Monte Carlo Likelihood Ratio Tests for Markov Switching Models *

Gabriel Rodriguez-Rondon† Jean-Marie Dufour‡

October 14, 2025

Abstract

Markov switching models are widely used in economics, finance, and related fields to capture nonlinearities arising from regime shifts. Most existing studies on testing the number of regimes focus on the null hypothesis of a single regime (i.e., a linear model) 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 that address these challenges and extend to more general settings, allowing one to test a null hypothesis with M_0 regimes against an alternative with $M_0 + m$ regimes, for any $M_0 \geq 1$ and $m \geq 1$. By applying Monte Carlo methods to the likelihood ratio statistic, we develop tests that remain valid in finite samples and are applicable to non-stationary processes, non-Gaussian errors, and multivariate models–scenarios that have received limited attention in the literature. A key contribution is the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT), an identification-robust procedure with both finite-sample and asymptotic validity. Importantly, the proposed tests are also applicable to testing for the synchronization of Markov processes and to Markov switching GARCH models. Simulation results demonstrate that the proposed tests effectively control the size and exhibit strong power across a range of empirically relevant scenarios. In an empirical application to U.S. GNP and GDP growth, we find support for a three-regime model that confirms the Great Moderation and indicates a return to the low-volatility regime following both the Great Recession and the COVID-19 recession. In a second application, we use Markov switching VAR models to test for international business cycle synchronization. The results suggest that the inclusion of COVID-era data weakens the previously observed 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, Russell Davidson, Marine Carrasco, Enrique Sentana, and 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: Bank of Canada, 234 Wellington St. W, Ottawa, ON K1A 0G9. e-mail: gabriel.rodriguezrondon@mail.mcgill.ca. Wed page: https://grodriguezrondon.com.

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

1 Introduction

Markov regime-switching models were first introduced by Goldfeld and Quandt (1973) and later popularized by Hamilton (1989). These models have since become widely used in economics and finance due to their ability to capture non-linear dynamics arising from discrete shifts in the underlying data-generating process. In such models, different regimes can represent distinct phases of the economy–for instance, in the case of U.S. GNP growth, one regime might correspond to positive growth during expansions and another to negative growth during recessions.

Due to this flexibility, Markov switching models have been applied extensively in macroeconomics and finance. Applications include business cycle identification (Chauvet, 1998; Chauvet and Hamilton, 2006; Chauvet et al., 2002; Diebold and Rudebusch, 1996; Hamilton, 1989; Kim and Nelson, 1999; Qin and Qu, 2021), interest rate modeling (Garcia and Perron, 1996), financial markets (Marcucci, 2005), volatility modeling (Augustyniak, 2014; Gray, 1996; Haas et al., 2004; Hamilton and Susmel, 1994; Klaassen, 2002), time-varying correlations (Pelletier, 2006), and state-dependent impulse response functions (Sims and Zha, 2006; Caggiano et al., 2017). They have also been used in identifying structural VAR models (Herwartz and Lütkepohl, 2014; Lanne et al., 2010; Lütkepohl et al., 2021), and more recently for modeling core inflation Rodriguez-Rondon (2024); Ahn and Luciani (2024); Le Bihan et al. (2024). For comprehensive reviews, see Hamilton (2010), Hamilton (2016), and Ang and Timmermann (2012).

Beyond macro and finance, Markov switching models have found applications in climate economics (Golosov et al., 2014; Dietz and Stern, 2015), environmental and energy studies (Cevik et al., 2021; Charfeddine, 2017), industrial organization (Aguirregabiria and Mira, 2007; Sweeting, 2013), health economics (Hernández and Ochoa, 2016; Anser et al., 2021), and other machine learning applications.

A fundamental challenge in using Markov switching models is determining the number of regimes, which is typically assumed *a priori*. Since the true number of regimes is unknown in practice, it is of interest to test a model with M_0 regimes against an alternative with $M_0 + m$ regimes. However, standard hypothesis testing procedures are not readily applicable in this context because key parameters are unidentified under the null, and the usual regularity conditions needed for standard asymptotic results are violated.

The literature on the asymptotic distribution of likelihood ratio (LR) tests for Markov switching models is rich (Carter and Steigerwald, 2012; Cho and White, 2007; Garcia, 1998; Hansen, 1992;

Kasahara and Shimotsu, 2018; Qu and Zhuo, 2021), with a particularly important contribution–for our LR setting–being the SupLR(Λ_ϵ) test of Qu and Zhuo (2021) and early study of Garcia (1998). However, most of these contributions, including the asymptotic LR tests, are limited to testing a linear null ($M_0 = 1$) against a two-regime alternative ($m = 1$). An exception is Kasahara and Shimotsu (2018), who establish asymptotic validity for a parametric bootstrap LR test when comparing M_0 and $M_0 + 1$ regime models for $M_0 \geq 1$, but only under strong and restrictive assumptions. Similarly, Qu and Zhuo (2021) derive validity results for broader classes of models, yet still within the setting of $M_0 = m = 1$, and again under restrictive conditions.

Meanwhile, other researchers have proposed alternative test procedures based on moments of least-squares residuals (Dufour and Luger, 2017), parameter stability (Carrasco et al., 2014), other moment-matching conditions (Antoine et al., 2022), and score-type tests (Amengual et al., 2025b,a). The parameter stability test of Carrasco et al. (2014) is powerful but primarily designed for testing linear models ($M_0 = 1$) against two-regime alternatives ($m = 1$). It is not suited for general tests where both $M_0 \geq 1$ and $m > 1$. Moreover, most of the tests discussed thus far are only valid asymptotically and require restrictive assumptions, such as stationarity, Gaussian errors, and constrained parameter spaces, and importantly, only consider the univariate setting. In contrast, Dufour and Luger (2017) adopt the finite-sample Monte Carlo (MC) testing framework of Dufour (2006) to develop moment-based tests that are valid without relying on asymptotic theory. Although their approach also focuses on the $M_0 = m = 1$ case, it demonstrates that finite-sample valid inference is possible without heavy distributional assumptions.

This paper builds on the MC framework of Dufour (2006) and develops two likelihood-based tests for Markov switching models: the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT). These procedures allow for testing hypotheses of the form $H_0 : M_0$ vs. $H_1 : M_0 + m$, where both $M_0 \geq 1$ and $m \geq 1$, and are applicable to both univariate and multivariate models–including Hidden Markov Models, Markov-switching VARs, and MS-GARCH models–areas largely ignored in prior literature. The MMC-LRT is an exact test valid in both finite samples and asymptotically, and it is robust to the identification problems that typically arise in regime-switching models. Both the LMC-LRT and MMC-LRT avoid the need for stationarity assumptions, Gaussianity, and constrained parameter spaces. 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 yet been established in the literature.

Notably, hypothesis testing in settings with $m > 1$, multivariate models, non-Gaussian errors, or non-stationary processes has received little attention in the literature, making this study a novel contribution. Simulation results confirm that the proposed LMC-LRT and MMC-LRT procedures maintain accurate size control and exhibit strong power across a wide range of empirically relevant scenarios–including those with multiple regimes, boundary parameters (e.g., absorbing regimes), and non-stationarity. In univariate settings, both tests outperform existing moment-based and asymptotic procedures, particularly when structural shifts occur in the mean or in both the mean and variance. The MMC-LRT delivers robust inference even in small samples, while the LMC-LRT remains computationally efficient and performs well even when a well-defined likelihood is unavailable.

While our simulations and empirical applications focus on changes in the conditional mean or variance of the outcome variable, the proposed testing procedures are equally applicable to more complex univariate settings, including Markov switching GARCH (MS-GARCH) models. These models feature regime-dependent volatility dynamics and are widely used in financial econometrics to capture shifts in conditional heteroskedasticity. Extending the testing framework to MS-GARCH models is straightforward so long as the likelihood function is available under both the null and alternative, thereby enabling formal inference on the number of volatility regimes. This builds on earlier work by Gray (1996), Haas et al. (2004), and Augustyniak (2014), and further broadens the scope of finite-sample valid testing in regime-switching contexts.

Our methodology also enables testing whether different model components are governed by the same or distinct regime-switching processes. In univariate models, for example, one can test whether the mean and variance follow a common Markov chain. In multivariate settings, such as VAR models, the framework allows testing whether individual equations share a synchronized regime structure. This is conceptually related to testing for common structural breaks in the structural change literature (Oka and Perron, 2018; Perron et al., 2020), but has not been feasible in the Markov switching context due to the technical challenges involved in testing models with multiple regimes. The framework developed here makes such testing feasible, and we illustrate this in an empirical application on international business cycle synchronization.

In the multivariate simulation evidence, the proposed tests prove valuable in detecting regime synchronization or independence. Their power depends on both the degree of misalignment between regimes and the sample size. While short-lived regime shifts may obscure partial independence in small samples, full independence can still be detected reliably under moderate conditions. These findings underscore the practical advantages of simulation-based testing methods that remain valid in finite samples, especially when analyzing complex regime dynamics.

Finally, we extend the moment-based test of Dufour and Luger (2017) to non-stationary processes and provide new insights into its performance under such conditions. All test procedures are implemented in the **MSTest** R package, detailed in the companion paper Rodriguez-Rondon and Dufour (2024). Although the focus of this study is on Markov switching models, the proposed tests are also applicable in Hidden Markov Model settings.

The remainder of the paper is organized as follows. Section 2 reviews the notation, the Markov switching model framework, and estimation strategies. Section 3 introduces the proposed testing procedures and required assumptions. Section 4 presents simulation results, comparing the proposed tests to existing ones in univariate settings and showcasing results for multivariate cases. Section 5 provides two empirical applications: one on U.S. GDP and GNP growth, and another using MS-VAR models to test for business cycle synchronization across countries. Finally, Section 6 concludes.

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 $S_t = \{1, \ldots, M\}$ 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 $\mathscr{Y}_{t-1} = \sigma\text{-field}\{\ldots, w_{t-1}, y_{t-2}, w_t, y_{t-1}\}$. 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 even allows one to consider a trend function within x_t or z_t. On the other hand, 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, Markov switching models are more commonly used the econometric literature and further, Hidden Markov models can be understood as a simplified version of Markov switching models. 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 including: computational molecular biology (Baldi et al., 1994; Krogh et al., 1994), handwriting and speech recognition (Jelinek, 1997; Nag et al., 1986; Rabiner and Juang, 1986, 1993), computer vision and pattern recognition (Bunke and Caelli, 2001), and other machine learning 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

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

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

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

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

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

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.

Typically, Markov switching and Hidden Markov models are estimated using the Expectation Maximization (EM) algorithm (Dempster et al., 1977), Bayesian methods, or through the use of the Kalman filter if using the state-space representation of the model. In very simple cases, Markov switching models can 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 make the following assumption:

Assumption 3.1 *Exchangeability of Likelihood Ratio Statistics Under the Null.*

$LR_T^{(0)}$ *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* $F[x\,|\,\theta_0]$*.*

Alternatively, since these models are often used in a time series framework, especially in macroeconomic and financial applications, it is often more convenient to work with the following assumption:

Assumption 3.2 *Independence and Identical Distribution of Simulated Statistics Under the Null.*

$LR_T^{(0)}$ *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 simulated statistics* $LR_T^{(1)}(\theta_0), \ldots, LR_T^{(N)}(\theta_0)$ *are independent and identically distributed (i.i.d.) with common distribution function* $F[x\,|\,\theta_0]$*, and are independent of* $LR_T^{(0)}$*.*

This assumption is stronger than exchangeability but is particularly appealing in time series models. While exchangeability requires that the joint distribution of the test statistics is invariant to permutations (allowing for potential dependence), the i.i.d. assumption explicitly rules out dependence between the simulated statistics. In Monte Carlo test procedures, i.i.d. simulations are typically obtained by generating independent sample paths of the data-generating process under the null. This is often easier to implement and verify in time series applications where one can simulate directly from the model for a fixed θ_0. Moreover, the i.i.d. structure simplifies the theoretical justification of the test and the computation of p-values.

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. $\hat{F}_N[x \mid \theta_0]$ is the sample distribution of the simulated statistics, and $\hat{G}_N[x \mid \theta_0]$ 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 3.3 *Measurability of Extremal Simulated Distributions under the Null.*

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*

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

Using assumptions 3.2 (or 3.1) and 3.3, 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*

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

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 (Dufour, 2006; Dufour and Neves, 2019; Rodriguez-Rondon and Dufour, 2024). 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

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

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 $\hat{F}_N[x \,|\, \theta_0]$ with $\hat{F}_{TN}[x \,|\, \theta_0] = \hat{F}_N[x; LR_T(N, \theta_0)]$ and $\hat{G}_N[x \,|\, \theta_0]$ with $\hat{G}_{TN}[x \,|\, \theta_0] = \hat{G}_N[x; LR_T(N, \theta_0)]$ 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 $L_T^0(\theta_0) = L_T(\theta_1)$ 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.

An important extension of this framework is its applicability to Markov switching GARCH (MS-GARCH) models. In such settings, the conditional variance evolves according to a GARCH process whose parameters switch across regimes. Estimation can be carried out using standard methods for MS-GARCH models, and the Monte Carlo test procedures proposed here remain valid so long as the likelihood function can be evaluated under both the null and alternative. This allows researchers to formally test the number of regimes in regime-dependent volatility models, a topic of growing empirical interest, especially in financial econometrics. Applications include detecting changes in volatility regimes during crises, testing for asymmetric responses to shocks, and identifying regime shifts in volatility persistence.

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

4.1 Univariate simulation results

This section presents simulation evidence on the performance of the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT) for univariate Markov switching models proposed in this chapter. Throughout, we consider data-generating processes (DGPs) of the form

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

where $\epsilon_t \sim \mathcal{N}(0, 1)$, and both the mean and variance can switch according to a Markov process S_t. Similar DGPs have been considered by Carrasco et al. (2014), Dufour and Luger (2017), and Qu and Zhuo (2021), among others. We adopt several of the same DGPs as in Dufour and Luger (2017) to evaluate performance across a wide range of scenarios, including low and high persistence, symmetric and asymmetric regimes, changes in mean only, variance only, and both simultaneously.

Given the generality of our proposed test procedures, we also consider cases where multiple regimes exist under the null (e.g., $M_0 > 1$ and $m = 1$), under the alternative (e.g., $M_0 = 1$ and $m > 1$), or under both (i.e., $M_0 > 1$ and $m > 1$). We further examine settings where the process is non-stationary (i.e., $\phi_1 = 1.00$) and where transition probabilities lie at the boundary of the parameter space (e.g., $p_{22} = 1$). Simulations are conducted for three sample sizes: $T \in \{100, 200, 500\}$. We believe these DGPs reflect many empirically relevant settings researchers may encounter. For instance, smaller sample sizes and asymmetric regimes are especially pertinent in macroeconomic applications, where quarterly data are used and some regimes are short-lived.

For cases involving a linear model under the null hypothesis (i.e., $H_0 : M_0 = 1$) versus a Markov switching model with two regimes under the alternative (i.e., $H_1 : M_0 + m = 2$), we compare the performance of our proposed tests with those of Dufour and Luger (2017) and Carrasco et al. (2014).

The tests proposed by Dufour and Luger (2017) are also based on the Monte Carlo methodology described in Dufour (2006), but avoid certain statistical issues associated with likelihood ratio tests by using the moments of residuals from the restricted model. These moments are designed to capture features of a normal mixture distribution. The test uses four moments of the residuals, producing four Monte Carlo (MC) p-values. To combine these p-values, two approaches are proposed–one based on the minimum and one on the product of the p-values. See Dufour et al. (2004) and Dufour et al. (2014) for further discussion on combining test statistics. As a result, Dufour and Luger (2017) propose four tests: LMC_{min}, LMC_{prod}, MMC_{min}, and MMC_{prod}. An advantage of these methods is that they only require estimating the linear model (without Markov switching) under the null. However, unlike the LMC-LRT and MMC-LRT, they are limited to testing a linear null against a two-regime switching alternative.

Carrasco et al. (2014) propose a test that is optimal for detecting inconsistencies in parameter estimates across random coefficient and Markov switching models. Their procedure is broadly designed to detect parameter heterogeneity, with the Markov switching model as a special case. Like the moment-based tests in Dufour and Luger (2017), a major benefit is that it only requires estimation under the null. However, as with Dufour and Luger (2017), it applies only when there is no regime switching under the null. To address the presence of nuisance parameters, the authors propose two alternatives: a Sup-type test, denoted supTS, following Davies (1987), and an Exponential-type test, denoted expTS, as in Andrews and Ploberger (1994). Below, 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 noted, the consistency of the parametric bootstrap procedure when $m = 1$ has been shown by Qu and Zhuo (2021) for the case $M_0 = 1$, and by Kasahara and Shimotsu (2018) for $M_0 > 1$, albeit under more restrictive assumptions than those required by our tests. In particular, these asymptotic procedures requires constraining the parameter space away from the boundary when simulating the null distribution. Moreover, their consistency has only been established in univariate, stationary, and Gaussian contexts–though Kasahara and Shimotsu (2018) consider some non-Gaussian cases as well. Kasahara and Shimotsu (2018) also impose additional constraints on variance parameters during estimation. Given the similarities between the LMC-LRT and the bootstrap approach–especially when the process is stationary and parameters are well within the interior–we do not report results from a parametric bootstrap procedure that imposes such constraints. However, we believe the LMC-LRT results presented below can shed light on the bootstrap's performance both when its assumptions hold and when they are violated. It is important to emphasize that the primary distinction between these approaches lies in how the null distribution is estimated and, more fundamentally, in their respective assumptions regarding the existence and approximation of an asymptotic distribution.

All test procedures discussed above, including those proposed in this paper, can be implemented using the `R` package **MSTest** (Rodriguez-Rondon and Dufour, 2024), available via the Comprehensive R Archive Network (CRAN) and described in a companion paper by Rodriguez-Rondon and Dufour (2024). All simulation results reported below were obtained using this package. For all simulations, the nominal significance level is set at $\alpha = 0.05$, and each result is based on 1,000 replications of the DGP.

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

Test	$\phi = 0.10$			$\phi = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500
			$H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$			
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
			$H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 3$			
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.

The results under the null hypothesis of no Markov switching (i.e., $H_0 : M_0 = 1$) are reported in Table 1. The table consists of two panels: the first evaluates the alternative hypothesis of a Markov switching model with two regimes, while the second considers a three-regime alternative. The rejection frequencies of the LMC-LRT proposed in this work are remarkably close to the nominal significance level. As expected from theory, the MMC-LRT exhibits empirical rejection rates at or below 5% under the null hypothesis.

The results for the moment-based tests proposed by Dufour and Luger (2017)–namely LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$–are consistent with those of our Monte Carlo likelihood ratio tests. The expTS test shows mild over-rejection in some cases with smaller sample sizes but performs well when $T = 500$. This behavior is anticipated, as expTS is an asymptotic procedure. In contrast, the supTS test demonstrates excellent size control across all sample sizes.

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: Here, we consider $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$. 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.

To study the power properties of the tests, 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 remaining transition probabilities are set as $p_{ij} = 1 - p_{ii}$ for $j \neq i$. In the first case, both regimes are symmetric and relatively persistent. Given the symmetry, the stationary distribution is $\boldsymbol{\pi} = (\pi_1, \pi_2) = (0.50, 0.50)$, implying that, on average, equal time is spent in each regime in the long run. In contrast, the second case features asymmetric regimes, with one regime being more persistent than the other. This results in $\boldsymbol{\pi} = (0.83, 0.17)$, indicating that one regime dominates in terms of long-run frequency.

Table 2 reports the empirical power of the tests. Since the MMC-LRT procedure accounts for a wider range of nuisance parameter values consistent with the null compared to the LMC-LRT, it consistently exhibits lower power across all settings. The same is true for the moment-based approaches. Specifically, the LMC$_{min}$, LMC$_{prod}$, MMC$_{min}$, and MMC$_{prod}$ procedures display the weakest power when only the mean changes and persistence is low. The supTS and expTS tests also exhibit very low power when only the mean changes and persistence is high. Qu and Zhuo

(2021) offers further discussion on why the supTS test performs poorly under high persistence.

In contrast, the LMC-LRT and MMC-LRT proposed here demonstrate higher power in both of these challenging scenarios involving changes in the mean only. When the variance changes, all tests exhibit improved power, although the LMC-LRT and MMC-LRT generally continue to outperform the others. This pattern remains when both the mean and variance change simultaneously, with our proposed tests maintaining a power advantage despite overall improvements across all procedures.

Overall, in the case where $H_0 : M_0 = 1$ and $H_1 : M_0 + m = 2$, the LMC-LRT and MMC-LRT maintain similar size properties to the alternative tests considered, while offering superior power. This is not surprising, as the moment-based procedures, supTS, and expTS are all derived primarily from the model under the null. As such, even in relatively simple settings where other test procedures are applicable, the methods proposed here may offer a more powerful alternative.

Table 3: Empirical performance of test when $M_0 = 1$, $m = 1$, & 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: Here, we consider $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$. 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.

As previously discussed, a notable feature of the LMC-LRT and MMC-LRT is their applicability even when the process is non-stationary or contains parameters on the boundary of the parameter space. As mentioned in Section 3, in such cases the likelihood function is not theoretically well-defined. Therefore, in these scenarios, our procedures are more accurately described as Local Monte Carlo and Maximized Monte Carlo *pseudo* Likelihood Ratio Tests. While this distinction is important, we continue to refer to them as LMC-LRT and MMC-LRT for consistency.

Table 3 reports the rejection frequencies under both the null and alternative hypotheses in the non-stationary case where $\phi_1 = 1.00$. We evaluate the performance of the LMC-LRT and MMC-LRT under unit-root DGPs. The results indicate that the supTS and expTS tests fail to maintain proper size control. Specifically, as sample size increases and the process more closely resembles a non-stationary one, these tests exhibit substantial over-rejection.

In contrast, the results suggest that Monte Carlo-based procedures exhibit remarkably accurate size properties in the non-stationary case. This includes the moment-based tests proposed by Dufour and Luger (2017). Under the alternative hypothesis, power improves when regimes are asymmetric—particularly when only the mean changes or when both the mean and variance change. All tests perform best with larger sample sizes and when the variance, or both the mean and variance, differ under the alternative.

To our knowledge, simulations for the moment-based procedures in this non-stationary setting were not reported in Dufour and Luger (2017). We are therefore the first to provide simulation evidence on the performance of moment-based approaches for non-stationary processes.

Table 4: Empirical power of test when $M_0 = 1$, $m = 1$, & parameter is at the boundary $(p_{11}, p_{22}) = (0.9, 1.0)$

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: Here, we consider $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$. 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 discussed case where the regimes are asymmetric, with one regime being absorbing–i.e., the transition probability lies on 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 $j \neq i$.

In this setting, we find that low persistence combined with changes in the mean leads to higher power for smaller sample sizes ($T = 100$ and $T = 200$). When the sample size increases to $T = 500$, power is high in all scenarios, except in the case of high persistence and changes in the mean only.

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: Here, we consider $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 3$. 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 5 reports the rejection frequencies of the LMC-LRT and MMC-LRT under the alternative hypothesis when $M_0 = 1$ and $m = 2$. That is, we consider a linear model under the null hypothesis (i.e., $H_0 : M_0 = 1$) versus a Markov switching model with three regimes under the alternative (i.e., $H_1 : M_0 + m = 3$). The results show consistently high power across all cases considered, which is expected given that the alternative is further from the null.

Table 6 presents results for the case where the null hypothesis involves two regimes (i.e., $H_0 : M_0 = 2$) and the alternative consists of three regimes (i.e., $H_1 : M_0 + m = 3$). Generally, detecting additional regimes when the null already involves multiple regimes is more challenging. Nonetheless, the size results indicate that both proposed test procedures maintain appropriate size control.

For the DGPs considered here, power appears to depend more heavily on the presence of changes in the mean. In particular, scenarios involving mean changes yield substantially higher power than those involving changes in variance only. This finding contrasts with earlier results, which suggested that variance changes were an important source of power for all test procedures. However, when multiple regimes are present, a high-variance regime may obscure lower-variance regimes and associated shifts in the mean. As a result, changes in location (mean) may become more informative for distinguishing regimes. These findings highlight a potentially important avenue for future research.

Table 6: Empirical size & power of test when $M_0 = 2$ & $m = 1$

Test	Empirical size											
	$(p_{11}, p_{22}) = (0.90, 0.90)$						$(p_{11}, p_{22}) = (0.90, 0.50)$					
	$\phi = 0.10$			$\phi = 0.90$			$\phi = 0.10$			$\phi = 0.90$		
	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	5.9	6.6	5.7	4.3	5.1	5.0	4.8	4.4	5.1	4.9	5.5	4.8
MMC-LRT	2.3	2.2	3.4	1.7	3.1	3.3	2.6	2.3	3.1	2.1	2.1	2.9
	$\Delta\sigma$											
LMC-LRT	4.6	4.5	5.5	5.3	4.9	4.3	4.5	5.8	5.7	4.2	4.6	5.9
MMC-LRT	2.2	2.1	3.1	2.0	3.2	2.8	2.6	2.9	3.0	2.6	2.4	3.8
	$\Delta\mu$ & $\Delta\sigma$											
LMC-LRT	5.8	4.4	5.1	4.2	5.2	5.3	4.2	4.8	5.4	4.6	5.6	5.2
MMC-LRT	2.5	2.8	4.1	2.1	2.3	3.8	2.9	3.4	4.0	2.4	2.5	3.4
	Empirical power											
	$(p_{11}, p_{22}, p_{33}) = (0.90, 0.90, 0.90)$						$(p_{11}, p_{22}, p_{33}) = (0.90, 0.50, 0.50)$					
	$\phi = 0.10$			$\phi = 0.90$			$\phi = 0.10$			$\phi = 0.90$		
	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	39.9	84.2	94.7	6.7	7.2	8.6	12.6	27.4	52.2	11.9	11.8	12.3
MMC-LRT	34.1	81.1	90.9	4.9	5.6	6.1	8.1	20.3	44.6	8.3	7.8	7.6
	$\Delta\sigma$											
LMC-LRT	8.5	24.0	57.6	10.0	22.2	56.2	6.4	9.0	20.7	5.8	8.9	21.1
MMC-LRT	6.5	19.2	52.7	6.2	17.7	49.9	4.1	6.9	18.3	4.1	5.6	14.2
	$\Delta\mu$ & $\Delta\sigma$											
LMC-LRT	40.4	88.4	100.0	14.2	26.8	56.8	15.4	32.4	79.2	9.4	16.2	30.0
MMC-LRT	35.2	74.3	93.7	11.6	21.4	50.2	11.1	28.6	74.4	7.0	11.5	24.5

Notes: Here, we consider $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$. 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 7: Empirical size of test when $M_0 = 2$ & $m = 1$ for alternative DGPs

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)

Table 7 presents results for an alternative set of DGPs, still within the context of testing a null hypothesis of two regimes (i.e., $H_0 : M_0 = 2$) against 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 DGPs considered in Kasahara and Shimotsu (2018), and we also report the Boot-LRT results from that paper, except for $T = 100$, as those were not provided by the authors.

As previously discussed, the parametric bootstrap and LMC-LRT procedures share many similarities. However, a key distinction is that the LMC-LRT does not require enforcing the additional assumptions needed to ensure the asymptotic validity of the bootstrap when estimating the null distribution. While larger sample sizes should improve the approximation quality of both procedures, the LMC-LRT does not rely on the existence of an asymptotic distribution, allowing for fewer simulations to achieve valid inference. In contrast, Kasahara and Shimotsu (2018) impose such constraints and an additional one on the variance parameters during model estimation and use $N = 199$ bootstrap simulations, whereas we use $N = 99$ and impose no such constraints. These differences help explain the discrepancies observed between the LMC-LRT and the parametric bootstrap test results.

Nonetheless, the results suggest that both procedures exhibit broadly similar patterns across these DGPs. Specifically, Table 7 shows that both the LMC-LRT and the parametric bootstrap tests exhibit some over-rejection for smaller sample sizes ($T = 100$ and $T = 200$), particularly in the case of the bootstrap test. As expected, the rejection frequencies approach the nominal level when $T = 500$. Meanwhile, the MMC-LRT performs as expected, maintaining rejection frequencies at or below 5% even for small samples. This highlights the strength of the MMC-LRT as a valid test procedure in both finite samples and asymptotic settings.

In summary, the univariate simulation results demonstrate that the proposed LMC-LRT and MMC-LRT procedures perform well across a wide range of empirically relevant scenarios. Both tests exhibit accurate size control, even in challenging cases involving multiple regimes, non-stationarity, and boundary parameters (absorbing regimes). Moreover, they deliver superior power relative to existing moment-based and asymptotic alternatives, particularly when changes in the mean or both mean and variance are present. While the MMC-LRT provides robust inference even in small samples, the LMC-LRT remains computationally efficient and exhibits favorable properties despite the lack of a formally defined likelihood in some settings. These findings underscore the flexibility and reliability of the proposed procedures for regime detection in univariate Markov switching models.

4.2 Multivariate simulation results and synchronization testing

The importance of modeling structural changes in macroeconomic and financial time series, within a multivariate context is well established. For instance, persistent shifts in mean or volatility–such as those observed during the Great Moderation or in financial market volatility regimes–are often common across many variables and well captured by Markov-switching models. Nonetheless, most multivariate applications restrict attention to two regimes for tractability, in part due to the lack of reliable tests for selecting the number of regimes in higher-dimensional settings (e.g., Primiceri (2005)).

We now turn to evaluating the performance of the proposed Local Monte Carlo (LMC-LRT) and Maximized Monte Carlo (MMC-LRT) likelihood ratio tests in multivariate settings, as discussed in Section 2. These models are particularly relevant in empirical applications involving multiple macroeconomic or financial time series, whether across countries, asset classes, or sectors. They are also suitable for testing hypotheses about common regime structures, including–but not limited to–synchronized dynamics between variables. We revisit such an application in the empirical section that follows.

We begin by assessing the size and power properties of the LMC-LRT and MMC-LRT when testing the null hypothesis of a single regime ($M = 1$) against the alternative of two regimes ($M = 2$) in a bivariate VAR(1) model. The experiments explore mean shifts ($\Delta\mu$), variance shifts ($\Delta\sigma$), and joint shifts. Results are summarized in Table 8.

Table 8: Empirical performance of test for bi-variate VAR model

	Empirical size					
Test	$\max(\lambda) = 0.10$			$\max(\lambda) = 0.90$		
	T=100	T=200	T=500	T=100	T=200	T=500
LMC-LRT	4.2	4.2	6.2	3.6	4.6	5.6
MMC-LRT	3.8	3.5	4.4	1.2	3.1	3.3

	Empirical Power											
	$(p_{11}, p_{22}) = (0.90, 0.90)$						$(p_{11}, p_{22}) = (0.90, 0.50)$					
Test	$\max(\lambda) = 0.10$			$\max(\lambda) = 0.90$			$\max(\lambda) = 0.10$			$\max(\lambda) = 0.90$		
	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	25.0	63.6	95.9	8.2	11.8	28.2	12.2	26.0	75.1	9.0	24.4	54.4
MMC-LRT	19.4	62.2	91.2	6.6	11.0	27.6	9.0	22.6	69.4	9.7	18.9	50.8
	$\Delta\sigma$											
LMC-LRT	46.4	90.6	100.0	53.0	91.8	100.0	34.4	68.6	100.0	38.2	73.6	96.7
MMC-LRT	45.6	82.4	100.0	51.0	85.2	99.8	31.8	64.6	100.0	34.4	70.1	92.8
	$\Delta\mu$ & $\Delta\sigma$											
LMC-LRT	87.0	100.0	100.0	60.4	93.4	99.7	73.6	98.2	100.0	62.0	93.6	100.0
MMC-LRT	82.0	100.0	100.0	53.6	86.1	96.1	63.4	93.9	99.8	54.7	81.0	96.0

Notes: The nominal level is 5%. Here, $\max(\lambda)$ is the largest eigenvalue of the autoregressive coefficient matrix Φ. 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.

The tests exhibit good size control and increasing power with sample size. The MMC-LRT performs particularly well in small samples and in the presence of strong persistence or heteroskedasticity. These results align closely with the univariate findings and confirm that the proposed testing procedures extend effectively to multivariate settings. To the best of our knowledge, this is the first simulation-based evidence for regime testing procedures in multivariate Markov-switching models.

Having established the accuracy of the tests for detecting regime-switching in multivariate systems, we now apply them in a related but distinct context: testing whether regime structures are synchronized across equations. While similar logic could be applied in univariate settings–e.g., where different coefficients such as the mean and variance follow separate regime paths as considered in Sims and Zha (2006)–the notion of (de)synchronization is particularly intuitive in a multivariate framework. For example, one might wish to test whether each equation in a system is governed by a common versus independent regime-switching process.

Related work in the structural break literature has proposed methods for detecting common breaks across equations (see Oka and Perron (2018)) or across coefficients within a single equation (see Perron et al. (2020)). However, such tools are not readily available in the Markov-switching framework, owing to the complexity of estimation and inference with multiple regimes. The framework proposed here fills this gap, enabling formal hypothesis tests for synchronized regime structures in multivariate systems.

To illustrate this, we simulate data from a bivariate MS-VAR model in which each equation may be governed by an independent Markov process. Our goal is to test whether the regime-switching processes are synchronized (i.e., governed by the same latent state) or independent. If both equations follow the same regime path, only two joint states are needed. But if they evolve independently, up to four distinct joint regimes may arise—one for each possible state combination.

For example, consider the following bivariate model for economies a and b:

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

We are interested in testing whether the regime-switching behavior of the two economies is governed by the same latent Markov process (i.e., $S_{a,t} = S_{b,t} = S_t$) or by independent processes (i.e., $S_{a,t} \neq S_{b,t}$). Suppose both $S_{a,t}$ and $S_{b,t}$ each take values in $\{1, 2\}$. Then the joint state S_t^* must account for up to four possible combinations:

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

Synchronized regimes imply only two of these occur (e.g., $S_t^* = 1$ if $S_{a,t} = 1$ and $S_{b,t} = 1$ and $S_t^* = 2$ if $S_{a,t} = 2$ $S_{b,t} = 2$), while desynchronization leads to three or four observable combinations, depending on the degree of offset between $s_{a,t}$ and $s_{b,t}$.

Interestingly, we can frame the problem as a test on the number of regimes in the MS-VAR model:

$$\begin{aligned} H_0 &: M_0 = 2 \quad \text{(synchronized cycles)} \\ H_{1a} &: M_0 + m = 3 \quad \text{(partial dependence)} \\ H_{1b} &: M_0 + m = 4 \quad \text{(independent cycles)} \end{aligned} \tag{29}$$

In the following, simulations are designed such that both series follow two-regime Markov processes, but desynchronization arises through lead-lag differences in regime transitions. The offset duration is governed by a parameter ξ, scaled by T, which determines the time spent in misaligned regimes.

Table 9: Empirical size & power of test for independent Markov processes

Test	$H_0 : M = 2$ vs. $H_0 : M = 3$												$H_0 : M = 2$ vs. $H_0 : M = 4$											
	Empirical size																							
	T=100				T=200				T=500				T=100				T=200				T=500			
LMC-LRT	1.2				2.6				4.0				3.0				4.0				5.8			
MMC-LRT	1.0				1.8				2.8				2.1				2.2				3.1			
	Empirical power																							
	$\xi = 0.02$						$\xi = 0.10$						$\xi = 0.02$						$\xi = 0.10$					
	T=100		T=200		T=500		T=100		T=200		T=500		T=100		T=200		T=500		T=100		T=200		T=500	
LMC-LRT	5.0		8.0		41.2		3.8		52.8		98.8		28.6		41.0		90.8		42.0		100.0		100.0	
MMC-LRT	3.4		7.3		34.6		2.0		44.9		92.6		20.9		37.1		85.2		38.2		98.9		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. Rejection frequencies are obtained using 500 replications. MC tests use $N = 99$ simulations. Each equation is generated $M = 2$ regimes and the parameter ξ along with the sample size T determine the duration the (lead or lag) third and/or fourth regimes (i. e., duration is $T \times \xi$).

Table 9 reports the size and power of the LMC-LRT and MMC-LRT. The data-generating process (DGP) allows both the mean and the covariance matrix in the bivariate MS-VAR model to vary according to a latent Markov process. As before, we evaluate performance across sample sizes $T \in \{100, 200, 500\}$.

The top panel presents empirical size results under the null hypothesis of $M = 2$ regimes, tested against alternatives with $M = 3$ (left) and $M = 4$ (right). The LMC-LRT maintains rejection frequencies close to the nominal 5% level, though some mild under-rejection is observed when testing against the $M = 3$ alternative. As expected, the MMC-LRT behaves consistently with the theoretical properties outlined in Dufour (2006), maintaining rejection frequencies at or below the nominal level in all cases.

To evaluate power, the lower panel reports rejection frequencies under DGPs with three or four regimes. These represent cases of partial (three regimes) or full (four regimes) independence between the two Markov processes. The parameter ξ controls the duration of the extra regime(s), determining the degree of temporal offset between the processes (e.g., lead-lag behavior). We consider two values, $\xi \in \{0.02, 0.10\}$, so that the regime duration scales proportionally with sample size: $T \times \xi$.

The power results suggest that detecting partial desynchronization (i.e., three regimes) is challenging in smaller samples unless the third regime is sufficiently persistent (e.g., $T = 200$ with $\xi = 0.10$). In contrast, both tests exhibit high power in detecting full independence (i.e., four regimes), even when regime durations are short. This reflects the greater distinctness of regime paths in the four-regime case, which facilitates detection.

In summary, the ability to detect synchronized versus independent regime structures using the proposed tests depends on both the extent of misalignment and the available sample size. Short-lived regimes can obscure partial independence unless a sufficiently long series is observed. However, full independence is easier to detect, even in moderate samples. These findings highlight the practical value of accurate finite-sample testing procedures when analyzing regime dynamics in multivariate time series.

5 Applications

In this section, we begin by analyzing U.S. GNP and GDP growth rates, with a focus on GDP, as it has become the more commonly used series in recent empirical studies. These series have been widely used in the literature to evaluate test procedures for Markov switching models, including in Hansen (1992), Carrasco et al. (2014), Dufour and Luger (2017), Qu and Zhuo (2021), and Kasahara and Shimotsu (2018), among others. All datasets used in this univariate application are available in the **MSTest** R package, which facilitates replication of the estimation and hypothesis testing results presented here.

We then turn to a second application in a multivariate setting. The objective is to apply the LMC-LRT and MMC-LRT procedures to assess the synchronization of international business cycles. This example illustrates the practical value of having a testing procedure that can accommodate multivariate settings–such as Markov switching VAR models–and hypotheses involving more than one additional regime (i.e., $m > 1$). These features are beyond the scope of previously proposed tests in the literature. Moreover, this application highlights how the proposed methods can be used to test for common breaks or shared regime structures–capabilities that have existed in the broader structural break literature for quite some time–but now within a Markov switching framework.

5.1 U.S. GNP and GDP growth

Many 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 in Hamilton (1989). Notable studies employing U.S. GNP data for regime testing include Hansen (1992), Carrasco et al. (2014), and Dufour and Luger (2017). Hansen (1992) examines the original quarterly sample from 1951:II to 1984:IV, as used in Hamilton (1989), with $p = 4$ lags and a specification where only the mean changes across regimes. In this case, the proposed test fails to reject the null hypothesis of a linear model (i.e., $M = 1$). Similarly, Carrasco et al. (2014) and Dufour and Luger (2017) also use this sample and reach the same conclusion.

These latter two studies also consider an extended sample from 1951:II to 2010:IV, which includes the Great Recession. They continue to use four lags ($p = 4$), but now also evaluate an alternative where both the mean and variance change across regimes, as suggested by Kim and Nelson (1999). Allowing for changes in variance is sensible for two reasons. First, the extended sample includes the structural decline in macroeconomic volatility during the mid-1980s, known as the Great Moderation. Second, since the objective is to capture recessionary episodes, it is reasonable to assume that volatility increases during such periods. For this extended sample, both Carrasco et al. (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, but only when the variance is allowed to change. As discussed in Qu and Zhuo (2021), when using GDP data, the inclusion of the Great Recession appears to be crucial for the supTS test of Carrasco et al. (2014) to reject linearity. However, when only the mean is allowed to change, the supTS and expTS tests continue to fail to reject the null. In contrast, Qu and Zhuo (2021), using GDP rather than GNP data, find stronger evidence in favor of a two-regime model even when only the mean is allowed to change.

To complement the existing literature, we consider the same two samples of U.S. GNP data as in prior studies, along with an extended sample spanning 1951:II to 2024:II. Results from the LMC-LRT and MMC-LRT applied to these three U.S. GNP growth rate samples are presented in Table 10. For the first two samples, our findings are broadly consistent with earlier tests. However, unlike Carrasco et al. (2014), we find evidence in favor of a two-regime model ($M = 2$) even when only the mean is allowed to change in the second sample that includes the Great Recession. This result is more in line with Qu and Zhuo (2021), who find similar evidence using U.S. GDP data.

Table 10: 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

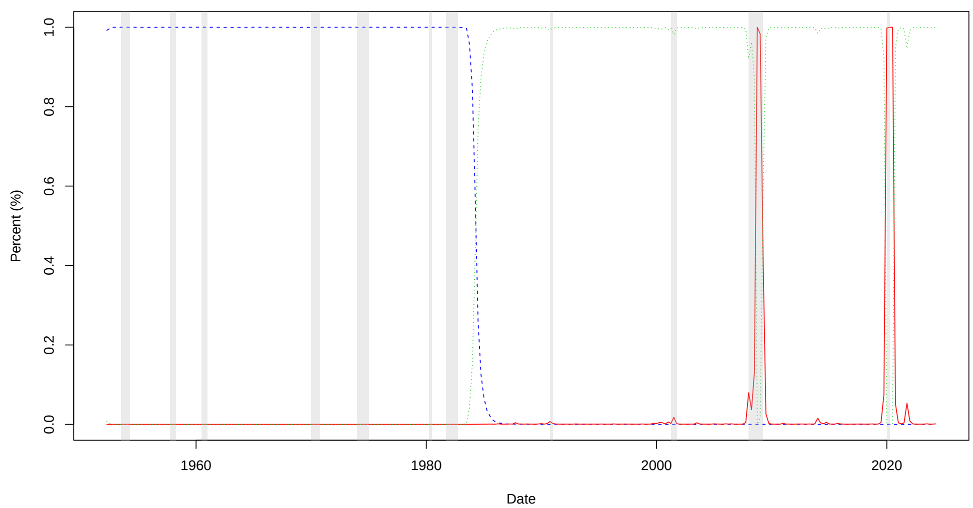
Notes: The GNP 1951:II-1984:IV series ($T = 135$) is the same as the one used in Hamilton (1989), Hansen (1992), and Carrasco et al. (2014). The GNP 1951:II-2010:IV series ($T = 239$) is the same as the one used in Carrasco et al. (2014) and Dufour and Luger (2017). The GNP 1951:II-2024:II series ($T = 293$) 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).

We extend the analysis by formally testing the null hypothesis of $M = 2$ regimes against the alternative of $M = 3$. To our knowledge, this is the first such test applied to U.S. GNP data for this sample. In both the mean-only and mean-and-variance-switching specifications, we fail to reject the $M = 2$ null, confirming that two regimes are sufficient. However, when we turn to the third, longer sample, we reject the null of $M = 2$ in favor of a Markov switching model with $M = 3$ regimes. We further test this three-regime model against a four-regime alternative and fail to reject the null, thereby supporting $M = 3$ as the preferred specification for the extended sample.

Figure 1 shows the smoothed regime probabilities for the $M = 3$ model in which both the mean and variance change. Smoothed probabilities for additional models are shown in Figures 4–7.

Parameter estimates for this model and others that include regime-dependent variance are provided in Table 15. In the $M = 3$ model, two regimes are expansionary with positive means, though one is characterized by substantially lower volatility ($\mu_1 = 1.87$, $\mu_2 = 1.31$, $\sigma_1 = 1.09$, $\sigma_2 = 0.49$). This reduction in volatility is consistent with the Great Moderation and is reflected in the smoothed regime probabilities, which shift in the mid-1980s. The third regime represents a deep recessionary state that captures both the Great Recession and the COVID-19 recession. Similar to the findings in Gadea et al. (2018) and Gadea et al. (2019), we observe that the low-volatility regime re-emerges after the Great Recession—and, in our case, again following the 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.

Given that much of the recent literature now relies on U.S. GDP data, we now shift our focus to this series. Both Qu and Zhuo (2021) and Kasahara and Shimotsu (2018), for example, use U.S. GDP instead of GNP when testing the number of regimes in Markov switching models. For this application, we consider the extended sample from 1951:II to 2024:II. This longer sample is particularly interesting because it includes the COVID-19 period, which poses challenges for macroeconomic modeling due to its stark departure from historical patterns.

Several approaches have been proposed to address the impact of the COVID-19 period. One strategy is to treat it as a known structural break. A benefit of this approach is that, by incorporating explanatory variables to account for the shock, one may justify a simpler model specification—potentially requiring fewer regimes to capture the non-linearities in the series. More generally, a key advantage of the testing procedure proposed here is its flexibility: users can include control variables to account for known structural features and then test whether the number of regimes can be reduced conditionally, given those controls.

To assess the robustness of our procedure, we evaluate the number of regimes in U.S. GDP growth under different treatments of structural breaks. Specifically, we include a dummy variable equal to 1 from 2020:I to 2021:IV and 0 otherwise, to control for the COVID period. We also include a second dummy equal to 1 from 1951:II to 1983:IV and 0 elsewhere, to control for the Great Moderation.[1] These simple mean-shift controls offer a baseline way to account for known structural breaks. However, they are likely insufficient in capturing the full dynamics of such episodes, which are typically driven by changes in both the conditional mean and conditional variance. As such, these specifications are intended as a starting point, rather than a comprehensive treatment.

Accordingly, we examine four models: Model 1 includes no dummy variables; Model 2 includes only the Great Moderation dummy; Model 3 includes only the COVID dummy; and Model 4 includes both dummies. For each specification, we test the number of regimes under two setups– one where only the mean changes and another where both the mean and variance change across regimes– leading to eight models in total. The results of these tests are reported in Table 11.

Table 11: 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).

As with the GNP growth data for the same sample period, we find evidence supporting a model with $M = 3$ regimes for U.S. GDP growth. Unlike the GNP models, we use one lag ($p = 1$) in

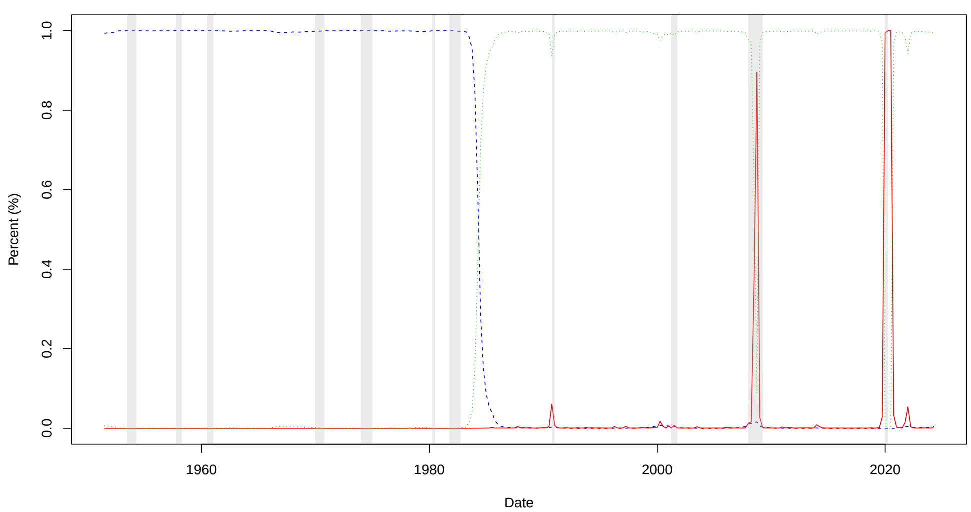
[1] We first estimate a Markov switching model without this dummy and find strong evidence that one of the regimes corresponds to the Great Moderation. We use the smoothed probabilities from this model to date the period.

this case, consistent with the specification in Qu and Zhuo (2021). Since our sample differs slightly from theirs, we first verified that a lag order of one remains appropriate and confirmed that this is indeed the case.

To assess which of the eight candidate models is preferred for this sample, we apply a likelihood ratio test (LRT) to evaluate the significance of the dummy variables. In this setting, the conventional regularity conditions are satisfied, allowing us to rely on standard LRT inference. Table 13 reports the estimates and log-likelihood values for each model. Across all specifications–whether only the mean changes or both the mean and variance–the inclusion of dummy variables does not significantly improve the log-likelihood, resulting in small LRT statistics and no statistical significance.

In addition, in all cases, the model that allows for changes in both the mean and variance is preferred to the corresponding model where only the mean changes. Taken together, these findings indicate that a model with $M = 3$ regimes, regime-dependent means and variances, and no dummy variables provides a sufficiently good fit to the data.

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

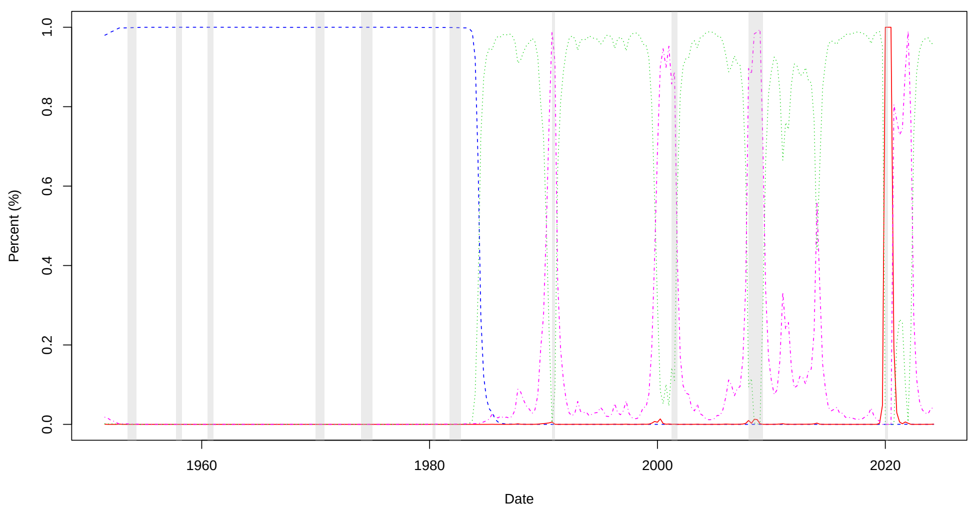


Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

The smoothed regime probabilities for this model, shown in Figure 2, closely resemble those obtained for the GNP case, though the parameter 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 $\sigma_3 = 6.5$). Notably, the inclusion of dummy variables did not alter the outcome of the hypothesis test for the number of regimes. This is likely because treating these episodes as known structural breaks in the conditional mean is insufficient to capture their full effects.

In contrast, Markov switching models that allow for regime-dependent variances are better equipped to account for such changes, which are central features of the Great Moderation and likely the COVID-19 period. Incorporating more sophisticated specifications–such as heteroskedastic error structures (e.g., GARCH) or structural breaks in the variance–may further improve model performance and can be considered within our regime testing framework. Regardless of the chosen specification, the testing procedures proposed here offer a useful tool for assessing whether a given model adequately captures such features, or whether additional regimes are required.

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.

It is also worth noting that, as shown in Figure 3, a model with $M = 4$ regimes can capture relatively milder recessions as distinct episodes. In this specification, the post-COVID recovery period is also identified as a separate regime. However, the log-likelihood value of the four-regime model, reported in Table 13, is very close to that of the three-regime model, which explains why the test fails to reject the null hypothesis of $M = 3$ regimes.

Although this difference is not statistically significant, the $M = 4$ specification may still be of interest if the primary goal is to more finely distinguish phases of the business cycle, such as separating deep from shallow recessions or identifying recovery periods explicitly.

In summary, the univariate empirical application highlights the effectiveness of the proposed LMC-LRT and MMC-LRT procedures in identifying regime structure in U.S. output growth data. For both GNP and GDP, we find consistent evidence supporting a three-regime model when both the mean and variance are allowed to change–especially in longer samples that include the Great Recession and COVID-19 period. Importantly, our results appear to be robust to the inclusion of control variables for known structural breaks in the conditional mean. These results underscore the flexibility and reliability of our proposed tests in practical settings and motivate their use in more complex multivariate applications, such as testing for synchronization in international business cycles.

5.2 Synchronization of business cycles

The synchronization of business cycles has re-emerged as a topic of renewed interest in light of recent global events, including the COVID-19 pandemic, persistent supply chain disruptions, and tariff disputes. These shocks have highlighted vulnerabilities associated with deep economic integration. Existing literature suggests that trade openness tends to amplify business cycle comovement across countries (e.g., Dées and Zorell (2012)). While a variety of methodologies have been proposed to measure business cycle synchronization, relatively few formal testing procedures exist. Moreover, many available approaches yield mixed results, rely on restrictive assumptions (e.g., linearity), focus primarily on correlations, or lack theoretical validation–particularly in finite samples. Examples include the procedures proposed by Phillips (1991a) and Camacho et al. (2006).

Phillips (1991a) analyzes business cycle transmission between two economies using a four-regime bivariate MS model. Regimes are defined by the joint states of both economies, and inference is conducted based on the estimated 4×4 transition matrix. However, this approach can lead to ambiguous conclusions: for example, the null hypotheses of both perfect correlation and full independence may fail to be rejected for certain country pairs (e.g., US–UK), making interpretation difficult.

Alternative strategies include estimating univariate MS models and correlating smoothed recession probabilities across countries (e.g., Guha and Banerji (1999), Artis (2004)). Yet Camacho and Perez-Quiros (2006) show that such correlations are biased downward, particularly when countries are in fact synchronized. To address this, Camacho and Perez-Quiros (2006) and Bengoechea et al. (2006) propose a bivariate MS model with a desynchronization parameter δ_{ab}, defined via a linear combination of synchronized and unsynchronized regimes. They introduce a simulation-based test of $H_0 : \delta_{ab} = 0$ (perfect synchronization) against $H_1 : \delta_{ab} = 1$ (full desynchronization). Though not explicitly framed as such, their procedure resembles that of the Local Monte Carlo (LMC) test of Dufour (2006).

More recent contributions include Bayesian state-space models (e.g., Leiva-Leon (2014b)) and frameworks allowing for time-varying synchronization (e.g., Leiva-Leon (2014a), Leiva-Leon (2017)). These approaches offer flexibility for measuring dynamic interdependence but often assume only two regimes and lack formal frequentist tests for the number of regimes. In contrast, the methods developed here are grounded in frequentist inference, allow for more than two regimes, and provide size-controlled tests of synchronization.

For our empirical application, we adopt a setup similar to Phillips (1991a), using quarterly industrial production (IP) data and, since IP captures only the supply side of the economy, we also include real GDP data following Camacho and Perez-Quiros (2006), for the United States, Canada, the United Kingdom, and Germany. Specifically, we apply the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT) to bivariate MS-VAR(1) models for three country pairs: (1) US–Canada, (2) US–UK, and (3) US–Germany. These tests allow us to evaluate the null hypothesis of perfectly synchronized business cycles against alternatives involving partial or full independence.

This approach offers two key advantages. First, it allows the data to determine the appropriate regime structure, which may include multiple types of recessionary or expansionary states that differ in timing or magnitude. Second, it accommodates various forms of desynchronization. For instance, if each economy follows a two-regime Markov process but transitions occur at slightly different times, a three-regime model may be more appropriate than a four-regime one. The methodology developed in this work is flexible enough to capture such patterns. In addition, the MMC variant provides exact inference in small samples—a critical advantage when using quarterly macroeconomic data.

We use seasonally adjusted quarterly data and examine two samples: 1985:I–2019:IV and 1985:I–2022:II, the latter of which includes the COVID-19 period. The starting point ensures data consistency and avoids earlier volatility shifts associated with the Great Moderation. For each country pair, we estimate a bivariate MS-VAR model with $p = 1$, as determined by a bottom-up likelihood ratio testing procedure. This specification is consistent with Phillips (1991b) and other studies that typically use one or no lags.

Figures 8 and 9 in the Appendix display the series for each country and sample. The extreme fluctuations during the COVID-19 episode justify showing the full and pre-COVID samples separately. The two GDP samples considered for each country are shown in Figure 9.

As evident in the full-sample figures, the COVID-19 shock introduces sharp but short-lived volatility. Although some literature recommends treating such shocks as outliers or structural breaks, we do not do so here. Instead, we use unadjusted data throughout and leave robustness checks for future work. Notably, in the previous subsection, we showed that controlling for COVID-19 as a structural break in the conditional mean has little effect on the estimated number of regimes in a univariate GDP model. Still, exploring alternative treatments for such shocks remains an important avenue for future research.

Table 12: 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.

Tables 12 here and 16 in the appendix report the results for real GDP and industrial production, respectively. Each table contains results for both the pre-COVID sample ending in 2019:IV (top panel) and the full sample ending in 2022:II (bottom panel).

We begin by testing $H_0 : M_0 = 1$ against $H_a : M_0 + m = 2$ to assess whether regime-switching dynamics are present. The first two columns of each table confirm significant regime-switching behavior across all country pairs, justifying $M_0 = 2$ as a reasonable baseline.

To test for synchronization, we next evaluate $H_0 : M_0 = 2$ against $H_{1a} : M_0 + m = 3$ and $H_{1b} : M_0 + m = 4$, which test for partial and full independence, respectively. Results using real GDP (Table 12) show that the US business cycle was synchronized with Canada and Germany prior to COVID-19, but not with the UK. In the full sample, all country pairs show signs of increased desynchronization, with particularly strong divergence post-COVID.

Industrial production results (Table 16) show broadly similar patterns with respect to the US and Canada relationship. Prior to COVID-19, the US was largely synchronized with all three economies. Post-COVID, evidence of desynchronization increases, especially between the US and Canada.

These findings provide preliminary evidence that international business cycle synchronization weakened following the COVID-19 shock. A plausible explanation is that national recoveries varied substantially in timing, policy response, and structural characteristics.

6 Conclusion

This paper proposes two new testing procedures for determining the number of regimes in Markov switching models: the Local Monte Carlo Likelihood Ratio Test (LMC-LRT) and the Maximized Monte Carlo Likelihood Ratio Test (MMC-LRT). Built on the Monte Carlo testing framework of Dufour (2006), these procedures address long-standing challenges in hypothesis testing for regime-switching models, particularly when standard tests, including the likelihood ratio test, are invalid due to unidentified parameters, nonstandard asymptotic distributions, or restrictive assumptions.

The proposed tests are applicable in a wide range of settings, including those with non-stationary processes, non-Gaussian errors, multivariate models, and cases where both the null and the alternative hypothesis involves more than one regime ($M_0 > 1$ and $m > 1$). To our knowledge, this is the first paper to introduce hypothesis testing procedures for multivariate Markov switching models and provide such a general framework. The MMC-LRT is both asymptotically valid and exact in finite samples, and is robust to identification issues–an important feature in Markov switching settings. Simulation results confirm that both the LMC-LRT and MMC-LRT control test size effectively and display strong power, even in empirically challenging environments. When compared to existing alternatives such as the moment-based test of Dufour and Luger (2017) and the parameter stability test of Carrasco et al. (2014), our tests perform comparably or better in standard cases ($M_0 = m = 1$), and are the only valid option in more general scenarios.

Two empirical applications demonstrate the practical value of the proposed tests. First, we sequentially test for the number of regimes in U.S. real GNP and GDP growth series. While a two-regime model suffices for U.S. GNP data through 2010, a three-regime specification better fits extended samples ending in 2024:II. These findings corroborate earlier evidence of the Great Moderation and suggest that the low-volatility regime reemerges following both the Great Recession and the COVID-19 downturn. Although we account for potential structural breaks in the conditional mean, our results indicate that a three-regime model remains preferable. Future work may explore whether treating the Great Moderation and COVID periods as breaks in variance could yield a more parsimonious model.

In the second application, we test for business cycle synchronization using MS-VAR models. Our results show that business cycles were highly synchronized between the U.S. and other major economies pre-COVID, but that the inclusion of data through the pandemic weakens this synchronization–especially between the U.S. and Canada. These findings illustrate the flexibility of our framework for addressing questions of regime coherence and synchronization in multivariate macroeconomic systems.

Overall, the proposed methodology provides a robust and flexible framework for regime testing in both univariate and multivariate contexts, offering significant advantages in terms of generality, finite-sample validity, and ease of implementation. All procedures are implemented in the `MSTest` `R` package, described in a companion paper Rodriguez-Rondon and Dufour (2024). An interesting avenue for future work involves applying these methods to Markov switching factor models, where both the regime process and the factors are latent. This is the subject of ongoing research.

References

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

Ahn, H. J. and Luciani, M. (2024). Common and idiosyncratic inflation. *Finance and Economics Discussion Series 2020-024r1. Washington: Board of Governors of the Federal Reserve System.*

Amengual, D., Bei, X., Carrasco, M., and Sentana, E. (2025a). Score-type tests for Markov switching models. *Work in progress.*

Amengual, D., Bei, X., Carrasco, M., and Sentana, E. (2025b). Score-type tests for normal mixtures. *Journal of Econometrics*, 248:105717.

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

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

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

Anser, M. K., Godil, D. I., Khan, M. A., Nassani, A. A., Zaman, K., and Abro, M. M. Q. (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, B., Khalaf, L., Kichian, M., and Lin, Z. (2022). Identification-robust inference with simulation-based pseudo-matching. *Journal of Business & Economic Statistics*, pages 1–18.

Artis, M. (2004). The European business cycle. *Oxford Economic Papers*, 56(1):1–44.

Augustyniak, M. (2014). Maximum likelihood estimation of the markov-switching garch model. *Computational Statistics & Data Analysis*, 76:61–75.

Baldi, P., Chauvin, Y., Hunkapiller, T., and McClure, M. A. (1994). Hidden markov models of biological primary sequence information. *Proceedings of the National Academy of Sciences*, 91(3):1059–1063.

Bengoechea, P., Camacho, M., and Perez-Quiros, G. (2006). A useful tool for forecasting the euro-area business cycle phases. *International Journal of Forecasting*, 22(4):735–749.

Bunke, H. and Caelli, T. M. (2001). *Hidden Markov models: applications in computer vision*, volume 45. World Scientific.

Caggiano, G., Castelnuovo, E., and Figueres, J. M. (2017). Economic policy uncertainty and unemployment in the united states: A nonlinear approach. *Journal of Applied Econometrics*, 32(2):281–298.

Camacho, M. and Perez-Quiros, G. (2006). A new framework to analyze business cycle synchronization. *Contributions to Economic Analysis*, 276:133–149.

Camacho, M., Perez-Quiros, G., and Saiz, L. (2006). Are European business cycles close enough to be just one? *Journal of Economic Dynamics and Control*, 30(9-10):1687–1706.

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

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

Cevik, E. I., Yıldırım, D. Ç., and Dibooglu, S. (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, L. (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, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International economic review*, pages 969–996.

Chauvet, M. and Hamilton, J. D. (2006). Dating business cycle turning points. *Contributions to Economic Analysis*, 276:1–54.

Chauvet, M., Juhn, C., and Potter, S. (2002). Markov switching in disaggregate unemployment rates. *Empirical Economics*, 27(2):205–232.

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

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

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

Dées, S. and Zorell, N. (2012). Business cycle synchronisation: disentangling trade and financial linkages. *Open Economies Review*, 23(4):623–643.

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

Diebold, F. X. and Rudebusch, G. D. (1996). Measuring business cycles: A modern perspective. *The Review of Economics and Statistics*, 78.

Dietz, S. and Stern, N. (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, J.-M. (2006). Monte Carlo tests with nuisance parameters: A general approach to finite-sample inference and nonstandard asymptotics in econometrics. *Journal of Econometrics*, 133(2):443–477.

Dufour, J.-M. and Khalaf, L. (2001). Monte carlo test methods in econometrics. *Companion to Theoretical Econometrics', Blackwell Companions to Contemporary Economics, Basil Blackwell, Oxford, UK*, pages 494–519.

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

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

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

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

Gadea, M. D., Gómez-Loscos, A., and Pérez-Quirós, G. (2018). Great moderation and great recession: From plain sailing to stormy seas? *International Economic Review*, 59(4):2297–2321.

Gadea, M. D., Gómez-Loscos, A., and Pérez-Quirós, G. (2019). The decline in volatility in the us economy. a historical perspective. *Oxford Economic Papers*, 72(1):101–123.

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

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

Goldfeld, S. M. and Quandt, R. E. (1973). A markov model for switching regressions. *Journal of econometrics*, 1(1):3–15.

Golosov, M., Hassler, J., Krusell, P., and Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, 82(1):41–88.

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

Guha, D. and Banerji, A. (1999). Testing for regional cycles: a Markov-switching approach. *Journal of Economic and Social Measurement*, 25(3-4):163–182.

Haas, M., Mittnik, S., and Paolella, M. S. (2004). A new approach to markov-switching garch models. *Journal of financial econometrics*, 2(4):493–530.

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

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

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

Hamilton, J. D. (2010). Regime switching models. In *Macroeconometrics and time series analysis*, pages 202–209. Palgrave Macmillan, London.

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

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

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

Hernández, P. and Ochoa, C. (2016). Switching models for the analysis of health care costs: An application to the spanish health care system. *Health Economics*, 25(6):830–843.

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

Jelinek, F. (1997). *Statistical methods for speech recognition*. MIT press.

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

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

Klaassen, F. (2002). Improving garch volatility forecasts with regime-switching garch. In *Advances in Markov-switching models*, pages 223–254. Springer.

Krogh, A., Mian, I. S., and Haussler, D. (1994). A hidden markov model that finds genes in e. coli dna. *Nucleic acids research*, 22(22):4768–4778.

Krolzig, H.-M. (1997). *Markov-Switching Vector Autoregressions*. Springer.

Lanne, M., Lütkepohl, H., and Maciejowska, K. (2010). Structural vector autoregressions with markov switching. *Journal of Economic Dynamics and Control*, 34(2):121–131.

Le Bihan, H., Leiva-León, D., and Pacce, M. (2024). Underlying inflation and asymmetric risks. *Review of Economics and Statistics*, pages 1–45.

Leiva-Leon, D. (2014a). A new approach to infer changes in the synchronization of business cycle phases. Technical report, Bank of Canada Working Paper.

Leiva-Leon, D. (2014b). Real vs. nominal cycles: a multistate Markov-switching bi-factor approach. *Studies in Nonlinear Dynamics & Econometrics*, 18(5).

Leiva-Leon, D. (2017). Measuring business cycles intra-synchronization in us: A regime-switching interdependence framework. *Oxford Bulletin of Economics and Statistics*, 79(4):513–545.

Lütkepohl, H., Meitz, M., Netšunajev, A., and Saikkonen, P. (2021). Testing identification via heteroskedasticity in structural vector autoregressive models. *The Econometrics Journal*, 24(1):1–22.

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

Nag, R., Wong, K., and Fallside, F. (1986). Script recognition using hidden markov models. In *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 11, pages 2071–2074. IEEE.

Oka, T. and Perron, P. (2018). Testing for common breaks in a multiple equations system. *Journal of Econometrics*, 204(1):66–85.

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

Perron, P., Yamamoto, Y., and Zhou, J. (2020). Testing jointly for structural changes in the error variance and coefficients of a linear regression model. *Quantitative Economics*, 11(3):1019–1057.

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

Phillips, K. L. (1991b). A two-country model of stochastic output with changes in regime. *Journal of international economics*, 31(1-2):121–142.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852.

Qin, A. and Qu, Z. (2021). Modeling regime switching in high-dimensional data with applications to u.s. business cycles. *Working paper*.

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

Rabiner, L. and Juang, B. (1986). An introduction to hidden markov models. *ieee assp magazine*, 3(1):4–16.

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

Rodriguez-Rondon, G. (2024). Underlying core inflation with multiple regimes. *arXiv preprint arXiv:2411.12845*.

Rodriguez-Rondon, G. and Dufour, J.-M. (2024). Mstest: An r-package for testing markov switching models. *arXiv preprint arXiv:2411.08188*.

Rodriguez-Rondon, G. and Dufour, J.-M. (2024). *MSTest: Hypothesis Testing for Markov Switching Models*. R package version 0.1.3.

Sims, C. A. and Zha, T. (2006). Were there regime switches in us monetary policy? *American Economic Review*, 96(1):54–81.

Sweeting, A. (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.

7 Appendix

7.1 Additional results for U.S. GNP and GDP growth

Table 13: 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 14: 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

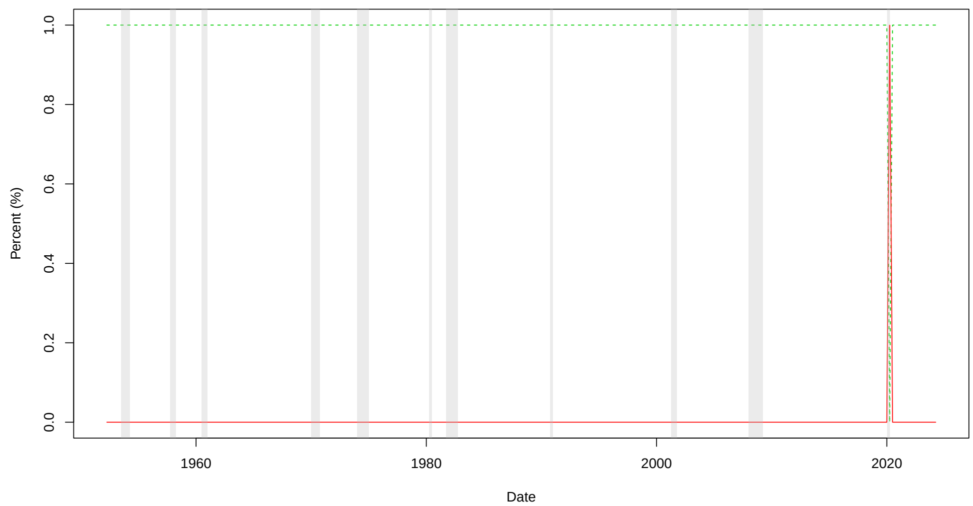
Notes: The GNP 1951:II-2024:II series ($T = 293$) is the GNP series from the St. Louis Fed (FRED) website. The models use $p = 4$ lags as in Hamilton (1989), Hansen (1992), Carrasco et al. (2014), and Dufour and Luger (2017).

Table 15: 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

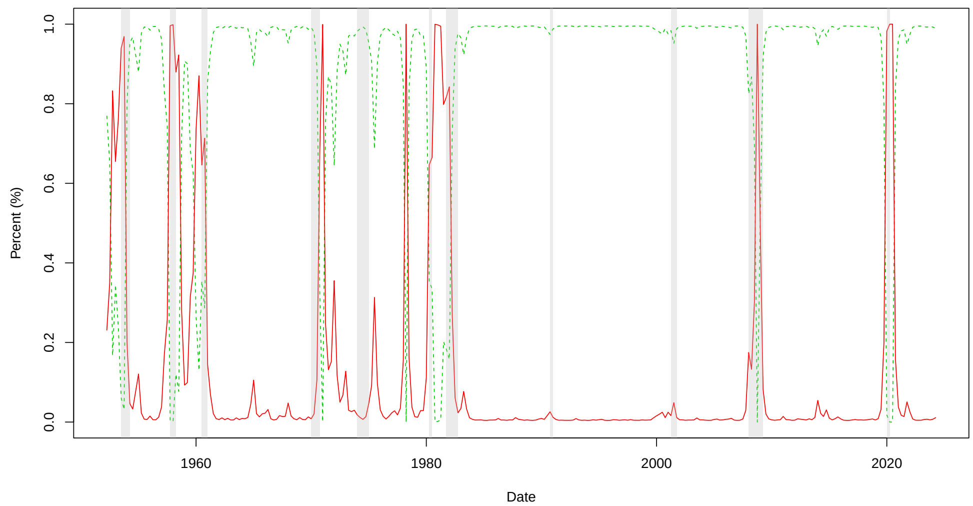
Notes: The GDP 1951:II-2024:II series ($T = 293$) is the GPC1 series from the St. Louis Fed (FRED) website. The models use $p = 1$ lags as in Qu and Zhuo (2021).

Figure 4: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ only and $M = 2$



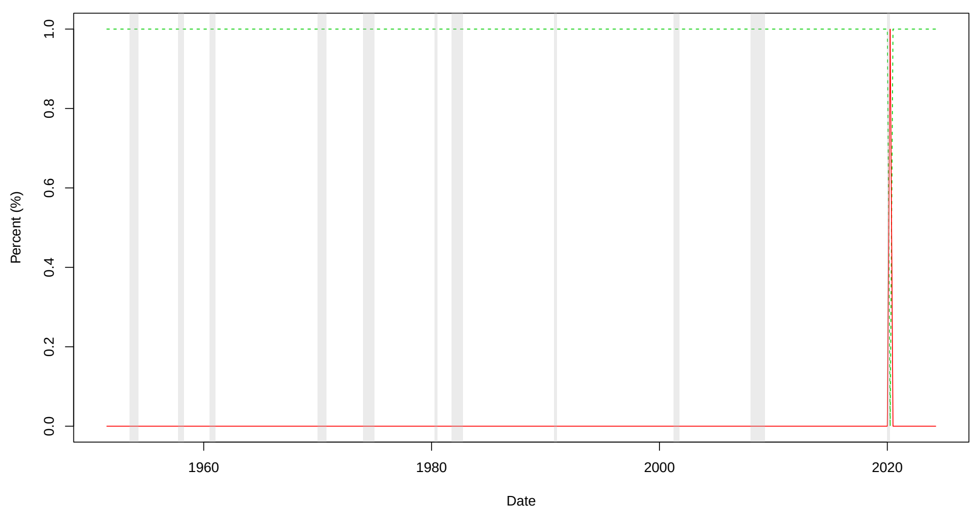
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure 5: Smoothed Probabilities of Regimes for US GNP when $\Delta\mu$ & $\Delta\sigma$ and $M = 2$



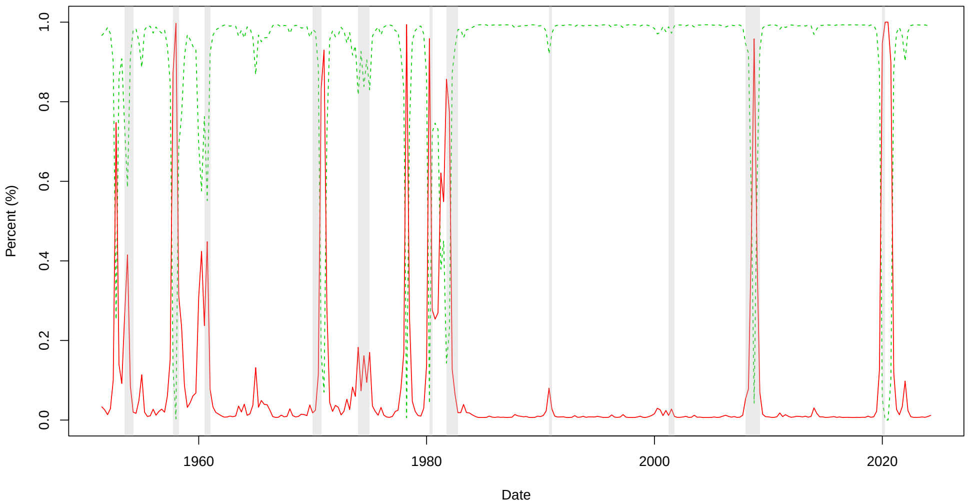
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

Figure 6: 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.

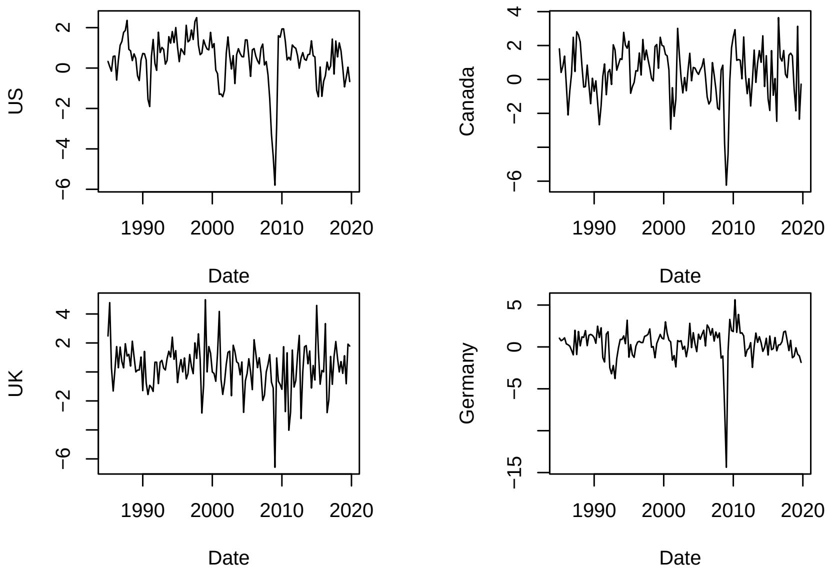
Figure 7: Smoothed Probabilities of Regimes for US GDP when $\Delta\mu$ & $\Delta\sigma$ and $M = 2$



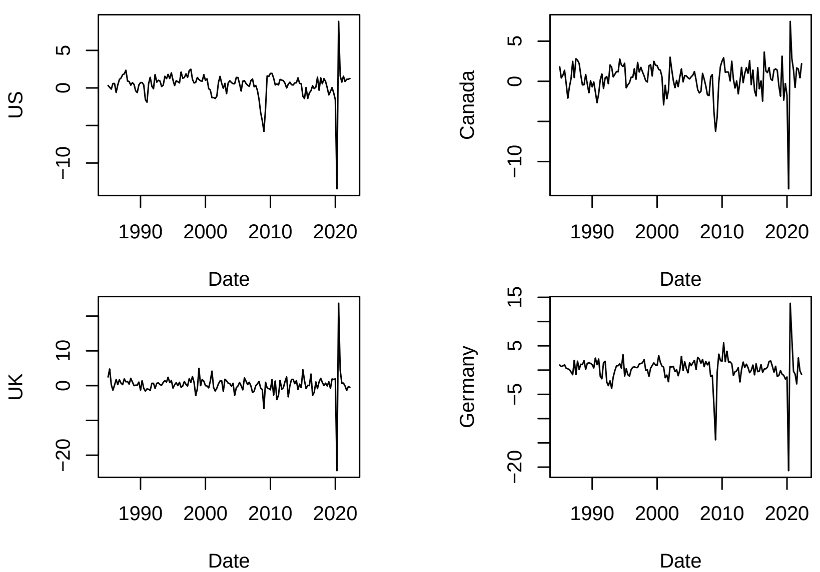
Notes: The sample is from 1951:III to 2024:II. The shaded areas correspond to the NBER recessions.

7.2 Additional results for synchronization of business cycles

Figure 8: Industrial production for four countries starting in 1985:I to 2019:IV (top) and 2022:II (bottom)

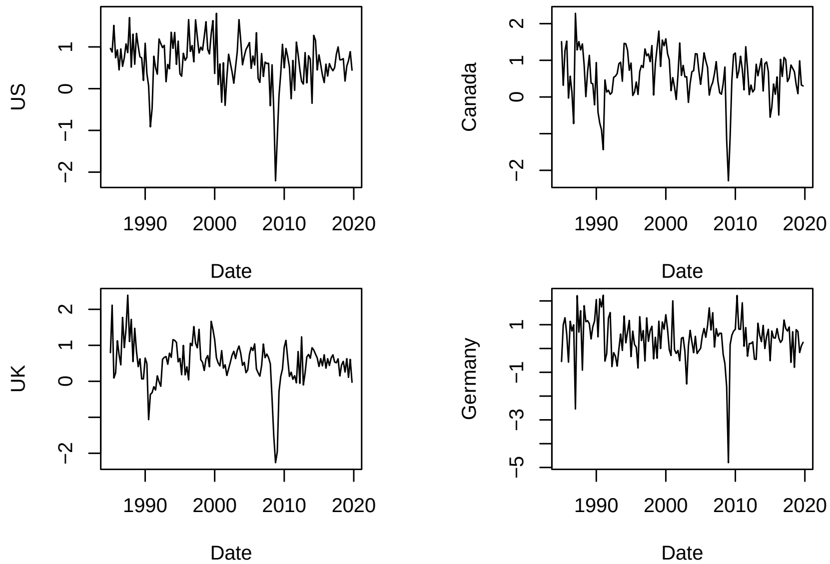


(a) Sample ending in 2019Q4

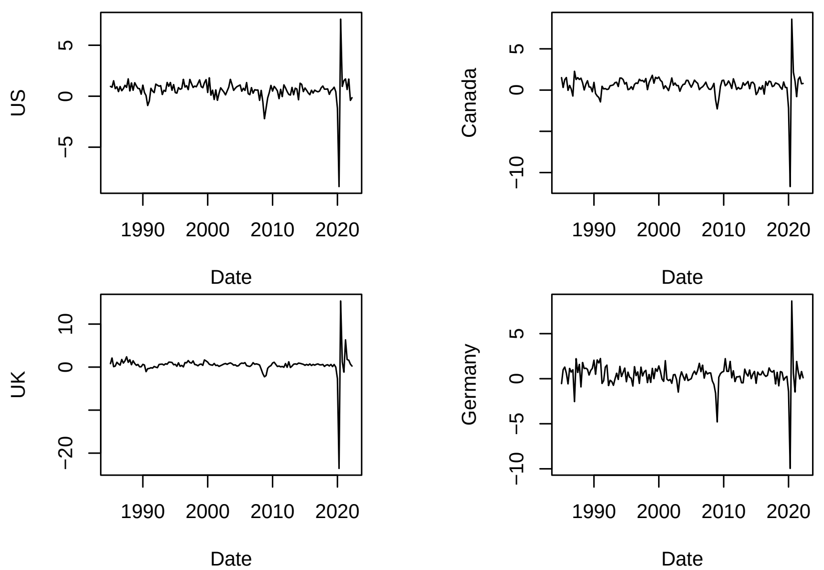


(b) Sample ending in 2022Q2

Figure 9: Real GDP for four countries starting in 1985:I to 2019:IV (top) and 2022:II (bottom)



(a) Sample ending in 2019Q4



(b) Sample ending in 2022Q2

Table 16: 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.