UC Santa Barbara
Departmental Working Papers

Title
Markov Regime-Switching Tests: Asymptotic Critical Values

Permalink
https://escholarship.org/uc/item/5rn986z6

Authors
Steigerwald, Douglas
Carter, Andrew

Publication Date
2011-08-12

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Markov Regime-Switching Tests: Asymptotic Critical Values

Andrew V. Carter
Department of Statistics
University of California, Santa Barbara

Douglas G. Steigerwald*
Department of Economics
University of California Santa Barbara

August 12, 2011

Abstract

Empirical research with Markov regime-switching models often requires the researcher not only to estimate the model but also to test for the presence of more than one regime. Despite the need for both estimation and testing, methods of estimation are better understood than are methods of testing. We bridge this gap by explaining, in detail, how to apply the newest results in the theory of regime testing, developed by Cho and White (2007). A key insight in Cho and White is to expand the null region to guard against false rejection of the null hypothesis due to a small group of extremal values. Because the resulting asymptotic null distribution is a function of a Gaussian process, the critical values are not obtained from a closed-form distribution such as the χ^2. Moreover, the critical values depend on the covariance of the Gaussian process and so depend both on the specification of the model and the specification of the parameter space. To ease the task of calculating critical values, we describe the limit theory and detail how the covariance of the Gaussian process is linked to the specification of both the model and the parameter space. Further, we show that for linear models with Gaussian errors, the relevant parameter space governs a standardized index of regime separation, so one need only refer to the tabulated critical values we present.

JEL Classification: Primary C15; Secondary C12
Key Words: mixture model, regime switching, numeric approximation

*Corresponding author: doug@econ.ucsb.edu

1 Introduction

Markov regime-switching models, in which the intercept varies over regimes, have many uses in applied econometrics. Researchers have used these models to describe the behavior of GDP, to detect multiple equilibria and to describe the behavior of asset prices. While estimation of these models is straightforward, testing for the possible presence of more than one regime is more difficult. Researchers are aware that test statistics could be based on a likelihood ratio, but are generally uncertain of how to obtain critical values from the asymptotic null distribution of the test statistics. Our goal is to enable researchers to obtain critical values from the asymptotic null distribution of the test statistic to provide valid inference regarding the presence of distinct regimes.

Cho and White (2007) provide an asymptotic null distribution that yields the critical values on which such a test should be based. Because the resulting asymptotic null distribution is a function of a Gaussian process, the critical values are not obtained from a closed-form distribution such as the χ^2. Further, because the Gaussian process depends upon both the specified model and the specified parameter space, the critical values differ across applications and cannot be obtained from a single reference calculation, such as is the case for the Dickey-Fuller distribution. In consequence, users face the daunting task of linking a general Gaussian process limit result to the specific structure of their model. We ease this task by detailing how the Gaussian process and, most importantly, how the covariance among the elements of the Gaussian process are linked to the specification of the model.

For the leading case of a linear model with Gaussian errors we bring forward three important points. First, the covariance of the Gaussian process does not depend on the presence of covariates, so the single analytic calculation we detail suffices for all such models. Second, the parameters of the model that characterize regime switching enter the covariance only through the standardized distance between regime means. In consequence a researcher does not need to specify the parameter space that contains the regime-specific intercepts, but only the number of standard deviations that separate the regime means. The first two points together imply that a researcher testing for regime switching under a linear model with Gaussian errors can refer to the tabulated critical values that appear in Section 4. Our third point is that these tabulated critical values can be used for a broader class of models. Specifically, for a system of two linear equations the same critical values apply, although the standardization of the distance between regimes must account for the error variance from each equation.

To frame the issues, consider the basic regime-switching model estimated by Cecchetti, Lam and Mark (1990), in which the growth rate of annual, per capita GNP, Y_t, is

$$Y_t = \theta_0 + \delta S_t + U_t, \tag{1}$$

where $U_t \sim i.i.d.N(0, \nu)$. The unobserved state variable $S_t \in \{0, 1\}$ indicates regimes, with $S_t = 0$ corresponding to a period of contraction in the

economy and $S_t = 1$ corresponding to a period of economic expansion. Further, the sequence $\{S_t\}_{t=1}^n$ is generated as a first-order Markov process with $\mathbb{P}(S_t = 1|S_{t-1} = 0) = p_0$ and $\mathbb{P}(S_t = 0|S_{t-1} = 1) = p_1$. The empirical feature that expansions tend to last longer than contractions is captured by $p_0 > p_1$.

A key issue is to test the null hypothesis of one regime against the alternative of Markov switching between two regimes. As $\delta = 0$ corresponds to only a single regime, it seems natural to base such a test on the t statistic for δ. Yet the fact that the unobserved sequence $\{S_t\}$ depends on parameters (p_0, p_1) that vanish from the model if $\delta = 0$, renders standard inference with the t statistic invalid. Tests based on the Lagrange Multiplier principle are also invalid, because the gradient of the likelihood function is identically zero when evaluated at null estimates. Valid tests of the null hypothesis of only a single regime are thus based on the likelihood ratio. Cecchetti, Lam and Mark estimate a likelihood-ratio test statistic and uncover evidence of multiple regimes but, absent a method to construct critical values from the asymptotic null distribution, use critical values that do not necessarily deliver valid inference.

To derive the asymptotic null distribution of the likelihood ratio statistic one additional non-standard feature must be considered. This feature, emphasized by Cho and White, is the presence of three regions in the null parameter space. To understand the importance of accounting for all three regions, it is helpful to present the regime-switching regression (1) in the form of conditional densities. Let θ_1 denote the mean of regime 1, so that $\theta_1 = \theta_0 + \delta$. The conditional densities for Y_t are:

$$\begin{aligned} f(Y_t, \theta_0) &= \frac{1}{\sqrt{2\pi\nu}} \exp\left[-\frac{1}{2\nu}(Y_t - \theta_0)^2\right] \text{ if } S_t = 0 \\ f(Y_t, \theta_1) &= \frac{1}{\sqrt{2\pi\nu}} \exp\left[-\frac{1}{2\nu}(Y_t - \theta_1)^2\right] \text{ if } S_t = 1. \end{aligned} \tag{2}$$

Under the null hypothesis of only a single regime with mean θ_*, three curves - which form the three regions of the null space - equivalently represent the population density $f(Y_t, \theta_*)$. The first curve corresponds to $p_0 > 0$ and $p_1 > 0$, so that both regimes are observed with positive probability, and $\theta_0 = \theta_1 = \theta_*$. For the remaining two curves, both regimes do not occur with positive probability. One curve corresponds to the boundary value $p_0 = 0$, so that regime 0 occurs with probability 1, and $\theta_0 = \theta_*$. The remaining curve corresponds to the boundary value $p_1 = 0$ and $\theta_1 = \theta_*$.

Ghosh and Sen (1985), who establish the importance of accounting for all three curves, note that when the null hypothesis is true the maximum of the likelihood will eventually be attained in a neighborhood of the union of all three curves that represent $f(Y_t, \theta_*)$. For this reason, attention cannot be confined to the single curve that corresponds to $\theta_0 = \theta_1 = \theta_*$. Moreover, the curves that correspond to the values $p_0 = 0$ and $p_1 = 0$ play an important role in empirical analysis. Observe that points in a neighborhood of $\theta_0 = \theta_1 = \theta_*$ correspond to a process in which there are two regimes, with slightly separated means that may occur with equal frequency. Points in a neighborhood of the values $p_0 = 0$

and $p_1 = 0$, in contrast, correspond to a process in which there are two widely separated regimes, one of which occurs infrequently. As false rejection of the null hypothesis is often thought to result from the misclassification of a small group of extremal values as a second regime, it is vital to include boundary values in the null parameter space to guard against this type of false rejection. The probability of this type of false rejection is indeed reduced, as enlarging the null space to include the boundary curves leads to an increase in critical values.

Cho and White find that when considering the likelihood for a Markov regime-switching process including $p_0 = 0$ and $p_1 = 0$ in the parameter space leads to difficulties in the asymptotic analysis of the likelihood ratio statistic. These difficulties lead Cho and White to analyze a quasi-likelihood ratio (QLR) statistic. In consequence they approximate the likelihood with a quasi-likelihood that corresponds to a process in which $\{S_t\}$ is a sequence of i.i.d. random variables with $\mathbb{P}(S_t = 1) = \pi$, where the stationary probability π equals $p_0/(p_0 + p_1)$. While the resulting quasi-likelihood ignores certain correlation properties implied by the Markov structure, it yields a tractable factorization of the likelihood and avoids the difficulties arising from the asymptotic null distribution of the score on the boundary of the parameter space.

Because $\pi = 1$ if and only if $p_1 = 0$ (and $\pi = 0$ if and only if $p_0 = 0$), the null hypothesis for test of one regime against two regimes is again expressed with three curves. The null hypothesis is, $H_0 : \theta_0 = \theta_1 = \theta_*$ (curve 1), $\pi = 0$ *and* $\theta_0 = \theta_*$ (curve 2), $\pi = 1$ *and* $\theta_1 = \theta_*$ (curve 3). The alternative hypothesis is $H_1 : \pi \in (0, 1)$ *and* $\theta_0 \neq \theta_1$. In Figure 1 we depict the null space together with local neighborhoods for two points in this space. The two neighborhoods illustrate the role of each curve in the null space. Points in the circular neighborhood surrounding the point on $\theta_1 - \theta_0 = 0$, have slightly separated regimes as they lie near $\theta_0 = \theta_1$. Points in the semicircular neighborhood around the point on $\pi = 1$, are infrequently drawn from the distribution with mean θ_0 as they lie near $\pi = 1$.

The two neighborhoods also illustrate the issues of identifiability. Under the alternative hypothesis switching occurs between two regimes, but the regimes are identified only up to labeling - as one could re-label $(\pi, \theta_0, \theta_1)$ as $(1 - \pi, \theta_1, \theta_0)$. Ignoring labeling, the parameters $(\pi, \theta_0, \theta_1)$ are identified under H_1. Under the null hypothesis the identification issues are more complex. On the curve $\theta_0 = \theta_1$, the parameter π is not identified. On the curve $\pi = 0$, θ_1 is not identified and on the curve $\pi = 1$ the parameter θ_0 is not identified. Further, each null distribution can be equivalently represented by a point on each of the three curves. It is these identification issues that give rise to the complex null distribution that Cho and White derive.

While Cho and White consider all three regions of the null space in deriving an asymptotic distribution, earlier researchers focused only on the region $\theta_1 - \theta_0 = 0$, together with the identifiability condition that $\pi \in (0, 1)$. As the boundary regions $\pi = 1$ and $\pi = 0$ do not appear, the likelihood, rather than the quasi-likelihood, is the object of analysis. Hansen (1992) obtains a bound on the asymptotic null distribution of a likelihood ratio statistic; this bound is a Gaussian process. Garcia (1998) obtains a χ^2 process as the asymptotic null

distribution of a likelihood ratio statistic, but to do so he requires that the matrix of second derivatives of the likelihood be non-singular when evaluated at the null estimates. As he notes (p. 764) this condition does not hold for the Markov regime-switching process he considers, which has Gaussian innovations with a regime-varying scale parameter. As we describe in Section 2, the presence of boundary values, together with a singular matrix of second derivatives, results in an asymptotic null distribution that is a function of a Gaussian process rather than a χ^2 process.

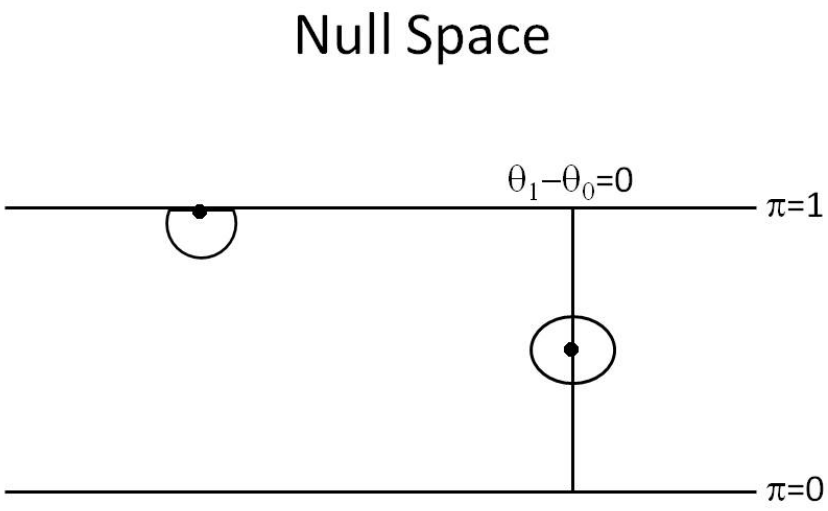

Figure 1 depicts all three regions of the null hypothesis $H_0 : \pi = 0$ *and* $\theta_0 = \theta_*$*;* $\pi = 1$ *and* $\theta_1 = \theta_*$*; or* $\theta_0 = \theta_1 = \theta_*$ together with local neighborhoods of $\pi = 1$ and $\theta_0 = \theta_1 = \theta_*$. Note that, in terms of the Markov model, $\pi = 1$ corresponds to $p_1 = 0$ and $\pi = 0$ corresponds to $p_0 = 0$.

We organize the results as follows. In Section 2 we present the single and multiple equation linear models we consider, together with the QLR statistic. We also present the asymptotic null distribution of the statistic, as derived by Cho and White, and detail how a Gaussian process enters the limit distribution. In doing so, we highlight the need to calculate the covariance between the random variables that enter the asymptotic null distribution. In Section 3 we derive the covariance structure of the Gaussian process that appears in the asymptotic null distribution and detail how to construct the structure for linear

models with Gaussian errors. Due to the covariance structure of the Gaussian process the critical values cannot be calculated directly, so in Section 4 we show how to numerically approximate the critical values. We focus on linear models with Gaussian errors and, for a set of standardized distances between regime means, we present a table of critical values. Finally, we link the simulation discussion to pseudo-code contained in the Appendix (and reference programs in Matlab, R and Stata) so that researchers are able to construct critical values for other sets of standardized distances.

2 A QLR Test for Regime Switching

The class of Markov regime-switching processes for which Cho and White establish consistency of a QLR test includes far more than the structure analyzed by Cecchetti, Lam and Mark. In this section we provide leading examples of allowable processes together with the asymptotic null distribution of the QLR statistic, deferring the formal conditions under which the distribution is derived to the Appendix. The process (1) can be augmented with covariates Z_t,

$$Y_t = \theta_0 + \delta S_t + Z_t'\beta + U_t. \tag{3}$$

There are two further generalizations of (3) that broaden the scope of application. First, the error density may be any element from the exponential family. Second, the dependent variable can be vector valued, although the difference between distributions in the mixture model must be in only one mean parameter. One example of such a system of equations is the structural model

$$\begin{aligned} Y_{t1} &= \theta_0 + \delta S_t + \alpha_{12} Y_{t2} + Z_{t1}'\beta_1 + U_{t1} \\ Y_{t2} &= \mu + \alpha_{21} Y_{t1} + Z_{t2}'\beta_2 + U_{t2}. \end{aligned} \tag{4}$$

For any of the allowable processes, let the conditional densities be $f\left(Y_t|Z^t;\gamma,\theta_j\right)$ with $j = 0, 1$ where $Z^t = (Z_t', \ldots, Z_1)$ and γ includes other parameters of the conditional density (e.g. $\gamma = \left(\nu, \beta'\right)$). The quasi-log-likelihood analyzed by Cho and White, which ignores the Markov structure and treats $\{S_t\}$ as i.i.d. with $\mathbb{P}\left(S_t = 1\right) = \pi$, is

$$L_n\left(\pi, \gamma, \theta_0, \theta_1\right) = \frac{1}{n}\sum_{t=1}^{n} l_t\left(\pi, \gamma, \theta_0, \theta_1\right),$$

where $l_t\left(\pi, \gamma, \theta_0, \theta_1\right) := \log\left(\left(1-\pi\right) f\left(Y_t|Z^t;\gamma,\theta_0\right) + \pi f\left(Y_t|Z^t;\gamma,\theta_1\right)\right)$. The use of this quasi-log-likelihood to form the quasi-maximum likelihood estimator (QMLE) leads to an important restriction on (3). Carter and Steigerwald (2010) establish that the QMLE is inconsistent in the presence of Markov switching if Z_t includes lagged values of Y_t. For this reason, the processes under study do not include autoregressions.

To describe the asymptotic null distribution of the QLR statistic, we first note that the null distribution is largely determined by the behavior of the

statistic in a neighborhood of the null region $\pi = 1$. The asymptotic null distribution is complicated by the fact that θ_0 is not identified if $\pi = 1$, so changes in the value of θ_0 do not alter the asymptotic null distribution. This stands in contrast to the identified parameters θ_1 and γ, for which changes in their value do alter the asymptotic null distribution of the QLR statistic. In consequence, if $\hat{\pi}$ is close to 1 we expect $\hat{\theta}_1$ and $\hat{\gamma}$ to be close to their population values, while there is no population value that $\hat{\theta}_0$ should be close to.

Define $\left(\hat{\pi}, \hat{\gamma}, \hat{\theta}_0, \hat{\theta}_1\right)$ as parameter values that maximize the L_n function. Let $\left(1, \tilde{\gamma}, \cdot, \tilde{\theta}_1\right)$ be parameter values that make L_n as large as possible over the null hypothesis that $\pi = 1$. The QLR statistic is

$$QLR_n = 2n\left(L_n\left(\hat{\pi}, \hat{\gamma}, \hat{\theta}_0, \hat{\theta}_1\right) - L_n\left(1, \tilde{\gamma}, \cdot, \tilde{\theta}_1\right)\right). \tag{5}$$

We investigate the behavior of the distribution of QLR_n in a neighborhood of the null region corresponding to $\pi = 1$, for which the alternative hypothesis is $\pi < 1$. Observe that, although π is a probability, it is possible that $\hat{\pi} > 1$. Thus $\hat{\pi}$ should be subject to a boundary condition.

At first we ignore the boundary condition on $\hat{\pi}$. If we fix θ_0 at θ_0', the regularity conditions imply that the asymptotic null distribution of QLR_n is χ^2, with one degree-of-freedom. As the value θ_0' is arbitrary, the distribution of QLR_n depends on the stochastic process formed from the sequence of χ^2 random variables, each indexed by a particular value of θ_0. Moreover, the elements of the χ^2 process are dependent upon each other. The dependence arises in the following way. For a fixed value θ_0', the maximum of the likelihood is $L_n\left(\hat{\pi}\left(\theta_0'\right), \hat{\gamma}\left(\theta_0'\right), \theta_0', \hat{\theta}_1\left(\theta_0'\right)\right)$. If we fix the value at $\theta_0^{''}$, then the estimates that maximize the likelihood are $\left(\hat{\pi}\left(\theta_0''\right), \hat{\gamma}\left(\theta_0''\right), \hat{\theta}_1\left(\theta_0''\right)\right)$. Because these two sets of estimates of (π, γ, θ_1) (at both θ_0' and θ_0'') are calculated from the same sample, the corresponding sequences $\chi^2\left(\theta_0'\right)$ and $\chi^2\left(\theta_0''\right)$ are dependent.

When we impose the boundary condition on $\hat{\pi}$, the asymptotic null distribution of QLR_n is no longer a χ^2 process.[1] To see this, note first that the boundary condition $\pi \leq 1$ implies that if $\hat{\pi} > 1$, then the estimate of π is truncated back to $\hat{\pi} = 1$ and $QLR_n = 0$. The event that $\hat{\pi} > 1$ is closely tied to the asymptotic null distribution for QLR_n. If θ_0 is fixed at θ_0', then the asymptotic null distribution of QLR_n that occurs in the absence of the boundary condition can be expressed as $\mathcal{G}\left(\theta_0'\right)^2$, where $\mathcal{G}\left(\theta_0'\right) \sim N(0,1)$, and the estimator $\hat{\pi}$ is asymptotically equal to $1 + c\mathcal{G}\left(\theta_0'\right)$, where c is a positive constant. In consequence, if $\mathcal{G}\left(\theta_0'\right) > 0$ then $\hat{\pi} > 1$. Thus, when the boundary condition is imposed the asymptotic null distribution of QLR_n has point mass at 0 and the remainder of the null distribution is governed by the negative part of the Gaussian process, $\mathcal{G}(\theta_0)$.

Let Θ define the set of possible values of θ_0. The procedure of first maximizing L_n for a fixed value of θ_0 and then obtaining the supremum over Θ,

[1] This is similar to the behavior of a one-sided likelihood ratio test (van der Vaart (1998) p. 235).

yields the asymptotic null distribution (Cho and White Theorem 6(a), p. 1692)

$$QLR_n \Rightarrow \sup_{\Theta} \left(\min\left[0, \mathcal{G}\left(\theta_0\right)\right]\right)^2. \tag{6}$$

The critical value corresponds to a quantile for the largest value, over Θ, of $\left[\mathcal{G}\left(\theta_0\right)_-\right]^2$, where $\mathcal{G}\left(\theta_0\right)_- := \min\left[0, \mathcal{G}\left(\theta_0\right)\right]$. As we show below for Gaussian error densities, the sign of $\mathcal{G}\left(\theta_0\right)$ switches at the origin, so the quantile exceeds 0 with probability 1 if Θ contains both positive and negative values.

One important wrinkle still remains. While (6) provides the asymptotic null distribution for many experiments, it does not provide the full distribution for all Gaussian experiments. If $U_t \sim i.i.d.N\left(0, \nu\right)$ the asymptotic null distribution of QLR_n is not determined solely by the behavior in a neighborhood of $\pi = 1$. If θ_0 is sufficiently close to θ_1 and $\pi = \frac{1}{2}$, then the asymptotic null distribution has an additional term (Cho and White Theorem 6(b), p. 1692)

$$QLR_n \Rightarrow \max\left[\left[\max\left(0, G\right)\right]^2, \sup_{\Theta}\left[\mathcal{G}\left(\theta_0\right)_-\right]^2\right]. \tag{7}$$

Here G is a standard Gaussian random variable that is correlated with $\mathcal{G}\left(\theta_0\right)$.

The critical value for a test based on QLR_n corresponds to a quantile for the largest value over $\max\left(0, G\right)^2$ and $\sup_{\Theta}\left[\mathcal{G}\left(\theta_0\right)_-\right]^2$. To determine this quantity one must account for the covariance among the elements of $\mathcal{G}\left(\theta_0\right)$ together with their covariance with G. Because the covariance among the elements of $\mathcal{G}\left(\theta_0\right)$ depends on the assumed process for Y_t, we show how to analytically calculate this covariance in the next section.

3 Gaussian Process Covariance

The first step in obtaining critical values from the asymptotic null distribution is to analytically derive the covariance function of $\mathcal{G}\left(\theta_0\right)$. To do so, we first present the Gaussian process, $\mathcal{G}\left(\theta_0\right)$, as a normalized score function, together with the expression for the covariance of the process across the values of θ_0. The subsections contain the explicit calculations of this covariance for the models (3) and (4).

Because the Gaussian process $\mathcal{G}\left(\theta_0\right)$ arises from the behavior of QLR_n in a neighborhood of the null region $\pi = 1$, the component of the gradient that determines $\mathcal{G}\left(\theta_0\right)$ is the score for π evaluated at $(1, \gamma, \theta_0, \theta_*)$ (which are the population values under the null hypothesis that $\pi = 1$)

$$\mathcal{S}\left(\theta_0\right) = \left.\frac{\partial}{\partial \pi} l_t\right|_{(1,\gamma,\theta_0,\theta_*)}.$$

Because $\mathcal{S}\left(\theta_0\right) \sim N\left(0, \mathcal{V}\left(\theta_0\right)\right)$, the standardized process $\mathcal{G}\left(\theta_0\right)$ is a scaled score function

$$\mathcal{G}\left(\theta_0\right) = \mathcal{V}\left(\theta_0\right)^{-\frac{1}{2}} \mathcal{S}\left(\theta_0\right).$$

The asymptotic variance of $\mathcal{S}\left(\theta_0\right)$ is

$$\mathcal{V}(\theta_0) = \mathcal{I}^{11}(\theta_0),$$

where $\mathcal{I}^{11}(\theta_0)$ is the (1,1) element of $\mathcal{I}(\theta_0)^{-1}$ and

$$\mathcal{I}(\theta_0) = \mathbb{E}\left[\left(\nabla_{\pi,\gamma,\theta_1} l_t(\theta_0)\right)\left(\nabla_{\pi,\gamma,\theta_1} l_t(\theta_0)\right)^T\right].$$

Here $\nabla_{\pi,\gamma,\theta_1} l_t$ denotes the gradient with respect to π, γ and θ_1 evaluated at $(\pi,\gamma,\theta_0,\theta_1) = (1,\gamma,\theta_0,\theta_*)$.[2] From the partitioned inverse formula (Theil 1971, p. 18), $\mathcal{V}(\theta_0)$ is

$$\mathcal{V}(\theta_0) = \left(\mathcal{I}_{11}(\theta_0) - \mathcal{I}_1(\theta_0)\left[\mathcal{I}_2(\theta_0)\right]^{-1}\mathcal{I}_1(\theta_0)^T\right)^{-1},$$

where $\mathcal{I} = \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_1 \\ \mathcal{I}_1^T & \mathcal{I}_2 \end{bmatrix}$.

Because the process $\mathcal{G}(\cdot)$ is a Gaussian process, the dependence among the elements of $\mathcal{G}(\cdot)$ is captured by the covariance among the elements of $\mathcal{G}(\cdot)$. If we let θ_0 and θ_0' denote two distinct elements of the process $\mathcal{G}(\cdot)$, then the covariance $\mathbb{E}\left[\mathcal{G}(\theta_0)\mathcal{G}(\theta_0')\right]$ is derived from the covariance $\mathbb{E}\left[\mathcal{S}(\theta_0)\mathcal{S}(\theta_0')\right]$ as

$$\mathbb{E}\left[\mathcal{G}(\theta_0)\mathcal{G}(\theta_0')\right] = \mathcal{V}(\theta_0)^{-\frac{1}{2}}\mathcal{V}(\theta_0')^{-\frac{1}{2}}\mathbb{E}\left[\mathcal{S}(\theta_0)\mathcal{S}(\theta_0')\right]. \tag{8}$$

The covariance $\mathbb{E}\left[\mathcal{S}(\theta_0)\mathcal{S}(\theta_0')\right]$ is the (1,1) element of

$$\mathcal{I}(\theta_0')^{-1}\mathcal{I}(\theta_0,\theta_0')\mathcal{I}(\theta_0)^{-1}.$$

The matrix $\mathcal{I}(\theta_0,\theta_0')$ is obtained by evaluating the gradient at distinct points: $\mathcal{I}(\theta_0,\theta_0') = \mathbb{E}\left[\left(\nabla_{\pi,\gamma,\theta_1} l_t(\theta_0)\right)\left(\nabla_{\pi,\gamma,\theta_1} l_t(\theta_0')\right)^T\right]$. We next show how to calculate these quantities for each class of data generating processes.

3.1 Single Equation Linear Model

For the single equation linear model (3) with $U_t \sim i.i.d.N(0,\nu)$, which excludes lagged values of Y_t as covariates, we show that $\mathbb{E}\left[\mathcal{G}(\theta_0)\mathcal{G}(\theta_0')\right]$ does not depend on Z_t. Thus whether one has an extensive set of covariates, or none as in Cecchetti, Lam and Mark (1), the following calculation is all that is needed. For this model

$$\mathcal{S}(\theta_0) = 1 - \exp\left[\frac{(\theta_0-\theta_*)}{\nu}\left(Y_t - Z_t'\beta - \frac{\theta_0+\theta_*}{2}\right)\right].$$

From the derivative calculations in the Appendix the asymptotic variance of $\mathcal{S}(\theta_0)$ is

$$\mathcal{V}(\theta_0) = \left(e^{\frac{1}{\nu}(\theta_0-\theta_*)^2} - 1 - \frac{(\theta_0-\theta_*)^2}{\nu} - \frac{(\theta_0-\theta_*)^4}{2\nu^2}\right)^{-1}.$$

[2] The element of the gradient corresponding to θ_0 is identically zero when evaluated at $\pi = 1$ and so is deleted from the vector that forms $\mathcal{I}(\theta_0)$ (Cho and White, Assumption A.6 p. 1678).

To obtain the largest value of $\left[\mathcal{G}\left(\theta_0\right)_-\right]^2$ over Θ, we must also know the covariance of $\mathcal{G}\left(\theta_0\right)$, which depends on the covariance of $\mathcal{S}\left(\theta_0\right)$. The covariance of the score, $\mathbb{E}\left[\mathcal{S}\left(\theta_0\right)\mathcal{S}\left(\theta_0'\right)\right]$, in turn requires

$$\mathcal{I}\left(\theta_0,\theta_0'\right)=\begin{bmatrix} e^{\frac{1}{\nu}(\theta_0-\theta_*)\left(\theta_0'-\theta_*\right)}-1 & -\frac{\left(\theta_0'-\theta_*\right)^2}{2\nu^2} & \mathbb{E}\left[Z_t'\right]\frac{\theta_0'-\theta_*}{\nu} & \frac{\theta_0'-\theta_*}{\nu} \\ -\frac{(\theta_0-\theta_*)^2}{2\nu^2} & \frac{1}{2\nu^2} & 0 & 0 \\ \mathbb{E}\left[Z_t\right]\frac{\theta_0'-\theta_*}{\nu} & 0 & \mathbb{E}\left[Z_t'Z_t\right]\frac{1}{\nu} & \mathbb{E}\left[Z_t'\right]\frac{1}{\nu} \\ \frac{\theta_0-\theta_*}{\nu} & 0 & \mathbb{E}\left[Z_t\right]\frac{1}{\nu} & \frac{1}{\nu} \end{bmatrix},$$

so $\mathcal{I}_{11}\left(\theta_0,\theta_0'\right)=e^{\frac{1}{\nu}(\theta_0-\theta_*)\left(\theta_0'-\theta_*\right)}-1$. Then $\mathbb{E}\left[\mathcal{S}\left(\theta_0\right)\mathcal{S}\left(\theta_0'\right)\right]$ equals $\mathcal{V}\left(\theta_0\right)\mathcal{V}\left(\theta_0'\right)$ times the following term

$$e^{\frac{1}{\nu}(\theta_0-\theta_*)\left(\theta_0'-\theta_*\right)}-1-\frac{\left(\theta_0-\theta_*\right)\left(\theta_0'-\theta_*\right)}{\nu}-\frac{\left(\theta_0-\theta_*\right)^2\left(\theta_0'-\theta_*\right)^2}{2\nu^2}.$$

Because neither $\mathbb{E}\left[\mathcal{S}\left(\theta_0\right)\mathcal{S}\left(\theta_0'\right)\right]$ nor $\mathcal{V}\left(\cdot\right)$ is a function of Z_t, the covariance of the Gaussian process, $\mathbb{E}\left[\mathcal{G}\left(\theta_0\right)\mathcal{G}\left(\theta_0'\right)\right]$ given by (8), is independent of the covariates that enter the model. Hence the calculations we detail here provide the covariance of the Gaussian process for all models of the form of (3).

Next observe that the regime-specific parameters θ_0 and θ_* enter $\mathbb{E}\left[\mathcal{S}\left(\theta_0\right)\mathcal{S}\left(\theta_0'\right)\right]$ and $\mathcal{V}\left(\cdot\right)$ only through $\eta=\frac{\theta_0-\theta_*}{\sqrt{\nu}}$. Hence the covariance of the Gaussian process is given by

$$\mathbb{E}\left[\mathcal{G}\left(\theta_0\right)\mathcal{G}\left(\theta_0'\right)\right]=\frac{e^{\eta\eta'}-1-\eta\eta'-\frac{\left(\eta\eta'\right)^2}{2}}{\left(e^{\eta^2}-1-\eta^2-\frac{\eta^4}{2}\right)^{\frac{1}{2}}\left(e^{(\eta')^2}-1-(\eta')^2-\frac{(\eta')^4}{2}\right)^{\frac{1}{2}}},\quad(9)$$

where $\eta'=\frac{\theta_0'-\theta_*}{\sqrt{\nu}}$. The quantity $\sup_\Theta\left[\mathcal{G}\left(\theta_0\right)_-\right]^2$ that appears in the asymptotic null distribution is determined by the covariance $\mathbb{E}\left[\mathcal{G}\left(\theta_0\right)\mathcal{G}\left(\theta_0'\right)\right]$. Because the regime-specific parameters enter (9) only through η, a researcher need only specify the set that contains η. That is, to calculate $\sup_\Theta\left[\mathcal{G}\left(\theta_0\right)_-\right]^2$ a researcher does not need to specify the parameter space Θ that contains the regime-specific intercepts, but need only specify the set H that contains the number of standard deviations that separate the regime means.

3.2 Simultaneous Equations Linear Model

For the simultaneous equations linear model (4), let (U_{t1},U_{t2}) be multivariate Gaussian random variables with zero mean, $Var\left(U_{ti}\right)=\nu_i$ and $Cov\left(U_{t1},U_{t2}\right)=\nu_{12}$. The (canonical) reduced form of the multivariate random variable $Y_t:=(Y_{t1},Y_{t2})'$ is

$$Y_t=A^{-1}\begin{pmatrix}\theta_0\\ \mu\end{pmatrix}+A^{-1}S_t\begin{pmatrix}\delta\\ 0\end{pmatrix}+A^{-1}\begin{pmatrix}Z_{t1}' & \mathbf{0}\\ \mathbf{0} & Z_{t2}'\end{pmatrix}\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}+A^{-1}\begin{pmatrix}U_{t1}\\ U_{t2}\end{pmatrix},$$

where $A = \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix}$. As we detail in the Appendix, the covariance of the Gaussian process takes the form

$$\mathbb{E}\left[\mathcal{G}(\theta_0)\,\mathcal{G}(\theta_0')\right] = \mathcal{V}(\eta)^{-\frac{1}{2}}\,\mathcal{V}(\eta')^{-\frac{1}{2}}\left[\exp(\eta\eta') - 1 - \eta\eta' - \frac{1}{2}(\eta\eta')^2\right],$$

with $\eta = \frac{\theta_0 - \theta_*}{\sqrt{\nu_1(1-\rho^2)}}$ and $\rho = Corr(U_{t1}, U_{t2})$. We see that the standardized distance between regimes is altered in a natural way, as $\nu_1\left(1-\rho^2\right)$ is the variance of Y_{1t} conditional on Y_{2t}, and that the form of the covariance function is identical to that of the single equation model. Moreover, the index of the standardized distance between regimes does not depend on A, so that the same calculations apply to a triangular system ($\alpha_{21} = 0$) and to a system of seemingly unrelated equations ($\alpha_{12} = \alpha_{21} = 0$). As in the case of the single equation model, calculation of the critical values only requires specification of the interval H that contains the standardized distance between regimes.

4 Quantile Simulation

The second step in obtaining a critical value is to construct the appropriate quantile from the asymptotic null distribution. For a QLR test with size 5 percent, the critical value corresponds to the .95 quantile of the limit distribution given on the right side of either (6) or (7). Because the dependence in the process $\mathcal{G}(\theta_0)$ renders numeric integration infeasible, we construct the quantile by simulating independent replications of a process. For the linear model with Gaussian errors, as the covariance of $\mathcal{G}(\theta_0)$ depends only on an index η, while $\mathcal{G}$ itself depends on $(\nu, \beta, \theta_0, \theta_*)$ through the score $\mathcal{S}(\theta_0)$, we do not simulate $\mathcal{G}(\theta_0)$ directly. Instead we simulate $\mathcal{G}^A(\eta)$, which has the same covariance structure as $\mathcal{G}(\theta_0)$ and so delivers the same quantile, but which depends only on the index.

To construct $\mathcal{G}^A(\eta)$ for the covariance structure in (9) recall that, by a Taylor-series expansion, $e^\eta = 1+\eta+\frac{\eta^2}{2!}+\cdots$. Hence, for $\{\epsilon_j\}_{j=0}^{\infty} \sim i.i.d.N(0,1)$:

$$\sum_{j=3}^{\infty} \frac{\eta^j}{\sqrt{j!}}\epsilon_j \sim N\left(0, e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right),$$

so $\mathcal{V}(\theta_0)\sum_{j=3}^{\infty}\frac{\eta^j}{\sqrt{j!}}\epsilon_j$ has the same covariance structure as $\mathcal{S}(\theta_0)$. The simulated process is

$$\mathcal{G}^A(\eta) = \left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-\frac{1}{2}} \sum_{j=3}^{J-1} \frac{\eta^j}{\sqrt{j!}}\epsilon_j,$$

where J determines the accuracy of the Taylor-series approximation. To capture the behavior of the limit distribution in (7), we must also account for the covariance between G and $\mathcal{G}(\theta_0)$. As this covariance is a function of η^4, whose

corresponding value is ϵ_4 in the expression for $\mathcal{G}^A(\eta)$, we set $G = \epsilon_4$ so that $Cov(G, \mathcal{G}(\theta_0)) = Cov\left(G, \mathcal{G}^A(\eta)\right)$.[3]

For each replication, we calculate $\mathcal{G}^A(\eta)$ at a fine grid of values over H. To do so, we must specify three quantities: the interval H, the grid of values over H (given by the grid mesh) and the number of terms in the Taylor-series approximation, J. To understand the interplay in specifying these three quantities, suppose that θ_0 is thought to lie within 3 standard deviations of θ_1. The interval is $H = [-3.0, 3.0]$ and, with a grid mesh of .01, the process is calculated at the points $(-3.00, -2.99, \ldots, 3.00)$. Because the process is calculated at only a finite number of values, while the maximum that appears in the limit distribution is obtained over a continuum of values, the accuracy of the calculated maximum increases as the grid mesh shrinks. For this reason we recommend a grid mesh of .01 (as do Cho and White, p. 1693).

To determine the value of J, let $\xi_{J,\eta} = \left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-\frac{1}{2}} \sum_{j=J}^{\infty} \frac{\eta^j}{\sqrt{j!}} \epsilon_j$ be the approximation error. Because $\{\epsilon_j\}$ is a mean zero process, it is the variance of $\xi_{J,\eta}$ that provides information about the magnitude of the approximation error. When $\eta > 0$, $e^{\eta} = 1 + \cdots + \frac{\eta^{J-1}}{(J-1)!} + e^{\eta^*} \frac{\eta^J}{J!}$ for some $0 < \eta^* < \eta$. The variance of $\xi_{J,\eta}$ is then bounded by $\left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-1} e^{\eta^2} \frac{\eta^{2J}}{J!}$ so, by Stirling's formula,

$$\log Var\left(\xi_{J,\eta}\right) \leq 2J \cdot \log \eta - \left(J + \frac{1}{2}\right) \log J + J - \log \sqrt{2\pi} + \log \left[\left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-1} e^{\eta^2}\right].$$

For large J, $Var\left(\xi_{J,\eta}\right)$ is governed by

$$Var\left(\xi_{J,\eta}\right) \leq e^{2J \log \eta - J \log J},$$

so $\frac{\eta^2}{J} << 1$ to ensure that the variance of $\xi_{J,\eta}$ declines rapidly to 0 as J grows. The value of J is then determined such that $\left(\max_H |\eta|\right)^2 << J$. In practice, we recommend that $\left(\max_H |\eta|\right)^2 / J \leq 1/2$.[4]

The critical value that corresponds to (7) for a test size of 5 percent is the .95 quantile of the simulated value

$$\max\left\{\left[\max\left(0, \epsilon_4\right)\right]^2, \max_{\eta \in H}\left[\min\left(0, \mathcal{G}^A(\eta)\right)\right]^2\right\}.$$

In Table 1 we present the critical value for specified intervals, which correspond to regime separations that range from 1 to 10 standard deviations. For the reported values we set $J = 150$ and use a grid mesh of .01. Both settings correspond to the values in Cho and White. We find that critical values are little impacted by use of a finer grid mesh of .001. A far more important

[3] Cho and White (p. 1693) show $Cov(G, \mathcal{G}(\theta_0)) = \left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-\frac{1}{2}} \eta^4$.

[4] Cho and White select $J = 150$ and consider a maximal value of $\eta = 5$, so $\eta^2 / J \leq 1/6$.

setting is the number of replications. Cho and White use 10,000 replications but we find that this number of replications does not produce stable critical values. To show this, we calculate the critical values from 10,000 simulations each based on 10,000 replications. The length of the interquartile range of 10,000 calculated critical values is on the order of .1 (e.g. for $H = [-3, 3]$ the length of the interquartile range is .11). If we increase the number of replications used to calculate the critical value to 100,000, the length of the interquartile range shrinks by an order of magnitude. Hence the critical values we report are based on 100,000 replications. To directly show the impact on calculated critical values, we also list the critical values reported by Cho and White in the row corresponding to 10,000 replications. Finally, as researchers may need critical values for other specified intervals, we present pseudo-code for the simulation in the Appendix. In addition, simulation programs in Matlab, R and Stata are available from the authors.

Table 1: Critical Values for Linear Models with Gaussian Errors[5]

H	$[-1,1]$	$[-2,2]$	$[-3,3]$	$[-4,4]$	$[-5,5]$	$[-10,10]$
Replications						
100,000	5.03	5.54	6.18	6.67	7.03	8.31
10,000	5.01	5.61	6.35	6.54	7.06	

To understand how to employ these critical values, we return to the study by Cecchetti, Lam and Mark. If we assume that the mean growth rate of annual, per capita GNP differs by no more than 5 standard deviations between expansions and contractions, then 7.03 is the critical value for a test with size 5 percent. (The estimated means differ by slightly less than 4 standard deviations.) As the estimated value of the test statistic is 28, the null hypothesis of only a single regime is clearly rejected for their analysis.

5 Remarks

The asymptotic null distribution that Cho and White establish provides valid inference for a test of more than one regime. The distribution depends both on the structure of the model and on the parameter space that contains the regime-specific intercepts. We show that for the class of linear models with Gaussian errors the dependence of the asymptotic null distribution on the parameter space and model structure is simplified. First, the regime-specific intercepts enter through an index that captures the standardized distance between regimes. Second, the presence of covariates does not affect the critical values. Together, these two points imply that the tabulated critical values we present deliver valid inference for all models within the class.

A question naturally arises: Can the QLR test proposed here be used if, in addition to the intercept, slope coefficients also vary over regimes? The answer

[5] Nominal level 5 percent; $J = 150$; grid mesh of .01, 100,000 replications; critical values corresponding to 10,000 replications are from Cho and White Table I, p.1694.

is yes. Because the null hypothesis is unaltered, the critical values we report deliver a test with correct size. Note that the test maximizes the QL statistic in which only the intercept varies over regimes, so the test may lose power against certain elements of the larger class of alternatives.

To obtain valid inference with the critical values in Table 1, a researcher must, prior to estimation, specify a set of values that contains the standardized distance between regimes. The specified set takes the form of an interval $[-c, c]$ in which the index η must lie, so the estimate $\hat{\eta}$ must also lie in the interval. If the population value of η lies outside the selected interval, then the estimated value of η will be constrained to lie on the boundary of the selected interval, which in turn leads to an increased estimate of the variance ν. The resultant upwardly biased estimate of ν reduces the power of the test to detect multiple regimes. To avoid the issues that arise when the estimate of η is constrained, a researcher can select a large value of c. Yet, as the critical value rises monotonically with c, a large selected interval also leads to a loss of power. This raises an interesting question: Can an alternative method be used to obtain asymptotically valid critical values for the QLR test of regime switching?

6 Appendix

6.1 Formal Conditions

We present the assumptions that define a class of processes to which the asymptotic theory presented in Section 2 applies. The two assumptions presented here combine A1-A2(i) from Cho and White with A2(ii) from Carter and Steigerwald (2010).

Assumption 1:

(i) *The observable random variables* $\left\{(Y_t', Z_t')' \in \mathbb{R}^d\right\}_{t=1}^n$, $d \in \mathbb{N}$, *are generated as a sequence of strictly stationary* β*-mixing random variables such that for some* $c > 0$ *and* $\rho \in [0,1)$ *the beta-mixing coefficient,* g_τ*, is at most* $c\rho^\tau$*.*

(ii) *The sequence of unobserved state variables that indicate regimes,* $\{S_t \in \{0,1\}\}_{t=1}^n$*, is generated as a first-order Markov process such that* $\mathbb{P}(S_t = 1|S_{t-1} = 0) = p_0$ *and* $\mathbb{P}(S_t = 0|S_{t-1} = 1) = p_1$ *with* $p_i \in [0,1]$ $(i = 0, 1)$*.*

(iii) *The given* $\left\{(Y_t', Z_t')'\right\}$ *is a Markov regime-switching process. That is, for some* $(\gamma, \theta_0, \theta_1) \in \mathbb{R}^{r_0+2}$*,*

$$Y_t|\mathcal{F}_{t-1} \sim \begin{cases} F(\cdot|Z^t; \gamma, \theta_0) & \text{if } S_t = 0 \\ F(\cdot|Z^t; \gamma, \theta_1) & \text{if } S_t = 1 \end{cases},$$

where $\mathcal{F}_{t-1} := \sigma\left(Y^{t-1}, Z^t, S^t\right)$ *is the smallest* σ*-algebra generated by* $\left(Y^{t-1}, Z^t, S^t\right) := \left(Y_{t-1}', \ldots, Y_1', Z_t', \ldots, Z_1', S_t, \ldots, S_1\right)$*;* $r_0 \in \mathbb{N}$*; and the conditional cumulative distribution function of* $Y_t|\mathcal{F}$*,* $F(\cdot|Z^t; \gamma, \theta_j)$ *has a probability density function* $f(\cdot|Z^t; \gamma, \theta_j)$ $(j = 0, 1)$*. Further, for* $(p_0, p_1) \in (0,1] \times (0,1] \setminus \{(1,1)\}$*,* $(\gamma, \theta_0, \theta_1)$ *is unique in* $\mathbb{R}^{r_0+2}$*.*

The vector γ captures all parameters of $F(\cdot)$, including the scale parameter, that do not vary across regimes. The point $p_0 = p_1 = 1$ is excluded from the parameter space to rule out a deterministically periodic process for $\{S_t\}$, which would imply that $\{Y_t\}$ is not strictly stationary.

The model for the data generating process specifies a compact parameter space.

Assumption 2:

(i) *A model for* $f(\cdot|Z^t; \gamma, \theta_j)$ *is* $\left\{f(\cdot|Z^t; \gamma, \theta_j) : \ (\gamma, \theta_j) \in \tilde{\Theta}\right\}$*, where* $\tilde{\Theta} := \Gamma \times \Theta \in \mathbb{R}^{r_0+1}$*, and* Γ *and* Θ *are convex and compact sets in* $\mathbb{R}^{r_0}$ *and* $\mathbb{R}$*, respectively. Further, for each* $(\gamma, \theta_j) \in \tilde{\Theta}$*,* $f(\cdot|Z^t; \gamma, \theta_j)$ *is a measurable probability density function, where the support of* $f(\cdot|Z^t; \gamma, \theta_j)$ *is the same for all* $\tilde{\Theta}$*, with cumulative distribution function* $F(\cdot|Z^t; \gamma, \theta_j)$ $(j = 0, 1)$*.*

(ii) *The covariates are exogenous in the sense that* $P(S_t = j|\mathcal{F}_{t-1})$ *is independent of* Z^t *for* $(j = 0, 1)$*.*

Additional continuity and bounded derivative conditions are needed to obtain (7) (Cho and White, p. 1687), which hold for a Gaussian density.

6.2 Gaussian Process Covariance

Single Equation - Derivative Calculations

For the process given by (3) with $U_t \sim i.i.d.N(0,\nu)$, the quasi-log-likelihood for observation t, l_t, equals

$$\log\left[(1-\pi)\exp\left(\frac{2\theta_0\left(Y_t - Z_t'\beta\right) - \theta_0^2}{2\nu}\right) + \pi\exp\left(\frac{2\theta_1\left(Y_t - Z_t'\beta\right) - \theta_1^2}{2\nu}\right)\right] - \frac{1}{2}\log(c\nu) - \frac{1}{2\nu}\left(Y_t - Z_t'\beta\right)^2,$$

where $c = 2 \cdot pi$.

The gradient of l_t evaluated at $(1, \gamma, \theta_0, \theta_*)$ contains

$$\frac{\partial}{\partial\pi} l_t = 1 - e^{b_t},$$

where $b_t = \left(Y_t - Z_t'\beta\right)\left(\frac{\theta_0 - \theta_*}{\nu}\right) - \left(\frac{\theta_0^2 - \theta_*^2}{2\nu}\right)$. The remaining elements of the gradient are

$$\frac{\partial}{\partial\nu} l_t = \frac{\left(Y_t - Z_t'\beta - \theta_*\right)^2}{2\nu^2} - \frac{1}{2\nu}, \frac{\partial}{\partial\beta} l_t = \frac{Z_t\left(Y_t - Z_t'\beta - \theta_*\right)^2}{\nu}, \frac{\partial}{\partial\theta_1} l_t = \frac{Y_t - Z_t'\beta - \theta_*}{\nu}.$$

We analyze the behavior of e^{b_t} in detail, as this forms the heart of the calculations for $\mathcal{I}(\theta_0) = \mathcal{I}(\theta_0, \theta_0)$. Further detail, covering the remaining calculations, can be found in Steigerwald and Carter (2011).

To determine the behavior of e^{b_t}, first note that because $\left(Y_t - Z_t'\beta\right) \sim N(\theta_*, \nu)$ the definition of a moment generating function yields $\mathbb{E}\left[\exp\left(\left(Y_t - Z_t'\beta\right)s\right)\right] = \exp\left(\theta_* s + \frac{1}{2}\nu s^2\right)$ for any real number s. Let $s = \frac{\theta_0 - \theta_*}{\nu}$, so $\exp\left(\theta_* s + \frac{1}{2}\nu s^2\right) = \exp\left(\frac{1}{2\nu}\left(\theta_0^2 - \theta_*^2\right)\right)$. Hence

$$\mathbb{E}\left[e^{b_t}\right] = \mathbb{E}\left[e^{\left(Y_t - Z_t'\beta\right)s - \frac{1}{2\nu}\left(\theta_0^2 - \theta_*^2\right)}\right] = 1.$$

In similar fashion, $2b_t = \left(Y_t - Z_t'\beta\right)\cdot 2s - \left(\frac{\theta_0^2 - \theta_*^2}{\nu}\right)$ and $\mathbb{E}\left[\exp\left(\left(Y_t - Z_t'\beta\right)\cdot 2s\right)\right] = \exp\left(\theta_* \cdot 2s + \frac{1}{2}\nu(2s)^2\right)$, hence

$$\mathbb{E}\left[e^{2b_t}\right] = e^{\frac{1}{\nu}\left(\theta_0^2 - \theta_*^2\right)}.$$

We also need to calculate $\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta\right)\right]$ and $\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta - \theta_*\right)^2\right]$. For the first quantity,

$$\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta\right)\right] = \int\left(y - Z_t'\beta\right)e^{b_t} c e^{-\frac{1}{2\nu}\left(y - Z_t'\beta - \theta_*\right)^2} dy,$$

where $c = (2pi\cdot\nu)^{-\frac{1}{2}}$. Note $e^{b_t} e^{-\frac{1}{2\nu}\left(y - Z_t'\beta - \theta_*\right)^2} = e^{-\frac{1}{2\nu}\left(y - Z_t'\beta - \theta_0\right)^2}$, so

$$\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta\right)\right] = \int\left(y - Z_t'\beta\right)c e^{-\frac{1}{2\nu}\left(y - Z_t'\beta - \theta_0\right)^2} dy = \theta_0.$$

For the second quantity,

$$\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta - \theta_*\right)^2\right] = \int\left(y - Z_t'\beta - \theta_*\right)^2 c e^{-\frac{1}{2\nu}\left(y - Z_t'\beta - \theta_0\right)^2} dy.$$

Because $\left(y - Z_t'\beta - \theta_*\right)^2 = \left(y - Z_t'\beta - \theta_0\right)^2 + 2\left(y - Z_t'\beta - \theta_0\right)\left(\theta_0 - \theta_*\right) + \left(\theta_0 - \theta_*\right)^2$,

$$\mathbb{E}\left[e^{b_t}\left(Y_t - Z_t'\beta - \theta_*\right)^2\right] = \nu + \left(\theta_0 - \theta_*\right)^2.$$

With these calculations in hand, the elements of the first row of $\mathcal{I}(\theta_0)$ are

$$\begin{array}{ll}
(1,1) & \mathbb{E}\left[1 - 2e^{b_t} + e^{2b_t}\right] = e^{\frac{1}{\nu}\left(\theta_0^2 - \theta_*^2\right)} - 1 \\
(1,2) & \mathbb{E}\left[\left(1 - e^{b_t}\right)\frac{1}{2\nu^2}\left(\left(Y_t - Z_t'\beta - \theta_*\right)^2 - \nu\right)\right] = -\frac{1}{2\nu^2}\left(\theta_0 - \theta_*\right)^2 \\
(1,3) & \mathbb{E}\left[\left(1 - e^{b_t}\right)\frac{1}{\nu}Z_t'\left(Y_t - Z_t'\beta - \theta_*\right)\right] = -\frac{Z_t'}{\nu}\left(\theta_0 - \theta_*\right) \\
(1,4) & \mathbb{E}\left[\left(1 - e^{b_t}\right)\frac{1}{\nu}\left(Y_t - Z_t'\beta - \theta_*\right)\right] = -\frac{1}{\nu}\left(\theta_0 - \theta_*\right).
\end{array}$$

Simultaneous Equations - Covariance Calculations

From the reduced form, the coefficient on the state variable, S_t, is $d = \delta A^{-1}\begin{pmatrix} 1 & 0 \end{pmatrix}^{\mathrm{T}}$ and the covariance matrix of the errors is $\Omega^{-1} = A^{-1}\Sigma\left(A^{-1}\right)^{\mathrm{T}}$ with $\Sigma = \begin{pmatrix} \nu_1 & \nu_{12} \\ \nu_{12} & \nu_2 \end{pmatrix}$. As detailed in Steigerwald and Carter, the covariance of the Gaussian process is

$$\mathbb{E}\left[\mathcal{G}(d_1)\mathcal{G}(d_2)\right] = \mathcal{V}(d_1)^{-\frac{1}{2}}\mathcal{V}(d_2)^{-\frac{1}{2}}\left[\exp\left(d_1^{\mathrm{T}}\Omega d_2\right) - 1 - d_1^{\mathrm{T}}\Omega d_2 - \frac{1}{2}\left(d_1^{\mathrm{T}}\Omega d_2\right)^2\right],$$

where $\mathcal{V}(d_1) = \exp\left(d_1^{\mathrm{T}}\Omega d_1\right) - 1 - d_1^{\mathrm{T}}\Omega d_1 - \frac{1}{2}\left(d_1^{\mathrm{T}}\Omega d_1\right)^2$.[6] The quantity $d_1^{\mathrm{T}}\Omega d_2$ simplifies as

$$\begin{aligned}
d_1^{\mathrm{T}}\Omega d_2 &= \delta_1\begin{pmatrix} 1 & 0 \end{pmatrix}\left(A^{-1}\right)^{\mathrm{T}}A^{\mathrm{T}}\Sigma^{-1}AA^{-1}\begin{pmatrix} 1 & 0 \end{pmatrix}^{\mathrm{T}}\delta_2 \\
&= \delta_1\begin{pmatrix} 1 & 0 \end{pmatrix}\Sigma^{-1}\begin{pmatrix} 1 & 0 \end{pmatrix}^{\mathrm{T}}\delta_2 = \frac{\delta_1\delta_2}{\nu_1\left(1 - \rho^2\right)}.
\end{aligned}$$

6.3 Pseudo-Code

Prior to the first iteration, the researcher must select the set H that contains η, the resolution of the grid of values in H (we recommend .001) and the number of normal random variables, J, used to approximate the Gaussian process covariance (we detail how to select J on page 12, for many applications $J = 150$ is sufficient). For each of $r = 1, \ldots, R$ iterations:

1. Generate $\{\epsilon_j\}_{j=0}^{J} \sim i.i.d.N(0,1)$
2. For each value of η in the grid mesh, construct $\mathcal{G}^A(\eta) = \left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-\frac{1}{2}}\sum_{j=3}^{J-1}\frac{\eta^j}{\sqrt{j!}}\epsilon_j$ (the equation for $\mathcal{G}^A(\eta)$ appears at the top of page 12)
3. Obtain $m_r = \max\left\{\left[\max\left(0, \epsilon_4\right)\right]^2, \max_\eta\left[\min\left(0, \mathcal{G}^A(\eta)\right)\right]^2\right\}$ (use of ϵ_4 is described at the top of page 11; the formula for m_r corresponds to the right side of (7))

This yields $\{m_r\}_{r=1}^{R}$. Let $\left\{m_{(r)}\right\}_{r=1}^{R}$ be the ordered values from which the critical value for a test with size 5 percent is $m_{[.95R]}$.

[6] If the errors are homoskedastic, so that $\nu_1 = \nu_2$, then the covariance contains an additional term, see Steigerwald and Carter for details.

References

[1] Carter, A. and D. Steigerwald, "Testing for Regime Switching: A Comment", *Economics Department Discussion Papers*, UC Santa Barbara 2010.

[2] Cecchetti, S., P. Lam and N. Mark, "Mean Reversion in Equilibrium Asset Prices", *American Economic Review*, 1990, **80**, 398-418.

[3] Cho, J. and H. White, "Testing for Regime Switching", *Econometrica*, 2007, **75**, 1671-1720.

[4] Garcia, R., "Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models", *International Economic Review*, 1998, **39**, 763-788.

[5] Ghosh, J. and P. Sen, "On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results", in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume 2, ed. by L. LeCam and R. Olshen, 1985, Belmont: Wadsworth Press, 789-806.

[6] Hansen, B., "The Likelihood Ratio Statistic Under Nonstandard Conditions: Testing the Markov-Switching Model of GNP", *Journal of Applied Econometrics,* 1992, **7**, S61-S82.

[7] Steigerwald, D. and A. Carter, "Technical Note to accompany Markov Regime-Switching Tests: Asymptotic Critical Values", *Economics Department Discussion Papers*, UC Santa Barbara, 2011.

[8] Theil, H., *Principles of Econometrics*, 1971, John Wiley: New York.

[9] van der Vaart, A., *Asymptotic Statistics*, 1998, Cambridge: New York.