The Stata Journal (2014)
14, Number 3, pp. 481–498

Obtaining critical values for test of Markov regime switching

Valerie K. Bostwick
Department of Economics
University of California, Santa Barbara
Santa Barbara, CA
vkbostwick@gmail.com

Douglas G. Steigerwald
Department of Economics
University of California, Santa Barbara
Santa Barbara, CA
doug@econ.ucsb.edu

Abstract. For Markov regime-switching models, a nonstandard test statistic must be used to test for the possible presence of multiple regimes. Carter and Steigerwald (2013, *Journal of Econometric Methods* 2: 25–34) derive the analytic steps needed to implement the Markov regime-switching test proposed by Cho and White (2007, *Econometrica* 75: 1671–1720). We summarize the implementation steps and address the computational issues that arise. We then introduce a new command to compute regime-switching critical values, `rscv`, and present it in the context of empirical research.

Keywords: st0347, rscv, Markov regime switching

1 Introduction

Markov regime-switching models are frequently used in economic analysis and are prevalent in fields such as finance, industrial organization, and business cycle theory. Unfortunately, conducting proper inference with these models can be exceptionally challenging. In particular, testing for the possible presence of multiple regimes requires the use of a nonstandard test statistic and critical values that may differ across model specifications.

Cho and White (2007) demonstrate that because of the unusually complicated nature of the null space, the appropriate measure for a test of multiple regimes in the Markov regime-switching framework is a quasi-likelihood-ratio (QLR) statistic. They provide an asymptotic null distribution for this test statistic from which critical values should be drawn. Because this distribution is a function of a Gaussian process, the critical values are difficult to obtain from a simple closed-form distribution. Moreover, the elements of the Gaussian process underlying the asymptotic null distribution are dependent upon one another. Thus the critical values depend on the covariance of the Gaussian process and, because of the complex nature of this covariance structure, are best calculated using numerical approximation. In this article, we summarize the steps necessary for such an approximation and introduce the new command `rscv`, which can be used to produce the desired regime-switching critical values for a QLR test of only one regime.

We focus on a simple linear model with Gaussian errors, but the QLR test and the `rscv` command are generalizable to a much broader class of models. This methodology can be applied to models with multiple covariates and non-Gaussian errors. It is also applicable to regime-switching models where the dependent variable is vector valued, although the difference between distributions must be in only one mean parameter. Although most regime-switching models are thought of in the context of time-series data, we provide an example in section 5 of how to use the QLR test in cross-section models. However, there is one notable restriction on the allowable class of regime-switching models. Carter and Steigerwald (2012) establish that the quasi-maximum likelihood estimator created using the quasi-log-likelihood is inconsistent if the covariates include lagged values of the dependent variable. Thus the QLR test should be used with extreme caution on autoregressive models.

The article is organized as follows. In section 2, we describe the unusual null space that corresponds to a test of only one regime versus the alternative of regime switching. In section 3, we present the QLR test statistic, as derived by Cho and White (2007), and the corresponding asymptotic null distribution. We also summarize the analysis in Carter and Steigerwald (2013) describing the covariance structure of the relevant Gaussian process. In section 4, we describe the methodology used by the `rscv` command to numerically approximate the relevant critical values. We also present the syntax and options of the `rscv` command and provide sample output. We illustrate the use of the `rscv` command with an application from the economics literature in section 5. Finally, we conclude in section 6 with some remarks on the general applicability of this command and the underlying methods.

2 Null hypothesis

Specifying a Markov regime-switching model requires a test to confirm the presence of multiple regimes. The first step is to test the null hypothesis of one regime against the alternative hypothesis of Markov switching between two regimes. If this null hypothesis can be rejected, then one can proceed to estimate the Markov regime-switching models with two or more regimes. The key to conducting valid inference is then a test of the null hypothesis of one regime, which yields an asymptotic size equal to or less than the nominal test size.

To understand how to conduct valid inference for the null hypothesis of only one regime, consider a basic regime-switching model,

$$y_t = \theta_0 + \delta s_t + u_t \tag{1}$$

where $u_t \sim$ i.i.d. $N\left(0,\sigma^2\right)$. The unobserved state variable $s_t \in (0,1)$ indicates that regime in state 0, y_t has mean θ_0, while regime in state 1, y_t has mean $\theta_1 = \theta_0 + \delta$. The sequence $(s_t)_{t=1}^n$ is generated by 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 key is to understand the parameter space that corresponds to the null hypothesis. Under the null hypothesis, there is one regime with mean θ_*. Hence, the null parameter space must capture all the possible regions that correspond to one regime. The first region corresponds to the assumption that $\theta_0 = \theta_1 = \theta_*$, which is the assumption that each of the two regimes is observed with positive probability: $p_0 > 0$

and $p_1 > 0$. The nonstandard feature of the null space is that it includes two additional regions, each of which also corresponds to one regime with mean θ_*. The second region corresponds to the assumption that only regime 0 occurs with positive probability, $p_0 = 0$, and that $\theta_0 = \theta_*$. In this second region, the mean of regime 1, θ_1 is not identified, so this region in the null hypothesis does not impose any value on $\theta_1 - \theta_0$. The third region is a mirror image of the second region, where now the assumption is that regime 1 occurs with probability 1: $p_1 = 0$ and $\theta_1 = \theta_*$. The three regions are depicted in figure 1. The vertical distance measures the value of p_0 and of p_1, and the horizontal distance measures the value of $\theta_1 - \theta_0$. Thus the vertical line at $\theta_1 = \theta_0$ captures the region of the null parameter space that corresponds to the assumption that $\theta_0 = \theta_1 = \theta_*$ together with $p_0, p_1 \in (0,1)$. The lower horizontal line captures the region of the null parameter space where $p_0 = 0$ and $\theta_1 - \theta_0$ is unrestricted. Similarly, the upper horizontal line captures the region of the null parameter space where $p_1 = 0$ and $\theta_1 - \theta_0$ is unrestricted.

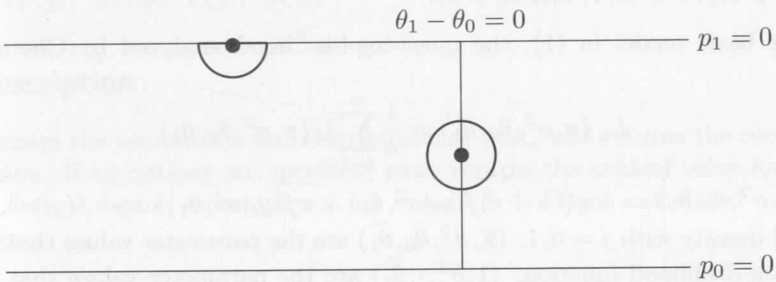

Figure 1. All three regions of the null hypothesis $H_0\colon p_0 = 0$ and $\theta_0 = \theta_*; p_1 = 0$ and $\theta_1 = \theta_*$; or $\theta_0 = \theta_1 = \theta_*$ together with local neighborhoods of $p_1 = 0$ and $\theta_0 = \theta_1 = \theta_*$

The additional curves that correspond to the values $p_0 = 0$ and $p_1 = 0$ help prevent one from misclassifying a small group of extremal values as a second regime. In figure 1, we depict the null space together with local neighborhoods for two points in this space. These two neighborhoods illustrate the different roles of the three curves in the null space. Points in the circular neighborhood of the point on $\theta_1 - \theta_0 = 0$ correspond to processes with two regimes that have only slightly separated means. Points in the semicircular neighborhood around the point on $p_1 = 0$ correspond to processes in which there are two regimes with widely separated means, one of which occurs infrequently. Because a researcher is often concerned that rejection of the null hypothesis of one regime is due to a small group of outliers rather than multiple regimes, including these boundary values reduces this type of false rejection. Consequently, a valid test of the null hypothesis of one regime must account for the entire null region and include all three curves.

3 QLR test statistic

To implement a valid test of the null hypothesis of one regime, a likelihood-ratio statistic is needed. When considering the likelihood-ratio statistic for a Markov regime-switching process, Cho and White (2007) find that including $p_0 = 0$ and $p_1 = 0$ in the parameter space creates significant difficulties in the asymptotic analysis. These difficulties lead them to consider a QLR statistic for which the Markov structure of the state variable is ignored and (s_t) is instead a sequence of independent and identically distributed (i.i.d.) random variables.

This i.i.d. restriction allows Cho and White (2007) to consider only the stationary probability, $\mathbb{P}(s_t = 1) = \pi$, where $\pi = p_0/(p_0 + p_1)$. Because $\pi = 1$ if and only if $p_1 = 0$ (and $\pi = 0$ if and only if $p_0 = 0$), the null hypothesis for a test of one regime based on the QLR statistic is 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), and $\pi = 1$ and $\theta_1 = \theta_*$ (curve 3). The alternative hypothesis is $H_1: \pi \in (0,1)$ and $\theta_0 \neq \theta_1$.

For our basic model in (1), the quasi-log-likelihood analyzed by Cho and White (2007) is

$$L_n(\pi, \sigma^2, \theta_0, \theta_1) = \frac{1}{n} \sum_{t=1}^{n} l_t(\pi, \sigma^2, \theta_0, \theta_1)$$

where $l_t(\pi, \sigma^2, \theta_0, \theta_1) := \log\{(1-\pi) f(y_t|\sigma^2, \theta_0) + \pi f(y_t|\sigma^2, \theta_1)\}$ and $f(y_t|\sigma^2, \theta_j)$ is the conditional density with $j = 0, 1$. $(\widehat{\pi}, \widehat{\sigma}^2, \widehat{\theta}_0, \widehat{\theta}_1)$ are the parameter values that maximize the quasi-log-likelihood function. $(1, \widetilde{\sigma}^2, \cdot, \widetilde{\theta}_1)$ are the parameter values that maximize L_n under the null hypothesis that $\pi = 1$. The QLR statistic is then

$$\mathrm{QLR}_n = 2n \left\{ L_n\left(\widehat{\pi}, \widehat{\sigma}^2, \widehat{\theta}_0, \widehat{\theta}_1\right) - L_n\left(1, \widetilde{\sigma}^2, \cdot, \widetilde{\theta}_1\right) \right\}$$

The asymptotic null distribution of QLR_n is (Cho and White 2007, theorem 6(b), 1692),

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

where $\mathcal{G}(\theta_0)$ is a Gaussian process, $\mathcal{G}(\theta_0)_- := \min\{0, \mathcal{G}(\theta_0)\}$, and G is a standard Gaussian random variable correlated with $\mathcal{G}(\theta_0)$. (For a more complete description of (2), see Bostwick and Steigerwald [2012]).

The critical value for a test based on the statistic QLR_n thus corresponds to a quantile for the largest value over $\max(0, G)^2$ and $\sup_\Theta \{\mathcal{G}(\theta_0)_-\}^2$. To determine this quantity, one must account for the covariance among the elements of $\mathcal{G}(\theta_0)$ as well as their covariance with G. The structure of this covariance, which is described in detail in Bostwick and Steigerwald (2012), is

$$\mathbb{E}\{\mathcal{G}(\theta_0)\mathcal{G}(\theta_0')\} = \frac{e^{\eta\eta'} - 1 - \eta\eta' - \frac{(\eta\eta')^2}{2}}{\left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{\frac{1}{2}} \left\{ e^{(\eta')^2} - 1 - (\eta')^2 - \frac{(\eta')^4}{2} \right\}^{\frac{1}{2}}} \tag{3}$$

where $\eta = (\theta_0 - \theta_*)/\sigma$ and $\eta' = (\theta_0' - \theta_*)/\sigma$. This covariance determines the quantity $\sup_\Theta\{\mathcal{G}(\theta_0)_-\}^2$ that appears in the asymptotic null distribution. Because the regime-specific parameters enter (3) only through η, a researcher does not need to specify the parameter space Θ to calculate $\sup_\Theta\{\mathcal{G}(\theta_0)_-\}^2$. The only requirement is to specify the set H that contains the number of standard deviations that separate the regime means. Finally, to fully capture the behavior of the asymptotic null distribution of QLR_n, we must also account for the covariance between G and $\mathcal{G}(\theta_0)$. Cho and White (2007) show that $\text{Cov}\{G, \mathcal{G}(\theta_0)\} = (e^{\eta^2} - 1 - \eta^2 - \eta^4/2)^{-1/2}\eta^4$.

4 The rscv command

4.1 Syntax

rscv [, ll(#) ul(#) r(#) q(#)]

4.2 Description

rscv simulates the asymptotic null distribution of QLR_n and returns the corresponding critical value. If no options are specified, rscv returns the critical value for a size 5% QLR test with a regime separation of ±1 standard deviation calculated over 100,000 replications.

4.3 Options

ll(#) specifies a lower bound on the interval H containing the number of standard deviations separating regime means, where $\eta \in H$. The default is ll(-1), meaning that the mean of regime 1 is no more than 1 standard deviation below the mean of regime 2.

ul(#) specifies an upper bound on the interval H containing the number of standard deviations separating regime means. The default is ul(1), meaning that the mean of regime 1 is no more than 1 standard deviation above the mean of regime 2.

r(#) specifies the number of simulation replications to be used in calculating the critical values. The default is r(100000), meaning that the simulation will be run 100,000 times.

q(#) specifies the quantile for which a critical value should be calculated. The default is q(0.95), which corresponds to a nominal test size of 5%.

4.4 Simulation process

For a QLR test with size 5%, the critical value corresponds to the 0.95 quantile of the limit distribution given on the right side of (2). Because the dependence in the process

$\mathcal{G}(\theta_0)$ renders numeric integration infeasible, we construct the quantile by simulating independent replications of the process. In this section, we describe the simulation process used to obtain these critical values and how each of the `rscv` command options affects those simulations.

Because the covariance of $\mathcal{G}(\theta_0)$ depends on only an index η, we do not need to simulate $\mathcal{G}(\theta_0)$ directly. Instead, we simulate $\mathcal{G}^A(\eta)$, which we will construct to have the same covariance structure as $\mathcal{G}(\theta_0)$. The process $\mathcal{G}^A(\eta)$ will therefore provide us with the correct quantile while relying solely on the index, η.

To construct $\mathcal{G}^A(\eta)$ for the covariance structure in (3), recall that by a Taylor-series expansion, $e^\eta = 1 + \eta + \eta^2/2! + \cdots$. Hence, for $(\epsilon_k)_{k=0}^{\infty} \sim$ i.i.d. $N(0,1)$,

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

Using this fact, our simulated process is constructed as

$$\mathcal{G}^A(\eta) = \left(e^{\eta^2} - 1 - \eta^2 - \frac{\eta^4}{2}\right)^{-\frac{1}{2}} \sum_{k=3}^{K-1} \frac{\eta^k}{\sqrt{k!}} \epsilon_k$$

where K determines the accuracy of the Taylor-series approximation. Note that the covariance of this simulated process, $\mathbb{E}\{\mathcal{G}^A(\eta)\mathcal{G}^A(\eta')\}$, is identical to the covariance structure of $\mathcal{G}(\theta_0)$ in (3).

We must also account for the covariance between G and $\mathcal{G}(\theta_0)$. Cho and White (2007) establish that this covariance corresponds to the term in the Taylor-series expansion for $k = 4$. Thus we set $G = \epsilon_4$ so that $\text{Cov}\{G, \mathcal{G}(\theta_0)\} = \text{Cov}\{G, \mathcal{G}^A(\eta)\}$. Therefore, the critical value that corresponds to (2) for a test size of 5% is the 0.95 quantile of the simulated value

$$\max\left(\{\max(0, \epsilon_4)\}^2, \max_{\eta \in H}\left[\min\{0, \mathcal{G}^A(\eta)\}\right]^2\right) \tag{4}$$

The `rscv` command executes the numerical simulation of (4) by first generating the series $(\epsilon_k)_{k=0}^{K} \sim$ i.i.d. $N(0,1)$. For each value in a discrete set of $\eta \in H$, it then constructs $\mathcal{G}^A(\eta) = (e^{\eta^2} - 1 - \eta^2 - \eta^4/2)^{-1/2} \sum_{k=3}^{K-1} \eta^k/\sqrt{k!}\epsilon_k$. The command then obtains the value $m_i = \max(\{\max(0, \epsilon_4)\}^2, \max_\eta[\min\{0, \mathcal{G}^A(\eta)\}]^2)$, corresponding to (2) for each replication (indexed by i). Let $(m_{[i]})_{i=1}^{r}$ be the vector of ordered values of m_i calculated in each replication. The command `rscv` returns the critical value for a test with size q from $m_{[(1-q)r]}$.

For each replication, `rscv` calculates $\mathcal{G}^A(\eta)$ at a fine grid of values over the interval H. To do so requires three quantities: the interval H (which must encompass the true value of η), the grid of values over H (given by the grid mesh), and the number of desired terms in the Taylor-series approximation, K. The user specifies the interval H using the `ll()` and `ul()` options. If θ_0 is thought to lie within 3 standard deviations of θ_1, the interval is $H = [-3.0, 3.0]$. Because the process is calculated at only a finite number of values, the accuracy of the calculated maximum increases as the grid mesh shrinks. Thus the command `rscv` implements a grid mesh of 0.01, as recommended in Cho and White (2007, 1693). For the interval $H = [-3.0, 3.0]$, and with a grid mesh of 0.01, the process is calculated at the points $(-3.00, -2.99, \dots, 3.00)$.

Given the grid mesh of 0.01 and the user-specified interval H, we must determine the appropriate value of K. To do so, we consider the approximation error, $\xi_{K,\eta} = (e^{\eta^2} - 1 - \eta^2 - \eta^4/2)^{-1/2} \sum_{k=K}^{\infty} \eta^k/\sqrt{k!}\epsilon_k$. We want to ensure that as K increases, the variance of $\xi_{K,\eta}$ decreases toward zero. Carter and Steigerwald (2013) show that for large K, $\text{var}(\xi_{K,\eta}) \leq e^{2J \log \eta - K \log K}$. Therefore, the command `rscv` implements a value of K such that for the user-specified interval H, $(\max_H |\eta|)^2/K \leq 1/2$.

The `rscv` command also allows the user to specify the number of simulation replications and the desired quantile. For large values of H and the default number of replications ($r = 100000$), the `rscv` command could require more memory than a 32-bit operating system can provide. In this case, the user may need to specify a smaller number of replications to calculate the critical values for the desired interval, H. Critical values derived using fewer simulation replications may be stable to only one significant digit. Table 1 depicts the results of `rscv` for a size 5% test over varying values of `ll()`, `ul()`, and `r()`.

Table 1. Critical values for linear models with Gaussian errors

	H	$(-1,1)$	$(-2,2)$	$(-3,3)$	$(-4,4)$	$(-5,5)$
Replications	100,000	4.9	5.6	6.2	6.7	7.0
	10,000	4.9	5.6	6.2	6.6	7.1

Nominal level 5%; grid mesh of 0.01.

5 Example

We demonstrate how to test for the presence of multiple regimes through an example from the economics literature. Unlike the simple model that we have considered until now, (1), the model in this example includes several added complexities that are commonly used in regime-switching applications. We describe how to construct the QLR test statistic for this more general model, how to use existing Stata commands to obtain the value of the test statistic, and, finally, how to use the new command, `rscv`, to obtain an appropriate critical value.

Our example is derived from Bloom, Canning, and Sevilla (2003), who test whether the large differences in income levels across countries are better explained by differences in intrinsic geography or by a regime-switching model where the regimes correspond to distinct equilibria. To this end, the authors use cross-sectional data to analyze the distribution of per capita income levels for countries with similar exogenous characteristics and test for the presence of multiple regimes.

Bloom, Canning, and Sevilla (2003) propose a model of switching between two possible equilibria. Regime 1 occurs with probability $p(x)$ and corresponds to countries that are in a poverty trap equilibrium.

$$y = \mu_1 + \beta_1 x + \epsilon_1 \ , \mathrm{Var}(\epsilon_1) = \sigma_1^2 \tag{5}$$

Regime 2 occurs with probability $1 - p(x)$ and corresponds to countries in a wealthy equilibrium.

$$y = \mu_2 + \beta_2 x + \epsilon_2 \ , \ \mathrm{Var}(\epsilon_2) = \sigma_2^2 \tag{6}$$

In both regimes, y is the log gross domestic product per capita, and x is the absolute latitude, which functions as a catchall for a variety of exogenous geographic characteristics. This model differs from a Markov regime-switching model in that the authors are looking at different regimes in a cross-section rather than over time. Thus the probability of being in either regime is stationary, and the unobserved regime indicator is an i.i.d. random variable. This modification corresponds exactly to that made by Cho and White (2007) to create the quasi-log-likelihood, so in this example, the log-likelihood ratio and the QLR are one and the same.

Note that this model is more general than the basic regime-switching model presented in section 2. Bloom, Canning, and Sevilla (2003) have allowed for three generalizations: covariates with coefficients that vary across regimes; error variances that are regime specific; and regime probabilities that depend on the included covariates. However, as Carter and Steigerwald (2013) discuss, the asymptotic null distribution (2) is derived under the following assumptions: that the difference between regimes be in only the intercept μ_j; that the variance of the error terms be constant across regimes; and that the regime probabilities do not depend on the exogenous characteristic, x. Thus, to form the test statistic, we must fit the following two-regime model: regime 1 occurs with probability p and corresponds to

$$y = \mu_1 + \beta x + \epsilon \tag{5'}$$

while regime 2, which occurs with probability $(1 - p)$, corresponds to

$$y = \mu_2 + \beta x + \epsilon \tag{6'}$$

where $\mathrm{Var}\,(\epsilon) = \sigma^2$.

Simplifying the model like this does not diminish the validity of the QLR as a one-regime test for the model in (5) and (6). Under the null hypothesis of one regime, there is necessarily only one error variance, only one coefficient for each covariate, and a regime probability equal to one. Thus, under the null hypothesis, the QLR test will necessarily have the correct size even if the data are accurately modeled by a more complex system.

Once the null hypothesis is rejected using this restricted model, the researcher can then fit a model with regime-specific variances and coefficients, if desired.[1]

For the restricted model in (5′) and (6′), the quasi-log-likelihood is

$$L_n\left(p, \sigma^2, \beta, \mu_1, \mu_2\right) = \frac{1}{n}\sum_{t=1}^{n} l_t\left(p, \sigma^2, \beta, \mu_1, \mu_2\right)$$

where $l_t(p, \sigma^2, \beta, \mu_1, \mu_2) := \log\{pf(y_t|x_t; \sigma^2, \beta, \mu_1) + (1-p)f(y_t|x_t; \sigma^2, \beta, \mu_2)\}$, and $f(y_t|x_t; \sigma^2, \beta, \mu_j)$ is the conditional density for $j = 1, 2$. It is common to assume, as Bloom, Canning, and Sevilla (2003) do, that ϵ is a normal random variable[2] so that $f(y_t|x_t; \sigma^2, \beta, \mu_j) = 1/(\sqrt{2\pi\sigma^2})e^{-(y_t-\mu_j-\beta x_t)^2/(2\sigma^2)}$. Let $(\widehat{p}, \widehat{\sigma}^2, \widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2)$ be the values that maximize L_n and let $(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot)$ be the values that make L_n as large as possible under the null hypothesis of one regime. The QLR statistic is then

$$\text{QLR}_n = 2n\left\{L_n\left(\widehat{p}, \widehat{\sigma}^2, \widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2\right) - L_n\left(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot\right)\right\}$$

To estimate QLR_n, we use the same Penn World Table and CIA World Factbook data as in Bloom, Canning, and Sevilla (2003).[3] First, we must determine the parameter values that maximize the quasi-log-likelihood under the null hypothesis, $(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot)$ and evaluate the quasi-log-likelihood at those values. To obtain these parameter values, we estimate a linear regression of y on x, which corresponds to maximizing

$$L_n\left(1, \sigma^2, \beta, \mu_1, \cdot\right) = \frac{1}{n}\sum_{t=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(y_t-\mu_1-\beta x_t)^2}\right)$$

While this can be achieved with a simple ordinary least-squares command, we also need the value of the log-likelihood, so we detail how to use Stata commands to obtain both the parameter estimates and this value.

1. With a more complex data-generating process, these restrictions could lead to an increased probability of failing to reject a false null hypothesis and, hence, a decrease in the power of the QLR test.
2. Bloom, Canning, and Sevilla (2003) assume normally distributed errors, but the QLR test allows for any error distribution within the exponential family.
3. Latitude data for countries appearing in the 1985 Penn World Tables and missing from the CIA World Factbook come from https://www.google.com/.

To find $(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot)$, we use the following code, which relies on the Stata command `ml`.

```
. program define llfsingle
  1. version 13
  2. args lnf mu beta sigma
  3. quietly replace `lnf´= (1/_N)*ln(((2*_pi*`sigma´^2)^(-1/2))*
> exp((-1/(2*`sigma´^2))*(lgdp-`mu´-`beta´*latitude)^2))
  4. end
. ml model lf llfsingle /mu /beta /sigma
. ml maximize
initial:       log likelihood =     -<inf>  (could not be evaluated)
feasible:      log likelihood =  -127.9261
rescale:       log likelihood = -31.297788
rescale eq:    log likelihood = -2.3397622
Iteration 0:   log likelihood = -2.3397622  (not concave)
Iteration 1:   log likelihood = -1.5884033  (not concave)
Iteration 2:   log likelihood = -1.2842957
Iteration 3:   log likelihood = -1.2479471
Iteration 4:   log likelihood = -1.1988284
Iteration 5:   log likelihood = -1.1982503
Iteration 6:   log likelihood = -1.1982487
Iteration 7:   log likelihood = -1.1982487
                                                Number of obs    =        152
                                                Wald chi2(0)     =          .
Log likelihood = -1.1982487                     Prob > chi2      =          .
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
mu						
_cons	6.927805	1.420095	4.88	0.000	4.144469	9.711141
beta						
_cons	.0408554	.049703	0.82	0.411	-.0565607	.1382714
sigma						
_cons	.8019654	.5670752	1.41	0.157	-.3094815	1.913412

```
. matrix gammasingle=e(b)
```

Then, using these estimates, we evaluate L_n at its maximum to find $L_n(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot)$.

```
. generate llf1regime=ln(((2*_pi*gammasingle[1,3]^2)^(-1/2))*
> exp((-1/(2*gammasingle[1,3]^2))*
> (lgdp-gammasingle[1,1]-gammasingle[1,2]*latitude)^2))
. quietly summarize llf1regime
. quietly replace llf1regime=r(sum)
. display "Final estimated quasi-log-likelihood for one regime: " llf1regime
Final estimated quasi-log-likelihood for one regime: -182.1338
```

Thus we have $n \times L_n(1, \widetilde{\sigma}^2, \widetilde{\beta}, \widetilde{\mu}_1, \cdot) = -182.1388$.

Second, we must determine the parameter values that maximize the quasi-log-likelihood under the alternative hypothesis of two regimes, $(\widehat{p}, \widehat{\sigma}^2, \widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2)$ and evaluate the quasi-log-likelihood at those values. Direct maximization is more difficult under the alternative hypothesis, because the quasi-log-likelihood involves the log of the sum of two terms.

$$L_n\left(p,\sigma^2,\beta,\mu_1,\mu_2\right)=\frac{1}{n}\sum_{t=1}^{n}\log\left\{pf\left(y_t|x_t;\sigma^2,\beta,\mu_1\right)+(1-p)\,f\left(y_t|x_t;\sigma^2,\beta,\mu_2\right)\right\}$$

The expectations-maximization (EM) algorithm is a method used to circumvent this difficulty. This algorithm requires iterative estimation of the latent regime probabilities, p, and maximization of the resultant log-likelihood function until parameter estimates converge. The EM algorithm proceeds as follows:

1. Choose starting guesses for the parameter values $p^{(0)},\sigma^{2(0)},\beta^{(0)},\mu_1^{(0)},\mu_2^{(0)}$.

2. For each observation, calculate $\eta_t=\mathbb{P}(s_t=1|y_t,x_t)$ such that

$$\widehat{\eta_t}=p^{(0)}\frac{f\left(y_t|x_t;\sigma^{2(0)},\beta^{(0)},\mu_1^{(0)}\right)}{p^{(0)}f\left(y_t|x_t;\sigma^{2(0)},\beta^{(0)},\mu_1^{(0)}\right)+\left(1-p^{(0)}\right)f\left(y_t|x_t;\sigma^{2(0)},\beta^{(0)},\mu_2^{(0)}\right)}$$

3. Use Stata's `ml` command to find the parameter values $p^{(1)},\sigma^{2(1)},\beta^{(1)},\mu_1^{(1)},\mu_2^{(1)}$ that maximize the complete log-likelihood.

$$\begin{aligned}L_n^C\left(p,\sigma^2,\beta,\mu_1,\mu_2\right)=\frac{1}{n}\sum_{t=1}^{n}&\left\{\widehat{\eta_t}\log f\left(y_t|x_t;\sigma^2,\beta,\mu_1\right)\right.\\&+\left(1-\widehat{\eta_t}\right)\log f\left(y_t|x_t;\sigma^2,\beta,\mu_2\right)\\&\left.+\left(1-\widehat{\eta_t}\right)\log(1-p)+\widehat{\eta_t}\log p\right\}\end{aligned}$$

4. To test for convergence, calculate

   a. $\max\left\{\left(p^{(1)},\sigma^{2(1)},\beta^{(1)},\mu_1^{(1)},\mu_2^{(1)}\right)-\left(p^{(0)},\sigma^{2(0)},\beta^{(0)},\mu_1^{(0)},\mu_2^{(0)}\right)\right\}$;

   b. $|L_n^C\left(p^{(1)},\sigma^{2(1)},\beta^{(1)},\mu_1^{(1)},\mu_2^{(1)}\right)-L_n^C\left(p^{(0)},\sigma^{2(0)},\beta^{(0)},\mu_1^{(0)},\mu_2^{(0)}\right)|$; and

   c. (using numeric derivatives) $\max(L_n^{C'})$.

5. If all 3 convergence criteria are less than some tolerance level (we use $1/n$), then quit and use $p^{(1)},\sigma^{2(1)},\beta^{(1)},\mu_1^{(1)},\mu_2^{(1)}$ as the final parameter estimates. Otherwise, repeat steps 2–5 with $p^{(1)},\sigma^{2(1)},\beta^{(1)},\mu_1^{(1)},\mu_2^{(1)}$ as the new starting guesses.

The following code illustrates the implementation of these steps to obtain ($\widehat{p}$, $\widehat{\sigma}^2$, $\widehat{\beta}$, $\widehat{\mu}_1$, and $\widehat{\mu}_2$).

```
. program define llfmulti
  1. version 13
  2. args lnf mu1 mu2 beta sigma p
  3. quietly replace `lnf´= (1/_N)*((1-etahat)*(ln((2*_pi*`sigma´^2)^(-1/2))+
> ((-1/(2*`sigma´^2))*(lgdp-`mu2´-`beta´*latitude)^2)+
> ln(1-`p´))+etahat*(ln((2*_pi*`sigma´^2)^(-1/2))+
> ((-1/(2*`sigma´^2))*(lgdp-`mu1´-`beta´*latitude)^2)+ln(`p´)))
  4. end
. generate error=10
. generate tol=1/_N
. while error>tol {
  2. quietly replace f1=((2*_pi*gammahat[1,4]^2)^(-1/2))*
> exp((-1/(2*gammahat[1,4]^2))*(lgdp-gammahat[1,1]-gammahat[1,3]*latitude)^2)
  3. quietly replace f2=((2*_pi*gammahat[1,4]^2)^(-1/2))*
> exp((-1/(2*gammahat[1,4]^2))*(lgdp-gammahat[1,2]-gammahat[1,3]*latitude)^2)
  4. quietly replace fboth=gammahat[1,5]*f1+(1-gammahat[1,5])*f2
  5. quietly replace etahat=gammahat[1,5]*f1/fboth
  6. ml model lf llfmulti /mu1 /mu2 /beta /sigma /p
  7. ml init gammahat, copy
  8. quietly ml maximize
  9. matrix gammanew=e(b)
 10. *Check for convergence using user-defined program nds
. nds
 11. quietly replace error=max(nd1,nd2,nd3,nd4,nd5)
 12. matrix gammahat=gammanew
 13. }
. ml display
                                                 Number of obs   =        152
                                                 Wald chi2(0)    =          .
Log likelihood = -1.4441013                      Prob > chi2     =          .
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
mu1						
_cons	6.532847	1.148891	5.69	0.000	4.281062	8.784632
mu2						
_cons	7.813265	1.45266	5.38	0.000	4.966102	10.66043
beta						
_cons	.0451607	.0374139	1.21	0.227	-.0281691	.1184905
sigma						
_cons	.5986278	.4232938	1.41	0.157	-.2310128	1.428268
p						
_cons	.7708245	.4203024	1.83	0.067	-.052953	1.594602

Using these estimates, we evaluate L_n at its maximum to find $L_n(\widehat{p}, \widehat{\sigma}^2, \widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2)$.

```
. quietly replace f1=((2*_pi*gammanew[1,4]^2)^(-1/2))*
> exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,1]-gammanew[1,3]*latitude)^2)
. quietly replace f2=((2*_pi*gammanew[1,4]^2)^(-1/2))*
> exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,2]-gammanew[1,3]*latitude)^2)
. generate lf2reg=gammanew[1,5]*f1+(1-gammanew[1,5])*f2
. generate llf2regime=ln(lf2reg)
. quietly summarize llf2regime
. quietly replace llf2regime=r(sum)
. display "Final estimated quasi-log-likelihood for two regimes: " llf2regime
Final estimated quasi-log-likelihood for two regimes: -179.9662
```

Thus we have $n \times L_n(\widehat{p}, \widehat{\sigma}^2, \widehat{\beta}, \widehat{\mu}_1, \widehat{\mu}_2) = -179.9662$. Then, to calculate the test statistic, QLR_n, we type

```
. generate QLR=2*(llf2reg-llf1reg)
. display "Quasi-likelihood-ratio test statistic of one regime: " QLR
Quasi-likelihood-ratio test statistic of one regime: 4.3352051
```

These estimates and the resulting QLR test statistic are summarized in table 2. For the complete Stata code used to create table 2, see the appendix.

Table 2. QLR test of one regime versus two regimes

	One regime	Two regimes	
		Regime I	Regime II
Constant (μ_1, μ_2)	6.928	6.533	7.813
Latitude (β)	0.041	0.045	
Standard deviation of error (σ)	0.802	0.599	
Probability of regime I (p)		0.771	
Log likelihood (L_n)	−182.1	−180.0	
QLR_n		4.3	

Finally, we use the `rscv` command to calculate the critical value for the QLR test of size 5%. We allow for the possibility that the two regimes are widely separated and set $H = (-5.0, 5.0)$. The command and output are shown below.

```
. rscv, ll(-5) ul(5) r(100000) q(0.95)
7.051934397
```

Given that this critical value of 7.05 exceeds the QLR statistic of 4.3, we cannot reject the null hypothesis of one regime.

This result is consistent with the findings of Bloom, Canning, and Sevilla (2003), although they use a different method to obtain the necessary critical values. They report a likelihood ratio and the corresponding critical values for a restricted version of their model where the regime probabilities are fixed (p does not depend on x). Using this restricted model, the authors do not reject the null hypothesis of one regime. At the time that Bloom, Canning, and Sevilla (2003) were published, researchers had yet to successfully derive the asymptotic null distribution for a likelihood-ratio test of regime switching. Therefore, the authors use Monte Carlo methods to generate their critical values using random data generated from the estimated relationship given by the model in (5) and (6). The primary disadvantage of this approach is that the derived critical values are then dependent upon the authors' assumptions concerning the underlying data-generating process.

Bloom, Canning, and Sevilla (2003) go on to report a likelihood-ratio test of a single regime model against the unrestricted model with latitude-dependent regime probabilities. With the unrestricted model, the authors can use the likelihood ratio and simulated critical values to reject the null hypothesis in favor of the alternative of two regimes. Because the null distribution derived by Cho and White (2007) applies to only the QLR constructed using the two-regime model given in (5′) and (6′), we cannot use the QLR test and, hence, the `rscv` command to obtain the critical values necessary to evaluate this unrestricted test statistic.

6 Discussion

We provide a methodology and a new command, `rscv`, to construct critical values for a test of regime switching for a simple linear model with Gaussian errors. Despite the complexity of the underlying methodology, `rscv` is relatively simple to execute and merely requires the researcher to provide a range for the standardized distance between regime means. In section 5, we demonstrate how these methods can be generalized to a very broad class of models, and we discuss the restrictions necessary to properly estimate the QLR statistic and use the `rscv` critical values.

7 References

Bloom, D. E., D. Canning, and J. Sevilla. 2003. Geography and poverty traps. *Journal of Economic Growth* 8: 355–378.

Bostwick, V. K., and D. G. Steigerwald. 2012. Obtaining critical values for test of Markov regime switching. Economics Working Paper Series qt3685g3qr, University of California, Santa Barbara. http://ideas.repec.org/p/cdl/ucsbec/qt3685g3qr.html.

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

———. 2013. Markov regime-switching tests: Asymptotic critical values. *Journal of Econometric Methods* 2: 25–34.

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

About the authors

Valerie Bostwick is currently completing a PhD in economics at the University of California, Santa Barbara.

Douglas G. Steigerwald joined the faculty of the Department of Economics at the University of California, Santa Barbara, after completing an MA in statistics and a PhD in economics at the University of California, Berkeley.

Appendix

The following Stata code was used to create table 2. The code fits the model in section 5 under the alternative hypothesis of two regimes using the EM algorithm and then under the null hypothesis of one regime using the Stata `ml` command. Finally, the QLR test statistic is calculated.

```
* Estimating QLR test statistic for Bloom, Canning, and Sevilla (2003)

* Log-likelihood function with two regimes
capture program drop llf
program define llf
version 13
args lnf theta1 theta0 delta sigma lambda
quietly replace `lnf'=(1/_N)*((1-etahat)*(ln((2*_pi*`sigma'^2)^(-1/2))      ///
   +((-1/(2*`sigma'^2))*(lgdp-`theta0'-`delta'*latitude)^2)+ln(1-`lambda')) ///
   +etahat*(ln((2*_pi*`sigma'^2)^(-1/2))+((-1/(2*`sigma'^2))*(lgdp-`theta1' ///
   -`delta'*latitude)^2)+ln(`lambda')))
end

* Log-likelihood function for one regime
capture program drop llfsingle
program define llfsingle
version 13
args lnf theta delta sigma
quietly replace `lnf'= (1/_N)*ln(((2*_pi*`sigma'^2)^(-1/2))* ///
   exp((-1/(2*`sigma'^2))*(lgdp-`theta'-`delta'*latitude)^2))
end

/**************************************************/
* First, estimate parameters and log likelihood for the case of two regimes:
* lgdp = theta0 + delta*latitude + u~N(0,sigma2) with probability (1-lambda)
* lgpp = theta1 + delta*latitude + u~N(0,sigma2) with probability lambda

/**************************************************/
* Start with initial guess for theta0, theta1, delta, sigma2, and lambda:
regress lgdp latitude
matrix beta=e(b)
svmat double beta, names(matcol)
scalar dhat=betalatitude
generate intercept=lgdp-dhat*latitude
summarize intercept
scalar t0hat=r(mean)-r(Var)
```

```
scalar t1hat=r(mean)+r(Var)
scalar shat=sqrt(r(Var))
scalar lhat=0.5
matrix gammahat=(t1hat, t0hat, dhat, shat, lhat)
display "Original guess for parameter values: "
matrix list gammahat

/************************************************/
* Start loop that continues until parameter estimates have converged
generate error1=10
generate error2=10
generate error3=10
generate tol=1/_N
generate count=0
generate count1=1
generate count2=1
generate count3=1
generate f1=0
generate f0=0
generate fboth=0
generate etahat=0
generate llfhat=0
generate llfnew=0
generate fdelta=0
generate fnew=0
generate lnllfnew=0
generate lnllfdelta=0
generate nd1=0
generate nd2=0
generate nd3=0
generate nd4=0
generate nd5=0

while error1>tol | error2>tol | error3>tol {

    * Calculate guess for eta_t=Pr(St=1|sample)
    * Calculate f(Yt|St=1, gammahat)
    quietly replace f1=((2*_pi*gammahat[1,4]^2)^(-1/2))*     ///
        exp((-1/(2*gammahat[1,4]^2))*(lgdp-gammahat[1,1]-gammahat[1,3]* ///
        latitude)^2)
    * Calculate f(Yt|St=0, gammahat)
    quietly replace f0=((2*_pi*gammahat[1,4]^2)^(-1/2))*     ///
        exp((-1/(2*gammahat[1,4]^2))*(lgdp-gammahat[1,2]-gammahat[1,3]* ///
        latitude)^2)
    * Calculate f(Yt|gammahat)
    quietly replace fboth=gammahat[1,5]*f1+(1-gammahat[1,5])*f0
    quietly replace etahat=gammahat[1,5]*f1/fboth

    /************************************************/
    * Now use etahat to create and maximize log-likelihood function

    ml model lf llf /theta1 /theta0 /delta /sigma /lambda
    ml init gammahat, copy
    ml maximize
    matrix gammanew=e(b)

    /************************************************/
    * Check whether the parameter estimates have converged
    mata: st_matrix("temp", max(abs(st_matrix("gammanew")-st_matrix("gammahat"))))
    quietly replace error1=temp[1,1]
```

```
    * Check whether the log likelihood has converged
    quietly replace llfnew=e(ll)
    quietly replace llfhat=(1/_N)*((1-etahat)                                  ///
        *(ln((2*_pi*gammahat[1,4]^2)^(-1/2))                                  ///
        +((-1/(2*gammahat[1,4]^2))                                             ///
        *(lgdp-gammahat[1,2]-gammahat[1,3]*latitude)^2)                       ///
        +ln(1-gammahat[1,5]))+etahat*(ln((2*_pi*gammahat[1,4]^2)^(-1/2))      ///
        +((-1/(2*gammahat[1,4]^2))                                             ///
        *(lgdp-gammahat[1,1]-gammahat[1,3]*latitude)^2)                       ///
        +ln(gammahat[1,5])))
    quietly summarize llfhat
    quietly replace llfhat=r(sum)
    quietly replace error2=abs(llfhat-llfnew)

    * Check whether the numeric derivative is zero
    * Recalculate incomplete log likelihood with new gamma estimates
    quietly replace f1=((2*_pi*gammanew[1,4]^2)^(-1/2))*      ///
        exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,1]-gammanew[1,3]*latitude)^2)
    quietly replace f0=((2*_pi*gammanew[1,4]^2)^(-1/2))*      ///
        exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,2]-gammanew[1,3]*latitude)^2)
    quietly replace fnew=gammanew[1,5]*f1+(1-gammanew[1,5])*f0
    quietly replace Inllfnew=log(fnew)
    quietly summarize Inllfnew
    quietly replace Inllfnew=r(sum)/_N
    * Calculate incomplete log likelihood for gamma + 0.0001
    forvalues i=1/5 {
        matrix gammadelta=gammanew
        matrix gammadelta[1,`i´]=gammadelta[1,`i´]+.0001
        quietly replace f1=((2*_pi*gammadelta[1,4]^2)^(-1/2)) ///
            *exp((-1/(2*gammadelta[1,4]^2))                     ///
            *(lgdp-gammadelta[1,1]-gammadelta[1,3]*             ///
            latitude)^2)
        quietly replace f0=((2*_pi*gammadelta[1,4]^2)^(-1/2)) ///
            *exp((-1/(2*gammadelta[1,4]^2))                     ///
            *(lgdp-gammadelta[1,2]-gammadelta[1,3]*             ///
            latitude)^2)
        quietly replace fdelta=gammadelta[1,5]*f1+(1-gammadelta[1,5])*f0
        quietly replace Inllfdelta=log(fdelta)
        quietly summarize Inllfdelta
        quietly replace Inllfdelta=r(sum)/_N
        quietly replace nd`i´=abs(Inllfdelta-Inllfnew)/.0001
    }
    quietly replace error3=max(nd1,nd2,nd3,nd4,nd5)

    /****************************************************/
    * Keep track of when each convergence criterion is met
    quietly replace count1=count1+1 if error1>tol
    quietly replace count2=count2+1 if error2>tol
    quietly replace count3=count3+1 if error3>tol

    * Update gammahat and overall iteration count
    matrix gammahat=gammanew
    quietly replace count=count+1

    * End of loop
}
```

```
/*************************************************/
* Calculate final log likelihood for two regimes
quietly replace f1=((2*_pi*gammanew[1,4]^2)^(-1/2))* ///
   exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,1]-gammanew[1,3]*latitude)^2)
quietly replace f0=((2*_pi*gammanew[1,4]^2)^(-1/2))* ///
   exp((-1/(2*gammanew[1,4]^2))*(lgdp-gammanew[1,2]-gammanew[1,3]*latitude)^2)
generate f2reg=gammanew[1,5]*f1+(1-gammanew[1,5])*f0
generate llf2reg=ln(f2reg)
quietly summarize llf2reg
quietly replace llf2reg=r(sum)
* Output final parameter estimates
display "Final estimated parameter values for two regimes: "
matrix list gammanew
display "Final estimated log likelihood for two regimes: " llf2reg
display "Total number of loop iterations: " count
display "Parameter values converged after " count1 " iterations"
display "Log likelihood value converged after " count2 " iterations"
display "Gradient of Log likelihood converged after " count3 " iterations"

/*************************************************/
* Second, estimate parameters and log likelihood for the case of only one regime:

* Maximize log likelihood with only one regime
* lgdp = theta + delta*lat + u~N(0,sigma2)
quietly summarize intercept
matrix gamma0=(r(mean), dhat, .1)
* Maximize to find new estimate of gamma
ml model lf llfsingle /theta /delta /sigma
ml init gamma0, copy
ml maximize
matrix gammasingle=e(b)

*Calculate log likelihood for one regime with estimated gamma
generate llf1reg=ln(((2*_pi*gammasingle[1,3]^2)^(-1/2))*                    ///
   exp((-1/(2*gammasingle[1,3]^2))*(lgdp-gammasingle[1,1]-gammasingle[1,2]* ///
   latitude)^2))
quietly summarize llf1reg
quietly replace llf1reg=r(sum)
* Output final parameter estimates
display "Final estimated parameter values for one regime: "
matrix list gammasingle
display "Final estimated log likelihood for one regime: " llf1reg

/*************************************************/
* Finally, calculate QLR test statistic:
generate QLR=2*(llf2reg-llf1reg)
display "Quasi-likelihood-ratio test statistic of one regime: " QLR
```