_0 : M_0 = 1$ and $H_1 : M_0+m = 2$, the LMC-LRT and MMC-LRT we propose have similar size properties to the other test procedures considered here but demonstrate better power properties. This is not surprising since the moment-based approaches, supTS, and expTS tests are based mainly on the model under the null. Therefore, even in simpler settings where other test procedures are available, the tests we propose may offer a better alternative due to their superior power properties.

As previously discussed, an interesting feature of the LMC-LRT and MMC-LRT is their applica-

Table 3: Empirical Performance of test when process is non-stationary

Test	Empirical size						
	T=100	T=200	T=500				
LMC-LRT	4.5	4.9	5.7				
MMC-LRT	2.2	2.3	4.5				
LMC_{min}	4.0	3.7	5.6				
LMC_{prod}	3.8	4.7	5.6				
MMC_{min}	1.4	1.5	3.1				
MMC_{prod}	1.5	2.0	2.6				
supTS	2.2	1.8	93.4				
expTS	2.6	38.3	98.2				
	Empirical Power						
	$(p_{11}, p_{22}) = (0.9, 0.9)$				$(p_{11}, p_{22}) = (0.9, 0.5)$		
	T=100	T=200	T=500		T=100	T=200	T=500
				$\Delta\mu$			
LMC-LRT	15.5	22.8	39.9		27.0	46.4	68.4
MMC-LRT	9.2	14.1	25.2		21.0	38.9	54.3
LMC_{min}	18.4	29.2	56.2		15.8	23.5	49.9
LMC_{prod}	19.2	30.4	57.8		16.9	25.3	52.2
MMC_{min}	7.0	16.3	44.0		6.5	14.2	38.4
MMC_{prod}	9.1	17.9	48.2		7.8	17.0	43.1
				$\Delta\sigma$			
LMC-LRT	41.8	76.3	99.1		36.2	61.2	93.9
MMC-LRT	23.5	41.3	91.2		25.2	48.9	91.8
LMC_{min}	38.9	63.1	94.8		45.6	71.6	95.4
LMC_{prod}	38.4	65.4	96.6		48.0	73.0	95.6
MMC_{min}	19.5	44.1	89.1		26.0	53.4	93.3
MMC_{prod}	21.8	46.8	90.1		27.4	54.4	93.3
				$\Delta\mu$ & $\Delta\sigma$			
LMC-LRT	29.7	54.4	77.3		49.7	76.9	90.4
MMC-LRT	21.7	43.1	63.8		34.4	67.9	88.1
LMC_{min}	32.7	57.1	92.6		61.2	88.4	99.5
LMC_{prod}	36.2	61.4	93.7		63.9	90.3	99.8
MMC_{min}	18.2	41.3	85.0		41.8	80.0	99.4
MMC_{prod}	20.7	47.8	87.7		46.6	83.3	99.6

Notes: The nominal level is 5%. Here, $\phi_1 = 1.00$ for all models so that we have a non-stationary (random-walk) process. LMC-LRT and MMC-LRT are the Local Monte Carlo and Maximized Monte Carlo Likelihood Ratio Tests proposed here, respectively. Rejection frequencies are obtained using $1,000$ replications. MC tests use $N = 99$ simulations.

bility even when the process is non-stationary or has parameters at the boundary of the parameter space. As mentioned in Section 3, in such cases, the likelihood function is not theoretically defined. Thus, in these scenarios, our tests may be better described as Local Monte Carlo and Maximized Monte Carlo *pseudo* Likelihood Ratio Tests. While this distinction is noteworthy, we continue to use the LMC-LRT and MMC-LRT acronyms for these settings. Table 3 reports the rejection frequencies under both the null and alternative hypotheses for the non-stationary case where $\phi_1 = 1.00$. Here, we consider unit-root DGPs and assess the performance of the LMC-LRT and MMC-LRT procedures. These results suggest that the supTS and expTS tests cannot control the size of the test. Specifically, when sample sizes get larger, and the process is more likely to exhibit more properties consistent with non-stationary processes, these tests have a substantial degree of over-rejection. In contrast, these results also suggest that use the Monte Carlo approach have remarkable size properties when the process is non-stationary. This includes the moment-based tests of Dufour and Luger (2017). Under the alternative, the power increases when regimes are asymmetric, particularly when only the mean changes or when both the mean and the variance change. The tests perform best with larger sample sizes and when variance or both the mean and variance change under the alternative hypothesis. Simulations for this setting for the moment-based approach were not provided in Dufour and Luger (2017) and so we are the first to show simulation evidence of the performance of the moment-based approach for non-stationary processes.

Table 4: Empirical Power of Test when $M_0 = 1$, $m = 2$, & $(p_{11}, p_{22}) = (0.9, 1.0)$

Test	$\phi = 0.10$			$\phi = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500
			$\Delta\mu$			
LMC-LRT	76.7	97.9	99.7	7.2	8.1	9.9
MMC-LRT	68.7	93.7	96.5	5.5	5.3	4.7
			$\Delta\sigma$			
LMC-LRT	30.8	56.0	91.9	27.8	52.1	93.5
MMC-LRT	24.6	50.3	86.4	23.3	48.8	82.7
			$\Delta\mu$ & $\Delta\sigma$			
LMC-LRT	49.9	83.8	99.5	19.5	41.5	90.1
MMC-LRT	40.7	81.0	96.0	11.2	34.0	84.0

Notes: The nominal level is 5%. LMC-LRT and MMC-LRT are the Local Monte Carlo and Maximized Monte Carlo Likelihood Ratio Tests proposed here, respectively. Rejection frequencies are obtained using $1,000$ replications. MC tests use $N = 99$ simulations.

Table 4 presents results for the previously mentioned interesting case where the regimes are asymmetric, but now one state being absorbing and so the transition probabilities of this regime lie at the boundary of the parameter space. Specifically, we consider $(p_{11}, p_{22}) = (0.9, 1.0)$, where, as before, $p_{ij} = (1 - p_{ii})$ for $j \neq i$. Here, we find that low persistence and changes in the mean contribute to higher power performance for smaller sample sizes of $T = 100$ and $T = 200$. When the sample size is $T = 500$, the power is very high in all cases except when there is high persistence and only changes in the mean.

Table 5 shows the rejection frequencies of the LMC-LRT and MMC-LRT proposed here under the alternative hypothesis when $M_0 = 1$ and $m = 2$. In other words, we consider a linear model under the null hypothesis (i.e., $H_0 : M_0 = 1$) against an alternative of a Markov switching model with three regimes (i.e., $H_1 : M_0 + m = 3$). The results indicate that the power is consistently high in all cases considered.

Now we consider a null hypothesis of a Markov switching model with two regimes (i.e., $H_0 : M_0 = 2$) and an alternative hypothesis of a Markov switching model with three regimes (i.e., $H_1 : M_0 + m = 3$). For this comparison, we include two classes of the DGPs used in Kasahara and

Table 5: Empirical Power of Test when $M_0 = 1$, $m = 2$

Test	$(p_{11}, p_{22}, p_{33}) = (0.9, 0.9, 0.9)$						$(p_{11}, p_{22}, p_{33}) = (0.9, 0.5, 0.5)$					
	$\phi = 0.10$			$\phi = 0.90$			$\phi = 0.10$			$\phi = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500	T=100	T=200	T=500	T=100	T=200	T=500
							$\Delta\mu$					
LMC-LRT	84.6	98.3	100.0	59.0	86.2	99.5	90.5	99.9	100.0	69.6	95.6	100.0
MMC-LRT	80.0	93.0	95.3	51.4	77.3	92.1	88.7	97.0	99.7	58.7	91.0	96.1
							$\Delta\sigma$					
LMC-LRT	71.6	95.6	100.0	67.7	95.4	100.0	86.7	99.3	99.2	84.7	98.9	99.2
MMC-LRT	62.5	84.0	92.4	59.0	86.3	93.4	58.4	80.7	94.4	54.7	78.0	93.5
							$\Delta\mu$ & $\Delta\sigma$					
LMC-LRT	85.5	99.9	100.0	77.1	95.9	100.0	99.6	100.0	100.0	84.9	99.2	100.0
MMC-LRT	79.4	90.1	98.1	60.6	92.0	94.3	99.1	93.3	96.1	74.0	97.0	100.0

Notes: The nominal level is 5%. LMC-LRT and MMC-LRT are the Local Monte Carlo and Maximized Monte Carlo Likelihood Ratio Tests proposed here, respectively. Rejection frequencies are obtained using 1,000 replications. MC tests use $N = 99$ simulations.

Table 6: Empirical Size of Test when $M_0 = 2$ & $m = 1$

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	3.80	3.70	3.30	3.10	3.60	2.70
Boot-LRT	-	7.16	4.43	-	6.07	4.20

Notes: *LMC-LRT* and *MMC-LRT* use $N = 99$ and are obtained using 1000 replications. *Boot-LRT* results are taken from Kasahara and Shimotsu (2018)

Shimotsu (2018). We also present the Boot-LRT results from Kasahara and Shimotsu (2018) for these DGPs, except for $T = 100$ as it was not reported. As previously discussed, the parametric bootstrap and LMC-LRT procedures share similarities, but an important distinction is that the LMC-LRT does not require enforcing assumptions previously discussed to obtain the asymptotic validity of the bootstrap procedure when estimating the null distribution. Furthermore, while larger sample sizes should better adhere to the asymptotic validity of both procedures, the LMC-LRT does not rely on the existence of an asymptotic distribution, so fewer simulations are sufficient. Kasahara and Shimotsu (2018) also apply other constraints on variance parameters during model estimation and use $N = 199$ simulations, whereas we use $N = 99$ simulations and impose no constraints. These features should explain the difference found LMC-LRT and parametric bootstrap test results. Nevertheless, the results suggest that the two procedures exhibit similar patterns for these DGPs. Specifically, Table 6 shows that both the LMC-LRT and parametric bootstrap procedures display some over-rejection for smaller sample sizes of $T = 100$ and $T = 200$, particularly for the bootstrap test. However, the rejection frequencies are much closer to the nominal level when $T = 500$, as the theory suggests. Meanwhile, the MMC-LRT procedure performs as expected, maintaining a rejection frequency of $\leq 5\%$, even with $T = 100$ or $T = 200$. This underscores the contribution of the MMC-LRT as a valid test procedure in both finite samples and asymptotically.

5 Applications

In this section, we first study U.S. GNP and GDP growth rates, with an emphasis on GDP data, as this series has been more commonly used in recent years. All samples used for this univariate application are included in the R package **MSTest**, which can easily reproduce all estimation and hypothesis testing results presented here. We also consider a second application in a multivariate setting. The goal here is to use the LMC-LRT and MMC-LRT tests to evaluate the synchronization of international business cycles. This example is meant to demonstrate the value of having a test procedure that can be applied to multivariate settings, such as Markov switching VAR models, and to test a hypothesis where $m > 1$—both of which are cases that could not be handled by previously proposed tests in the literature. A more detailed analysis of testing the synchronization of international business cycles is provided in Rodriguez-Rondon and Dufour (2024).

5.1 U.S. GNP and GDP growth

Many of the procedures for testing the number of regimes in Markov switching models have used U.S. GNP growth data, as it was one of the original applications of these models in Hamilton (1989). Studies that have utilized U.S. GNP data for regime testing include Hansen (1992), Carrasco, Hu, and Ploberger (2014), and Dufour and Luger (2017). Hansen (1992) examines the original quarterly sample from 1951:II to 1984:IV used by Hamilton (1989), with $p = 4$ and allowing only the mean to change across regimes, as in the original model. In this case, the proposed test fails to reject the null hypothesis of a linear model (i.e., $M = 1$). Similarly, Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) also use the same sample and fail to reject the null of a linear model for this sample.

These studies also consider an extended sample from 1951:II to 2010:IV, which includes the Great Recession. They continue to use a model with four lags ($p = 4$), but now also consider an alternative where both the mean and variance change across regimes, as suggested by Kim and Nelson (1999). Allowing variance to change is a sensible feature for two reasons. First, the second period includes the structural decline in variance around the mid-1980s, known as the Great Moderation. Second, since the goal is to capture recessionary periods, it is reasonable to assume that variance would increase during such periods of distress. Both Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017) reject the null hypothesis of a linear model in favor of a Markov switching model with $M = 2$ regimes for this second sample when the variance is allowed to change. As discussed in Qu and Zhuo (2021), the inclusion of the Great Recession appears to be crucial for the supTS test of Carrasco, Hu, and Ploberger (2014) to reach this conclusion. However, if only the mean is allowed to change, the supTS and expTS tests continue to fail to reject the null hypothesis of a linear model. On the other hand, in Qu and Zhuo (2021), using GDP data, the authors find more evidence for a model with $M = 2$ regimes, even when only the mean is allowed to change.

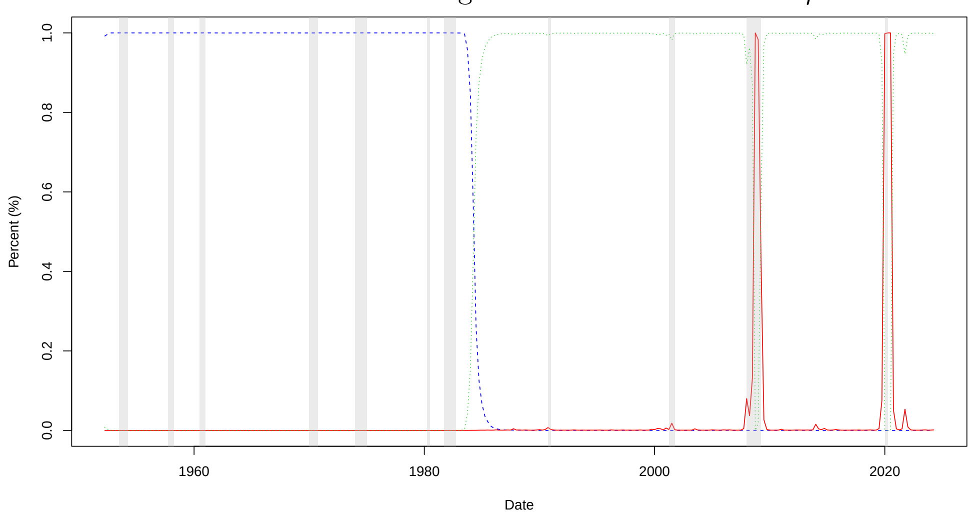
Table 7: Results For U.S. GNP Growth Series Hypothesis Tests

Series	$H_0 : M = 1$ vs. $H_1 : M = 2$		$H_0 : M = 2$ vs. $H_1 : M = 3$		$H_0 : M = 3$ vs. $H_1 : M = 4$	
	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT
$\Delta\mu$						
GNP 1951:II-1984:IV	0.35	0.93	-	-	-	-
GNP 1951:II-2010:IV	0.03	0.05	0.06	0.23	-	-
GNP 1951:II-2024:II	0.01	0.01	0.01	0.01	0.52	1.00
$\Delta\mu$ & $\Delta\sigma$						
GNP 1951:II-1984:IV	0.38	0.85	-	-	-	-
GNP 1951:II-2010:IV	0.01	0.01	0.58	1.00	-	-
GNP 1951:II-2024:II	0.01	0.01	0.02	0.04	0.70	1.00

Notes: The GNP 1951:II-1984:IV series ($T = 135$) is the same as the one used in Hamilton (1989), Hansen (1992), and Carrasco, Hu, and Ploberger (2014). The GNP 1951:II-2010:IV series ($T = 239$) is the same as the one used in Carrasco, Hu, and Ploberger (2014) and Dufour and Luger (2017). The GNP 1951:II-2024:II series ($T = 293$) is the GNP series from the St. Louis Fed (FRED) website. All MC test results are obtained using $N = 99$. The MMC-LRT procedure uses a particle swarm optimization algorithm. Models for GNP use $p = 4$ lags as in Hamilton (1989) while models for GDP use $p = 1$ lags as in Qu and Zhuo (2021).

To complement this part of the literature, we also consider these two samples of US GNP data, along with an extended sample ranging from 1951:II to 2024:II. The results of the LMC-LRT and MMC-LRT for these three US GNP growth rate samples are presented in 7. For the first two samples, our results mostly align with other proposed tests. However, unlike Carrasco, Hu, and Ploberger (2014), we find evidence supporting a model with $M = 2$ regimes, even when only changes in the mean are considered in the second sample that includes the Great Recession. This finding is more in line with the results of Qu and Zhuo (2021) for similar samples of US GDP data. We further extend this analysis by comparing the preferred $M = 2$ model under the null hypothesis to a model with $M = 3$ regimes under the alternative. To the best of our knowledge, this is the first time this hypothesis has been tested for this sample of US GNP. In doing so, we confirm that a model with $M = 2$ regimes is indeed sufficient to explain this sample, both when only the mean changes and when both the mean and variance change. On the other hand, when we consider the third, larger sample of US GNP, we reject the null hypothesis of $M = 2$ regimes in favor of a Markov switching model with $M = 3$ regimes. We confirm this model by considering an alternative with four regimes but fail to reject the null hypothesis of a Markov switching model with $M = 3$ regimes when doing so. Figure 1 shows the smoothed regime probabilities for the model with $M = 3$ when both the mean and the variance can change. The smoothed probabilities for some of the other models being considered here are found in Figures A1 - A3. The estimates for this model and others that allow the variance to change are reported in A3. For here, we observe that two regimes are expansionary, with positive means, though the second has significantly lower volatility (i.e., $\mu_1 = 1.87$, $\mu_2 = 1.31$, $\sigma_1 = 1.09$, and $\sigma_2 = 0.49$). This reduction in volatility is consistent with the Great Moderation and this is confirmed by examining the smoothed probabilities of these two regimes, which switch around the mid-1980s. In this case, the third regime is a deep recessionary regime, capturing only the Great Recession and COVID recession. Like Gadea, Gómez-Loscos, and Pérez-Quirós (2018) and Gadea, Gómez-Loscos, and Pérez-Quirós (2019), we find that the low-volatility period returns after the Great Recession. Here, we also find that the low-volatility period returns after the recent COVID recession.

Figure 1: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ & $\Delta\sigma$ and $M = 3$



Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Since most of the recent literature uses US GDP data, we now turn to this series. For testing the number of regimes in a Markov switching model, Qu and Zhuo (2021) and Kasahara and Shimotsu (2018) also use US GDP data instead of GNP data. When using US GDP data, we only consider the larger sample from 1951:II to 2024:II. This larger sample is especially interesting as it includes the more recent COVID-19 period, which is known to be problematic when working with many macroeconomic variables due to its stark difference from the rest of the sample. Various approaches have been proposed for dealing with this period. One approach involves treating it as a known structural break. One advantage of treating it as a known structural break, by for example incorporating explanatory variables that account for it, is that such an approach can justify using a simpler model, potentially requiring fewer regimes to capture all the non-linearities in the series. For this reason, we evaluate the robustness of the resulting number of regimes proposed by our method when testing US GDP growth data by treating this period as a known structural break in the mean. This is done by including a dummy variable that takes a value of 1 from 2020:I to 2021:IV and 0 elsewhere. We also consider a dummy variable that takes a value of 1 for the period from 1951:II to 1983:IV and 0 elsewhere to control for the Great Moderation.[1] Hence, we consider a model with no dummy variables (Model 1), a model that treats only the Great Moderation as a known structural break (Model 2), a model that treats only the COVID period as a known structural break (Model 3), and a model that includes both the Great Moderation and the COVID period as known structural breaks in the mean (Model 4). The hypothesis testing results for the number of regimes in all four of these models, both where only changes in the mean or both changes in the mean and variance are considered. Table 8 presents these results.

Table 8: Results For U.S. GDP Growth Series Hypothesis Tests With Known Breaks

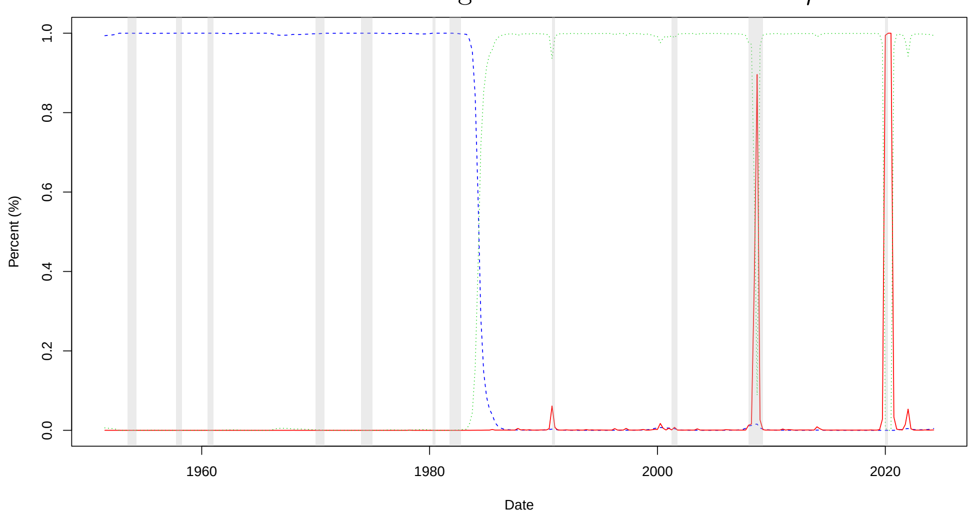
Series	$H_0 : M = 1$ vs. $H_1 : M = 2$		$H_0 : M = 2$ vs. $H_1 : M = 3$		$H_0 : M = 3$ vs. $H_1 : M = 4$	
	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT
			$\Delta\mu$			
Model 1	0.01	0.01	0.01	0.01	0.76	1.00
Model 2	0.01	0.01	0.01	0.01	0.76	1.00
Model 3	0.01	0.01	0.01	0.01	0.94	1.00
Model 4	0.01	0.01	0.01	0.01	0.59	1.00
			$\Delta\mu$ & $\Delta\sigma$			
Model 1	0.01	0.01	0.01	0.01	0.44	1.00
Model 2	0.01	0.01	0.01	0.01	0.35	1.00
Model 3	0.01	0.01	0.01	0.01	0.27	1.00
Model 4	0.01	0.01	0.01	0.01	0.24	1.00

Notes: The GDP 1951:II-2024:II series ($T = 293$) is the GPC1 series from the St. Louis Fed (FRED) website. Model 1: no fixed exogenous regressors, Model 2: includes dummy variable treating Great Moderation as known structural break, Model 3: includes dummy variable treating COVID period as known multiple structural breaks, and Model 4: includes dummy variables treating Great Moderation and COVID period as known multiple structural breaks. Specifically, the dummy variable for the Great Moderation takes values of 1 for the period 1951:II to 1983:IV, and 0 elsewhere. Similarly, dummy variable for the COVID period takes values of 1 for the period 2020:I to 2021:IV, and 0 elsewhere. the All MC test results are obtained using $N = 99$. The MMC-LRT procedure uses a particle swarm optimization algorithm. Models GDP use $p = 1$ lags as in Qu and Zhuo (2021).

1. We first fit a Markov switching model with no dummy variable for the Great Moderation and find clear evidence suggesting this to be one of the regimes. The smoothed probabilities of this model are then used to date this period.

As with the GNP growth data for the same sample, we find evidence of a model with $M = 3$ regimes for the US GDP data. Unlike the US GNP models, here we use one lag ($p = 1$). This is also the case in Qu and Zhuo (2021). Since our sample is slightly different from theirs, we first verified that this lag order is still appropriate here and confirmed that this was indeed the case. To determine which of the eight resulting models is preferred for this sample, we consider an LRT approach to test the significance of our dummy variables. In this case, the conventional regularity conditions are met, so we can rely on the conventional procedure. Table A1 reports the estimates of these models and their log likelihood functions. From here, we can see that both when only the mean changes and when both the mean and variance change, the log likelihood values of the models with dummy variables are not significantly different from those of the corresponding model with no dummy variables, resulting in very small LRT statistics. Hence, it is no surprise that these are not statistically significant. In all cases, we can also compare the model with changes in both the mean and variance against the models where only the mean changes. The model that also includes changes in variance is preferred in all cases. As a result,we find that the model with $M = 3$ regimes, changes in mean and variance, and no dummy variables is a sufficiently good model for this data. The smoothed regime probabilities for this model, as shown in Figure 2, are very similar to the GNP

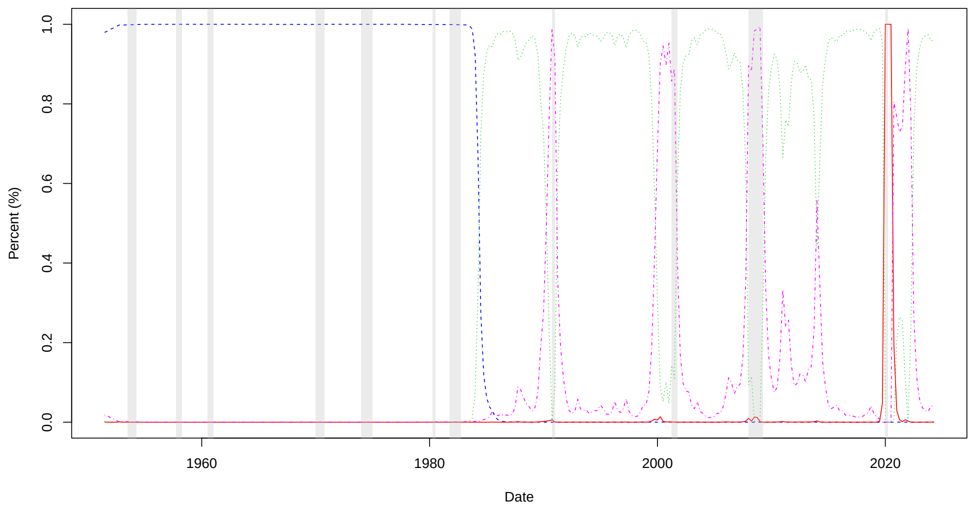
Figure 2: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ & $\Delta\sigma$ and $M = 3$



Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

case, but the estimates differ slightly (i.e., $\mu_1 = 0.79$, $\mu_2 = 0.72$, $\mu_3 = -0.50$, $\sigma_1 = 1.06$, $\sigma_2 = 0.45$, and $\sigma_3 = 6.5$). It is worth highlighting that the fact that dummy variables did not change the outcome of the hypothesis test for the number of regimes is likely because treating these periods as known structural breaks in the mean is not sufficient. The Markov switching models with changes in variance are likely especially preferred here because they consider such changes in variance, which are likely important features of these known structural changes. Considering more sophisticated models with heteroskedastic functions or known structural breaks in the variance may prove useful. In either case, our test proposed here can be used to determine if the models being used capture such features or if a model with more regimes is needed to do so. It may also be worth noting that,

Figure 3: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ & $\Delta\sigma$ and $M = 4$



Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

as shown in Figure 3, a model with $M = 4$ regimes can capture the relatively shallower recessions. In this case, the recovery period following the COVID recession is also included. However, if we look at the log likelihood of this model in Table A1, it is very close to that of the model with $M = 3$ regimes, so it is no surprise that this model was rejected. Although not statistically significant, this model may still be worth considering if the objective is the identification of business cycles.

5.2 Synchronization of business cycles

The synchronization of business cycles has again become a topic of interest after restrictions due to the global pandemic and supply chain issues highlighted some risks of integration. The business cycle literature includes different methodologies to determine the degree of synchronization of business cycles across various economies. However, few business cycle synchronization hypothesis testing procedures are available. Those that do can sometimes yield mixed results, require assumptions about the functional form of the relationship, or are not formally shown to be valid testing procedures, especially in finite samples. Examples of this include the procedures proposed in Phillips (1991) and Camacho, Perez-Quiros, and Saiz (2006). Here, we identify business cycles using Markov switching models, as in the related literature, but propose using the Monte Carlo likelihood ratio testing procedures introduced here to test the hypothesis of synchronized business cycles against different alternatives of non-synchronized business cycles. We test the hypothesis of synchronized business cycles between the United States and Canada, the United Kingdom, or Germany using real GDP and industrial production data while considering two samples: one from 1985:I to 2019:IV and another from 1985:I to 2022:IV, which notably includes the COVID period. These economies were considered in Phillips (1991) and here, we also consider seasonally adjusted data at the quarterly frequency. The application presented here is only meant to showcase the value of having the LMC-LRT and MMC-LRT procedures, which are useful in multivariate settings and in cases where $m > 1$, as we will see, is empirically relevant in this context. Further, we also observe that the MMC-LRT procedure is valid in finite samples, which is especially important here, where we use quarterly data and only have sample sizes of $T = 140$ in the first sample and $T = 155$ in the second sample.

The general idea is as follows. We propose testing for the synchronization of two business cycles by identifying the appropriate number of regimes in a bivariate Markov switching VAR model that includes both economies being considered. Business cycles are defined as intervals of recessions and expansions, so we should consider a Markov process with at least two regimes. If the business cycles of both economies in the system are synchronized, we should expect that one Markov process with two states is sufficient to explain our bivariate data. However, if the business cycles of these economies are not synchronized, then each requires its own independent Markov process to model the states of each economy. In the latter case, we can still use a single Markov process, but with more states to summarize the state of each underlying Markov process. For example, consider the following bivariate model with economies a and b:

$$
\begin{aligned}
y_{a,t} &= \mu_{a,s_{a,t}} + \sum_{k=1}^{p} \phi_{aa,k}\left(y_{a,t-k} - \mu_{a,s_{a,t-k}}\right) + \sum_{k=1}^{p} \phi_{ab,k}\left(y_{b,t-k} - \mu_{b,s_{b,t-k}}\right) + \sigma_{a,s_{a,t}}\epsilon_{a,t} \\
y_{b,t} &= \mu_{b,s_{b,t}} + \sum_{k=1}^{p} \phi_{ba,k}\left(y_{a,t-k} - \mu_{a,s_{a,t-k}}\right) + \sum_{k=1}^{p} \phi_{bb,k}\left(y_{b,t-k} - \mu_{b,s_{b,t-k}}\right) + \sigma_{b,s_{b,t}}\epsilon_{b,t}
\end{aligned}
$$

Here, we are interested in knowing if the Markov processes $S_{a,t}$ and $S_{b,t}$ are perfectly dependent such that $S_{a,t} = S_{b,t} = S_t$ or if they are independent such that $S_{a,t} \neq S_{b,t}$. If we suppose that

$S_{a,t} = \{1, 2\}$ and $S_{b,t} = \{1, 2\}$ then we should consider up to four cases:

$$S_t^* = 1 \;\text{ if }\; S_{a,t} = 1 \;\;\&\;\; S_{b,t} = 1$$
$$S_t^* = 2 \;\text{ if }\; S_{a,t} = 1 \;\;\&\;\; S_{b,t} = 2$$
$$S_t^* = 3 \;\text{ if }\; S_{a,t} = 2 \;\;\&\;\; S_{b,t} = 1$$
$$S_t^* = 4 \;\text{ if }\; S_{a,t} = 2 \;\;\&\;\; S_{b,t} = 2$$

and if the Markov processes are perfectly dependent, then we have the following two cases

$$S_t^* = 1 \;\text{ if }\; S_{a,t} = 1 \;\;\&\;\; S_{b,t} = 1$$
$$S_t^* = 2 \;\text{ if }\; S_{a,t} = 2 \;\;\&\;\; S_{b,t} = 2$$

We can also consider a specific type of dependence where one of the Markov processes, say $S_{a,t}$ ($S_{b,t}$) is leading (lagging) the other. Here we have the following three cases for example

$$S_t^* = 1 \;\text{ if }\; S_{a,t} = 1 \;\;\&\;\; S_{b,t} = 1$$
$$S_t^* = 2 \;\text{ if }\; S_{a,t} = 2 \;\;\&\;\; S_{b,t} = 1$$
$$S_t^* = 3 \;\text{ if }\; S_{a,t} = 2 \;\;\&\;\; S_{b,t} = 2$$

where alternatively, $S_t^* = 2$ if $S_{a,t} = 1$ and $S_{b,t} = 2$. As a result, testing for the synchronization of business cycles boils down to testing the null hypothesis of a Markov switching model with two regimes (i.e., business cycles are synchronized) against the alternative hypothesis of three or four regimes (i.e., not synchronized). That is, we are interested in testing

$$H_0 : M_0 = 2 \quad \text{vs.}$$
$$H_{1a} : M_0 + m = 3 \quad \text{or} \quad H_{1b} : M_0 + m = 4$$

where M_0 is the number of regimes for a bivariate Markov switching model under the null hypothesis and $M_0 + m$ is the number of regimes under the alternative.

It is worth noting that, even though we presented this in a setting where two economies are governed by independent Markov processes, the same could be applied in a univariate setting where we believe the two coefficients—such as the mean and the variance—are governed by independent

Markov processes (see Sims and Zha (2006) for example).

Table 9: Results For Synchronization of Business Cycle Hypothesis Tests using GDP series

Series	$H_0: M=1$ vs. $H_1: M=2$		$H_0: M=2$ vs. $H_1: M=3$		$H_0: M=2$ vs. $H_1: M=4$	
	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT
1985:I - 2019:IV ($T=140$)						
US-CA	0.02	0.04	0.20	0.65	0.17	0.67
US-UK	0.01	0.01	0.01	0.01	0.01	0.01
US-GR	0.03	0.05	0.27	0.54	0.11	0.51
1985:I - 2022:IV ($T=155$)						
US-CA	0.01	0.01	0.08	0.43	0.03	0.05
US-UK	0.01	0.01	0.13	0.21	0.01	0.01
US-GR	0.01	0.01	0.21	0.53	0.04	0.06

Notes: This table includes results when $\Delta\mu$ & $\Delta\sigma$ as it is a statistically preferred model over a model where only $\Delta\mu$. The GDP series are OECD Main Economic Indicator Releases obtained from the St. Louis Fed (FRED) website. All MC test results are obtained using $N=99$. The MMC-LRT procedure uses a particle swarm optimization algorithm.

Our results shown in Table 9 suggest that, when considering real GDP data up to 2019:IV (the pre-COVID period), the US business cycle is perfectly synchronized with that of Canada and Germany. However, when we include the COVID period, our results suggest that the US business cycle may no longer synchronized with all three economies. We provide similar results when using industrial production data in Table A4. Here, the results also suggest that the US is no longer synchronized with Canada. These findings provide preliminary evidence suggesting that these economies are no longer synchronized. One possible explanation may be that the recovery following the COVID period was different for these economies compared to that of the US economy. A more thorough analysis is presented in Rodriguez-Rondon and Dufour (2024) for interested readers.

6 Conclusion

Using the Monte Carlo procedures described in Dufour (2006) in a likelihood ratio test setting for Markov switching models, we propose the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT) and the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) for Markov switching models. These tests help determine the number of regimes required to capture nonlinearities in the data, whether using Markov switching models or Hidden Markov models. 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 $M_0,\ m \geq 1$. Notably, the case where $m>1$ is one for which other tests are unavailable. Moreover, these tests can be applied to non-stationary processes, models with non-Gaussian errors, and even multivariate settings. To the best of our knowledge, we are the first to consider hypothesis testing of multivariate Markov switching models. 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. Simulations suggest that both versions of the Monte Carlo likelihood ratio test effectively control the level of the test. Another important contribution is the MMC-LRT, which is an identification-robust procedure that is valid in finite samples and asymptotically. Finite-sample validity can be important for many macroeconomic applications where quarterly data are used and being robust to identification issues is especially important as such issues can often arise when dealing with Markov switching models. Simulations also show that both tests have good power. When compared to the moment-based approach of Dufour and Luger (2017) and the parameter stability test of Carrasco, Hu, and Ploberger (2014) in the $M_0 = m = 1$ setting, where these tests are available, our tests often exhibit higher power when only the mean changes and comparable or better power when both the mean and variance change.

Finally, we present two empirical applications. First, we use the proposed test procedures sequentially to determine the number of regimes for modeling real U.S. GNP and GDP growth data. For the GNP series from 1951 to 2010, we confirm that a two-regime model fits best, while a three-regime model is more suitable for the extended sample, ending in 2024:II, for both GNP and GDP. Our results confirm evidence about the Great Moderation and the return of the low volatility regime after the Great Recession, as shown by Gadea, Gómez-Loscos, and Pérez-Quirós (2018) and Gadea, Gómez-Loscos, and Pérez-Quirós (2019). Additionally, we find that the low-volatility regime also returns after the COVID-19 recession. Although we consider treating these periods as known structural breaks in the mean, we still find three regimes are needed. We conjecture that treating them as breaks in the variance may simplify the model, but leave this for future research. In the second application, we use our test procedures with Markov switching VAR models to test business cycle synchronization. Preliminary evidence suggests that adding COVID data weakens the synchronization between the U.S. and Canada that was previously present.

References

Aguirregabiria, Victor, and Pedro Mira. 2007. "Sequential estimation of dynamic discrete games." *Econometrica* 75 (1): 1–53.

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.

Caggiano, Giovanni, Efrem Castelnuovo, and Juan M. Figueres. 2017. "Economic Policy Uncertainty and Unemployment in the United States: A Nonlinear Approach." *Journal of Applied Econometrics* 32 (2): 281–298. https://doi.org/10.1002/jae.2587. https://onlinelibrary.wiley.com/doi/10.1002/jae.2587.

Camacho, Maximo, Gabriel Perez-Quiros, and Lorena Saiz. 2006. "Are European business cycles close enough to be just one?" *Journal of Economic Dynamics and Control* 30, nos. 9-10 (September): 1687–1706.

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.

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.

Dietz, Simon, and Nicholas Stern. 2015. "Endogenous growth, convexity of damages and climate risk: How Nordhaus' framework supports deep cuts in carbon emissions." *The Economic Journal* 125 (583): 574–620.

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.

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.

Golosov, Mikhail, John Hassler, Per Krusell, and Aleh Tsyvinski. 2014. "Optimal taxes on fossil fuel in general equilibrium." *Econometrica* 82 (1): 41–88.

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.

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.

Hernández, Pablo, and Carlos Ochoa. 2016. "Switching models for the analysis of health care costs: An application to the Spanish health care system." *Health Economics* 25 (6): 830–843. https://doi.org/10.1002/hec.3284.

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.

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

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).

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.

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

Phillips, Kerk L. 1991. "A two-country model of stochastic output with changes in regime." *Journal of International Economics* 31 (1-2): 121 –142.

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.

Rodriguez-Rondon, Gabriel, and Jean-Marie Dufour. 2024. "MSTest: An R-package for Testing Markov-Switching Models." *Discussion paper, McGill University (Economics Department).*

Rodriguez-Rondon, Gabriel, and Jean-Marie Dufour. 2024. *MSTest: Hypothesis Testing for Markov Switching Models.* R package version 0.1.3. https://github.com/roga11/MSTest.

Rodriguez-Rondon, Gabriel, and Jean-Marie Dufour. 2024. "Testing for the Synchronization of International Business Cycles." *Discussion paper, McGill University (Economics Department).*

Sims, Christopher A, and Tao Zha. 2006. "Were there regime switches in US monetary policy?" *American Economic Review* 96 (1): 54–81.

Sweeting, Andrew. 2013. "Dynamic product positioning in differentiated product markets: The effect of fees for musical performance rights on the commercial radio industry." *Econometrica* 81 (5): 1763–1803.

Appendix

Table A1: Comparison of Models with Dummy Variables for Known Structural Breaks

	μ_1	μ_2	μ_3	ϕ_1	GMd	CVd	σ_1	σ_2	σ_3	LogLike	AIC	BIC
						$\Delta\mu$						
Model 1	7.473	0.748	-8.220	0.329	-	-	0.819	-	-	-362.771	753.543	805.017
Model 2	7.473	0.748	-8.220	0.323	0.118	-	0.817	-	-	-362.032	754.063	809.214
Model 3	7.473	0.748	-8.220	0.329	-	0.084	0.819	-	-	-362.731	755.461	810.612
Model 4	7.473	0.748	-8.220	0.323	0.125	0.141	0.817	-	-	-361.919	755.838	814.666
						$\Delta\mu$ & $\Delta\sigma$						
Model 1	0.794	0.718	-0.459	0.262	-	-	1.07	0.449	6.499	-337.072	706.145	764.973
Model 2	0.795	0.717	-0.463	0.261	0.027	-	1.07	0.450	6.502	-337.020	708.039	770.544
Model 3	0.794	0.717	-0.442	0.260	-	0.185	1.07	0.450	6.437	-337.016	708.033	770.538
Model 4	0.800	0.717	-0.447	0.260	0.022	0.158	1.07	0.451	6.449	-336.984	709.967	776.149

Notes: The GDP 1951:II-2024:II series ($T = 293$) is the GPC1 series from the St. Louis Fed (FRED) website. Model 1: no fixed exogenous regressors, Model 2: includes dummy variable treating Great Moderation as known structural break and is labeled *GMd*, Model 3: includes dummy variable treating COVID period as known multiple structural breaks and is labeled *CVd*, and Model 4: includes dummy variables treating Great Moderation and COVID period as known multiple structural breaks. Specifically, the dummy variable for the Great Moderation takes values of 1 for the period 1951:II to 1983:IV, and 0 elsewhere. Similarly, dummy variable for the COVID period takes values of 1 for the period 2020:I to 2021:IV, and 0 elsewhere. All MC test results are obtained using $N = 99$. The MMC-LRT procedure uses a particle swarm optimization algorithm. Models GDP use $p = 1$ lags as in Qu and Zhuo (2021).

Table A2: Estimates Models for US GNP series

	μ_1	μ_2	μ_3	ϕ_1	ϕ_2	ϕ_3	ϕ_4	σ_1	σ_2	σ_3	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	p_{31}	p_{32}	p_{33}	LogLike
M=1	1.51	-	-	0.14	0.18	0.03	0.02	1.19	-	-	-	-	-	-	-	-	-	-	-	-458.38
M=2	1.56	1.21	-	0.31	0.24	0.01	0.00	0.62	3.03	-	0.96	0.04	-	0.32	0.68	-	-	-	-	-363.79
M=3	1.87	1.31	-0.77	0.23	0.23	0.06	0.00	1.09	0.49	5.74	0.99	0.01	0.00	0.00	0.99	0.01	0.00	0.37	0.63	-341.57

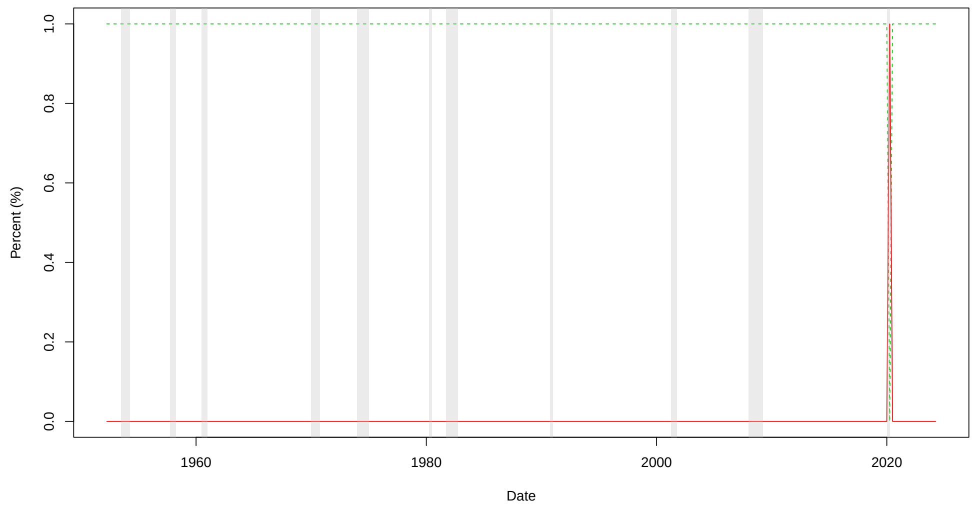
Notes: The GNP 1951:II-2024:II series ($T = 293$) is the GNP series from the St. Louis Fed (FRED) website. The models use $p = 4$ lags as in Hamilton (1989), Hansen (1992), Carrasco, Hu, and Ploberger (2014), and Dufour and Luger (2017).

Table A3: Estimates of Preferred Models for US GDP series

	μ_1	μ_2	μ_3	ϕ_1	σ_1	σ_2	σ_3	p_{11}	p_{12}	p_{13}	p_{21}	p_{22}	p_{23}	p_{31}	p_{32}	p_{33}	LogLike
M=1	0.74	-	-	0.10	1.09	-	-	-	-	-	-	-	-	-	-	-	-437.54
M=2	0.80	0.11	-	0.30	0.68	3.00	-	0.96	0.04	-	0.47	0.53	-	-	-	-	-368.08
M=3	0.79	0.72	-0.46	0.26	1.06	0.45	6.50	0.97	0.03	0.00	0.01	0.98	0.01	0.32	0.00	0.68	-337.07

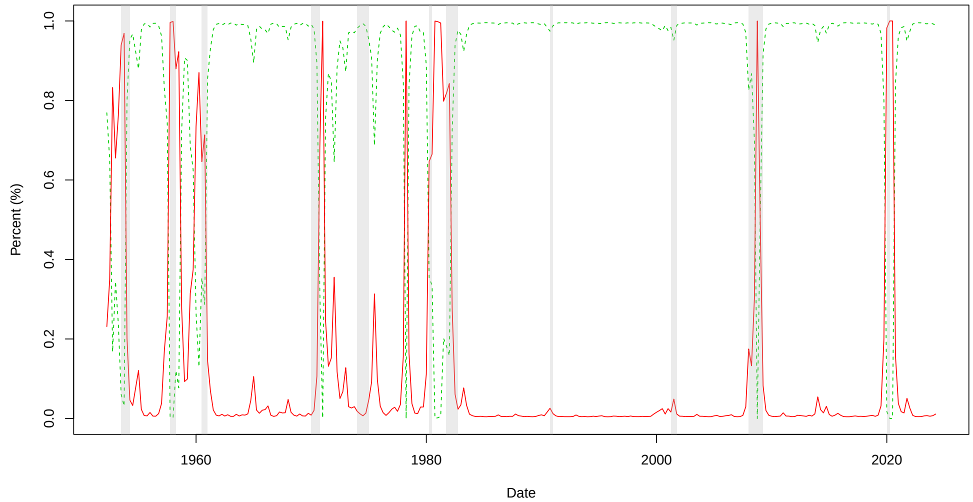
Notes: The GDP 1951:II-2024:II series ($T = 293$) is the GPC1 series from the St. Louis Fed (FRED) website. The models use $p = 1$ lags as in Qu and Zhuo (2021).

Figure A1: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ only and $M = 2$



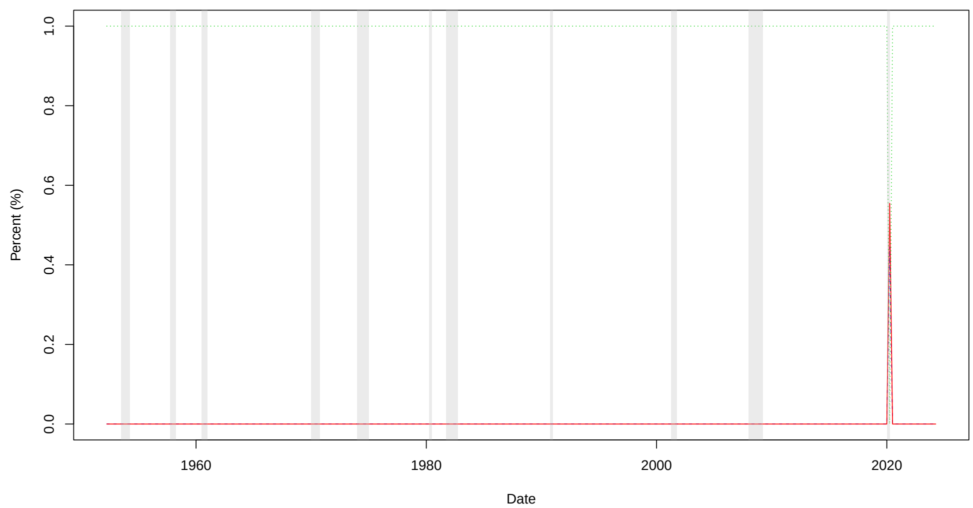
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure A2: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ & $\Delta\sigma$ and $M = 2$



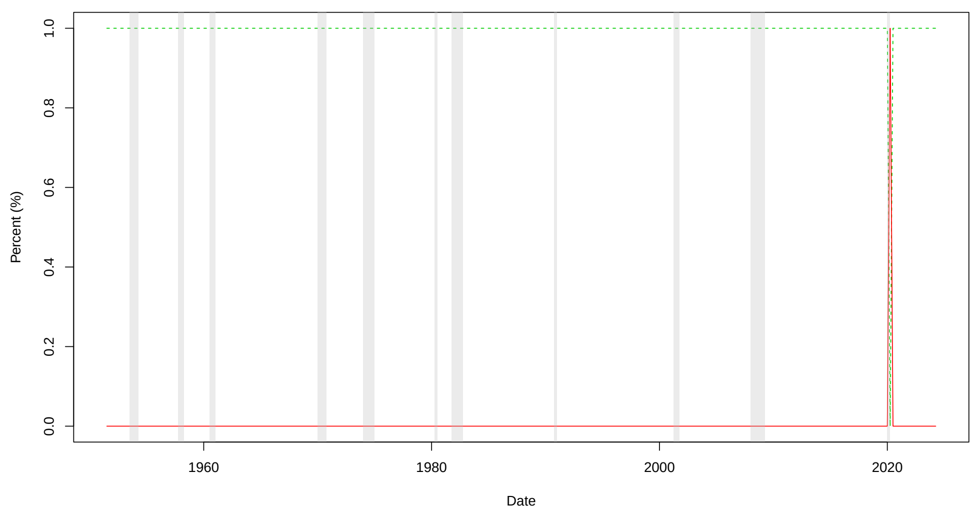
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure A3: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ only and $M = 3$



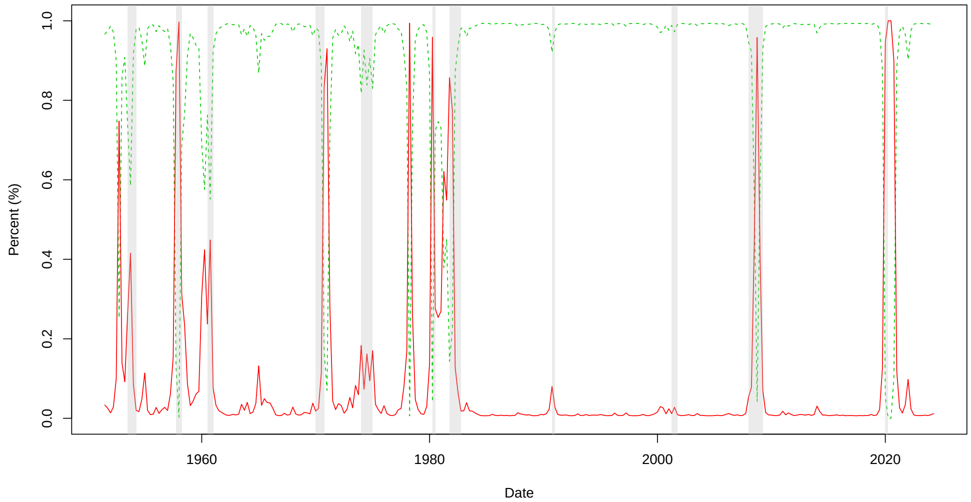
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure A4: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ only and $M = 2$



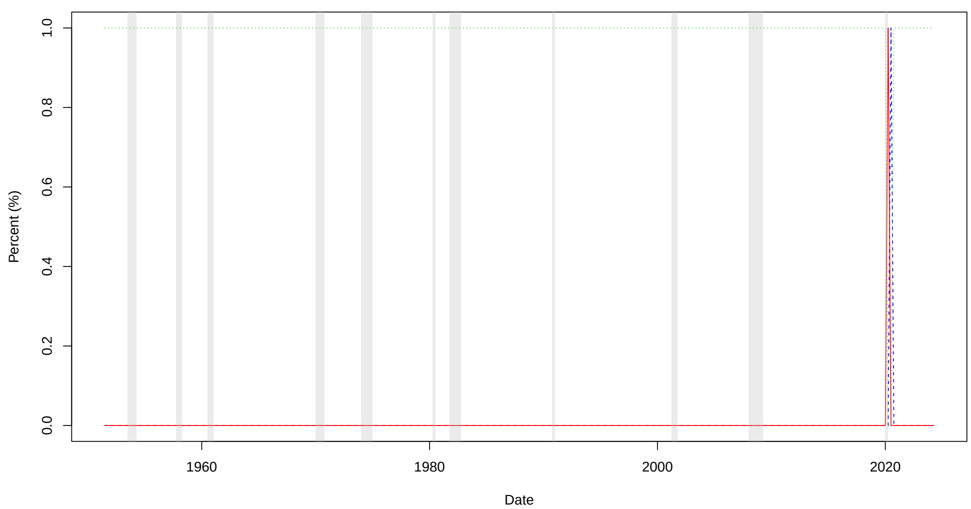
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure A5: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ & $\Delta\sigma$ and $M = 2$



Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure A6: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ only and $M = 3$



Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Table A4: Results For Synchronization of Business Cycle Hypothesis Tests using IP series

Series	$H_0 : M = 1$ vs. $H_1 : M = 2$		$H_0 : M = 2$ vs. $H_1 : M = 3$		$H_0 : M = 2$ vs. $H_1 : M = 4$	
	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT	LMC-LRT	MMC-LRT
1985:I - 2019:IV ($T = 140$)						
US-CA	0.01	0.01	0.19	0.73	0.23	0.65
US-UK	0.01	0.01	0.18	0.61	0.21	0.68
US-GR	0.01	0.01	0.58	1.00	0.76	1.00
1985:I - 2022:IV ($T = 155$)						
US-CA	0.01	0.01	0.05	0.05	0.03	0.04
US-UK	0.01	0.01	0.18	0.48	0.12	0.37
US-GR	0.01	0.01	0.19	0.51	0.14	0.44

Notes: This table includes results when $\Delta\mu$ & $\Delta\sigma$ as it is a statistically preferred model over a model where only $\Delta\mu$. The IP series are OECD Main Economic Indicator Releases obtained from the St. Louis Fed (FRED) website. All MC test results are obtained using $N = 99$. The MMC-LRT procedure uses a particle swarm optimization algorithm.