Monte Carlo Likelihood Ratio Tests for Markov Switching Models

Gabriel Rodriguez-Rondon[1] Jean-Marie Dufour[2]

[1]Bank of Canada

[2]McGill University

2025 World Congress of the Econometric Society

Augsust 18, 2025

Disclaimer: The views expressed in this presentation are those of the author and do not necessarily reflect the views of the Bank of Canada.

Objective

Use Monte Carlo test procedures to identify the number of regimes in Markov switching models using a likelihood ratio-based approach while dealing with issues related to:

- violation of regularity conditions
- non-stationary processes
- non-Gaussian errors
- multivariate setting

Overall, to propose a test that performs better and is more general than alternatives (applicable & valid in settings not previously available)

Motivation

- Markov Switching models (MSM) are now used in many macroeconomic and financial applications including
  - ⋆ Identification of Business Cycle to provide probabilistic statement about state of the economy; see Chauvet (1998), Diebold and Rudebusch (1996), Kim and Nelson (1999), Chauvet and Hamilton (2006) and Qin and Qu (2021)
  - Stock market volatility using Markov switching ARCH, GARCH & Stochastic Volatility models; see Hamilton (1994), Gray (1996), Klaassen (2002), Haas et al. (2004), Pelletier (2006) and So et al. (1998)
  - State-dependent IRFs; see Sims and Zha (2006) and Caggiano et al. (2017)
  - Identification of structural shocks in SVAR; see Lanne et al. (2010), Herwartz and Lütkepohl (2014), Lütkepohl et al. (2021)
  - ⋆ Measuring core inflation with multiple inflation regimes; Rodriguez-Rondon (2024)

Motivation

Issues with determining number of regimes M:

- Number of regimes must be specified *a priori*
- Conventional hypothesis testing procedures are not valid due to violation of regularity conditions
- Consistency of the information criterion (e.g., AIC & BIC) for selecting M has not been established in the literature
- Empirically, some authors use AIC and BIC for model comparison but these can lead to mixed results (e.g., Herwartz and Lütkepohl (2014) and Kasahara and Shimotsu (2018)).

MSM Hypothesis Testing Literature

Most approaches for determining number of regimes were limited to comparing linear models (one regime) to models with two regimes under the alternative:

- Moment based test: Dufour and Luger (2017)
- Parameter homogeneity vs. heterogeneity: Carrasco et al. (2014)
- Likelihood Ratio based tests: Hansen (1992), Garcia (1998), Cho and White (2007), and Qu and Zhuo (2021).

Kasahara and Shimotsu (2018) consider the more general case where $H_0 : M_0$ and $H_1 : M_0 + 1$ where $M_0 \geq 1$ in the context of LRT.

In all cases, only univariate settings are considered and with the exception of the Moment-based test of Dufour and Luger (2017), all procedures are only valid asymptotically.

Methodology

In this paper, we propose using the Maximized Monte Carlo (MMC) and Local Monte Carlo (LMC) test procedures to develop Monte Carlo Likelihood ratio tests that:

- Deal transparently with violations of regularity conditions
- Work with sample distribution of test statistic instead of asymptotic distribution allowing us to relax assumptions typically used in this literature
- deal with nuisance parameters
  - *MMC-LRT*: by searching nuisance parameter space
  - *LMC-LRT* : using consistent estimates (like parametric bootstrap)

Contributions

- *MMC-LRT* is
  - Identification robust
  - an Exact test (type I error cannot be larger than the nominal level)
  - valid in finite samples or asymptotically
- Both tests work in cases where validity of parametric bootstrap isn't available
- Improved power in many settings where alternative tests are available
- More general settings where $H_0 : M_0$ vs $H_1 : M_0 + m$ where $M_0, m \geq 1$
- Applicable to multivariate models (e.g. MS-VAR models)
- Relax assumptions used in the LRT stream of the literature
- Useful for testing common regime structure/breaks
- Tests are available in R package **MSTest** described in companion paper

Contributions

Table 1: Contribution & Literature

	$H_0 : M_0 = 1$ vs. $H_1 : M_0 = 2$	$H_0 : M_0$ vs. $H_1 : M_0 + 1$	$H_0 : M_0$ vs. $H_1 : M_0 + m$
Available tests	RD, DL, QZ, CHP, KS, CW, G, H	RD, KS	RD
Non-constrained param. space	RD, DL, CHP	RD	RD
Non-stationary	RD, DL	RD	RD
Non-Gaussian errors	RD, DL	RD, KS	RD
Multivariate	RD	RD	RD
MS-GARCH	RD, CHP	RD	RD
Valid in finite samples	RD, DL	RD	RD
Identification robust	RD, DL	RD	RD
Test common breaks	-	-	RD

Notes: In the above, RD refers to the tests proposed here, DL is Dufour and Luger (2017), QZ is Qu and Zhuo (2021), CHP is Carrasco et al. (2014), KS is Kasahara and Shimotsu (2018), CW is Cho and White (2007), G is Garcia (1998), and H is Hansen (1992)

Markov Switching Model

Markov switching model (MSM) allow some coefficients and variance to depend on a Markov process S_t (δ_{s_t} & $\sigma^2_{s_t}$) while others can remain constant (β)

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

The Markov process S_t takes values in $\{1, ..., M\}$ where M is the number of regimes.

When y_t depends on lags (e.g., $\{y_{t-1}, \ldots, y_{t-p}\}$), it is referred to as a Hidden MSM or simply MSM (see An et al. (2013)). Lags of y_t can be included in x_t or z_t depending on whether we want to allow the autoregressive coefficients to depend on the regimes or not. This setting also allows us to consider a trend function within x_t or z_t.

MSM Example: $M = 2$ Regimes

If the MSM has $M = 2$ regimes then $S_t = \{1, 2\}$ with

$$Pr(S_t = j) = \sum_{i=1}^{2} p_{ij} Pr(S_{t-1} = i) \qquad (2)$$

where $p_{ij} = Pr(S_t = j | S_{t-1} = i)$ are the one-step transition probabilities. These transition probabilities can be collected in a matrix:

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

and the ergodic probabilities can be obtained as

$$\pi_1 = \frac{p_{21}}{(p_{12} + p_{21})} \qquad \pi_2 = 1 - \pi_1$$

For example, π_1 tells us, on average in the long-run, the proportion of time spent in regime 1.

Hypothesis Test

Consider

$$y_t = \mu_{s_t} + \sum_{i=1}^{p} \phi_i (y_{t-i} - \mu_{s_{t-i}}) + \sigma \epsilon_t$$

such that the mean is a function of the latent Markov process S_t and $\epsilon_t \sim \mathcal{N}(0,1)$. Here, $\delta_{s_t} = \mu_{s_t}$, $\beta = (\phi_1, \ldots, \phi_p)'$, and $\theta_1 = (\delta_{s_t}, \beta, \sigma^2, vec(P))'$.

Now, suppose we are interested in

$$\begin{aligned} H_0 &: \ \delta_1 = \delta_2 = \delta^* && \text{for some unknown } \delta^* = \mu^* \\ H_1 &: \ (\delta_1, \delta_2) = (\delta_1^*, \delta_2^*) && \text{for some unknown } \delta_1^* \neq \delta_2^* \end{aligned}$$

That is $H_0 : M_0 = 1$ (linear model) vs. $H_1 : M_0 + m = 2$.

Hypothesis Test

The log likelihood function is given by

$$L_T(\theta_i) = \sum_{t=1}^{T} log\ f(y_t|y_{-p+1}^{t-1};\theta_i) \tag{3}$$

where $i = \{0, 1\}$. Let

$$\bar{L}_T(H_1) = \sup\{L_T(\theta_1) : \theta_1 \in \Omega\}, \tag{4}$$

$$\bar{L}_T(H_0) = \sup\{L_T^0(\theta_0) : \theta_0 \in \bar{\Omega}_0\} = \sup\{L_T(\theta_1) : \theta_1 \in \Omega_0\}. \tag{5}$$

so that our test statistic is

$$LR_T = 2[\bar{L}_T(H_1) - \bar{L}_T(H_0)] \tag{6}$$

and the null distribution of (6) depends only on $\theta_0 \in \bar{\Omega}_0$. Note that we could easily include ϕ_{i,s_t}, σ_{s_t}, or other model parameters in δ_{s_t} and simply change θ_i accordingly.

Methodology

Violation of regularity conditions:

- Parameter values may be at the boundary; see Andrews (1999) & Andrews (2001)
- Score function is equal to 0 when evaluated at restricted MLE
- Unidentified nuisance parameters under null; see Davies (1977) & Andrews and Ploberger (1994))

In this paper:

- Work with sample distribution of test statistic
  - do not need regularity conditions to derive asymptotic distribution of the test statistic
  - replace "theoretical" null distribution $F(x)$ of LR_T with its simulation-based "estimate" $\hat{F}(x)$
  - allows us to deal with more general cases directly (i.e., non-stationary, non-Gaussian, parameter boundary, multivariate, among others)
- Use MMC and LMC procedures to deal with presence of nuisance parameters
- Valid even when asymptotic distribution does not exist

Maximized Monte Carlo Likelihood Ratio Test

Monte Carlo p-value is given by

$$\hat{p}_N[LR_T^{(0)}|\theta_0] = \frac{N + 1 - R_{LR}[LR_T^{(0)}; N]}{N + 1} \tag{7}$$

where $R_{LR}[LR_T^{(0)}; N] = \sum_{i=1}^{N} \mathbb{1}\{LR_T^{(0)} \geq LR_T^i\}$.

In proposition 3.1 we extend Dufour (2006) to show that, under the null hypothesis, the LRT statistic for Markov switching models has

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

That is, we have a valid test procedure.

MMC-LRT - Consistent Set C_T

Can also search a smaller consistent set of the parameter space C_T. For example, let $\hat{\theta}$ be the consistent point estimate of θ_0. Then, we can define

$$C_T = \{\theta_0 \in \bar{\Omega}_0 : \left\|\hat{\theta} - \theta_0\right\| < c\} \tag{8}$$

where c is a fixed constant (doesn't depend on T) and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^k$

To search over the parameter space $\bar{\Omega}_0$ or C_T, we can use:

- Generalized Simulated Annealing
- Genetic Algorithms
- Particle Swarm

Local Monte Carlo Likelihood Ratio Test

Alternatively, 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*).

- Like parametric bootstrap
  - $\hat{\theta}_0$ is an asymptotically efficient estimator of θ_0 (want $T \to \infty$)
- Unlike parametric bootstrap
  - Do not need $N \to \infty$ ($N = 19$ sufficient for $\alpha = 0.05$)
  - Unnecessary as we do not try to approximate asymptotic distribution of test statistic
  - Valid even if asymptotic distribution does not exist

Tests For Comparison

- Use R-package **MSTest** described in Rodriguez-Rondon and Dufour (2024)
- For $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$, we consider
    - Moment based test: LMC_{min}, LMC_{prod}, MMC_{min} & MMC_{prod} of Dufour and Luger (2017) for $H_0 : M_0 = 1$ vs. $H_0 : M_0 = 2$ only
    - Parameter homogeneity vs. heterogeneity: supTS & expTS of Carrasco et al. (2014) for $H_0 : M_0 = 1$ vs. $H_0 : M_0 = 2$ only
- When $M_0 \geq 1$ and $m \geq 1$ we only consider tests proposed here
- Simulation results are obtained using 1000 replications of the DGP.

Empirical size of test for H_0: $M = 1$

Table 2: Empirical size of test when $H_0 : M_0 = 1$

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

Notes: Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Empirical Power of test for H_0: $M = 1$ vs. H_1: $M = 2$

Table 3: Empirical Power of Test when $M_0 = 1$, $m = 1$, & $(p_{11}, p_{22}) = (0.9, 0.9)$

Test	$\phi = 0.10$			$\phi = 0.90$		
	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
MMC-LRT	58.0	81.7	90.0	7.5	14.7	31.3
LMC_{min}	5.3	5.4	3.7	14.5	20.9	42.1
LMC_{prod}	4.8	4.3	4.3	16.2	22.3	43.0
MMC_{min}	1.1	2.3	1.9	6.7	13.2	33.8
MMC_{prod}	0.9	2.4	2.0	6.9	14.5	34.2
supTS	36.4	64.0	96.5	5.5	3.9	6.1
expTS	35.6	60.9	95.4	5.4	3.9	6.4
			$\Delta\mu$ & $\Delta\sigma$			
LMC-LRT	81.2	98.7	100.0	39.5	78.0	98.7
MMC-LRT	78.0	94.5	100.0	25.6	66.0	96.0
LMC_{min}	53.1	80.9	99.4	35.3	60.7	92.6
LMC_{prod}	46.1	74.1	98.7	38.7	63.9	95.3
MMC_{min}	37.2	69.6	99.0	22.9	49.3	89.4
MMC_{prod}	34.2	66.0	98.1	26.3	55.5	92.7
supTS	74.0	96.0	100.0	34.0	62.9	95.4
expTS	73.3	92.0	100.0	45.6	76.0	97.0

Notes: Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Empirical Power of test for H_0: $M = 1$ vs. H_1: $M = 3$

Table 4: Empirical Power of Test when $M_0 = 1$, $m = 2$, & $(p_{11}, p_{22}, p_{33}) = (0.9, 0.9, 0.9)$

Test	$\phi = 0.10$			$\phi = 0.90$		
	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	99.6
MMC-LRT	80.0	93.0	95.3	51.4	77.3	92.1
			$\Delta\mu$ & $\Delta\sigma$			
LMC-LRT	85.5	99.9	100.0	77.1	95.9	100.0
MMC-LRT	79.4	90.1	98.1	60.6	92.0	94.3

Notes: Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Empirical Performance of test when process is non-stationary

Table 5: Empirical Performance of test when process is non-stationary

Test	$\phi = 1.00$		
	T=100	T=200	T=500
	Empirical size		
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: $\Delta\mu$		
LMC-LRT	15.5	22.8	39.9
MMC-LRT	9.2	14.1	25.2
	Empirical power: $\Delta\mu$ & $\Delta\sigma$		
LMC-LRT	29.7	54.4	77.3
MMC-LRT	21.7	43.1	63.8

Notes: Here $H_0 : M_0 = 1$ vs. $H_1 : M_0 + m = 2$. For alternative model, $p_{22} = 0.90$. Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Empirical Power of test when parameter is at boundary

Table 6: Empirical Power of Test when $M_0 = 1$, $m = 1$, & $(p_{11}, p_{22}) = (0.9, 1.0)$

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
			$\Delta\mu$ & $\Delta\sigma$			
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: Rejection frequencies are obtained using 1000 replications. MC tests use $N = 99$ simulations.

Empirical Size of Test: H_0: $M = 2$ vs. H_1: $M = 3$ Regimes

Table 7: Empirical Size of Test when $M_0 = 2$ & $m = 1$

Test	$(p_{11}, p_{22}) = (0.5, 0.5)$			$(p_{11}, p_{22}) = (0.7, 0.7)$		
	T=100	T=200	T=500	T=100	T=200	T=500
$(\phi, \mu_1, \mu_2, \sigma) = (0.5, -1, 1, 1)$						
LMC-LRT	6.80	6.30	4.60	6.00	6.00	4.80
MMC-LRT	3.80	3.70	3.30	3.10	3.60	2.70
Boot-LRT	-	7.16	4.43	-	6.07	4.20

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

U.S. GNP & GDP Growth Series

- **U.S. GNP growth series**
  - from 1951:II to 1984:IV
    - Hansen (1992), CHP, and DL consider Hamilton's original sample
    - All studies fail to reject null hypothesis of linear model (even when allowing $\Delta\sigma$)
  - from 1951:II to 2010:IV
    - CHP and DL consider this extended sample and reject null hypothesis of linear model
    - Here, we confirm $M = 2$ regime model (when $\Delta\sigma$) by considering $H_1 : M = 3$
    - Unlike CHP, we find $M = 2$ even when only $\Delta\mu$
  - from 1951:II to 2024:II
    - Here, we also consider this more recent sample which includes COVID period
    - In all U.S. GNP models, model (**??**) is used
    - Find evidence for a model with $M = 3$ regimes
- **U.S. GDP growth series from 1951:II to 2024:II**
  - Following Qu and Zhuo (2021) and more recent literature, we focus on GDP data
  - Consider controlling for known structural breaks such as great moderation and COVID period

U.S. GNP & GDP Growth

Table 8: Results For U.S. GNP & GDP 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
GDP 1951:II-2024:II	0.01	0.01	0.01	0.01	0.47	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
GDP 1951:II-2024:II	0.01	0.01	0.01	0.01	0.68	1.00

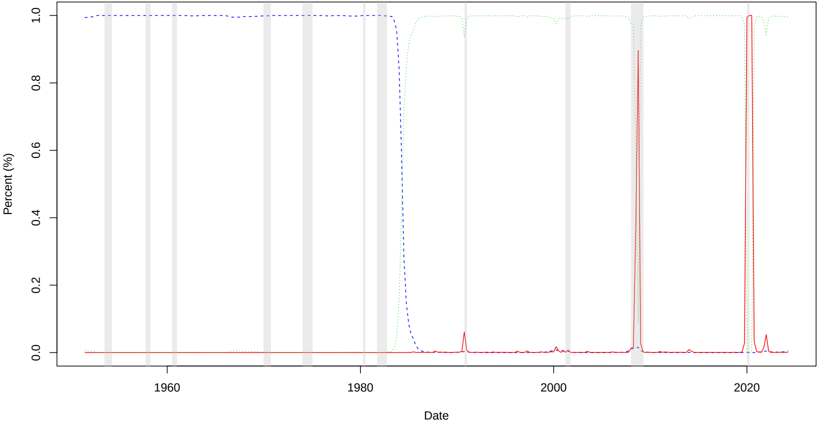
Notes: The GNP 1951:II-1984:IV series ($T = 135$) is the same as the one used in Hamilton (1989), Hansen (1992), and Carrasco et al. (2014). The GNP 1951:II-2010:IV series ($T = 239$) is the same as the one used in Carrasco et al. (2014) and Dufour and Luger (2017). The GNP 1951:II-2024:II and GDP 1951:II-2024:II series ($T = 293$) are the GNP and GPC1 series respectively 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).

$\leftarrow$

U.S. GDP Growth

- Regime 1 (blue): expansionary state ($\mu_1 = 0.79$) with high vol. ($\sigma_1 = 1.06$)
- Regime 2 (green): expansionary state ($\mu_2 = 0.72$) with low vol. ($\sigma_2 = 0.45$)
- Regime 3 (red): recessionary states ($\mu_3 = -0.50$) with very high vol. ($\sigma_2 = 6.5$)

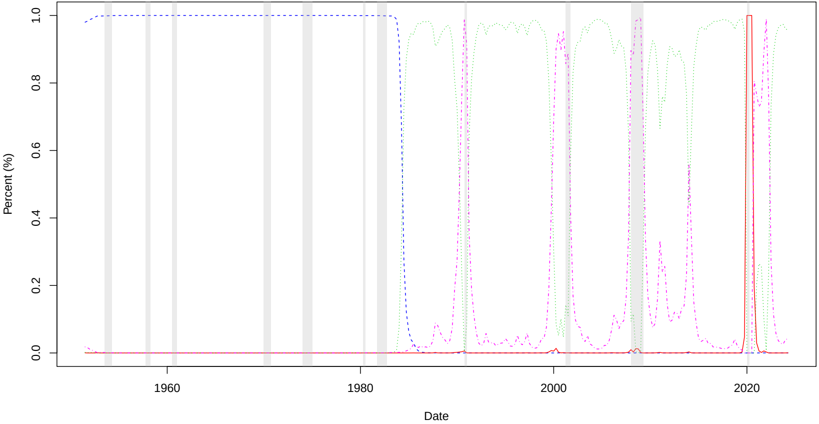
Figure 1: Smoothed Probabilities of Regimes from 1951:III - 2024:II ($M = 3$)



U.S. GDP Growth

- Here, we have two recessionary regimes

Figure 2: Smoothed Probabilities of Regimes from 1951:III - 2024:II ($M = 4$)



Testing the Synchronization of Business Cycles

Consider the following models

$$\Delta \text{GDP}_{a,t} = \mu_{a,s_{a,t}} + x_t\beta_a + \sigma_a\epsilon_{a,t}$$

and

$$\Delta \text{GDP}_{b,t} = \mu_{b,s_{b,t}} + x_t\beta_b + \sigma_b\epsilon_{b,t}$$

and suppose, we are interested in knowing if the Markov processes $S_{a,t}$ and $S_{b,t}$ are perfectly dependent (synchronized) such that $S_{a,t} = S_{b,t} = S_t$.

Testing the Synchronization of Business Cycles

If $S_{a,t} = \{1, 2\}$ and $S_{b,t} = \{1, 2\}$, then up to four cases are possible in a joint MSVAR model:

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

From here, we can see that

- if perfectly synchronized, then $S_{a,t} = S_{b,t} = S^*_t$ and $S^*_t = \{1, 2\}$ such that $M^* = 2$
- if leading (lagging), then then $S^*_t = \{1, 2, 3\}$ such that $M^* = 3$
- if perfectly independent, then then $S^*_t = \{1, 2, 3, 4\}$ such that $M^* = 4$

Hence, testing the synchronization of BCs (or common regime structure/breaks) can be tested in bi-variate MSVAR by considering

$$H_0 : M^* = 2 \text{ vs. } H_1 : M^* = 3 \qquad \text{or} \qquad H_0 : M^* = 2 \text{ vs. } H_1 : M^* = 4$$

Synchronization of Business Cycles: GDP series

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

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

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

Conclusion

Propose Monte Carlo Likelihood Ratio-based tests (i.e., *LMC-LRT* and *MMC-LRT*) to determine appropriate number of regimes for MSM

- *MMC-LRT* is identification robust and valid even in finite samples
- tests are applicable when dealing with (1) non-stationary process, (2) non-Gaussian errors, (3) at boundary of parameter space, and (4) multivariate settings
- Can be used to compare more MSM models (i.e., when interested in $H_0 : M_0$ vs. $H_1 : M_0 + m$ where $M_0, m \geq 1$)
- Simulation results suggest tests are able to control level of test and have favourable power performance
- Used in simple applications to determine number of regimes in US GDP growth and test the synchronization of international business cycles

Thank you!

References I

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

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

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

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

Andrews, D. W. 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.

References II

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.

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.

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.

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

References III

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.

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

References IV

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

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. *Journal of Econometrics*, 133(2):443–477.

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.

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

References V

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.

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. (1994). *Time series analysis*. Princeton university press.

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

References VI

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.

References VII

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.

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

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.

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.

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

References VIII

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. *Discussion paper, McGill University (Economics Department)*.

Rodriguez-Rondon, G. and Dufour, J.-M. (2024). MSTest: An R-package for testing Markov-switching models. *Discussion paper, McGill University (Economics Department)*.

References IX

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

So, M. E. P., Lam, K., and Li, W. K. (1998). A stochastic volatility model with markov switching. *Journal of Business & Economic Statistics*, 16(2):244–253.

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.

Motivation

- Other applications of MSM in economics:
  - Climate change; see Golosov et al. (2014) and Dietz and Stern (2015)
  - Environmental & energy economics; see Cevik et al. (2021) and Charfeddine (2017)
  - Industrial Organization; see Aguirregabiria and Mira (2007) and Sweeting (2013)
  - Health economics; see Hernández and Ochoa (2016) and Anser et al. (2021)

- Alternative but related Hidden Markov Model (HMM) have applications in:
  - Computational molecular biology (Krogh et al. (1994) and Baldi et al. (1994))
  - Handwriting and speech recognition (Rabiner and Juang (1986), Nag et al. (1986), Rabiner and Juang (1993) and Jelinek (1997))
  - Computer vision and pattern recognition (see Bunke and Caelli (2001)) and other machine learning applications

MSTest R package

- Has 6500+ downloads
- Package includes:
  - LMC-LRT and MMC-LRT procedures proposed here and
  - Moment-based test of Dufour and Luger (2017)
  - Parameter stability test of Carrasco et al. (2014)
  - Conservative LRT of Hansen (1992)
  - Model estimation (EM & MLE)
  - Simulation functions
- See page on CRAN and GitHub for more details
- Package is also described in Rodriguez-Rondon and Dufour (2024)

←

MC Test Procedure

MMC LRT

1. Use $\hat{\theta}_1$ from observed data but consider different $\theta_0 \in C_T$ (or $\bar{\Omega}_0$), to obtain $LR_T^{(0)}$
2. Repeat steps 2 - 5 using different $\theta_0 \in C_T$ to obtain $\sup \left\{ \hat{p}_N[LR_T^{(0)} | \theta_0] : \theta_0 \in \bar{\Omega}_0 \right\}$
   $\rightarrow$ using numerical optimization to search over C_T

MSM Hypothesis Testing Literature - Bootstrap

Asymptotic validity of parametric bootstrap procedure:

- Qu and Zhuo (2021) for $H_0 : M_0 = 1$ vs. $H_1 : M_0 = 2$ and a board set of models (i.e., doesn't include models with weakly exogenous regressors)
- Kasahara and Shimotsu (2018) for $H_0 : M_0$ vs. $H_1 : M_0 + 1$ but for limited set of models (i.e., fixed regressors only).

In general, these results require:

- stationarity
- constrained parameter spaces
- Gaussian errors
- univariate settings

and are only valid asymptotically.

Log-likelihood Function of Linear Model

$$L_T(\theta_0) = logf(y_1^T|y^0_{-p+1};\theta_0) = \sum_{t=1}^{T} logf(y_t|y^{t-1}_{-p+1};\theta_0) \quad (9)$$

where $\theta_0 = \{\mu_1, \sigma_1^2, \phi_1, \dots \phi_p\}$ and

$$f(y_t|y^{t-1}_{-p+1};\theta_0) = \frac{1}{\sqrt{2\pi\sigma_1^2}} exp\left\{\frac{-[y_t - \mu_1 - \sum_{k=1}^{p}\phi_k(y_{t-k} - \mu)]^2}{2\sigma_1^2}\right\} \quad (10)$$

Log-likelihood Function of MSM

$$L_T(\theta_1) = logf(y_1^T|y_{-p+1}^0;\theta_1) = \sum_{t=1}^{T} logf(y_t|y_{-p+1}^{t-1};\theta_1) \tag{11}$$

where $\theta_1 = (\mu_1, \mu_2, \sigma_1, \sigma_2, \phi_1, \ldots, \phi_p, p_{11}, p_{22})$ and

$$f(y_t|y_{-p+1}^{t-1};\theta_1) = \sum_{s_t=1}^{2}\sum_{s_{t-1}=1}^{2}\cdots\sum_{s_{t-p}=1}^{2} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots, S_{t-p} = s_{t-p}|y_{-p+1}^{t-1};\theta_1) \tag{12}$$

$$\begin{aligned} f(y_t, S_t = s_t, \ldots, S_{t-p} = s_{t-p}|y_{-p+1}^{t-1};\theta_1) = {} & \frac{P(S_t^* = s_t^*|y_{-p+1}^{t-1};\theta_1)}{\sqrt{2\pi\sigma_{s_t}^2}} \\ & \times exp\left\{\frac{-[y_t - \mu_{s_t} - \sum_{k=1}^{p}\phi_k(y_{t-k} - \mu_{s_{t-k}})]^2}{2\sigma_{s_t}^2}\right\} \end{aligned} \tag{13}$$

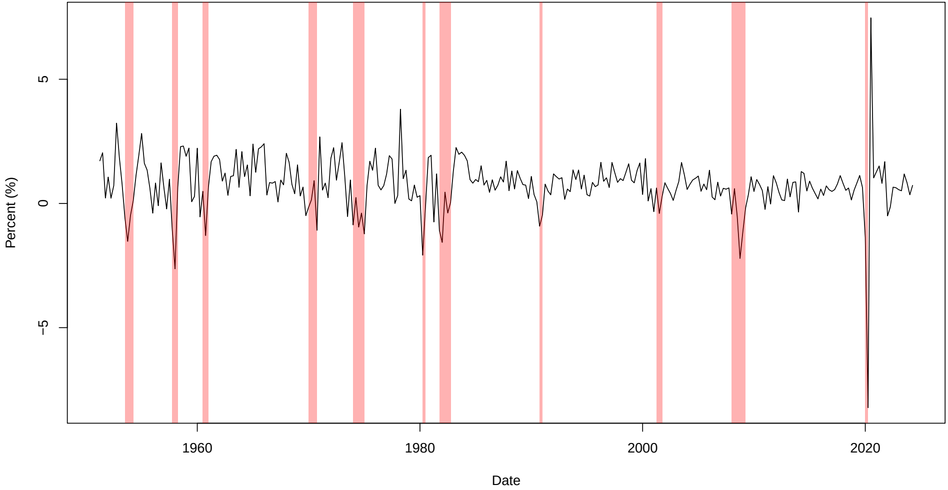
where we let

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

and $P(S_t^* = s_t^*|y_{-p+1}^{t-1};\theta_1)$ is the probability that this occurs.

U.S. GDP Growth

Figure 3: U.S. real GDP growth from 1951:II - 2024:II



U.S. GDP Growth - Parameters Estimates

Table 10: Estimates of Preferred Model ($\Delta\mu$ & $\Delta\sigma$)

	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). Here, we allow for changes in mean and variance (i.e., $\Delta\mu$ & $\Delta\sigma$).

U.S. GDP Growth

Table 11: Results For U.S. GDP Growth Series Hypothesis Tests With Known Breaks

Series	$H_0: M = 1$ vs. $H_1: M = 2$ LMC-LRT	$H_0: M = 1$ vs. $H_1: M = 2$ MMC-LRT	$H_0: M = 2$ vs. $H_1: M = 3$ LMC-LRT	$H_0: M = 2$ vs. $H_1: M = 3$ MMC-LRT	$H_0: M = 3$ vs. $H_1: M = 4$ LMC-LRT	$H_0: M = 3$ vs. $H_1: M = 4$ 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. 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).

U.S. GDP Growth - Comparison with Dummy Variables

Table 12: Comparison of models with dummy variables

	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	-337.016	708.033	770.538

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. 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).

Synchronization of Business Cycles: IP series

Table 13: 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.