Abstract

Markov regime switching models are very common in economics and finance. Despite persisting interest in them, the asymptotic distributions of likelihood ratio based tests for detecting regime switching remain unknown. This study examines such tests and establishes their asymptotic distributions in the context of nonlinear models, allowing multiple parameters to be affected by regime switching. The analysis addresses three difficulties: (i) some nuisance parameters are unidentified under the null hypothesis; (ii) the null hypothesis yields a local optimum; and (iii) the conditional regime probabilities follow stochastic processes that can only be represented recursively. Addressing these issues permits substantial power gains in empirically relevant settings. This study also presents the following results: (1) a characterization of the conditional regime probabilities and their derivatives with respect to the model’s parameters; (2) a high order approximation to the log likelihood ratio; (3) a refinement of the asymptotic distribution; and (4) a unified algorithm to simulate the critical values. For models that are linear under the null hypothesis, the elements needed for the algorithm can all be computed analytically. Furthermore, the above results explain why some bootstrap procedures can be inconsistent, and why standard information criteria can be sensitive to the hypothesis and the model structure. When applied to US quarterly real GDP growth rate data, the methods detect relatively strong evidence favoring the regime switching specification. Lastly, we apply the methods in the context of dynamic stochastic equilibrium models, and obtain similar results as the GDP case.

Keywords: Hypothesis testing, likelihood ratio, Markov switching, nonlinearity.

JEL codes: C12, C22, E32.
1 Introduction

Markov regime switching models are widely used in economics and finance. Hamilton (1989) is a seminal contribution, which provides not only a framework for describing economic recessions, but also a general algorithm for filtering, smoothing, and estimation. Surveys of this voluminous body of literature are available in Hamilton (2008, 2016) and Ang and Timmermann (2012).

Three approaches have been considered for detecting regime switching. The first approach involves treating the issue as testing for parameter homogeneity against heterogeneity. To detect heterogeneity, Neyman and Scott (1966) studied the $C(\alpha)$ test, while Chesher (1984) derived a score test and showed that it is related to the information matrix test of White (1982). Carrasco, Hu, and Ploberger (2014) further developed this approach by allowing the parameters to follow flexible weakly dependent processes. They analyzed a class of tests, showing that they are asymptotically locally optimal against a specific alternative characterized in their study. The above tests have two features. First, they require estimation under the null hypothesis only. Second, they are designed to detect general parameter heterogeneity, not Markov regime switching specifically. Their power can be substantially lower than that achievable if the parameters follow a finite state Markov chain.

The second approach, proposed by Hamilton (1996), involves conducting generic tests for the hypothesis that a $K$-regime model (e.g., $K = 1$) adequately describes the data. If the hypothesis holds, then the score function should have mean zero, and follow a martingale difference sequence. Otherwise, the model should be enriched to allow for additional features, possibly by introducing an additional regime. Hamilton (1996) demonstrated the implementation of such tests as a by-product of calculating the smoothed regime probability, making them widely applicable. However, it remains desirable to have testing procedures designed to detect Markov switching alternatives.

The third approach proceeds under the (quasi) likelihood ratio principle, where the (quasi) likelihood functions are constructed assuming a single regime under the null hypothesis, and two regimes under the alternative hypothesis. The analysis faces three challenges. (i) The transition probabilities ($p$ and $q$) are unidentified under the null hypothesis. This is known as the Davies’ (1977) problem. In other words, some parameters are identified only under the alternative hypothesis, and consequently, they are not consistently estimable under the null hypothesis. (ii) The null hypothesis yields a local optimum (cf., Hamilton, 1990); that is, the score function is equal to zero when evaluated at the restricted MLE. Because the resulting unrestricted MLE converges slower than $T^{-1/2}$, the second order Taylor approximation is insufficient for studying the asymptotic properties of the likelihood ratio. (iii) The conditional regime probability (the probability of occupying
a particular regime at time $t$, given the observed information up to $t - 1$) follows a stochastic process that can only be represented recursively. The first two difficulties are also present when testing for mixtures. The simultaneous occurrence of all three difficulties challenges the study of the likelihood ratio in the current context. For example, when studying a high order expansion of the likelihood ratio, it is necessary to consider high order derivatives of the conditional regime probability with respect to the model’s parameters. Thus far, their statistical properties have remained elusive. Consequently, the asymptotic distribution of the log likelihood ratio remains unknown.

Important progress is documented in Hansen (1992), Garcia (1998), Cho and White (2007), and Carter and Steigerwald (2012). Hansen (1992) explained why difficulties (i) and (ii) cause the conventional approximation to the likelihood ratio to break down. Furthermore, he treated the likelihood function as a stochastic process, and obtained a bound for its asymptotic distribution. His result provides a platform for conducting conservative inference. Garcia (1998) suggested an approximation to the log likelihood ratio that would follow if the score had a positive variance at the restricted MLE. Results in this current study will show that this distribution is, in general, different from the actual limiting distribution. Recently, Cho and White (2007) made significant progress. They suggested a quasi likelihood ratio (QLR) test against a two-component mixture alternative, a model in which the regime arrives independently of its past values. The difficulty (iii) is avoided because the conditional regime probability is reduced to a constant, which can be treated as an unknown parameter. Later, Carter and Steigerwald (2012) discussed a consistency issue related to the QLR test. Our study uses several important techniques in Cho and White (2007). Simultaneously, it goes beyond their framework to directly confront Markov switching alternatives. As we will show, the power gains from doing so can be quite substantial.

Specifically, this study considers a family of likelihood ratio based tests and establishes their asymptotic distributions in the context of nonlinear models, allowing multiple parameters to be affected by regime switching. The analysis under the null hypothesis, presented in Sections 4 and 5, takes five steps. Step 1 provides a characterization of the conditional regime probability and its derivatives with respect to the model’s parameters. When evaluated under the null hypothesis, this probability reduces to a constant, and its derivatives can all be represented as linear first order difference equations with lagged coefficients $p + q - 1$. Because $0 < p, q < 1$, the equations are stable and amenable to the application of uniform laws of large numbers and functional central limit theorems. This novel characterization is a critical step that makes the subsequent analysis feasible. Step 2 examines a fourth order Taylor approximation of the likelihood ratio for fixed $(p, q)$. This step builds on Cho and White (2007), but goes beyond to account for the effect on the likelihood
ratio of the time variation in the conditional regime probability. Step 3 derives an approximation of the likelihood ratio, as an empirical process over \((p, q): \epsilon \leq p, q \leq 1 - \epsilon\) and \(p + q \geq 1 + \epsilon\), where \(\epsilon\) is a small positive constant. The empirical process perspective follows several existing studies, including Hansen (1992), Garcia (1998), Cho and White (2007), and Carrasco, Hu, and Ploberger (2014). Step 4 provides a finite sample refinement, motivated by the observation that while the limiting distribution in Step 3 is adequate for a broad class of models, it can lead to over-rejections when a singularity (specified later) is present. This problem is addressed by examining a sixth order expansion of the likelihood ratio along the line \(p + q = 1\), and an eighth order expansion at \(p = q = 0.5\). The leading terms are then incorporated into the asymptotic distribution to safeguard against their effects. The resulting refined distribution delivers reliable approximations in all of our experiments. Finally, Step 5 presents an algorithm that simulates the refined approximation. For linear models, the elements of this algorithm can all be computed analytically.

The null asymptotic distribution has several unusual features. First, the nuisance parameters, though constrained not to switch, can affect the limiting distribution. Second, the properties of the regressors (e.g., whether they are strictly or weakly exogenous) also affect the distribution. Third, this distribution depends on which parameter is allowed to switch. These features imply that some bootstrap procedures may be inconsistent and that the standard information criteria, such as the BIC, can be sensitive to the hypothesis and the model structure; see Section 7 for details.

Next, we study the likelihood ratio under the alternative hypothesis. The results explain the potential local power difference between the likelihood ratio test and the tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) in an empirically important setting.

We conduct simulations using a data generating process (DGP) considered in Cho and White (2007). The results show that the power difference can indeed be quite large when the regimes are persistent, a situation that is common in practice. We also apply the testing procedure to US quarterly real GDP growth rate data for the period 1960:I-2014:IV and to a range of subsamples. The results consistently favor the regime switching specification. In addition, we apply the methods in the context of dynamic stochastic equilibrium models, and obtain results similar to those of the GDP case. To the best of our knowledge, this is the first time such consistent evidence for regime switching in the US business cycle dynamics has been documented using hypothesis testing.

Empirical studies have estimated regime switching models on a wide range of time series, including exchange rates, output growth, interest rates, debt-output ratios, bond prices, equity returns, and consumption and dividend processes (Hamilton, 2008). Regime switching has also been incorporated into DSGE models; see Schorfheide (2005), Liu, Waggoner, and Zha (2011), Bianchi
(2013), and Lindé, Smets, and Wouters (2016). However, because of the lack of methods with good power properties, the regime switching hypothesis is rarely formally tested from a frequentist perspective. The methods proposed here can potentially narrow this gap in the literature.

This study contributes to the literature on hypothesis testing when some regularity conditions fail to hold. Some related studies are as follows. Davies (1987), King and Shively (1993), Andrews and Ploberger (1994, 1995), and Hansen (1996) considered tests when a nuisance parameter is unidentified under the null hypothesis. Hartigan (1985), Lindsay (1995), Liu and Shao (2003), and Kasahara and Shimotsu (2012, 2015) tackled the issues of zero score and/or unidentified nuisance parameters in the context of mixture models. Rotnitzky, Cox, Bottai, and Robins (2000) developed a theory for deriving the asymptotic distribution of the likelihood ratio statistic when the rank of the information matrix is one less than full. Our study is the first in the hypothesis testing literature to simultaneously address the difficulties (i) to (iii) described earlier. We conjecture that the techniques and results will have implications for hypothesis testing in other contexts that involve hidden Markov structures.

This paper is structured as follows. Section 2 presents the model and hypotheses. Section 3 introduces the test statistics. Section 4 examines the log likelihood ratio for fixed $p$ and $q$, while Section 5 presents the limiting distribution, a finite sample refinement, and an algorithm for simulating the critical values. Section 6 examines a boundary issue and the local power properties. Section 7 discusses some implications of the theory for bootstrapping and information criteria. Section 8 examines the finite sample properties. Section 9 considers an application to US real GDP growth rate data and several other applications in the context of DSGE models. Section 10 concludes the paper, and the online appendix contains all proofs. Readers interested in the empirical applications can first read Sections 2, 3, and 9 and then return to Sections 4 and 5.

The following notation is used throughout. $||x||$ is the Euclidean norm of a vector $x$. $||X||$ is the vector induced norm of a matrix $X$. $x^\otimes k$ and $X^\otimes k$ denote the $k$-fold Kronecker product of $x$ and $X$, respectively. $\text{vec}(A)$ is the vectorization of an array $A$. For example, for a three dimensional array $A$, with $n$ elements along each dimension, $\text{vec}(A)$ returns an $n^3$-vector with the $(i + (j - 1)n + (k - 1)n^2)$-th element equal to $A(i, j, k)$. $\mathbf{1}_{\{\cdot\}}$ is the indicator function. For a real valued function $f(\theta)$ of $\theta \in \mathbb{R}^p$, $\nabla_\theta f(\theta_0)$ denotes a $p$-by-1 vector of partial derivatives evaluated at $\theta_0$, $\nabla_\theta^t f(\theta_0)$ is its transpose, and $\nabla_{\theta_{j_1}} \nabla_{\theta_{j_2}} \cdots \nabla_{\theta_{j_k}} f(\theta_0)$ is the $k$-th order partial derivative of $f(\theta)$ with respect to the $j_1, j_2, \ldots, j_k$-th element of $\theta$ at $\theta_0$. The symbols $\Rightarrow$, $\rightarrow^d$, and $\rightarrow^p$ denote weak convergence under the Skorohod topology, convergence in distribution, and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.
2 Model, likelihood functions, and hypotheses

The model is as follows. Let \( \{y_t, x_t'\} \) be a sequence of random vectors, where \( y_t \) is a scalar, and \( x_t \) is a finite dimensional vector. Let \( s_t \) be a latent binary variable, the value of which determines the regime at time \( t \). Define the information set at time \( t-1 \) as \( \Omega_{t-1} = \sigma\text{-field}\{\ldots, x'_{t-1}, y_{t-2}, x'_t, y_{t-1}\} \).

Let \( f(\cdot|\Omega_{t-1}; \beta, \delta) \) denote the conditional density of \( y_t \), given \( \Omega_{t-1} \), and assume that it satisfies
\[
y_t|\Omega_{t-1}, s_t \sim \begin{cases} f(\cdot|\Omega_{t-1}; \beta, \delta_1), & \text{if } s_t = 1, \\ f(\cdot|\Omega_{t-1}; \beta, \delta_2), & \text{if } s_t = 2, \end{cases} \quad \text{(1)}
\]
where \( \delta_1, \delta_2, \text{ and } \beta \) are unknown parameters. Henceforth, we abbreviate the densities on the right hand side of (1) as \( f_t(\beta, \delta_1) \) and \( f_t(\beta, \delta_2) \), respectively. The regimes are assumed to be Markovian, with \( P(s_t = 1|\Omega_{t-1}, s_{t-1} = 1, s_{t-2}, \ldots) = P(s_t = 1|s_{t-1} = 1) = p \) and \( P(s_t = 2|\Omega_{t-1}, s_{t-1} = 2, s_{t-2}, \ldots) = P(s_t = 2|s_{t-1} = 2) = q \). The stationary probability for \( s_t = 1 \) is thus given by
\[
\xi_s(0, q) = \frac{1 - q}{2 - p - q}. \quad \text{(2)}
\]

Under (1), the log likelihood function, evaluated at \( 0 < p, q < 1 \), is
\[
\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2) = \sum_{t=1}^{T} \log \left\{ f_t(\beta, \delta_1)\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2) + f_t(\beta, \delta_2)(1 - \xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)) \right\}, \quad \text{(3)}
\]
where \( \xi_{t|v}(\cdot) \) denotes the conditional probability of \( s_t = 1 \) given \( \Omega_v \); that is,
\[
\xi_{t|v}(p, q, \beta, \delta_1, \delta_2) = P(s_t = 1|\Omega_v; p, q, \beta, \delta_1, \delta_2) \quad \text{(4)}
\]
which satisfies
\[
\xi_{t|t}(p, q, \beta, \delta_1, \delta_2) = \frac{f_t(\beta, \delta_1)\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)}{f_t(\beta, \delta_1)\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2) + f_t(\beta, \delta_2)(1 - \xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2))}, \quad \text{(5)}
\]
\[
\xi_{t+1|t}(p, q, \beta, \delta_1, \delta_2) = p\xi_{t|t}(p, q, \beta, \delta_1, \delta_2) + (1 - q)(1 - \xi_{t|t}(p, q, \beta, \delta_1, \delta_2)). \quad \text{(6)}
\]
Without loss of generality, we set \( \xi_{1|0} = \xi_s \) throughout the paper. When \( \delta_1 = \delta_2 = \delta \), the log likelihood (3) reduces to
\[
\mathcal{L}^N(\beta, \delta) = \sum_{t=1}^{T} \log f_t(\beta, \delta). \quad \text{(7)}
\]

Note that (2) and (3)-(6) are implied by the derivations in Sections 2 and 4 of Hamilton (1989).

In this paper, we study tests based on (7) and (3). The null and alternative hypotheses are
\[
H_0 : \delta_1 = \delta_2 = \delta_s \text{ for some unknown } \delta_s;
\]
\[
H_1 : (\delta_1, \delta_2) = (\delta_1^*, \delta_2^*) \text{ for some unknown } \delta_1^* \neq \delta_2^* \text{ and } (p, q) \in (0, 1) \times (0, 1).
\]
To proceed, we impose the following assumptions on the DGP and the parameter space. Let \( n_\beta \) and \( n_\delta \) denote the dimensions of \( \beta \) and \( \delta \), respectively.

5
Assumption 1 (i) The random vector \((x_t', y_t)\) is strictly stationary, ergodic, and absolutely regular, with mixing coefficients \(b_r\) satisfying \(b_r \leq c p^r\) for some \(c > 0\) and \(p \in [0, 1)\). (ii) Under the null hypothesis, \(y_t\) is generated by \(f(\cdot | \Omega_{t-1}; \beta_s, \delta_s)\), where \(\beta_s\) and \(\delta_s\) are interior points of \(\Theta \subset \mathbb{R}^{n_A}\) and \(\Delta \subset \mathbb{R}^{n_\delta}\), respectively, and \(\Theta\) and \(\Delta\) are compact.

Assumption 1(i) is the same as Assumption A.1(i) in Cho and White (2007). It allows regime switching in \(x_t\) under the null hypothesis. Assumption 1(ii) specifies the true parameter values, where the interior point condition ensures that the expansions considered later are well-defined.

Assumption 2 Under the null hypothesis: (i) \((\beta_s, \delta_s)\) uniquely solves \(\max_{(\beta, \delta) \in \Theta \times \Delta} E[\mathcal{L}^N(\beta, \delta)]\); and (ii) for any \(0 < p, q < 1\), \((\beta_s, \delta_s)\) uniquely solves \(\max_{(\beta, \delta_1, \delta_2) \in \Theta \times \Delta \times \Delta} E[\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)]\).

Part (i) of the assumption implies that \((\beta, \delta)\) is globally identified at \((\beta_s, \delta_s)\) under the null hypothesis. Part (ii) rules out observational equivalence, that is, no two-regime specification with some \(\delta_1 \neq \delta_2\) is observationally equivalent to the single-regime specification with \(\delta_1 = \delta_2 = \delta_s\).

Assumption 3 Under the null hypothesis: (i) \(T^{-1}[\mathcal{L}^N(\beta, \delta) - E\mathcal{L}^N(\beta, \delta)] = o_p(1)\) holds uniformly over \((\beta, \delta) \in \Theta \times \Delta\), and \(T^{-1} \sum_{t=1}^T \nabla_{(\beta', \delta')} \log f_t(\beta, \delta) \nabla_{(\beta', \delta')} \log f_t(\beta, \delta)\) is positive definite in an open neighborhood of \((\beta_s, \delta_s)\) with probability tending to one; (ii) for any \(0 < p, q < 1\), \(T^{-1}[\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2) - E\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)] = o_p(1)\) holds uniformly over \((\beta, \delta_1, \delta_2) \in \Theta \times \Delta \times \Delta\).

Assumption 3 requires (7) and (3) to satisfy the uniform law of large numbers, while allowing (3) to have multiple local maxima. Under the null hypothesis and Assumptions 2 and 3, the maximizers of (7) and (3) for \(0 < p, q < 1\) converge in probability to \((\beta_s, \delta_s)\) and \((\beta_s, \delta_s, \delta_s)\), respectively.

Below, we introduce a model that we will use throughout the paper to illustrate the theory.

An illustrative model. A prominent application of regime switching in macroeconomics is to linear models with normal errors:

\[ y_t = z_t' \alpha + w_t' \gamma_1 1_{\{s_t=1\}} + w_t' \gamma_2 1_{\{s_t=2\}} + u_t \quad \text{with} \quad u_t \sim i.i.d. N(0, \sigma^2). \]  

This model encompasses finite order AR and ADL models as special cases. In relation to (1), we have \(\Omega_{t-1} = \sigma\)-field \(\{\ldots, z_{t-1}', w_{t-1}', y_{t-2}, z_t', w_t', y_{t-1}\}\) and \(x_t' = (z_t', w_t')\). We now use this model to illustrate Assumptions 1-3. For Assumption 1, the absolute regularity of \((x_t', y_t)\) is satisfied if \(x_t\) follows a stationary VARMA(P,Q) process \(\sum_{j=0}^P B_j x_{t-j} = \sum_{j=0}^Q A_j \varepsilon_{t-j}\), where \(\varepsilon_t\) is an i.i.d. random vector with mean zero and density that is absolutely continuous with respect to Lebesgue measure.
on $\mathbb{R}^{\dim(x_t)}$; see Mokkadem (1988). Other processes that are absolutely regular with a geometric rate of decay, as reviewed in Chen (2013), include those generated by threshold autoregressive models, functional coefficient autoregressive models, and GARCH and stochastic volatility models. For Assumption 2, its part (i) is satisfied if $E x_t x_t'$ has full rank, and its part (ii) is satisfied if the single-regime specification and the two-regime specification with $\delta_1 \neq \delta_2$ are not observationally equivalent. Finally, Assumption 3 requires that $T^{-1} \sum_{t=1}^{T} x_t x_t'$ is positive definite in large samples and that the uniform laws of large numbers hold. Because $\xi_{tlt-1}(p, q, \beta, \delta_1, \delta_2)$ is bounded between zero and one, they hold under Assumption 1 and mild conditions on the moments of $y_t$ and $x_t$. ■

3 The test statistic

We first introduce a family of test statistics based on the likelihood ratio, and then examine empirically relevant values of the transition probabilities $p$ and $q$. The second issue is important, not only for making the tests empirically relevant, but also for the technical analysis presented later.

Let $\tilde{\beta}$ and $\tilde{\delta}$ denote the restricted MLE, i.e., $(\tilde{\beta}, \tilde{\delta}) = \arg\max_{\beta, \delta} \mathcal{L}^{N}(\beta, \delta)$. The log likelihood ratio, evaluated at some $0 < p, q < 1$, is equal to

$$LR(p, q) = 2 \left[ \max_{\beta, \delta_1, \delta_2} \mathcal{L}^{A}(p, q, \beta, \delta_1, \delta_2) - \mathcal{L}^{N}(\tilde{\beta}, \tilde{\delta}) \right]. \quad (9)$$

This leads to the following test statistic:

$$\text{SupLR}(\Lambda_e) = \sup_{(p, q) \in \Lambda_e} LR(p, q),$$

where $\Lambda_e$ is a compact set, specified below, and the supremum is taken to obtain the strongest evidence against the null hypothesis. The supremum can be replaced by other operators. For example, following Andrews and Ploberger (1994), one can consider $\text{ExpLR}(\Lambda_e) = \int_{\Lambda_e} LR(p, q) \, dJ(p, q)$, with $J(p, q)$ being a weight function. Such considerations lead to a family of tests based on $LR(p, q)$ and $\Lambda_e$. We focus on $\text{SupLR}(\Lambda_e)$; the results extend immediately to other tests including $\text{ExpLR}(\Lambda_e)$.

Hamilton (2008) reviews twelve articles that apply regime switching models in a wide range of contexts. Of these, ten articles consider two-regime specifications with constant transition probabilities, related to exchange rates (Jeanne and Masson, 2000), output growth (Hamilton, 1989; Chauvet and Hamilton, 2006), interest rates (Hamilton, 1988, 2005; Ang and Bekaert, 2002b), debt-output ratio (Davig, 2004), bond prices (Dai, Singleton, and Yang, 2007), equity returns (Ang and Bekaert, 2002a), and consumption and dividend processes (Garcia, Luger, and Renault, 2003). Eighteen sets of estimates are reported, where the values of the transition probabilities range between 0.855 and 0.998 for the more persistent regime, and between 0.740 and 0.997 for the other
regime. These estimates have two features: (i) $p + q$ is substantially above one; and (ii) $p$ (or $q$) is close to one. Motivated by these two features, we suggest specifying $\Lambda_{\epsilon}$ as

$$\Lambda_{\epsilon} = \{(p,q) : p + q \geq 1 + \epsilon \text{ and } \epsilon \leq p, q \leq 1 - \epsilon \text{ with } \epsilon > 0\}. \quad (10)$$

In our Monte Carlo experiments, we experiment with small values of $\epsilon$, such that the features (i) and (ii) are consistent with the analysis. The results suggest that $\epsilon = 0.02$ is a reasonable choice.

The set in (10) can be generalized to allow for different trimming proportions, leading to $\{(p,q) : p + q \geq 1 + \epsilon_1 \text{ and } \epsilon_2 \leq p, q \leq 1 - \epsilon_3 \text{ with } \epsilon_1, \epsilon_2, \epsilon_3 > 0\}$. It can also be modified to incorporate additional information. For example, if $p$ and $q$ are known to be higher than 0.5, we can use

$$\{(p,q) : 0.5 + \epsilon \leq p, q \leq 1 - \epsilon \text{ with } \epsilon > 0\}. \quad (11)$$

In this paper, we focus on (10). The results hold for the latter two specifications of $\Lambda_{\epsilon}$, provided that the $\Lambda_{\epsilon}$ in the limiting distribution is modified accordingly. As a limitation, (10) and (11) exclude some of the estimates reported above when they are very close to one. To fully resolve this issue, we would need to allow $p$ or $q$ to approach one as $T \rightarrow \infty$, and study the asymptotic distribution of the likelihood ratio. We leave this topic for future research.

4 Log likelihood ratio under given $p$ and $q$

The time-varying conditional regime probability $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ represents the key difference between the Markov switching model and the mixture model. In this section, we first study this probability and its partial derivatives with respect to $\beta, \delta_1$, and $\delta_2$ (Subsection 4.1), and then apply the result to develop an asymptotic expansion of the concentrated log likelihood under the null hypothesis (Subsection 4.2). All results presented in this section are uniform with respect to $(p,q) \in [\epsilon, 1-\epsilon] \times [\epsilon, 1-\epsilon]$, where $\epsilon$ can be any constant satisfying $0 < \epsilon < 1/2$.

4.1 The conditional regime probability

The following two observations are key to our analysis. First, (5) and (6) can be combined to produce a first order difference equation that relates $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ to $\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2)$:

$$\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$$

$$= p + (p + q - 1)\frac{f_t(\beta,\delta_2)(\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2) - 1)}{f_t(\beta,\delta_1)\xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2) + f_t(\beta,\delta_2)(1 - \xi_{t|t-1}(p,q,\beta,\delta_1,\delta_2))}. \quad (12)$$

From this representation, it is clear that the partial derivatives of $\xi_{t+1|t}(p,q,\beta,\delta_1,\delta_2)$ with respect to $\beta, \delta_1$, and $\delta_2$ all follow first order difference equations. Second, although these difference equations
are nonlinear when $\delta_1$ and $\delta_2$ are unrestricted, they simplify greatly if $\delta_1 = \delta_2$. Because the asymptotic expansions here are around the restricted MLE, focusing on $\delta_1 = \delta_2$ is sufficient.

The next lemma characterizes $\xi_{t+1|t}(p, q, \beta, \delta_1, \delta_2)$ and its derivatives evaluated at $(\beta, \delta_1, \delta_2) = (\beta, \delta, \delta)$, with $\beta$ and $\delta$ denoting generic values in the parameter space. Define

$$
\theta = (\beta', \delta_1', \delta_2')',
$$

(13)

$I_0 = \{1, ..., n_\beta\}$, $I_1 = \{n_\beta + 1, ..., n_\beta + n_\delta\}$, and $I_2 = \{n_\beta + n_\delta + 1, ..., n_\beta + 2n_\delta\}$, where $I_0$, $I_1$, and $I_2$ are index sets that refer to the elements of $\beta$, $\delta_1$, and $\delta_2$, respectively, which are needed forLemma 1. Define $\tilde{\xi}_{t+1|t} = \xi_{t+1|t}(p, q, \beta, \delta, \delta)$ and $\tilde{f}_t = f_t(\beta, \delta)$. Let $\nabla_{\theta_{j1}} ... \nabla_{\theta_{jk}} \tilde{\xi}_{t|t-1}$, $\nabla_{\theta_{j1}} ... \nabla_{\theta_{jk}} \tilde{f}_t$, and $\nabla_{\theta_{j1}} ... \nabla_{\theta_{jk}} \tilde{f}_{2t}$ denote the $k$-th order partial derivatives of $\xi_{t|t-1}(p, q, \beta, \delta_1, \delta_2)$, $f_t(\beta, \delta_1)$, and $f_t(\beta, \delta_2)$ with respect to the $j_1$-th,...,$j_k$-th elements of $\theta$, evaluated at generic $\beta$ and $\delta_1 = \delta_2 = \delta$.

**Lemma 1** Let $\rho = p + q - 1$ and $r = \rho(1 - \xi_*)$, with $\xi_*$ defined in (2). Then, for any $t \geq 1$:

1. $\tilde{\xi}_{t+1|t} = \xi_*$.
2. $\nabla_{\theta_j} \tilde{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \tilde{\xi}_{t|t-1} + \tilde{\xi}_{jt,t}$, where $\tilde{\xi}_{jt,t}$ is equal to zero if $j \in I_0$, $r \nabla_{\theta_j} \log \tilde{f}_t$ if $j \in I_1$, and $-r \nabla_{\theta_j} \log \tilde{f}_t$ if $j \in I_2$.
3. $\nabla_{\theta_j} \nabla_{\theta_k} \tilde{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \nabla_{\theta_k} \tilde{\xi}_{t|t-1} + \tilde{\xi}_{jk,t}$, where $\tilde{\xi}_{jk,t}$ is equal to (Let $(I_a, I_b)$ denote the case $j \in I_a$ and $k \in I_b; a, b = 0, 1, 2.$)

\[
\begin{align*}
(0, I_0) & : 0 \\
(0, I_1) & : -\frac{\nabla_{\theta_j} \tilde{f}_{2t}}{\tilde{f}_t} \nabla_{\theta_k} \tilde{f}_t + \frac{\nabla_{\theta_j} \nabla_{\theta_k} \tilde{f}_{2t}}{\tilde{f}_t} \\
(0, I_2) & : \frac{\nabla_{\theta_j} \tilde{f}_{2t}}{\tilde{f}_t} \nabla_{\theta_k} \tilde{f}_t - \frac{\nabla_{\theta_j} \nabla_{\theta_k} \tilde{f}_{2t}}{\tilde{f}_t} \\
(I_1, I_1) & : \rho(1 - 2\xi_*) \nabla_{\theta_j} \tilde{\xi}_{t|t-1} \nabla_{\theta_k} \tilde{f}_t + \rho(2\xi_* - 1) \nabla_{\theta_j} \tilde{\xi}_{t|t-1} \nabla_{\theta_k} \tilde{f}_t + \frac{\nabla_{\theta_j} \nabla_{\theta_k} \tilde{f}_{2t}}{\tilde{f}_t} \\
(I_1, I_2) & : \rho(2\xi_* - 1) \nabla_{\theta_j} \tilde{\xi}_{t|t-1} \nabla_{\theta_k} \tilde{f}_t + \frac{\nabla_{\theta_j} \nabla_{\theta_k} \tilde{f}_{2t}}{\tilde{f}_t} \\
(I_2, I_2) & : \rho(2\xi_* - 1) \nabla_{\theta_j} \tilde{\xi}_{t|t-1} \nabla_{\theta_k} \tilde{f}_t + \frac{\nabla_{\theta_j} \nabla_{\theta_k} \tilde{f}_{2t}}{\tilde{f}_t}
\end{align*}
\]

4. $\nabla_{\theta_j} \nabla_{\theta_k} \tilde{\xi}_{t+1|t} = \rho \nabla_{\theta_j} \nabla_{\theta_k} \tilde{\xi}_{t|t-1} + \tilde{\xi}_{jk,l,t}$, where the expressions of $\tilde{\xi}_{jk,l,t}$, with $j, k, l \in \{I_a, I_b, I_c\}$ and $a, b, c = 0, 1, 2$, are given in the appendix.

**Remark 1** Lemma 1 holds in finite samples. When $\delta_1 = \delta_2$, the conditional regime probability reduces to a constant that is simply its stationary probability, whereas its derivatives up to the third order all follow linear first order difference equations. Because the lagged coefficients equal $p + q - 1$ with $0 < p, q < 1$, these difference equations are all stable, and therefore are amenable to the application of uniform laws of large numbers and functional central limit theorems. Lemma 1 is the key result that makes our subsequent analysis feasible.
4.2 Concentrated log likelihood and its expansion

To obtain the limiting distribution of the log likelihood ratio (9), a standard approach is to expand $\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)$ around the restricted MLE, $(\tilde{\beta}, \tilde{\delta}, \tilde{\delta})$, and study the supremum of this expansion over the parameter space. Unfortunately, this approach is infeasible for the current problem because $\mathcal{L}^A(p, q, \beta, \delta_1, \delta_2)$ can have multiple local optima. In particular, when computing the supremum, the first order condition with respect to $(\beta, \delta_1, \delta_2)$ can produce multiple zeros, which makes subsequent analysis difficult. Cho and White (2007) encountered a similar problem, and proceeded by working with the concentrated likelihood. We follow their insightful strategy. This allows us to break the analysis into two steps. In the first step, we quantify the relationship between $(\beta, \delta_1)$ and $\delta_2$ using the first order conditions that define the concentrated likelihood (Lemma ??). This removes $\beta$ and $\delta_1$ from the subsequent analysis, reducing the dimension of the problem by half or more. This step uses the property that, once $\delta_2$ and $(p, q)$ are fixed, the likelihood has a unique maximum. In the second step, we expand the concentrated likelihood around $\delta_2 = \tilde{\delta}$ (Lemma 2), and obtain an approximation to $LR(p, q)$. Because the conditional regime probability is time varying, the analysis here is substantially more challenging than that in Cho and White (2007).

Let $\hat{\beta}(\delta_2)$ and $\hat{\delta}_1(\delta_2)$ denote the maximizer of the log likelihood for a generic $\delta_2 \in \Delta$ (the dependence of $\hat{\beta}$ and $\hat{\delta}_1$ on $p$ and $q$ is suppressed to shorten the expressions):

$$((\hat{\beta}(\delta_2), \hat{\delta}_1(\delta_2)) = \arg \max_{\beta, \delta_1} \mathcal{L}^A(p, q, \beta, \delta_1, \delta_2).$$

(14)

Let $\mathcal{L}(p, q, \delta_2)$ be the concentrated log likelihood: $\mathcal{L}(p, q, \delta_2) = \mathcal{L}^A(p, q, \hat{\beta}(\delta_2), \hat{\delta}_1(\delta_2), \delta_2)$. Because (9) satisfies $\max_{\beta, \delta_1, \delta_2} \mathcal{L}^A(p, q, \beta, \delta_1, \delta_2) = \max_{\delta_2} \mathcal{L}(p, q, \delta_2)$ and $\mathcal{L}^N(\tilde{\beta}, \tilde{\delta}) = \mathcal{L}(p, q, \tilde{\delta})$, we have

$$LR(p, q) = 2 \max_{\delta_2} \left[ \mathcal{L}(p, q, \delta_2) - \mathcal{L}(p, q, \tilde{\delta}) \right].$$

(15)

For any $k \geq 1$, let $\mathcal{L}_{i_1,...,i_k}^{(k)}(p, q, \delta_2)$ ($i_1, ..., i_k \in \{1, ..., n_\delta\}$) denote the $k$-th order derivative of $\mathcal{L}(p, q, \delta_2)$ with respect to the $i_1$-th, ..., $i_k$-th elements of $\delta_2$. Let $d_j$ ($j \in \{1, ..., n_\delta\}$) denote the $j$-th element of $(\delta_2 - \tilde{\delta})$. A fourth order Taylor expansion of $\mathcal{L}(p, q, \delta_2)$ around $\tilde{\delta}$ is

$$\mathcal{L}(p, q, \delta_2) = \mathcal{L}(p, q, \tilde{\delta}) + \sum_{j=1}^{n_\delta} \mathcal{L}_j^{(1)}(p, q, \tilde{\delta}) d_j + \frac{1}{2!} \sum_{j=1}^{n_\delta} \sum_{k=1}^{n_\delta} \mathcal{L}_{jk}^{(2)}(p, q, \tilde{\delta}) d_j d_k + \frac{1}{3!} \sum_{j=1}^{n_\delta} \sum_{k=1}^{n_\delta} \sum_{l=1}^{n_\delta} \mathcal{L}_{jkl}^{(3)}(p, q, \tilde{\delta}) d_j d_k d_l + \frac{1}{4!} \sum_{j=1}^{n_\delta} \sum_{k=1}^{n_\delta} \sum_{l=1}^{n_\delta} \sum_{m=1}^{n_\delta} \mathcal{L}_{jklm}^{(4)}(p, q, \tilde{\delta}) d_j d_k d_l d_m.$$

(16)
where $\tilde{\delta} = \delta + c(\delta_2 - \tilde{\delta})$ for some $c \in (0, 1)$ by Taylor’s theorem in several variables.

In a standard testing problem, a second order Taylor expansion around the restricted MLE is sufficient to establish the asymptotic distribution of the likelihood ratio. This is because the score vector and the Hessian matrix are $O_p(T^{1/2})$ and $O_p(T)$, respectively, and the unrestricted MLE converges at rate $T^{-1/2}$. In the current problem (see Lemmas 2 and ??), the score vector in (16) is identically zero, and the Hessian matrix and the fourth order derivatives are $O_p(T^{1/2})$ and $O_p(T)$, respectively, while the unrestricted MLE converges at rate $T^{-1/4}$ over $\Lambda_c$. The third order derivatives are $O_p(T^{1/2})$; therefore, their effect on the expansion is dominated by that of the second and fourth order derivatives. Consequently, the second and fourth order derivatives play the roles of the first and second order derivatives in the standard problem, and a fourth order Taylor expansion is needed to derive the distribution of the log likelihood ratio under the null hypothesis.

**Assumption 4** There exists an open neighborhood of $(\beta_*, \delta_*)$, denoted by $B(\beta_*, \delta_*)$, and a sequence of positive, strictly stationary, and ergodic variables \{v_t\} satisfying $E v_t^{1+c} < L < \infty$ for some $c > 0$, such that $\sup_{(\beta, \delta_1) \in B(\beta_*, \delta_*)} |[\nabla_{\theta_1} \ldots \nabla_{\theta_q} f_t(\beta, \delta_1)]/f_t(\beta, \delta_1)|^{\alpha(k)/k} < v_t$ for all $i_1, \ldots, i_k \in \{1, \ldots, n_\beta + n_\delta\}$, where $1 \leq k \leq 5$, $\alpha(k) = 6$ if $k = 1, 2, 3$, and $\alpha(k) = 5$ if $k = 4, 5$.

This assumption is slightly stronger than Assumption A5 (iii) in Cho and White (2007). There, instead of $\alpha(k)/k$, the values are 4, 2, 2, and 1 for $k = 1, 2, 3$, and 4, respectively.

**Assumption 5** There exists $\eta > 0$ such that $\sup_{p,q \in [\epsilon, 1-\epsilon]} \sup_{|\beta-\tilde{\beta}| < \eta} T^{-1} |L_{jklm}^{(5)}(p, q, \delta)| = O_p(1)$ for all $j, k, l, m, n \in \{1, \ldots, n_\delta\}$, where $\epsilon$ is an arbitrarily small constant satisfying $0 < \epsilon < 1/2$.

In a standard problem, we need $L_{jk}^{(2)}(p, q, \delta)$ to be continuous in $\delta$, or $L_{jk}^{(3)}(p, q, \delta)$ to be $O_p(T)$ around $\tilde{\delta}$, to ensure that a quadratic expansion is an adequate approximation of the log likelihood ratio. Here, because the fourth order derivatives play the role of the usual second order derivatives, we need to impose Assumption 5 on the fifth order derivatives.

Let $\tilde{\xi}_{t+1|t}$ and $\tilde{f}_t$ denote $\xi_{t+1|t}(p, q, \tilde{\beta}, \tilde{\delta})$ and $f_t(\tilde{\beta}, \tilde{\delta})$. Let $\nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{\xi}_{t|t-1}$ and $\nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{f}_t$ denote the $k$-th order derivatives of $\xi_{t+1|t}(p, q, \beta, \delta_1, \delta_2)$ and $f_t(\beta, \delta_1)$ with respect to the $i_1$-th, ..., $i_k$-th elements of $\delta_1$, evaluated at $(\tilde{\beta}, \tilde{\delta})$. Define

\[
U_{jk,t} = \frac{1}{f_t} \left\{ \left( 1 - \frac{\xi_t}{\xi_*} \right) \nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{f}_t + \frac{1}{\xi_*} \nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{\xi}_{t|t-1} \nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{f}_t + \frac{1}{\xi_*^{k+1}} \nabla_{\delta_{i_1}} \ldots \nabla_{\delta_{i_k}} \tilde{\xi}_{t|t-1} \right\}, \tag{17}
\]

\[
D_{jk,t} = \frac{\nabla_{(\beta', \delta_{i_1}')} \tilde{f}_t}{f_t} U_{jk,t}, \quad \tilde{t}_t = \frac{\nabla_{(\beta', \delta_{i_1}')} \tilde{f}_t}{f_t} \tilde{t}_t = \frac{\nabla_{(\beta', \delta_{i_1}')} \tilde{f}_t}{f_t},
\]

\[
\tilde{V}_{jklm} = T^{-1} \sum_{t=1}^T U_{jk,t} \tilde{U}_{lm,t}, \quad \tilde{D}_{lm} = T^{-1} \sum_{t=1}^T D_{lm,t,t}, \quad \tilde{I} = T^{-1} \sum_{t=1}^T \tilde{t}_t.
\]
Lemma 2  Under the null hypothesis and Assumptions 1-5, for any \( j, k, l, m \in \{1, \ldots, n\} \), we have
\[
L^{(1)}_j(p, q, \tilde{\delta}) = 0, \quad T^{-1/2} L^{(2)}_j(p, q, \tilde{\delta}) = T^{-1/2} \sum_{t=1}^T \tilde{U}_{jkt} + o_p(1), \quad T^{-3/4} L^{(3)}_{jk}(p, q, \tilde{\delta}) = O_p(T^{-1/4}), \quad \text{and} \quad T^{-1} L^{(4)}_{jklm}(p, q, \tilde{\delta}) = - \{ \tilde{V}_{jklm} - \tilde{D}'_{jk} \tilde{I}^{-1} \tilde{D}_{lm} + \tilde{V}_{jmkl} - \tilde{D}'_{jm} \tilde{I}^{-1} \tilde{D}_{kl} + \tilde{V}_{jikm} - \tilde{D}'_{ji} \tilde{I}^{-1} \tilde{D}_{km} \} + o_p(1).
\]

This lemma characterizes how various terms in the expansion in (16) affect the limiting distribution of the likelihood ratio. The score vector is identically zero; thus, it has no effect whatsoever. The Hessian matrix is nonzero. However, because it converges to a random matrix after division by \( T^{1/2} \), its effect is very different to that in the standard situation. The third order derivatives do not affect the limiting distribution, but we will need to study them when establishing a finite sample refinement. Finally, the fourth order derivatives are \( O_p(T) \). Their leading terms provide consistent estimates for the limiting variances and covariances of \( T^{-1/2} \sum_{t=1}^T \tilde{U}_{jkt} \), for \( j, k \in \{1, \ldots, n\} \).

Remark 2  Of the three components of \( \tilde{U}_{jkt} \), the first component \( (1 - \xi)/\xi \nabla_{\delta_j} \nabla_{\delta_k} \tilde{f}_t/\tilde{f}_t \) is familiar in the mixture literature as a measure of dispersion. For example, it is in the test of Cho and White (2007) against mixtures. The remaining two components are new, and are the result of the Markov switching structure. They can be rewritten as \( (1 - \xi)/\xi \sum_{t=1}^{T-1} \rho^t \nabla_{\delta_j} \log \tilde{f}_1(t-s) \nabla_{\delta_k} \log \tilde{f}_t \) and \((1 - \xi)/\xi \sum_{t=1}^{T-1} \rho^t \nabla_{\delta_k} \log \tilde{f}_1(t-s) \nabla_{\delta_j} \log \tilde{f}_t \), respectively, and as such, they measure the serial dependence caused by the regime switching. Their relative magnitudes increase with \( |p| \). This suggests that the power difference between testing against Markov switching alternatives and mixture alternatives can be substantial when the regimes are persistent, that is, when \( p \) is close to one.

The illustrative model (cont’d). Suppose that in model (8), only the regression coefficients are allowed to switch. Then, \( \tilde{U}_{jkt} \) and \( \tilde{D}_{jkt} \) in Lemma 2 are equal to

\[
(1 - \xi) \left\{ \frac{w_{kt}}{\sigma^2} \left( \frac{\tilde{u}_t^2}{\sigma^2} - 1 \right) + \sum_{s=1}^{T-1} \rho^s \left( \frac{w_{kt}(t-s)\tilde{u}_{t-s}}{\sigma^2} \right) \left( \frac{w_{kt}\tilde{u}_t}{\sigma^2} \right) \right\} \tag{18}
\]

and

\[
\left[ \frac{z'_{1} \tilde{u}_t}{\sigma^2} \left( \frac{\tilde{u}_t^2}{\sigma^2} - 1 \right) - \frac{1}{2} \frac{w_{kt} \tilde{u}_t}{\sigma^2} \right]' \tilde{U}_{jkt}, \tag{19}
\]

respectively, where \( \tilde{u}_t \) are the OLS residuals and \( \tilde{\sigma}^2 = T^{-1} \sum_{t=1}^T \tilde{u}_t^2 \). Thus, \( \tilde{U}_{jkt} \) and \( \tilde{D}_{jkt} \) depend on the regressors and the OLS residuals only, and the covariance function of \( T^{-1/2} L^{(2)}_{jk}(p, q, \tilde{\delta}) \) is consistently estimable. This feature is useful for computing the critical values for the test.

5  Asymptotic approximations

This section presents three sets of results: (1) the limiting distribution of \( \text{SupLR}(\Lambda) \); (2) a finite sample refinement to further improve the aforementioned limiting distribution for an important
special case; and (3) a unified algorithm to compute the critical values.

5.1 Limiting distribution of SupLR($\Lambda_\epsilon$)

Let $T^{-1/2} \mathcal{L}^{(2)}(p, q, \tilde{\delta})$ denote an $n_\delta$-by-$n_\delta$ matrix whose $(j, k)$-th element is equal to $T^{-1/2} \mathcal{L}^{(2)}_{jk}(p, q, \tilde{\delta})$. Under Assumptions 1-5, $T^{-1/2} \mathcal{L}^{(2)}(p, q, \tilde{\delta})$ converges weakly to a vector of Gaussian processes over $\epsilon \leq p, q \leq 1 - \epsilon$ (Lemma ??). Below, we express the limiting distribution of SupLR($\Lambda_\epsilon$) in terms of this Gaussian process. For any $0 < p_r, q_r, p_s, q_s < 1$ and $j, k, l, m \in \{1, 2, ..., n_\delta\}$, define

$$\omega_{jklm}(p_r, q_r; p_s, q_s) = V_{jklm}(p_r, q_r; p_s, q_s) - D_{jk}(p_r, q_r) I^{-1} D_{lm}(p_s, q_s),$$

with $V_{jklm}(p_r, q_r; p_s, q_s) = E[U_{jk,t}(p_r, q_r) U_{im,t}(p_s, q_s)]$, $D_{jk}(p_r, q_r) = ED_{jk,t}(p_r, q_r)$, and $I = EI_t$, where $U_{jk,t}(p_r, q_r)$, $D_{jk,t}(p_r, q_r)$ and $I_t$ equal $\tilde{U}_{jk,t}$, $\tilde{D}_{jk,t}$, and $I_t$ in (17), respectively, but are evaluated at $(p_r, q_r, \beta_s, \vec{\delta})$ instead of $(p_r, q_r, \bar{\beta}, \vec{\delta})$. Let $\Omega(p_r, q_r; p_s, q_s)$ be an $n_\delta^2$-by-$n_\delta^2$ matrix, with the $(j + (k - 1)n_\delta, l + (m - 1)n_\delta)$-th element equal to $\omega_{jklm}(p_r, q_r; p_s, q_s)$. Let $\Omega(p, q) = \Omega(p, q; p, q)$.

**Assumption 6**

\[ \min_{x \in \mathbb{R}^{n_\delta}, \|x\|_2 = 1} (x \otimes^2)' \Omega(p, q) (x \otimes^2) > L \text{ for some } L > 0 \text{ and all } (p, q) \in \Lambda_\epsilon. \]

This assumption is not restrictive because $\Lambda_\epsilon$ is bounded away from $p + q = 1$. In the appendix (pp. A26-A29), we illustrate this assumption in two ways. First, we consider some cases for which $\Omega(p, q)$ can be computed analytically. In addition to showing that this assumption holds, the results also reveal that $\Omega(p_r, q_r; p_s, q_s)$ is affected by the following: (i) the model’s dynamic properties (e.g., whether the regressors are weakly or strictly exogenous); (ii) which parameters are allowed to switch (e.g., the regressions coefficients or the variance of the errors); and (iii) whether nuisance parameters are present in the model. Next, we consider $\Omega(p, q)$ in the context of the model given in (8), and explain intuitively why this assumption is expected to hold in general cases.

**Proposition 1** Under the null hypothesis and Assumptions 1-6, for $\Lambda_\epsilon$ given by (10), we have

$$\text{SupLR}(\Lambda_\epsilon) \implies \sup_{(p, q) \in \Lambda_\epsilon} \sup_{\eta \in \mathbb{R}^{n_\delta}} \mathcal{W}^{(2)}(p, q, \eta)$$

where $\mathcal{W}^{(2)}(p, q, \eta) = (\eta \otimes^2)' \text{vec} G(p, q) - (1/4)(\eta \otimes^2)' \Omega(p, q) (\eta \otimes^2)$, and $\text{vec} G(p, q)$ is an $n_\delta^2$-vector of zero-mean continuous Gaussian processes such that $E[\text{vec} G(p_r, q_r) \text{vec} G(p_s, q_s)]' = \Omega(p_r, q_r; p_s, q_s)$ for any $(p_r, q_r), (p_s, q_s) \in \Lambda_\epsilon$.

In Proposition 1 and its proof, the weak convergence is in the space of continuous functions defined on a compact set. If $n_\delta = 1$, the optimization over $\eta$ can be solved analytically, leading to $\text{SupLR}(\Lambda_\epsilon) \Rightarrow \max[0, \sup_{(p, q) \in \Lambda_\epsilon} G(p, q) / \sqrt{\Omega(p, q)}]^2$. If $n_\delta > 1$, it needs to be solved numerically. Because $\mathcal{W}^{(2)}(p, q, \eta)$ is a quadratic function of $\eta \otimes^2$ with $\eta$ unrestricted, the computation is standard.
The illustrative model (cont’d). We use the following special case of (8) to illustrate the limiting distribution in (21), and compare it with the finite sample distribution:

\[ y_t = \mu + \alpha y_{t-1} + u_t, \]  

where \( u_t \sim i.i.d. \ N(0, \sigma^2); \ (\mu, \alpha, \sigma^2) = (0, 0.5, 1); \ T = 250; \) and \( \Lambda_e \) is given by (11) with \( \epsilon = 0.05 \).

The results for testing \( \alpha \) and \( \mu \) are reported in Figures 1(a) and 1(b), respectively. The distributions in panel (a) are substantially different from those in panel (b), which confirms that the distribution of \( \text{SupLR}(\Lambda_e) \) is not invariant to which parameter is allowed to switch. Meanwhile, although the curves are close to each other in panel (a), a gap is observed in (b). In the latter case, applying the asymptotic approximation will lead to an over-rejection of the null hypothesis.

The gap is due to the structure of \( T^{-1/2} \sum_{t=1}^{T} \tilde{U}_{jk,t} \). When testing \( \mu \), we have

\[ T^{-1/2} \sum_{t=1}^{T} \tilde{U}_{jk,t} = \frac{1}{\sqrt{T}} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{\hat{y}_t^2}{\hat{\sigma}^2} - 1 \right) + \frac{2}{\sqrt{T}} \sum_{t=1}^{T} \left( \sum_{s=1}^{t-1} \rho^s \frac{y_{t-s}}{\hat{\sigma}} \frac{\hat{y}_t}{\hat{\sigma}} \right) \right\}, \]  

where the first summation is equal to zero because \( \hat{\sigma}^2 = T^{-1} \sum_{t=1}^{T} \hat{y}_t^2 \), and the second is small when \( p + q \) is close to one. Thus, although asymptotically, \( T^{-1/2} \sum_{t=1}^{T} \tilde{U}_{jk,t} \) is the leading term in the likelihood ratio expansion over \( \Lambda_e \), in finite samples, its value can be too small to dominate the omitted higher order terms such as \( T^{-3/4} \mathcal{L}_{jk}^{(3)}(p, q, \delta) \) when \( p + q \) is close to one. This omission produces the gap in panel (b). The situation is different when testing for switching in \( \alpha \), where

\[ T^{-1/2} \sum_{t=1}^{T} \tilde{U}_{jk,t} = \frac{1}{\sigma^2} \left\{ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{\hat{y}_t^2}{\hat{\sigma}^2} - 1 \right) y_{t-1}^2 + \frac{2}{\sqrt{T}} \sum_{t=1}^{T} \left( \sum_{s=1}^{t-1} \rho^s y_{t-s} \frac{\hat{y}_t}{\hat{\sigma}} \frac{\hat{y}_t}{\hat{\sigma}} \right) \right\}. \]

Because the first term in the braces has a positive variance, this complication does not arise. 

5.2 A refinement

This subsection presents a refined approximation to the distribution of \( \text{SupLR}(\Lambda_e) \) under the null hypothesis. We begin with the following assumption, which, in practice, is used to determine whether a refinement is needed for a particular testing problem.

**Assumption 7** The following linear relationship holds for some \( i_1, i_2 \in \{1, \ldots, n_\delta\} \) and all \( t \):

\[ \nabla_\delta_{i_1} \nabla_\delta_{i_2} \tilde{f}_t = \alpha_{1i_1}^{(1)} \nabla_\beta \tilde{f}_t + \alpha_{1i_2}^{(2)} \nabla_\delta_{i_1} \tilde{f}_t, \]  

where \( \alpha_{1i_2}^{(1)} \) and \( \alpha_{1i_2}^{(2)} \) are \( n_\beta \)- and \( n_\delta \)-dimensional known vectors of constants, respectively.

If (24) holds, then certain elements of the second order derivatives of the log likelihood cancel out because of the linear dependencies (see, e.g., (23)), implying that a refinement is needed.
For the model in (8), checking (24) can lead to three outcomes, depending on which parameter is allowed to switch. If only the intercept $\gamma$ is allowed to switch, then (24) is satisfied with $\nabla_\gamma \nabla_\gamma \tilde{f}_{1t} = 2\nabla_\gamma^2 \tilde{f}_{1t}$. If the intercept is not allowed to switch, then (24) is violated for all $i_1, i_2 \in \{1, \ldots, n_\delta\}$. If the intercept and some other parameters are allowed to switch, then (24) is satisfied when $\delta_{i_1}$ and $\delta_{i_2}$ both represent the intercept, but not otherwise. In the first and third cases, Assumption 7 holds, and a refinement to (21) is needed. In the second case, no complication arises, and thus no refinement is needed. For other models, checking (24) is expected to remain simple, because it requires computing only the first and second order derivatives of the density under the null hypothesis. For example, it is simple to verify that (24) is satisfied when testing $\gamma$ in the generalized linear model $y_t = g(\gamma + z_t^\prime \alpha + u_t)$, where $g(\cdot)$ is a smooth invertible function and $u_t \sim N(0, \sigma^2)$, and that it is satisfied when testing the intercepts in Gaussian vector autoregressions. The relevant computational steps for the above three models are presented in the appendix (pp. A37-A40), which illustrates how to check this assumption in various situations.

The next assumption strengthens Assumption 4. It is similar to A.5(iv) in Cho and White (2007). The subsequent analysis makes heavy use of their results in Section 2.3.2.

**Assumption 8** There exists an open neighborhood of $(\beta^*_s, \delta^*_s)$, $B(\beta^*_s, \delta^*_s)$, and a sequence of positive, strictly stationary, and ergodic random variables $\{v_t\}$, satisfying $Ev_t^{1+c} < \infty$ for some $c > 0$, such that the supremums of the following quantities over $B(\beta^*_s, \delta^*_s)$ are bounded from above by $v_t$: $$\left| \nabla_{\theta_{i_1}} \cdots \nabla_{\theta_{i_k}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1) \right|^4, \left| \nabla_{\theta_{i_1}} \cdots \nabla_{\theta_{i_m}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1) \right|^2, \left| \nabla_{\theta_{i_1}} \cdots \nabla_{\theta_{i_8}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1) \right|, \left| \nabla_{\theta_{i_1}} \nabla_{\theta_{i_2}} \nabla_{\theta_{i_3}} \cdots \nabla_{\theta_{i_6}} f_t(\beta, \delta_1) / f_t(\beta, \delta_1) \right|,$$ where $k = 1, 2, 3, 4$, $m = 5, 6, 7, i_1, \ldots, i_7 \in \{1, \ldots, n_\beta + n_\delta\}$, and $j_1, j_2 \in \{1, \ldots, n_\beta\}$.

Because the refinement of the order is to adequately account for the effects of the higher order terms when $(p, q)$ is close to $p + q = 1$, in Lemma 3 below, we study a sixth order Taylor expansion of the log likelihood ratio along $p + q = 1$ and an eighth order expansion at $p = q = 1/2$. To approximate the third and sixth order derivatives of the concentrated log likelihood, define

$$\tilde{s}_{jkl,t}(p, q) = \frac{(1 - p)(p - q)}{(1 - q)^2} \frac{\nabla_{\delta_1} \cdots \nabla_{\delta_6} \nabla_{\delta_1} \tilde{f}_{1t}}{f_t},$$

and let $G_{jkl}^{(3)}(p, q)$ be a zero-mean continuous Gaussian process, satisfying

$$\omega_{jklmnu}^{\text{(3)}}(p_r, q_r; p_s, q_s) = \text{Cov}(G_{jkl}^{(3)}(p_r, q_r), G_{mnu}^{(3)}(p_s, q_s))$$

$$= E[s_{jkl,t}(p, q)s_{mnu,t}(p, q)] - E \left[ \frac{\nabla(\beta' \delta') f_{1t}}{f_t} s_{jkl,t}(p, q) \right] I^{-1} \left[ \frac{\nabla(\beta' \delta') f_{1t}}{f_t} s_{mnu,t}(p, q) \right],$$
where \( s_{ijkl}(p, q) \) is equal to \( \tilde{s}_{ijkl}(p, q) \), but is evaluated at the true parameter values. Let \( \omega_{ijklmn}^{(3)}(p, q) = \omega_{ijklmn}^{(3)}(p, q) \). To approximate the fourth and eighth order derivatives, define

\[
\tilde{k}_{ijkl,m}(p, q) = \frac{1 - p}{2 - p - q} \left( 1 + \left( \frac{1 - p}{1 - q} \right)^3 \right) \left( \nabla_1 \delta_{ik} \nabla_1 \delta_{ij} \nabla_1 \delta_{im} \tilde{f}_{1t} \right) + \frac{1}{1 - q} \sum_{(i_1, i_2, i_3, i_4) \in S} \frac{1}{\tilde{f}_{1t}} \left\{ -\nabla_{\delta_{i11}} \nabla_{\delta_{i12}} \nabla_{\delta_{i13}} \tilde{f}_{1t} \alpha_{i3i4}^{(1)} - \left( \frac{1 - p}{1 - q} \right) \nabla_{\delta_{i11}} \nabla_{\delta_{i12}} \nabla_{\delta_{i13}} \tilde{f}_{1t} \alpha_{i3i4}^{(2)} \right\},
\]

and let \( G_{ijklm}^{(4)}(p, q) \) denote a zero-mean continuous Gaussian process, satisfying

\[
\omega_{i1i2...i8}^{(4)}(p, q) = \text{Cov} \left( G_{i1i2i3i4}^{(4)}(p, q), G_{k5i6i7i8}^{(4)}(p, q) \right) = E \left[ k_{i1i2i3i4,t}(p, q) k_{i5i6i7i8,t}(p, q) \right] - E \left[ \frac{\nabla_{(\beta', \beta)} f_{1t}}{\tilde{f}_{1t}} k_{i1i2i3i4,t}(p, q) \right] \left[ \frac{\nabla_{(\beta', \beta)} f_{1t}}{\tilde{f}_{1t}} k_{i5i6i7i8,t}(p, q) \right],
\]

where \( S = \{ jklm, jklm, jkmnl, jkmkl, jmlkn, jmlkj \} \), and \( k_{i1i2i3i4,t}(p, q) \) is equal to \( \tilde{k}_{i1i2i3i4,t}(p, q) \), but is evaluated at the true parameter values. Let \( \omega_{i1i2...i8}(p, q) = \omega_{i1i2...i8}(p, q) \).

**Lemma 3** Under the null hypothesis and Assumptions 1-8 with (24) satisfied for all \( i_1, i_2 \in \{ 1, ..., n_\delta \} : \)

1. The following results hold uniformly over \( \{ (p, q) : \epsilon \leq p, q \leq 1 - \epsilon, p + q = 1 \} : T^{-1/2} \mathcal{L}_{ijkl}^{(3)}(p, q, \tilde{\delta}) = T^{-1/2} \sum_{t=1}^{T} \tilde{s}_{ijkl,t}(p, q) + o_p(1) \Rightarrow G_{ijkl}^{(3)}(p, q); T^{-1/2} \mathcal{L}_{ijkl}^{(4)}(p, q, \tilde{\delta}) = O_p(1); T^{-1/2} \mathcal{L}_{ijklmn}^{(5)}(p, q, \tilde{\delta}) = O_p(1); T^{-1/2} \mathcal{L}_{jklmn}^{(6)}(p, q, \tilde{\delta}) = -\sum_{(i_1, i_2, ..., i_8) \in IND_1} \omega_{i1i2...i8}^{(3)}(p, q) + o_p(1), where IND_1 = \{ jklnmr, jklnmr, jklnmr, jkmnr, jkmnr, jkmnr, jkmnr, jkmnr, jkmnr, jkmnr, jkmnr, jkmnr \}.

2. The following results hold at \( p = q = 1/2 : T^{-1/2} \mathcal{L}_{ijkl}^{(3)}(p, q, \tilde{\delta}) = o_p(1); T^{-1/2} \mathcal{L}_{ijkl}^{(4)}(p, q, \tilde{\delta}) = T^{-1/2} \sum_{t=1}^{T} \tilde{k}_{ijkl,m}(p, q) + o_p(1) \Rightarrow G_{ijkl}^{(4)}(p, q); T^{-1/2} \mathcal{L}_{ijklm}^{(5)}(q, \tilde{\delta}) = -\sum_{(i_1, i_2, ..., i_8) \in IND_2} \omega_{i1i2...i8}^{(4)}(p, q) + o_p(1), where the elements of IND_2 are defined as follows: \( i_1 = j \), each triplet \((i_2, i_3, i_4)\) corresponds to one of the 35 outcomes of picking three elements from \{k, l, m, n, r, s, u\} (the ordering does not matter), and \( i_5, i_6, i_7, \) and \( i_8 \) correspond to the remaining elements.

When \( p + q = 1 \) and \( p \neq 1/2, T^{-1/2} \sum_{t=1}^{T} \tilde{s}_{ijkl,t}(p, q) \) serves as the leading term of the Taylor expansion. As a result, a sixth order expansion is needed to approximate the likelihood ratio. When \( p = q = 1/2, T^{-1/2} \sum_{t=1}^{T} \tilde{k}_{ijkl,m}(p, q) \) becomes the leading term, and an eighth order expansion is
needed. Lemma 3 assumes that (24) is satisfied for all \( i_1, i_2 \in \{1, ..., n_3\} \). If this relationship holds for a subset of derivatives, then we set \( \alpha_{i_1 i_2}^{(1)} = 0 \) and \( \alpha_{i_1 i_2}^{(2)} = 0 \) for the cases that do not satisfy (24).

Proposition 1 and Lemma 3 lead to three different approximations to \( LR(p, q) \). The approximation in Proposition 1 is expected to perform well when \( (p, q) \) is not close to the boundary \( p + q = 1 \). The approximation implied by Lemma 2.1, which is based on the limit of \( T^{-1/2} \sum_{t=1}^{T} \tilde{z}_{jkl,t}(p, q) \), is expected to perform well when \( p + q = 1 \) but \( p \neq 1/2 \), as well as when \( (p, q) \) is local to such a point because of the continuity of \( LR(p, q) \) with respect to \( p \) and \( q \). Finally, the approximation implied by Lemma 2.2, which is based on the limit of \( T^{-1/2} \sum_{t=1}^{T} \hat{e}_{jkl,t}(p, q) \), is expected to perform well when \( (p, q) \) is equal to, or is local to \((1/2, 1/2)\). These three approximations complement each other. By merging their leading terms in a proper way, according to how they appear in the Taylor expansion, it is potentially possible to obtain an approximation that performs well over a wide range of transition probabilities. This is the intuition behind our refined approximation. We present this approximation below, and examine it further in Subsection 6.1.

Let \( G_{ij}^{(3)}(p, q) \) be an \( n_3^2 \)-dimensional vector, with the \((j + (k - 1)n_3 + (l_1 - 1)n_3^2)\)-th element given by \( G_{jkl}^{(3)}(p, q) \), and \( \Omega_{ij}^{(3)}(p, q) \) an \( n_3^3 \)-by-\( n_3^3 \) matrix, with the \((j + (k - 1)n_3 + (l_1 - 1)n_3^2, m + (n_3 - 2)n_3^2 + (r_1 - 1)n_3^3)\)-th element given by \( \omega_{jklmn}^{(3)}(p, q) \). Define

\[
W^{(3)}(p, q, \eta) = T^{-1/4} \left( \frac{1}{3} \right) (\eta^{(3)})' \text{vec} G^{(3)}(p, q) - T^{-1/2} \left( \frac{1}{36} \right) (\eta^{(3)})' \Omega^{(3)}(p, q) (\eta^{(3)})
\]

Similarly, let \( G_{ij}^{(4)}(p, q) \) be an \( n_4^2 \)-dimensional vector, with the \((j + (k - 1)n_4 + (l_1 - 1)n_4^2 + (m_1 - 1)n_4^3)\)-th element equal to \( G_{jkl,m}^{(4)}(p, q) \), and \( \Omega^{(4)}(p, q) \) an \( n_4^3 \)-by-\( n_4^3 \) matrix, with the \((j + (k - 1)n_3 + (l_1 - 1)n_4^2 + (m_1 - 1)n_4^3, n + (r_1 - 1)n_3 + (s_1 - 1)n_4^2 + (u_1 - 1)n_4^3)\)-th element equal to \( \omega_{jklmnrsu}^{(4)}(p, q) \). Define

\[
W^{(4)}(p, q, \eta) = T^{-1/4} \left( \frac{1}{12} \right) (\eta^{(4)})' \text{vec} G^{(4)}(p, q) - T^{-1/2} \left( \frac{1}{576} \right) (\eta^{(4)})' \Omega^{(4)}(p, q) (\eta^{(4)})
\]

We propose using \( S_\infty(\Lambda_\epsilon) \) as the refined approximation to \( \text{SupLR}(\Lambda_\epsilon) \), where

\[
S_\infty(\Lambda_\epsilon) \equiv \sup \left\{ \sup \left\{ \frac{1}{2^{p}} \right\} \right\}
\]

Corollary 1 Under the null hypothesis and Assumptions 1-8 with \( \Lambda_\epsilon \) equal to (10), we have

\[
\Pr \left( \text{SupLR}(\Lambda_\epsilon) \leq s \right) - \Pr \left( S_\infty(\Lambda_\epsilon) \leq s \right) \rightarrow 0 \text{ for any } s \in R.
\]

The illustrative model (cont'd). For the model in (22), when testing \( \mu \), (25) and (26) equal

\[
\frac{(1-p)(p-q)}{(1-q)^2} \frac{1}{s^2} \left\{ \left( \frac{\hat{u}_2}{s} \right)^3 - 3 \left( \frac{\hat{u}_2}{s} \right) \right\} \text{ and } \left[ \frac{1-p}{2-p-q} \left( 1 + \left( \frac{1-p}{1-q} \right)^3 \right) - 3 \left( \frac{1-p}{1-q} \right)^2 \right] \frac{1}{s^2} \left\{ \left( \frac{\hat{u}_2}{s} \right)^4 - 6 \left( \frac{\hat{u}_2}{s} \right)^2 + 3 \right\},
\]

respectively. The refined approximation is reported in Figure 1(b). The improvement over the original approximation is substantial. When testing \( \alpha \), by Assumption 7, no refinement is needed.
5.3 Critical values

This section presents an algorithm to produce the critical values of $S_{\infty}(\Lambda_e)$ in (27). The idea is to sample from the estimated distribution of the vector Gaussian process

$$[\text{vec} \ G (p, q)' , \text{vec} \ G^{(3)} (p, q)' , \text{vec} \ G^{(4)} (p, q)'],$$

and then, for each draw, solve the maximization problem (27). Similar algorithms are considered in Hansen (1992) and Garcia (1998). The main steps are as follows.

**STEP 1** (Estimate the parameters). Estimate the model after imposing the null hypothesis to obtain $\tilde{\beta}$ and $\tilde{\delta}$. Create an equidistant grid over $\Lambda_e$, and denote the grid points by $\{\pi_i, q_i\}_{i=1}^n$.

**STEP 2** (Estimate the covariance function of the Gaussian process). Compute $\tilde{U}_{jk, t}(\pi_i, q_i)$, $\tilde{s}_{jklt}(\pi_i, q_i)$, and $\tilde{k}_{jkltm}(\pi_i, q_i)$ using (17), (25), and (26), respectively, where $j, k, l, m \in \{1, \ldots, n_\delta\}$ and $i \in \{1, \ldots, n\}$. For each $i$, store their values in a vector as

$$\tilde{G}_t (\pi_i, q_i) = \begin{bmatrix} \tilde{U}^{(2)}_t (\pi_i, q_i) \\ \tilde{U}^{(3)}_t (\pi_i, q_i) \\ \tilde{U}^{(4)}_t (\pi_i, q_i) \end{bmatrix},$$

where $\tilde{U}^{(2)}_t (p, q)$ is an $n_\delta^2$-vector, with the $(j + (k-1)n_\delta)$-th element equal to $\tilde{U}_{jk, t}(p, q)$, $\tilde{U}^{(3)}_t (p, q)$ is an $n_\delta^3$-vector, with the $(j + (k-1)n_\delta + (l-1)n_\delta^2)$-th element equal to $\tilde{s}_{jklt}(p, q)$, and $\tilde{U}^{(4)}_t (p, q)$ is an $n_\delta^4$-vector, with the $(j + (k-1)n_\delta + (l-1)n_\delta^2 + (m-1)n_\delta^3)$-th element equal to $\tilde{k}_{jkltm}(p, q)$.

For each $i, s \in \{1, \ldots, n\}$, compute

$$\tilde{\Omega} (\pi_i, q_i; p_s, q_s) = \begin{bmatrix} \tilde{G}_t (\pi_i, q_i) \tilde{G}_t (p_s, q_s)' \\ \tilde{G}_t (\pi_i, q_i) \tilde{G}_t (p_s, q_s)' \end{bmatrix},$$

$$= T^{-1} \sum_{t=1}^T \tilde{G}_t (\pi_i, q_i) \tilde{G}_t (p_s, q_s)'$$

$$- \left\{ T^{-1} \sum_{t=1}^T \tilde{G}_t (\pi_i, q_i) \frac{\nabla (\beta', \delta') \tilde{f}_t}{\tilde{f}_t} \right\} \left( T^{-1} \sum_{t=1}^T \tilde{G}_t (p_s, q_s) \frac{\nabla (\beta', \delta') \tilde{f}_t}{\tilde{f}_t} \right)'.$$

Let $\tilde{\Omega}(\pi_i, q_i)$, $\tilde{\Omega}^{(3)}(\pi_i, q_i)$, and $\tilde{\Omega}^{(4)}(\pi_i, q_i)$ denote the three consecutive diagonal blocks of $\tilde{\Omega} (\pi_i, q_i; p_i, q_i)$ of dimensions $n_\delta^2$, $n_\delta^3$, and $n_\delta^4$, respectively.

**STEP 3** (Sampling). Generate an $n$-by-$1$ zero-mean normal random vector, with covariance equal to (30), and repeat $B$ times. Save the values as $[\text{vec} \ G_b (\pi_i, q_i)', \text{vec} \ G^{(3)}_b (\pi_i, q_i)', \text{vec} \ G^{(4)}_b (\pi_i, q_i)'],$ where $i = 1, \ldots, n$, and $b = 1, \ldots, B$.

**STEP 4** (Optimization). For each $b \in \{1, \ldots, B\}$, solve

$$S_b (\Lambda_e) \equiv \sup_{\{p_i, q_i\}_{i=1}^n} \sup_{\eta \in R^{n_\delta}} \sum_{j=2}^4 W^{(j)}_b (p_i, q_i, \eta),$$

where $W^{(j)}_b (p_i, q_i, \eta)$ is the $j$-th moment of the estimated model after imposing the null hypothesis.
where

\[ W_b^{(2)}(p_i, q_i, \eta) = (\eta \otimes 2)' \text{vec} \, G_b(p_i, q_i) - \frac{1}{4} (\eta \otimes 2)' \overline{G}(p_i, q_i) (\eta \otimes 2), \]

\[ W_b^{(3)}(p_i, q_i, \eta) = T^{-1/4} \frac{1}{3} (\eta \otimes 3)' \text{vec} \, G_b^{(3)}(p_i, q_i) - T^{-1/2} \frac{1}{36} (\eta \otimes 3)' \overline{G}^{(3)}(p_i, q_i) (\eta \otimes 3), \]

\[ W_b^{(4)}(p_i, q_i, \eta) = T^{-1/2} \frac{1}{12} (\eta \otimes 4)' \text{vec} \, G_b^{(4)}(p_i, q_i) - T^{-1} \frac{1}{576} (\eta \otimes 4)' \overline{G}^{(4)}(p_i, q_i) (\eta \otimes 4). \]

Sort the values of \( S_b(\Lambda_e) \) \((b=1, ..., B)\) to obtain the desired critical value.

**The illustrative model (cont’d).** When testing the intercept of (22), we have

\[ \bar{U}_t^{(2)}(p, q) = \frac{2}{\sigma^2} \left( \frac{1-p}{1-q} \right) \sum_{s=1}^{t-1} (p+q-1)^s \bar{u}_t \bar{u}_t, \quad \bar{U}_t^{(3)}(p, q) = \frac{(1-p)(p-q)}{(1-q)^2} \frac{1}{\sigma^4} \left\{ \left( \frac{\bar{u}_t}{\sigma} \right)^3 - 3 \left( \frac{\bar{u}_t}{\sigma} \right) \right\}, \]

\[ \bar{U}_t^{(4)}(p, q) = \left[ \frac{1-p}{2-2q} \left( 1 + \left( \frac{1-p}{1-q} \right)^2 \right) - 3 \left( \frac{1-p}{1-q} \right)^2 \right] \frac{1}{\sigma^4} \left\{ \left( \frac{\bar{u}_t}{\sigma} \right)^4 - 6 \left( \frac{\bar{u}_t}{\sigma} \right)^2 + 3 \right\}, \]

\[ \bar{u}_t^{(8)}(p, q) = \left[ \frac{1}{\sigma^2} \left( \frac{\bar{u}_t^2}{\sigma^2} - 1 \right) - \frac{\bar{u}_t}{\sigma^2} \right]. \]

The covariance function in (30) follows from (33) and (29). When testing the same hypothesis in AR\((p)\) or ADL\((p, q)\) models, the covariance function can be computed in the same way, except that \(y_{t-1}\) in (33) needs to be replaced by \((y_{t-1}, ..., y_{-p})\) and \((y_{t-1}, ..., y_{-p}, x_{t-1}, ..., x_{t-q})\), respectively.

### 6 The boundary issue and local power properties

In this section, we study two issues. First, we study the likelihood ratio under the null hypothesis when the transition probabilities are close to the boundary:

\[ \{(p, q) : p + q = 1 \text{ and } \epsilon \leq p, q \leq 1 - \epsilon \text{ for some } 0 < \epsilon < 0.5\}. \]

The results further justify the refined approximation presented and implemented in Subsections 5.2-5.3. Next, we study the likelihood ratio under the alternative hypothesis. The results explain the potential local power difference between the likelihood ratio test and the tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) in an empirically important setting.

#### 6.1 The boundary issue

We allow the transition probabilities to depend on the sample size, such that they approach (34) as \(T \to \infty\). The analysis below is based on an eighth order expansion of \(L(p_T, q_T, \delta_2)\) around \(\delta\):

\[ L(p_T, q_T, \delta_2) - L(p_T, q_T, \overline{\delta}) = \frac{1}{k!} \sum_{k=2}^{7} \sum_{i_k=1}^{n_k} L_{i_1...i_k}(p_T, q_T, \overline{\delta}) \, d_{i_1}...d_{i_k} + \frac{1}{8!} \sum_{i_1...i_8=1}^{n_8} L_{i_1...i_8}(p_T, q_T, \overline{\delta}) \, d_{i_1}...d_{i_8}, \]
where $d_j$ is the $j$-th element of $(\delta_2 - \tilde{\delta})$ and $\tilde{\delta} = \delta + c(\delta_2 - \tilde{\delta})$ for some $c \in (0, 1)$. We assume (24) is satisfied for all, not just some $i_1, i_2 \in \{1, \ldots, n_\delta\}$. Relaxing this assumption is left for future work.

**Assumption 9**

(a) $\sup_{(\beta, \delta) \in B(\beta_*, \delta_*)} |\nabla_{\theta_1} \cdots \nabla_{\theta_q} f_1(\beta, \delta_1) / f_1(\beta, \delta_1)|^{\alpha(k) / k} < \nu_t$ holds for $k = 1, 2, 5, 8$, where $i_1, \ldots, i_8 \in \{1, ..., n_\beta + n_\delta\}$, $\alpha(k) = 12$ if $k \in \{1, 2, 5\}$, $\alpha(8) = 8$, and $B(\beta_*, \delta_*)$ and $\nu_t$ are defined in Assumption 8; (b) $E(\nabla_{\theta_1} \cdots \nabla_{\theta_q} f_t(\beta_*, \delta_*) \nabla_{\theta} f_t(\beta_*, \delta_*) / f_t(\beta_*, \delta_*)^2 | \Omega_{t-1} = 0$ holds for $k = 3, 4$, where $i_1, \ldots, i_4 \in \{n_\beta + 1, \ldots, n_\beta + n_\delta\}$, and $\beta_*$ and $\delta_*$ denote the true parameter values.

Assumption 9(a) strengthens Assumptions 4 and 8, enabling us to use the CLT and LLN to study (35). Assumption 9(b) requires that the score and $\nabla_{\theta_1} \cdots \nabla_{\theta_q} f_t(\beta_*, \delta_*) / f_t(\beta_*, \delta_*)$ are uncorrelated for $k = 3, 4$, which holds in the model in (8), regardless of which parameters are allowed to switch.

We first study the refined approximation under the following two drifting sequences of $(p_T, q_T)$:

**SEQ1**: $p_T = p + c_1 T^{-a_1}$ and $q_T = q + c_2 T^{-a_2}$ for some $(p, q) \neq (0.5, 0.5)$,

**SEQ2**: $p_T = 0.5 + c_1 T^{-a_1}$ and $q_T = 0.5 + c_2 T^{-a_2}$, (36)

where $(p, q)$ denotes a point in (34), $a_1, a_2 \geq 0$, and $c_1, c_2 \neq 0$. Let $a = \min(a_1, a_2)$. We assume $c_1 \neq -c_2$ when $a_1 = a_2$; otherwise, $\rho_T$ is zero for any $a$. Define

$$\tilde{\mathcal{V}}^{(2)}(p_T, q_T, \eta) = (\eta \otimes 2)' \left( T^{-1/2} \sum_{t=1}^T \tilde{U}_t^{(2)}(p_T, q_T) \right) - \frac{1}{4} (\eta \otimes 2)' \tilde{\Omega}(p_T, q_T)(\eta \otimes 2),$$

$$\tilde{\mathcal{V}}^{(3)}(p_T, q_T, \eta) = T^{-1/4} \frac{1}{3} (\eta \otimes 3)' \left( T^{-1/2} \sum_{t=1}^T \tilde{U}_t^{(3)}(p_T, q_T) \right) - T^{-1/2} \frac{1}{36} (\eta \otimes 3)' \tilde{\Omega}(p_T, q_T)(\eta \otimes 3),$$

$$\tilde{\mathcal{V}}^{(4)}(p_T, q_T, \eta) = T^{-1/2} \frac{1}{12} (\eta \otimes 4)' \left( T^{-1/2} \sum_{t=1}^T \tilde{U}_t^{(4)}(p_T, q_T) \right) - T^{-1} \frac{1}{576} (\eta \otimes 4)' \tilde{\Omega}(p_T, q_T)(\eta \otimes 4).$$

Let $S_b(p_T, q_T)$ denote the output from STEP 4 of the algorithm when $(p_i, q_i) = (p_T, q_T)$, that is, $S_b(p_T, q_T) = \sup_{\eta \in R^{n_\delta}} \sum_{j=2}^4 W_b^{(j)}(p_T, q_T, \eta)$; see (31)-(32).

**Proposition 2**

Suppose that the null hypothesis and Assumptions 1-9 hold, with (24) satisfied for all $i_1, i_2 \in \{1, \ldots, n_\delta\}$. Then:

1. Under SEQ1, if $\min_{\|x\otimes 2\|=1} \frac{(x \otimes 2)' \tilde{\mathcal{V}}(p_T, q_T)(x \otimes 2)}{p_T^2} > L$ and $\min_{\|x\otimes 3\|=1} \frac{(x \otimes 3)' \tilde{\mathcal{V}}(p_T, q_T)(x \otimes 3)}{(p_T - q_T)^2} > L$ for some $L > 0$ in probability, and $\sup_{\|\delta - \tilde{\delta}\|\ll 1} T^{-1} |\mathcal{L}_{i_1 \ldots i_T}(p_T, q_T, \delta)| = O_p(1)$ for some $c > 0$ and any $i_1, \ldots, i_T \in \{1, \ldots, n_\delta\}$, then $Pr(LR(p_T, q_T) \leq x) - Pr(S(p_T, q_T) \leq x) \rightarrow 0$ for $x \in R$, where

$$\tilde{S}(p_T, q_T) = \begin{cases} 
\sup_{\eta \in R^{n_\delta}} \tilde{\mathcal{V}}^{(2)}(p_T, q_T, \eta) & \text{if } a < 1/6 \\
\sup_{\eta \in R^{n_\delta}} \{\tilde{\mathcal{V}}^{(2)}(p_T, q_T, \eta) + \tilde{\mathcal{V}}^{(3)}(p_T, q_T, \eta)\} & \text{if } a = 1/6 \\
\sup_{\eta \in R^{n_\delta}} \tilde{\mathcal{V}}^{(3)}(p_T, q_T, \eta) & \text{if } a > 1/6
\end{cases}.$$
2. Under SEQ2, if \( \min_{\|x\|^2=1} \frac{(x)\tilde{\Omega}(p_T,q_T)(x)}{p_T^2} > L \) and \( \min_{\|x\|^2=1} \frac{(x)\tilde{\Omega}(p_T,q_T)(x)}{p_T^2} > L \) for some \( L > 0 \) in probability, and \( \sup_{\|\delta-\tilde{\delta}\|<c} T^{-1}\ell_{\tilde{\delta}i_1\ldots i_n}(p_T,q_T,\delta) = O_p(1) \) for some \( c > 0 \) and any \( i_1,\ldots,i_n \in \{1,\ldots,n_\delta\} \), then \( \Pr(LR(p_T,q_T) \leq x) - \Pr(S(p_T,q_T) \leq x) \to 0 \) for \( x \in R \), where

\[
\tilde{S}(p_T,q_T) = \begin{cases} 
\sup_{\eta \in R^n_\delta} \tilde{W}(2)(p_T,q_T,\eta) & \text{if } a < 1/4 \\
\sup_{\eta \in R^n_\delta} \{\tilde{W}(2)(p_T,q_T,\eta) + \tilde{W}(4)(p_T,q_T,\eta)\} & \text{if } a = 1/4 \\
\sup_{\eta \in R^n_\delta} \tilde{W}(4)(p_T,q_T,\eta) & \text{if } a > 1/4 
\end{cases} .
\]

3. The distribution of \( S_b(p_T,q_T) \) is a weakly consistent estimator of the limiting distribution of \( LR(p_T,q_T) \) under SEQ1 and SEQ2 if the above conditions in 2.1 and 2.2 are satisfied.

The first two results of Proposition 2 reveal how the distribution of \( LR(p_T,q_T) \) differs between SEQ1 and SEQ2, and how it changes with \( a \). The third result shows that the refined approximation is a consistent estimator of the limiting distribution of \( LR(p_T,q_T) \) under SEQ1 and SEQ2, for any \( a \geq 0 \). The consistency holds because the refined approximation encompasses all terms that potentially matter for the limiting distribution of \( LR(p_T,q_T) \) under SEQ1 and SEQ2. For example, although \( W_b(4)(p_T,q_T,\eta) \) is asymptotically negligible when \( (p_T,q_T) \) follows SEQ1, including it ensures consistency under SEQ2. In fact, if any \( W_b(k)(p_T,q_T,\eta) \) \( (k = 2,3,4) \) is removed, the resulting approximation will become inconsistent under SEQ1 or SEQ2 for some \( a > 0 \). The assumptions on \( \ell_{i_1\ldots i_k}(p_T,q_T,\delta) \) \( (k = 7,9) \) serve a similar purpose as Assumption 5. The condition on \( (x)\tilde{\Omega}(p_T,q_T)(x)\ell_T/\rho_T^2 \) is illustrated using a linear model in Lemma 24 of the appendix.

The next result allows \( (p_T,q_T) \) to follow sequences that are more general than (36).

**Corollary 2** Assume \( \rho_T \to 0 \) as \( T \to \infty \). Suppose that the conditions in Proposition 2 hold, and that \( E(\nabla_{\theta_{i_1}}\ldots\nabla_{\theta_{i_4}}f_1(\beta_*,\delta_*)\nabla_{\theta_{i_1}}\ldots\nabla_{\theta_{i_4}}f_1(\beta_*,\delta_*)|\Omega_{t-1}) = 0 \) for \( i_1,\ldots,i_4 \in \{n_\beta+1,\ldots,n_\beta+n_\delta\} \). Then, \( \Pr(LR(p_T,q_T) \leq x) - \Pr(\sup_{\eta \in R^n_\delta} \sum_{j=2}^4 \tilde{W}(j)(p_T,q_T,\eta) \leq x) \to 0 \) for any \( x \in R \).

Corollary 2 implies that the refined approximation is valid under general drifting sequences. Because \( (p_T,q_T) \) may not converge to any point, we do not obtain a consistency result, as in Proposition 2.3. The assumption on \( \nabla_{\theta_{i_1}}\ldots\nabla_{\theta_{i_4}}f_1(\beta_*,\delta_*) \) \( (k = 3,4) \) ensures that \( \ell_{i_1\ldots i_r}(p_T,q_T,\delta) = O_p(T^{-1/2} + |p_T| + |p_T - q_T|^2) \). This holds in the model in (8) if the variance is not allowed to switch.

In the above analysis, we have allowed the transition probabilities of a Markov switching model to approach those of a mixture model. The results show that the distribution of \( LR(p_T,q_T) \) is path dependent and that the number of terms needed for a consistent approximation to it varies accordingly. Because the refined approximation encompasses all terms that are potentially nonnegligible, it provides an adequate approximation over a wide range of transition probabilities.
6.2 Local power properties

We first derive a refined approximation to the distribution of $LR(p, q)$ under a simple DGP. Then, we apply it to study the power properties of the tests. By focusing on a simple DGP, we lose some generality, but we gain analytical results that clearly show the differences between the tests.

The DGP is

$$y_t = \mu_* + A_T 1_{\{s_t=2\}} + e_t,$$

where $e_t \sim i.i.d. N(0, \sigma_e^2)$; $A_T = c_* T^{-1/4}$; $\mu_*$, $\sigma_*^2$, and $c_*$ are independent of $T$; and the transition probabilities $p_*$ and $q_*$ are fixed and belong to $\Lambda_\circ$. The local alternatives are $O(T^{-1/4})$ (proved in Lemma ?? in the appendix), as in Carrasco, Hu, and Ploberger (2014).

We consider two estimated models in order to reflect the empirical literature and to quantify the effect of the model specification on the testing power:

$$(\text{Static Model}) \quad y_t = \delta_1 1_{\{s_t=1\}} + \delta_2 1_{\{s_t=2\}} + u_t,$$

$$(\text{Dynamic Model}) \quad y_t = \delta_1 1_{\{s_t=1\}} + \delta_2 1_{\{s_t=2\}} + \alpha y_{t-1} + u_t,$$

where $u_t \sim N(0, \sigma^2)$; and $\delta_1, \delta_2, \alpha$, and $\sigma^2$ are unknown parameters. The static model is considered in Cecchetti, Lam, and Mark (1990), Chauvet and Hamilton (2006), and Hamilton (2016); and the dynamic model is considered in Davig (2004), Hansen (1992), and Cho and White (2007). To simplify the analysis and focus on the main issue, we assume $(p_*, q_*)$ is known.

To present the results, let $p_*=p_*+q_*-1$, and define the following noncentrality parameters:

$$A_{2s}(p_*, q_*) = \frac{2c_*^2}{\sigma_*^4} \left( \frac{1-p_*}{1-q_*} \right) \sum_{j=1}^{\infty} p_*^j \text{Cov} \left( 1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}} \right),$$

$$A_{2d}(p_*, q_*) = \frac{2c_*^2}{\sigma_*^4} \left( \frac{1-p_*}{1-q_*} \right) \sum_{j=2}^{\infty} p_*^j \text{Cov} \left( 1_{\{s_t=2\}}, 1_{\{s_{t-j}=2\}} \right),$$

$$A_3(p_*, q_*) = \frac{c_*^3}{3\sigma_*^6} \left( \frac{1-p_*}{1-q_*} \right) \frac{(p_* - q_*)}{(1-q_*)^2} E \left( 1_{\{s_t=2\}} - E 1_{\{s_t=2\}} \right)^3,$$

$$A_4(p_*, q_*) = \frac{c_*^4}{12\sigma_*^8} \left( \frac{1-p_*}{2-p_*-q_*} \right) \left( 1 + \left( \frac{1-p_*}{1-q_*} \right)^3 \right)^{-3} \left( \frac{1-p_*}{1-q_*} \right)^2 \left( 1 - E \left( E 1_{\{s_t=2\}} - E 1_{\{s_t=2\}} \right)^4 - 3E \left( 1_{\{s_t=2\}} - E 1_{\{s_t=2\}} \right)^2 \right).$$

The above quantities depend only on $(p_*, q_*)$ and $c_*/\sigma_*^2$. Their analytical expressions are given in Lemma ???. Let $\sup_{\eta \in \mathcal{R}} \left( \sum_{j=2}^{4} \mathcal{W}_j^{(j)}(p_*, q_*, \eta) \right)$ denote the refined approximation to $LR(p_*, q_*)$ under the null hypothesis, i.e., with $c_* = 0$; see (27) and its simulated version in (32).
Proposition 3 Under DGP (38) and Assumption 5, \( \text{Pr}(LR(p_\ast, q_\ast) \leq s) - \text{Pr}(S(p_\ast, q_\ast) \leq s) \to 0 \) for any \( s \in R \), where \( S(p_\ast, q_\ast) \) equals

\[
\begin{align*}
\sup_{\eta \in R} \{ & \sum_{j=2}^{4} W^{(j)}(p_\ast, q_\ast, \eta) + A_{2s}(p_\ast, q_\ast)\eta^2 + T^{-1/2}A_3(p_\ast, q_\ast)\eta^3 + T^{-1}A_4(p_\ast, q_\ast)\eta^4 \}, \\
\sup_{\eta \in R} \{ & \sum_{j=2}^{4} W^{(j)}(p_\ast, q_\ast, \eta) + A_{2d}(p_\ast, q_\ast)\eta^2 + T^{-1/2}A_3(p_\ast, q_\ast)\eta^3 + T^{-1}A_4(p_\ast, q_\ast)\eta^4 \},
\end{align*}
\]

in the static and dynamic cases, respectively.

Proposition 3 shows that, under (38)-(39), regime switching affects the local power of \( LR(p_\ast, q_\ast) \) through three channels: (i) serial dependence, measured by \( A_{2s}(p_\ast, q_\ast) \) or \( A_{2d}(p_\ast, q_\ast) \); (ii) asymmetry, measured by \( T^{-1/2}A_3(p_\ast, q_\ast) \); and (iii) tail behavior, measured by \( T^{-1}A_4(p_\ast, q_\ast) \). Although (ii) and (iii) appear as high order terms in the refined approximation, their values can exceed (i) in finite samples. For example, \( T^{-1/2}A_3(p_\ast, q_\ast) \) is greater than \( A_{2s}(p_\ast, q_\ast) \) and \( A_{2d}(p_\ast, q_\ast) \) when \((T, c_\ast, \sigma_\ast, p_\ast, q_\ast) = (500, 6, 1, 0.9, 0.2)\), and it remains greater than \( A_{2d}(p_\ast, q_\ast) \) after \( q_\ast \) increases to 0.4. Therefore, neglecting (ii) or (iii) can lead to a poor, even misleading representation of the power properties of the likelihood ratio test. Proposition 3 also shows that, in the dynamic case, the first order dependence caused by the regime switching no longer helps the power. This result explains why detecting regime switching in the mean is more difficult when allowing for linear dynamics.

The QLR test of Cho and White (2007) captures the mixture properties, but not the serial correlation caused by the regime switching. When the mixing probability \( \pi \) is set to \( \xi_\ast = (1 - q_\ast)/(2 - p_\ast - q_\ast) \), a refined approximation to its distribution is given by

\[
\begin{align*}
\sup_{\eta \in R} \left\{ W^{(3)}(p_\ast, q_\ast, \eta) + W^{(4)}(p_\ast, q_\ast, \eta) + T^{-1/2}A_3(p_\ast, q_\ast)\eta^3 + T^{-1}A_4(p_\ast, q_\ast)\eta^4 \right\}.
\end{align*}
\]

The \( TS(\rho_\ast) \) test of Carrasco, Hu, and Ploberger (2014) captures the serial correlation, but not the mixture properties. From their analysis, it follows that \( 2TS(\rho_\ast) \) converges to

\[
\begin{align*}
\sup_{\eta \in R} \left\{ W^{(2)}(p_\ast, q_\ast, \eta) + A_{2s}(p_\ast, q_\ast)\eta^2 \right\} \quad \text{and} \quad \sup_{\eta \in R} \left\{ W^{(2)}(p_\ast, q_\ast, \eta) + A_{2d}(p_\ast, q_\ast)\eta^2 \right\}
\end{align*}
\]

in the static and dynamic cases, respectively. Therefore, under (38)-(39), (i) does not contribute to the local power of the QLR test, and (ii) and (iii) do not affect the \( TS(\rho_\ast) \) test.

We now use simulations to quantify these differences. We set \((T, \sigma_\ast, p_\ast) = (500, 1, 0.9)\) and \((q_\ast, c_\ast) = (0.2, 7.3), (0.4, 6.4), (0.6, 4.7), (0.8, 3.0)\), where the values of \( c_\ast \) are chosen such that the power of \( LR(p_\ast, q_\ast) \) is about 50% in the static case at the 5% nominal level. For \( c_\ast = 7.3 \), the change between the regimes is 1.54, while for \( c_\ast = 3.0 \), it is 0.63. As the standard deviation is 1.0, these values imply that the change needs to be big for the test to have good power and that the power
increases with the persistence of the regimes. Table 1 reports the rejection frequencies computed using (41), (42), and (43). The distributions that produce these values are reported in Figures S2-S9, along with the finite sample distributions to reflect the adequacy of the approximations.

In the static case (Panel (a) of Table 1), the power of $LR(p_*, q_*)$ is substantially higher than that of $QLR(\pi_*)$ when $q_*=0.6,0.8$, and substantially higher than that of $TS(\rho_*)$ when $q_*=0.2,0.4$, where the differences reach 47.36% and 44.90%, respectively. This confirms that serial dependence is important for power when the regimes are persistent, and that the asymmetry and tail behavior are important otherwise. The dynamic case (Panel (b)) shows a similar pattern, where the differences reach 31.31% and 48.71%, respectively. Figures S2-S9 show that, overall, the approximations are close to their finite sample distributions. The approximation improves further when $T$ is increased to 1000; see Figures S10-S17.

Proposition 3 implies that the local power difference between $LR(p_*, q_*)$ and $TS(\rho_*)$ should decrease as the sample size increases. To examine this further, we repeat the above analysis using $T=1000,5000,20000$, and 50000. We focus on $(q_*, c_*)=(0.2,7.3)$, because in this case, the power difference is the largest. For the static model, the power differences at the 5% level computed using (41) and (43) for the four sample sizes are 35.58%, 17.75%, 7.35%, and 3.12%, respectively. Although the value decreases, the rate is slow, and it remains significant for sample sizes that are enormous from an empirical perspective. The pattern is similar for the dynamic model, with the rate of the decrease being even slower, where the corresponding values are 41.60%, 21.90%, 12.24%, and 7.02%, respectively. The results also confirm that the refined approximations accurately represent their finite sample distributions in all cases; see Figures S18-S21. Therefore, the asymptotic of the LR test kicks in only in very large samples, and for many applications, the finite sample correction term $T^{-1/2}A_3(p_*, q_*)\eta^3 + T^{-1}A_4(p_*, q_*)\eta^4$ is necessary.

In summary, we have examined how regime switching affects the local power of the likelihood ratio test by altering the serial dependence, symmetry, and tail behavior of a time series. The results show that these three channels are all potentially important for power. The tests of Cho and White (2007) and Carrasco, Hu, and Ploberger (2014) turn off some channels and, as a result, their power can be lower than that achievable.

7 Implications for bootstrap procedures and information criteria

The results in the previous sections can be used to evaluate the consistency, or the lack thereof, of various bootstrap procedures. We illustrate some important aspects using the linear model in (8).
Bootstrap procedures. We begin with the important special case where the model specifies a stationary AR(p) process with normal errors. A standard parametric bootstrap procedure is as follows. (1) Estimate the model under the null hypothesis. (2) Sample from a normal distribution with mean zero and variance equal to the sample variance of the residuals. Use the sampled values and the estimated coefficients to generate a new AR(p) series. (3) Compute the test using this series. (4) Repeat steps (1)-(3). This procedure preserves the normality of the errors and the autoregressive structure. The covariance function in the bootstrap world is thus in agreement with that in Proposition 1. As a result, the procedure is asymptotically valid.

Next, we consider the more general situation where a second variable is present among the regressors; for example, an autoregressive distributed lags (ADL) model. Because this model does not specify the joint distribution of the dependent variable and the regressors, the bootstrap procedure described above is no longer applicable. Two alternative approaches deserve consideration. The first approach involves keeping the regressors fixed at their original values when generating the data; that is, we use the fixed regressor bootstrap. This procedure is asymptotically valid in the context of testing for structural breaks (Hansen, 2000). However, in the current context, it is, in general, inconsistent. In contrast to the original DGP, the regressors are strictly, but not weakly exogenous in the bootstrap world. As a result, the covariance function in the bootstrap world differs from that in Proposition 1. We now illustrate the potential severity of the size distortion using the setting in (22) with \( T = 250 \). The finite sample distribution and the bootstrap distribution for testing regime switching in the intercept are reported in Figure S22. Using the critical values from the fixed regressor bootstrap, the rejection rates at the 10% and 5% levels are 21.8% and 10.0%, respectively. The overrejection does not decrease when the sample size is increased to 500.

The second approach involves specifying the joint distribution of the data. For example, if we have an ADL model with normal errors, we specify a full model that is a Gaussian vector autoregression and apply the parametric bootstrap to the augmented model. This bootstrap procedure is consistent if it reproduces the covariance function in Proposition 1 asymptotically.

Information criteria. The asymptotic results imply that the performance of conventional information criteria, such as the BIC, can be sensitive to the structure of the model and to the choice of which parameters are allowed to switch. This is because the distribution of the likelihood ratio depends on which parameter is allowed to switch, whereas the penalty term in the BIC depends only on the dimension of the model and the sample size. We illustrate this sensitivity using the model in (22) for two cases. In the first case, we apply the BIC to determine whether there is regime
switching in the intercept. The second case is the same as the first, except that the slope parameter is allowed to switch instead. In the simulated data, no regime switching is present; $\mu = 0, \alpha = 0.5,$ and $\sigma^2 = 1$. The set $\Lambda_\epsilon$ is specified as (11) with $\epsilon = 0.05$. The sample size is 250. Of the 5000 realizations, the BIC falsely classifies 12.5% in the first case, while only 2.4% in the second case. Because the penalty terms in the Akaike information criterion and the Hannan–Quinn information criterion have the same structure, they are expected to exhibit the same sensitivity.

8 Monte Carlo

We examine the finite sample properties of the $\text{SupLR}(\Lambda_\epsilon)$ test, and compare these properties with those of the tests of Cho and White (2007) and Carrasco, Hu and Ploberger (2014). The DGP is

$$y_t = \mu_1 \cdot 1_{s_t=1} + \mu_2 \cdot 1_{s_t=2} + \alpha y_{t-1} + e_t \quad \text{with} \quad e_t \sim i.i.d. N(0, \sigma^2),$$

where $P(s_t = 1|s_{t-1} = 1) = p, P(s_t = 2|s_{t-1} = 2) = q, \alpha = 0.5,$ and $\sigma^2 = 1$. This DGP is considered in Cho and White (2007), and it is a sensible approximation to the postwar U.S. quarterly real GDP growth series, as shown in Section 9. Throughout this section, $\Lambda_\epsilon$ is given by (10) with $\epsilon = 0.05, 0.02$. For the $\text{supTS}$ of Carrasco, Hu, and Ploberger (2014), the supremum is taken over $\rho \in [0.05, 0.90]$ or $\rho \in [0.02, 0.96]$. Because $\rho = p + q - 1$, these two sets for $\rho$ are consistent with $\Lambda_{0.05}$ and $\Lambda_{0.02}$, specified above. The resulting tests are denoted by $\text{supTS}_1$ and $\text{supTS}_2$, respectively. The rejection frequencies are based on 5000 replications.

The results under the null hypothesis are reported in Table 2. The rejection frequencies of $\text{SupLR}(\Lambda_\epsilon)$ are overall close to the nominal levels, although some mild over-rejections do exist. When $T = 200$, the rejection rates at the 5% and 10% levels are 6.78% and 14.36% for $\epsilon = 0.05$, and 6.86% and 14.58% for $\epsilon = 0.02$. Similar rejection rates are observed when $T = 500$. The results confirm that the QLR and supTS tests exhibit excellent size properties.

For power properties, following Cho and White (2007), we let $\mu_1 = -\mu_2$ with $\mu_2 = 0.2, 0.6, 1.0$. Motivated by the estimates discussed in Section 3, we set $(p, q)$ to $(0.70, 0.70)$, $(0.70, 0.90)$, and $(0.90, 0.90)$. The rejection rates at the 5% nominal level are reported in Table 3.

Because the alternatives are not mixtures, the power of the $\text{SupLR}(\Lambda_\epsilon)$ test is higher than that of the QLR. For example, when $(p, q) = (0.7, 0.7)$, the rejection rates of the $\text{SupLR}(\Lambda_{0.05})$ test are 18.38% and 96.38% for $\mu_2 = 0.6$ and 1.0, and the corresponding values of the QLR are 9.46% and 68.83%. When $(p, q) = (0.9, 0.9)$, the values become 59.26% and 100% for $\text{SupLR}(\Lambda_{0.05})$, and 7.06% and 7.30% for the QLR. Therefore, although the QLR test can be valuable for detecting mixtures, the $\text{SupLR}(\Lambda_\epsilon)$ test can offer substantial power gains when the regimes are dependent.
The power of $\text{SupLR}(\Lambda_\epsilon)$ is substantially higher than that of the $\text{supTS}$ test. In addition to the explanation in Subsection 6.2, the power difference also arises from the following channel. Note that a key element of $\text{supTS}$ is $\mu_{2,t}(\rho) = (1/(2\sigma^4)) \sum_{s < t} \rho^{t-s} \tilde{e}_t \tilde{e}_s$, which measures the serial correlation in the residuals ($\tilde{e}_t$) computed under the null hypothesis. When the parameters are estimated under the null hypothesis, the regime switching is removed from the model and forced into the residuals, which causes $\tilde{e}_t$ to be positively serially correlated because $p + q - 1 > 0$. At the same time, the autoregressive coefficient $\tilde{\alpha}$ is upward biased, and the bias is stronger when the data are more persistent. The bias in $\tilde{\alpha}$ leads to overdifferencing the data and, consequently, making $\tilde{e}_t$ negatively serially correlated. In finite samples, these two opposite effects can offset each other, making the value of $\mu_{2,t}(\rho)$ insensitive to the departure from the null hypothesis. (A similar phenomenon is studied in Perron (1990, 1991) in a structural change context.) This finding is consistent with the simulation results in Carrasco, Hu, and Ploberger (2014, Table II), which show that the test can have good power properties when a lagged dependent variable is not allowed in the model.

Next, we examine the situation where the DGP is a mixture model with $(p, q) = (0.5, 0.5)$. The results (the last five rows of Table 3) show that the power of the QLR test is higher than that of the $\text{SupLR}(\Lambda_\epsilon)$ test. However, the maximum difference is only 8.86% for the cases considered.

9 Applications

We first study the quarterly US real GDP growth rate, and then consider additional applications in the context of dynamic stochastic equilibrium models. The model in (44) is used for the analysis. The samples contain quarterly observations for the period 1960:I–2014:IV, unless it is stated otherwise. All tests are evaluated at the 5% nominal level. The relevant p-values are also reported.

9.1 US GDP growth

Following the influential work of Hamilton (1989), a large body of literature has modeled US real output growth as a regime switching process. Here, we apply the $\text{SupLR}(\Lambda_\epsilon)$ test to assess the empirical evidence for this specification.

**Testing results.** When applied to the full sample, the $\text{SupLR}(\Lambda_{0.02})$ test is equal to 8.75, and the p-value is 0.033 (Table 4), implying that the null hypothesis is rejected at the 5% level. To exclude the influence of the Great Recession, we consider the subsample 1960:I–2006:IV. Now, the test is equal to 8.57, and the p-value is 0.035, implying the same conclusion. To take the analysis further, we use 1960:I–1980:I as the first subsample and incorporate observations quarter by quarter. This
leads to 140 subsamples of increasing sizes; see Figure 2(a), in which the null hypothesis is rejected in 102 subsamples. Therefore, there is consistent evidence favoring the regime switching specification. The results of the QLR and supTS tests over the same subsamples are shown in Figures 2(b)-(c). The null hypothesis is rejected only after the Great Recession is included. Therefore, the empirical evidence for regime switching is substantially weaker when viewed through these two tests.

Recession probability. Figures 3(a) and 3(b) display the smoothed recession probabilities when the model in (44) is applied to 1960:I–2006:IV and 1960:I–2014:IV, respectively. The NBER business cycle indicators are also included, with the shaded areas corresponding to recessions. The results show that the model is informative. For the subsample, the recession probabilities implied by the model closely track the NBER’s recession indicators. For the full sample, they are broadly similar, with the main difference being that the model assigns low recession probabilities to the relatively shallow recessions of 1969:IV-1970:IV and 2001:I-2001:IV. This difference arises because the estimated mean growth rate in recessions decreases from $-0.18$ to $-0.67$ when the Great Recession is included. Therefore, it reflects the unusual nature of the Great Recession. The parameter estimates are reported in Table 4.

Robustness check. In practice, the order of the autoregression under the null hypothesis is often determined by some information criterion. To reflect this practice, we estimate the lag order using the BIC for each subsample, and repeat the analysis. The minimum and maximum orders are set to 1 and 4, respectively. The null hypothesis is rejected for 92 of the 140 subsamples. Therefore, the evidence of regime switching remains considerable. We also repeat the analysis using reverse recursive subsamples. We let 1994:IV–2014:IV be the first subsample, and then incorporate additional observations backward quarter by quarter. The lag order is determined using the BIC for each subsample. The null hypothesis is rejected in 120 of the 140 subsamples. Finally, we exclude the Great Recession (i.e., we let 1986:IV-2006:IV be the first subsample), and then incorporate additional observations backward quarter by quarter. The null hypothesis is rejected in 47 out of the 108 subsamples. Therefore, although the evidence is weaker in this case, it remains considerable and fairly consistent across the subsamples.

Calibrated simulations. We generate data from the model in (44), using the empirical estimates in Table 4. The sample sizes are set to those implied by the subsample (1960:I–2006:IV) and the full sample (1960:I–2014:IV), respectively. Under the null hypothesis, in the subsample case, the rejection rates at the 5% level are 6.86% for SupLR($A_{0.05}$) and 7.16% for SupLR($A_{0.02}$), respectively,
while in the full sample case, they are 6.90% and 7.14%. These values are consistent with those in Table 2. Under the alternative hypothesis, at the 5% level, the rejection rates of SupLR($\Lambda_{0.02}$) are 66% and 65% in the two cases, respectively. In contrast, the rejection rates of the QLR test are 14% and 25%, and that of the supTS$_2$ test are 24% and 10%. The power differences are substantial.

We also compute the p-value of the full-sample SupLR($\Lambda_{0.02}$) test using the parametric bootstrap in Section 7. The resulting value is 0.041, slightly above the value 0.033 reported in Table 4.

9.2 Other applications

Here, we consider three sets of applications in the context of dynamic stochastic equilibrium models. A description of the relevant data sets can be found in the footnote of Table 5.

**Hours worked and capital utilization.** Regime switching in real output growth has testable implications when viewed through the lens of medium scale DSGE models. In such models, the real output is usually modeled as a Cobb-Douglas function of capital stock $K_t$, capital utilization $U_t$, hours worked $H_t$, and some exogenous productivity process $Z_t$, that is, $Y_t = Z_t^{1-\alpha}(U_tK_{t-1})^{\alpha}H_t^{\beta}$; see Smets and Wouters (2007) and Schmitt-Grohé and Uribe (2012). In terms of growth rates (i.e., $y_t = \log(Y_t/Y_{t-1})$), this implies $y_t = \alpha (u_t + k_{t-1}) + \beta h_t + (1 - \alpha)z_t$. Because of the linearity, at least one endogenous variable among $u_t, k_{t-1},$ and $h_t$ must show regime switching. Otherwise, we arrive at a contradiction, implying that the production function is misspecified for the data under consideration, or the regime switching conclusion with regard to $y_t$ should be revisited.

Motivated by this observation, we examine the regime switching hypothesis for $u_t$ and $h_t$. The $k_t$ series is not considered, because it is difficult to measure, and its official data are available only at the annual frequency. The SupLR($\Lambda_{0.02}$) test rejects the null hypothesis in both cases (see Table 5). The smoothed recession probabilities, see Figures S23-S24, are consistent with those in Figure 3 based on the GDP series. Overall, the results show internal consistency, and they support using the above production function as a building block of DSGE models.

**Unemployment.** The aggregate unemployment is a focal point of business cycle analysis. The empirical findings obtained thus far, particularly those related to hours worked, suggest that the dynamics of this variable may also exhibit regime switching. Table 5 shows that the null hypothesis is indeed rejected by the SupLR($\Lambda_{0.02}$) test. Furthermore, the estimated parameter values reveal an important asymmetry: the unemployment rate rises sharply at 1.03% per quarter in recessions, while it decreases slowly at only 0.10% per quarter in expansions. The asymmetry becomes invisible
under the linear model in (44). Interestingly, the recession probabilities computed based on the
change in the unemployment rate correlate closely with the NBER’s indicators (Figure S25). There-
therefore, although the level of unemployment is a lagging indicator of the business cycle, the change of
this variable should be viewed as a coincident indicator. This appears to be a new finding.

Consumption. Some dynamic stochastic equilibrium models have modeled the mean growth
rate of aggregate consumption as a regime switching process in order to reproduce the observed
properties of asset returns, including those related to the riskless rate and the equity premium;
see Cecchetti, Lam, and Mark (1990), Kandel and Stambaugh (1991), and Ju and Miao (2012).
However, partly because of the technical difficulty, formal statistical testing of this regime switching
hypothesis is rare, particularly after allowing for linear dynamics. After applying the SupLR(0.02)
test, we find that the null hypothesis of no regime switching is rejected for the nondurable consump-
tion expenditure series; see row D of Table 5. Also, the estimated recession probabilities (Figure
S26) are consistent with those based on the other series. These findings provide formal statistical
support for allowing for regime switching in aggregate consumption in these models.

Finally, we apply the supTS2 and QLR tests to the same four time series. The SupTS2 does
not reject any null hypothesis. Its values in the four cases are (critical values in parentheses): 1.28
(2.65), 2.22 (2.64), 2.07 (2.64), and 2.02 (2.34). The QLR test rejects the null hypothesis only for
the capacity utilization and the unemployment series. Its values are: 4.84 (6.18), 19.00 (6.34), 24.11
(6.28), and 6.06 (6.11). Therefore, similarly to the GDP case, the evidence of regime switching is
weaker when viewed through these two tests.

10 Conclusion

We have examined a family of likelihood ratio based tests for detecting Markov regime switching.
In addition to obtaining the limiting distribution under the null hypothesis and a finite sample
refinement, thus resolving a long standing problem in the literature, we provide a unified algorithm
for simulating the relevant critical values. Working with a simple DGP, we show analytically
why these tests can be more powerful than some other tests that are based on alternative testing
principles. When applied to the US real GDP growth data and four other time series in the context
of dynamic stochastic equilibrium models, the proposed methods detect consistent evidence favoring
the regime switching specification. We conjecture that the techniques and results presented here
can have implications for hypothesis testing in other contexts, such as testing for Markov switching
in state space models and multivariate regressions. Such investigations are currently in progress.
Data availability.

The code underlying this article is available on Zenodo, at https://dx.doi.org/10.5281/zenodo.3908270.
References


Table 1: Local power under a simple DGP

<table>
<thead>
<tr>
<th></th>
<th>Static model</th>
<th>Dynamic model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a)</td>
<td>(b)</td>
</tr>
<tr>
<td>$(p_<em>, q_</em>, c_*)$</td>
<td>(0.9, 0.2, 7.3)</td>
<td>(0.9, 0.2, 7.3)</td>
</tr>
<tr>
<td></td>
<td>(0.9, 0.4, 6.4)</td>
<td>(0.9, 0.4, 6.4)</td>
</tr>
<tr>
<td></td>
<td>(0.9, 0.6, 4.7)</td>
<td>(0.9, 0.6, 4.7)</td>
</tr>
<tr>
<td></td>
<td>(0.9, 0.8, 3.0)</td>
<td>(0.9, 0.8, 3.0)</td>
</tr>
<tr>
<td>LR</td>
<td>55.14</td>
<td>54.44</td>
</tr>
<tr>
<td>QLR</td>
<td>53.69</td>
<td>54.57</td>
</tr>
<tr>
<td>TS</td>
<td>10.24</td>
<td>5.73</td>
</tr>
</tbody>
</table>

Note. LR: the likelihood ratio test; QLR: Cho and White’s (2007) test; TS: Carrasco, Hu, and Ploberger’s (2014) test. The transition probabilities are set to $(p_*, q_*)$ for all three tests to ensure a fair comparison. For the LR test, the refined approximation under the null hypothesis is obtained using the algorithm in Subsection 5.3. The refined approximation under the alternative hypothesis is computed in the same way, except that the noncentrality parameters $A_{2s}(p_*, q_*)$ (or $A_{2d}(p_*, q_*)$), $T^{-1/4}A_3(p_*, q_*)$, and $T^{-1/2}A_4(p_*, q_*)$ are added to the Gaussian random vectors $G_b(p_*, q_*)$, $G_b^{(3)}(p_*, q_*)$, $G_b^{(4)}(p_*, q_*)$ before the optimization. The sample size is 500. The critical values and rejection frequencies are based on 10,000 realizations.

Table 2: Rejection frequencies under the null hypothesis

<table>
<thead>
<tr>
<th>Nominal size</th>
<th>2.50</th>
<th>5.00</th>
<th>7.50</th>
<th>10.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T=200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SupLR($\Lambda_{0.05}$)</td>
<td>3.42</td>
<td>6.78</td>
<td>10.44</td>
<td>14.36</td>
</tr>
<tr>
<td>SupLR($\Lambda_{0.02}$)</td>
<td>3.50</td>
<td>6.86</td>
<td>10.62</td>
<td>14.58</td>
</tr>
<tr>
<td>QLR</td>
<td>2.43</td>
<td>5.30</td>
<td>7.50</td>
<td>10.00</td>
</tr>
<tr>
<td>supTS$_1$</td>
<td>2.86</td>
<td>5.06</td>
<td>7.58</td>
<td>10.14</td>
</tr>
<tr>
<td>supTS$_2$</td>
<td>2.84</td>
<td>5.06</td>
<td>7.62</td>
<td>10.06</td>
</tr>
<tr>
<td>$T=500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SupLR($\Lambda_{0.05}$)</td>
<td>3.50</td>
<td>7.04</td>
<td>10.36</td>
<td>14.04</td>
</tr>
<tr>
<td>SupLR($\Lambda_{0.02}$)</td>
<td>3.46</td>
<td>7.08</td>
<td>10.46</td>
<td>14.58</td>
</tr>
<tr>
<td>QLR</td>
<td>2.33</td>
<td>5.43</td>
<td>7.53</td>
<td>10.20</td>
</tr>
<tr>
<td>supTS$_1$</td>
<td>3.04</td>
<td>5.80</td>
<td>8.06</td>
<td>10.86</td>
</tr>
<tr>
<td>supTS$_2$</td>
<td>3.02</td>
<td>5.72</td>
<td>8.12</td>
<td>10.70</td>
</tr>
</tbody>
</table>

Note. The values corresponding to the QLR test are taken from Table II in Cho and White (2007). The values corresponding to the supTS tests are obtained using the accompanying code of Carrasco, Hu and Ploberger (2014). The number of replications: 5000. In each replication, the critical values of the SupLR test are computed using the algorithm in Subsection 5.3 with 199 realizations.
<table>
<thead>
<tr>
<th>$(p, q)$</th>
<th>$\mu_2 = 0.20$</th>
<th>$\mu_2 = 0.60$</th>
<th>$\mu_2 = 1.00$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.70, 0.70)$</td>
<td>SupLR($\Lambda_{0.05}$) 6.38 18.38 96.38</td>
<td>SupLR($\Lambda_{0.02}$) 6.60 17.48 95.62</td>
<td>QLR 6.16 9.46 68.83</td>
</tr>
<tr>
<td></td>
<td>supTS$_1$ 5.70 10.22 32.98</td>
<td>supTS$_2$ 5.50 10.10 32.70</td>
<td></td>
</tr>
<tr>
<td>$(0.70, 0.90)$</td>
<td>SupLR($\Lambda_{0.05}$) 6.80 36.58 99.72</td>
<td>SupLR($\Lambda_{0.02}$) 7.28 35.16 99.72</td>
<td>QLR 6.14 13.40 60.56</td>
</tr>
<tr>
<td></td>
<td>supTS$_1$ 4.84 5.54 18.72</td>
<td>supTS$_2$ 4.82 5.46 18.46</td>
<td></td>
</tr>
<tr>
<td>$(0.90, 0.90)$</td>
<td>SupLR($\Lambda_{0.05}$) 8.56 59.26 100.00</td>
<td>SupLR($\Lambda_{0.02}$) 8.96 57.54 100.00</td>
<td>QLR 5.76 7.06 7.30</td>
</tr>
<tr>
<td></td>
<td>supTS$_1$ 6.62 12.12 4.84</td>
<td>supTS$_2$ 6.56 12.16 4.84</td>
<td></td>
</tr>
<tr>
<td>$(0.50, 0.50)$</td>
<td>SupLR($\Lambda_{0.05}$) 6.84 10.54 76.14</td>
<td>SupLR($\Lambda_{0.02}$) 7.16 10.20 76.78</td>
<td>QLR 6.03 11.33 85.10</td>
</tr>
<tr>
<td></td>
<td>supTS$_1$ 4.90 5.52 5.68</td>
<td>supTS$_2$ 4.88 5.46 5.50</td>
<td></td>
</tr>
</tbody>
</table>

Note. The values corresponding to the QLR test are taken from Table III in Cho and White (2007). Note that there the values in the rows of 0.1 and 0.9 in their table should be exchanged. The values related to the supTS tests are obtained using the accompanying code of Carrasco, Hu and Ploberger (2014). The number of replications: 5000. In each replication, the critical values of the SupLR test are computed using the algorithm in Subsection 5.3 with 199 realizations. Nominal level: 5%. Sample size: 500.
### Table 4: Results for the GDP growth series

<table>
<thead>
<tr>
<th>Tests</th>
<th>SupLR($\Lambda_{0.02}$)</th>
<th>5% critical value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960:I–2014:IV</td>
<td>8.75</td>
<td>7.62</td>
<td>0.033</td>
</tr>
<tr>
<td>1960:I–2006:IV</td>
<td>8.57</td>
<td>7.61</td>
<td>0.035</td>
</tr>
</tbody>
</table>

Estimates under $H_0$

<table>
<thead>
<tr>
<th>Series</th>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960:I–2014:IV</td>
<td>0.51</td>
<td>0.33</td>
<td>0.64</td>
</tr>
<tr>
<td>1960:I–2006:IV</td>
<td>0.60</td>
<td>0.28</td>
<td>0.65</td>
</tr>
</tbody>
</table>

Estimates under $H_1$

<table>
<thead>
<tr>
<th>Series</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\alpha$</th>
<th>$\sigma^2$</th>
<th>$p$</th>
<th>$q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1960:I–2014:IV</td>
<td>-0.54</td>
<td>0.75</td>
<td>0.19</td>
<td>0.49</td>
<td>0.66</td>
<td>0.96</td>
</tr>
<tr>
<td>1960:I–2006:IV</td>
<td>-0.16</td>
<td>0.97</td>
<td>0.09</td>
<td>0.48</td>
<td>0.77</td>
<td>0.94</td>
</tr>
</tbody>
</table>

Note. The data series is GDPC1, retrieved from the St. Louis Fed website. The critical values of the SupLR test are obtained using the algorithm in Subsection 5.3 with 5000 realizations.

### Table 5: Results for the other applications

<table>
<thead>
<tr>
<th>Tests</th>
<th>SupLR($\Lambda_{0.02}$)</th>
<th>5% critical value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series A</td>
<td>12.86</td>
<td>7.80</td>
<td>0.009</td>
</tr>
<tr>
<td>Series B</td>
<td>29.95</td>
<td>7.15</td>
<td>0.000</td>
</tr>
<tr>
<td>Series C</td>
<td>32.12</td>
<td>7.96</td>
<td>0.000</td>
</tr>
<tr>
<td>Series D</td>
<td>10.08</td>
<td>7.56</td>
<td>0.022</td>
</tr>
</tbody>
</table>

Estimates under $H_0$

<table>
<thead>
<tr>
<th>Series</th>
<th>$\mu$</th>
<th>$\alpha$</th>
<th>$\sigma^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series A</td>
<td>0.12</td>
<td>0.63</td>
<td>0.40</td>
</tr>
<tr>
<td>Series B</td>
<td>-0.01</td>
<td>0.58</td>
<td>1.54</td>
</tr>
<tr>
<td>Series C</td>
<td>0.00</td>
<td>0.65</td>
<td>0.07</td>
</tr>
<tr>
<td>Series D</td>
<td>0.27</td>
<td>0.20</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Estimates under $H_1$

<table>
<thead>
<tr>
<th>Series</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\alpha$</th>
<th>$\sigma^2$</th>
<th>$p$</th>
<th>$q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Series A</td>
<td>-1.31</td>
<td>0.21</td>
<td>0.52</td>
<td>0.32</td>
<td>0.51</td>
<td>0.98</td>
</tr>
<tr>
<td>Series B</td>
<td>-3.66</td>
<td>0.13</td>
<td>0.45</td>
<td>1.05</td>
<td>0.43</td>
<td>0.98</td>
</tr>
<tr>
<td>Series C</td>
<td>0.51</td>
<td>-0.06</td>
<td>0.44</td>
<td>0.04</td>
<td>0.68</td>
<td>0.96</td>
</tr>
<tr>
<td>Series D</td>
<td>-0.55</td>
<td>0.44</td>
<td>0.03</td>
<td>0.41</td>
<td>0.80</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Note. Series A: hours worked (HOANBS), percent change, 1960:I-2014:IV. Series B: capacity utilization (TCU), percent change, 1967:IV-2014:IV. This sample is shorter because of data availability. Series C: unemployment (UNRATE), change, percent, 1960:I-2014:IV. Series D: Consumption (A796RX0Q048SBEA), percent change, 1960:I-2014:IV. All data series are at the quarterly frequency and are retrieved from the St. Louis Fed website. The critical values of the SupLR test are obtained using the algorithm in Subsection 5.3 with 5000 realizations.
Note. The figure displays the finite sample distributions of the SupLR(Λₜ) test and their approximations for detecting regime switching in an AR(1) model: \( y_t = \mu + \alpha y_{t-1} + u_t \) with \( u_t \sim i.i.d.N(0, \sigma^2) \), where \( \mu = 0, \alpha = 0.5, \sigma^2 = 1 \), and \( T = 250 \). The original approximation and the refined approximation are given in Proposition 1 and Corollary 1, respectively. All results are based on 5000 replications.
Figure 3. Smoothed recession probabilities

Note. The solid lines represent the recession probabilities implied by the regime switching model. The shaded areas correspond to the NBER recessions.