On the Power of Tests for Regime Switching

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

Benjamin Hansen
Department of Economics
University of Oregon

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

August 25, 2015

Abstract

This paper is concerned with the power of tests for regime switching. Asymptotic results are currently available only for the null distribution, and this distribution is dependent on the underlying parameter space. This paper addresses the lack of asymptotic results. The asymptotic behavior of the test statistic is determined under a full range of drifting sequences of true distributions. The results are based on a construction of the test statistic in terms of Hermite polynomials. The number of terms in the polynomial construction depends on the size of the parameter space. Using this relation, a collection of power curves, indexed by the sample space, is obtained. The finite sample properties of the test statistic are analyzed via Monte Carlo simulation.

JEL Classification: C12
Key Words: mixture model, parameter support, regime switching, test consistency

1

1 Introduction

The main contributions of this paper are as follows. (i) We provide a complete description of the asymptotic power for a constrained likelihood ratio test in (unobserved) regime switching models. (ii) We introduce a polynomial representation of the limiting Gaussian process. The polynomial terms correspond to moments of the underlying residuals and provide a natural interpretation of the standard series approximation. (iii) We analyze the behavior of the polynomial representation as the constrained parameter space changes. This provides a link between the size of the parameter space and the influence of higher order moments on the test statistic. (iv) We examine the finite sample performance of the test and compare it with standard tests based on skewness and kurtosis and with a test based on serial correlation.

The main technical innovations of the paper are the following. (i) To obtain the asymptotic distribution, we use a Hermite polynomial expansion. The effect of local alternatives is to shift the mean of the random variables that form the expansion. (ii) For the asymptotic power results, we do an expansion that varies over local alternatives. In consequence, for a sample size of n the test only has the power to detect local alternatives that are $n^{-\frac{1}{4}}$ distant from the null in certain neighborhoods.

We study models in which the null hypothesis is characterized not by a point but by two lines in the plane. Consider, for example, the case where the two regimes correspond to the Normal distributions, $\mathcal{N}\left(\mu, \sigma^2\right)$ and $\mathcal{N}\left(\mu + \delta, \sigma^2\right)$, which occur with respective probabilities $1 - \pi$ and π.[1] The null hypothesis is characterized by the set of all pairs (π, δ) such that $\pi\delta = 0$. The lack of uniqueness under the null hypothesis implies, for example, that π is not identified if $\delta = 0$.

We suppose $\theta = \pi\delta$ is estimated by maximizing a likelihood function $L_n\left(\theta, \gamma\right)$ over a parameter space $\Theta \times \Gamma$. For the above model, $\gamma = \left(\mu, \sigma^2\right)$. The null hypothesis of only a single regime is the following.

$$H_0 : \theta = 0.$$

Let Λ_n denote the likelihood ratio statistic.

In a local neighborhood of the null hypothesis, $L_n\left(\theta, \gamma\right)$ is relatively flat: (i) with respect to δ when $\pi = \frac{1}{2}$ and (ii) with respect to both π and δ when

---

[1] Carter and Steigerwald (2013) establish that the likelihood ratio statistic is a function of the scaled separation of the means. This implies that δ is the number of standard deviations that separate the means of the two regimes in models with unknown variance.

(π, δ) are both local to 0. This causes difficulties with standard asymptotic expansions because the terms corresponding to the second derivative are identically zero.[2]

This lack of uniqueness has ramifications for the likelihood ratio statistic. Hartigan (1985) argued that, if the parameter space for δ is the whole real line, then not only did the statistic not have a proper asymptotic distribution but that the statistic diverged to infinity with n, although at the slow rate of $\log \log n$. Hall and Stewart (2005) make a direct connection between the nonstandard behavior under the null hypothesis and loss of statistical power. We follow this logic in the model we study, for which the estimator $\widehat{\theta}_n$ is $n^{\frac{1}{2}}$ consistent over part of the parameter space, but converges more slowly (at the rate $n^{\frac{1}{3}}$ or $n^{\frac{1}{4}}$) in another part of the parameter space. In consequence, when we establish the large sample properties of a constrained likelihood ratio statistic over the full range of local alternatives we find that the asymptotic power of the associated test depends crucially on the curvature of $L_n(\theta, \gamma)$.

Our results apply to the constrained likelihood ratio test where $\delta \in [a, b]$. Although the constrained test does not have asymptotic optimality properties, the asymptotic null distribution of the statistic is established for the general case in which γ is unknown. The unconstrained test, in contrast, has an established asymptotic distribution only for the case in which γ is known.[3]

We introduce methods that link $[a, b]$ to the polynomial representation of the asymptotic distribution. As the magnitude of a or b increases, the number of terms in the polynomial expansion must increase in order to maintain a given level of approximation accuracy. As higher order terms in the expansion correspond to higher order moments of the data, increasing the parameter space increases the influence on the test statistic of moments beyond the measures of skewness and kurtosis. Moreover, the relation between the magnitude of a and b and the influence of higher order moments on the test statistic, provides information on selection of $[a, b]$ as a function of the sample size.

The remainder of the paper is organized as follows. In Section 2 we establish the asymptotic properties of the constrained likelihood ratio sta-

---

[2]Because the first derivative is identically 0 over the entire parameter space, the second derivative forms the leading term of the expansion. Because the second derivative vanishes in parts of the parameter space, the results of Andrews and Cheng (2012) do not apply.

[3]Liu and Shao (2004), who focus on the case where the mean and variance of the null distribution are known, establish that $\Lambda_n - \log \log n$ has an asymptotic null distribution that is an extreme value type.

tistic under local alternatives. In Section 3 we introduce the link between the parameter space and the polynomial approximation. Section 4 provides asymptotic and finite sample numerical results for common models of regime switching. Our results may show that the imposition of the boundary may lead to inconsistent estimators and a loss of power. Section 5 explains how to employ the method in an empirical setting explored in Hamilton (2014).

2 Asymptotic Power

To fix ideas, consider the model in which

$$Y_t = X_t^{\mathrm{T}} \beta + \delta S_t + U_t, \tag{1}$$

where X_t is a vector of suitably exogenous covariates (the first column of which corresponds to the intercept μ), $\{U_t\}$ is a sequence of $i.i.d.$ Gaussian random variables with mean 0 and variance σ^2, and the unobserved state variable $S_t \in \{0, 1\}$ indicates regimes. 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 $\mathbb{P}(S_t = 0|S_{t-1} = 1) = p_1$. Inclusion of the boundary values $p_0 = 0$ and $p_1 = 0$ in the null parameter space plays an important role by guarding against falsely classifying a small group of extremal values as a second regime. Their inclusion, however, when there is a Markov process for S_t, renders analysis of the likelihood intractable. To render asymptotic analysis tractable, we construct the test statistic from a quasi-likelihood ratio for which the Markov structure of the state variable is ignored and $\{S_t\}_{t=1}^n$ is a sequence of i.i.d. random variables. Under the i.i.d. restriction we need only consider the stationary probability $\mathbb{P}(S_t = 1) = \pi.$[4]

The key parameters for test of $H_0 : \theta = 0$ are (π, δ), the remaining parameters are (β, σ). One of the lines that characterizes the null hypothesis corresponds to $\delta = 0$, which represents two regimes that have the same mean and both of which can occur with positive probability. The second line corresponds to $\pi = 0$, which represents two regimes with different means, but one of which occurs with probability 1. There is a third case, which corresponds to the value $\pi = 1$, but as the asymptotic behavior of the test statistic when $\pi = 0$ is symmetric to the asymptotic behavior when $\pi = 1$, the full asymptotic behavior is captured by H_0.

---

[4]Kasahara, Okimoto and Shimotsu (2014) construct a modified likelihood-ratio test with a simpler asymptotic null distribution, but the simplification comes at the expense of ignoring the boundary values that guard against the false classification of extremal values.

The (quasi-) log-likelihood is thus a function of β, σ, π, and δ:

$$L\left(\beta,\sigma,\pi,\delta\right) = \sum_{t=1}^{n} \log\left[\phi\left(v_t\right)\right] + \sum_{t=1}^{n} \log\left[1 + \pi\left(e^{v_t\delta - \frac{1}{2}\delta^2} - 1\right)\right],$$

where $\phi\left(\cdot\right)$ is the standard Gaussian density function and $v_t = \frac{y_t - x_t^{\mathrm{T}}\beta}{\sigma}$ is the residual under the null hypothesis. To express the log-likelhood as a function of θ, which captures the full null hypothesis, we note that $\pi = \theta/\delta$ and follow Chen and Chen (2001) to write

$$L\left(\beta,\sigma,\theta,\delta\right) = \sum_{t=1}^{n} \log\left[\phi\left(v_t\right)\right] + \sum_{t=1}^{n} \log\left[1 + \theta Z_\delta\left(v_t\right)\right],$$

where $Z_\delta\left(v_t\right) := \frac{1}{\delta}\left(e^{v_t\delta - \frac{1}{2}\delta^2} - 1\right)$. The function $Z_\delta\left(\cdot\right)$, through which δ enters the log-likelihood, is akin to a sufficient statistic for δ.

If $\theta = 0$, then the MLE is the OLS estimator with standardized residual $\widetilde{v}_t$ and mean-squared error s^2. Thus

$$\max_{\beta,\sigma} L\left(\beta,\sigma,0,\delta\right) = -\frac{n}{2}\log\left(s^2\right) - \frac{1}{2}\sum_{t=1}^{n}\widetilde{v}_t^2 = -\frac{n}{2}\log\left(s^2\right) - \frac{n}{2}.$$

The test statistic is the likelihood ratio

$$
\begin{aligned}
Q_n &= 2\left[\max_{\Theta} L\left(\beta,\sigma,\theta,\delta\right) - L\left(b,s,0,\delta\right)\right] & (2)\\
&= \max n\log\left(\frac{s^2}{\sigma^2}\right) + \sum_{t=1}^{n}\left(1 - v_t^2\right) + 2\sum_{t=1}^{n}\log\left[1 + \theta Z_\delta\left(v_t\right)\right].
\end{aligned}
$$

We assume $\Theta = \Delta \times [0,1]$, where $\delta \in \Delta = [a,b]$.

Let $q_n\left(\beta,\sigma,\theta,\delta\right)$ be the value of the likelihood expression (2) as a function of $\left(\beta,\sigma,\theta,\delta\right)$. In a neighborhood of the null hypothesis, $q_n\left(\beta,\sigma,\theta,\delta\right)$ has three local maxima: the first occurs when maximizing over π for fixed δ, the second occurs when maximizing over δ for $\pi = \frac{1}{2}$, and the third occurs when maximizing over δ for fixed $\pi \neq \frac{1}{2}$. The three local maxima, and the associated behavior of the estimator $\widehat{\theta}$, can be understood through analysis of the derivatives of the likelihood. First consider

$$\frac{\partial}{\partial\pi}L = \sum_{t=1}^{n} \frac{e^{v_t\delta - \delta^2/2} - 1}{1 + \pi\left(e^{v_t\delta - \delta^2/2} - 1\right)}.$$

When evaluated at $\delta = 0$ the derivative is identically 0, reflecting the lack of identification of π when $\delta = 0$. For fixed $\delta \neq 0$, the derivative is not identically zero in a local neighborhood of $\pi = 0$, so standard asymptotics imply that $\widehat{\theta}$ is $n^{\frac{1}{2}}$-consistent for local alternatives where $\pi_n = \frac{h}{\sqrt{n}}$ and $\delta = \delta_*$. Next consider

$$\frac{\partial}{\partial \delta} L = \sum_{t=1}^{n} \frac{\pi (v_t - \delta) e^{v_t \delta - \delta^2/2}}{1 + \pi \left(e^{v_t \delta - \delta^2/2} - 1 \right)}.$$

When evaluated at $\pi = 0$ the derivative is identicaly 0, reflecting the lack of identification of δ when $\pi = 0$. For fixed $\pi \neq 0$, the derivative is identically zero in a local neighborhood of $\delta = 0$, because $\frac{\partial}{\partial \delta} L \big|_{\delta=0} = \pi \sum_{t=1}^{n} v_t$ and $\sum_{t=1}^{n} v_t = 0$, so standard asymptotics do not apply. Indeed, if $\pi \in \left[0, \frac{1}{2} \right)$ then $\widehat{\theta}$ is $n^{\frac{1}{3}}$-consistent and if $\pi = \frac{1}{2}$, then $\widehat{\theta}$ is $n^{\frac{1}{4}}$-consistent.

2.1 Local Maxima 1

For the local maxima in which $\widehat{\theta}$ is $n^{\frac{1}{2}}$-consistent,

$$H_{1,n} : \pi_n = \frac{h}{\sqrt{n}} \quad \delta = \delta_*.$$

As detailed in Carter and Steigerwald (2013) this local neighborhood captures the empirically relevant case in which a researcher is trying to avoid mistakenly classifying a small group of outliers as a second regime.

Under $H_{1,n}$, a Taylor expansion yields

$$q_n (\beta, \sigma, \theta, \delta) = n \log \left(\frac{s^2}{\sigma^2} \right) + \sum_{t=1}^{n} (1 - v_t)^2 + 2\theta \sum_{t=1}^{n} Z_\delta (v_t) - \theta^2 \sum_{t=1}^{n} Z_\delta (v_t)^2 + o_{\mathbb{P}} \left(n\theta^3 \right).$$

$$(3)$$

Lemma 1: *For each fixed value of δ we have that under both H_0 and $H_{1,n}$:*

$$\max_{\beta, \sigma, \theta} q_n (\beta, \sigma, \theta, \delta) = \frac{\left(\frac{\delta}{\sqrt{n}} \sum_{t=1}^{n} Z_\delta (\tilde{v}_t) \right)^2}{\exp \left(\delta^2 \right) - 1 - \delta^2 - \frac{\delta^4}{2}} + o_{\mathbb{P}} (1).$$

PROOF: See Appendix.

The behavior of $q_n (\cdot)$ is determined by the behavior of $Z_\delta (v_t)$. A key to our results is the representation of $Z_\delta (v_t)$ as an expansion of Hermite

6

polynomials $\{H_j(\cdot)\}_{j=1}^J$, which are defined in the Appendix. Importantly, the expansion holds under both the null and under a sequence of local alternatives.

Lemma 2: *If as $n \to \infty$, $J_n \to \infty$, and $\frac{J_n}{n} \to 0$, then for each δ:*

$$Z_\delta(v_t) = \sum_{j=3}^{J_n} \frac{\delta^{j-1}}{j!} H_j(v_t) + o_P(1).$$

PROOF: See Appendix.

A multivariate central limit theorem for triangular arrays then yields the asymptotic distribution for the scaled sum $\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(\tilde{v}_t)$ under both the null and a sequence of local alternatives.

Lemma 3: *If $v_t \sim \mathcal{N}(0,1)$, then under both H_0 and $H_{1,n}$:*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(\tilde{v}_t) \rightsquigarrow \mathcal{N}(m, V),$$

where $V = \delta^{-2}\left(\exp(\delta^2) - 1 - \delta^2 - \frac{\delta^4}{2}\right)$ and
i) under H_0, $m = 0$;
ii) under $H_{1,n}$, $m = \frac{h}{\delta}\left(\exp(\delta\delta_*) - 1 - \delta\delta_* - \frac{(\delta\delta_*)^2}{2}\right)$.

PROOF: See Appendix.

Here $Q_n = \sup_{\delta \in \Delta} q_n(\delta)$. Our main result yields the asymptotic distribution of the test statistic under both the null and a sequence of local alternatives.

Theorem 1: *If $v_t \sim \mathcal{N}(0,1)$, then:*

$$Q_n \rightsquigarrow \max\left((\max(0, G))^2, \sup_{\delta \in \Delta}(\min[0, \mathcal{G}(\delta)])^2\right),$$

where $V = \delta^{-2}\left(\exp(\delta^2) - 1 - \delta^2 - \frac{\delta^4}{2}\right)$ and
i) under H_0,
    a) $\mathcal{G}(\delta)$ is a correlated zero-mean Gaussian process with $\mathcal{G}(\delta) \sim \mathcal{N}(0,1)$,
    b) $G \sim \mathcal{N}(0,1)$ and is correlated with $\mathcal{G}(\delta)$;
ii) under $H_{1,n}$,
    a) $\mathcal{G}(\delta)$ is a correlated nonzero-mean Gaussian process with $\mathcal{G}(\delta) \sim \mathcal{N}\left(\frac{h}{\delta}\left(\exp(\delta\delta_*) - 1 - \delta\delta_* - \frac{(\delta\delta_*)^2}{2}\right), 1\right)$,

b) $G \sim \mathcal{N} \left(\frac{h \delta_*^3}{\sqrt{6}}, 1 \right)$ and is correlated with $\mathcal{G}(\delta)$.

PROOF: See Appendix.

Remarks: The asymptotic behavior is uniform over all δ outside an $\epsilon-$neighborhood of 0. Inside the neighborhood, the behavior depends on the value of π. If $\pi \neq \frac{1}{2}$, then the third Hermite polynomial remains the leading term in the expansion, but the rate of convergence of $\hat{\theta}$ slows from $n^{\frac{1}{2}}$ to $n^{\frac{1}{6}}$. If $\pi = \frac{1}{2}$, then the third Hermite polynomial vanishes (because the skewness is 0) and the fourth Hermite polynomial is the leading term in the expansion. Moreover, for this case, the rate of convergence of $\hat{\theta}$ slows further to $n^{\frac{1}{8}}$. The portion of the null space with the irregular behavior may not play a vital role in empirical applications as it corresponds to two regimes with nearly identical means. Also, with known γ, the score for δ evaluated under the null hypothesis, $\sum_{t=1}^{n} \left(y_t^2 - 1 \right)$, is not identically zero. Hence the likelihood has sufficient curvature to be locally approximated by a quadratic function and the results on weak identification in Andrews and Cheng (2012) hold. Importantly this means that the asymptotic distribution is described completely by considering only neighborhoods of π.

3 Parameter Space Selection

From the central limit theorem in Lemma 3, it follows that

$$\mathcal{G}(\delta) = \nu_\delta \sum_{j=3}^{\infty} \frac{\delta^j}{\sqrt{j!}} \eta_j \qquad \nu_\delta = \left(e^{\delta^2} - 1 - \delta^2 - \frac{\delta^4}{2} \right)^{-\frac{1}{2}}.$$

The random variable η_j has the same limiting distribution as H_j, where the Hermite polynomial H_j is the jth central moment of $\tilde{v}_t$. Importantly

$$\eta_j \overset{H_0}{\sim} \mathcal{N}(0,1) \qquad \eta_j \overset{H_{1,n}}{\sim} \mathcal{N} \left(h \delta_*^j, 1 \right).$$

In a local neighborhood of $\delta = 0$, the leading terms of the expansion dominate and

$$\mathcal{G}(\delta) \approx \nu_\delta \frac{\delta^3}{\sqrt{3!}} \eta_3 \quad \text{if } \pi \neq \frac{1}{2},$$

$$\mathcal{G}(\delta) \approx \nu_\delta \frac{\delta^4}{\sqrt{4!}} \eta_4 \quad \text{if } \pi = \frac{1}{2}.$$

Because $\mathcal{G}(\delta)$ is the asymptotic distribution of $q_n(\delta)$, we find that in a local neighborhood of $\delta = 0$ that the test statistic is equivalent to the skewness in

$\tilde{v}$ if $\pi \neq \frac{1}{2}$, and is equivalent to the kurtosis in $\tilde{v}$ if $\pi = \frac{1}{2}$. Thus in this local neighborhood, $q_n(\delta)$ is approximately the maximum of the sample excess skewness and excess kurtosis. As δ increases in magnitude, additional terms in the Hermite polynomial expansion grow in importance. As these additional terms correspond to higher-order moments, the behavior of $q_n(\delta)$ depends on higher-order moments for larger values of δ. To understand how these moments enter, Figure 1 contains a graph of the weights attached to each term in the expansion for $\mathcal{G}(\delta)$. These weights, which are given by $\nu_\delta \frac{\delta^j}{\sqrt{j!}}$, are graphed as a function of δ from which one can see that the peaks are at $\sqrt{j}$ for moderate values of j. Thus the leading curve corresponds to H_3 and reveals that near the origin, the skewness of the data is the dominant component.

The relationship between the values of δ and the weights on the terms in the expansion for the limiting process $\mathcal{G}(\delta)$, can be used to inform a researcher on how to select Δ The logic is, as Δ is increased, the maximizing value of $\hat{\delta}_n$ may be larger. The adequacy of the asymptotic approximation at these larger values increasingly depends on higher-order moments, which rely on larger samples for variance control. Thus the larger is Δ, the larger the sample size to ensure the adequacy of the asymptotic approximation across Δ. We note that even with a smaller value of Δ, say one that is adequately approximated by the first three terms in the expansion, the test statistic can still have substantial power to reject as one component that shifts the entire process away from the origin itself is shifted by $h\delta_*^3$.

4 Finite Sample Performance

We provide Monte Carlo evidence regarding two points: (i) the adequacy of the asymptotic approximation for finite samples and (ii) the power to detect alternatives with Q_n when the full LR test is not available. For (i) we employ the model

$$y_t = \mu + \delta s_t + u_t \qquad s_t \sim Bernoulli(\pi).$$

The key parameter values are set to resemble estimates for the growth rate of postwar quarterly real U.S. GDP as reported in Hamilton (2011), where $s_t = 1$ indicates a contraction: $\delta = -2$ and $\pi = .24$.[5] We also consider

---

[5]The average growth rate of expansions is 2 standard deviations above that of contractions, so $\mu = 0$ and $\sigma = 1$. The stationary probability corresponds to the estimated Markov probabilities: $\mathbb{P}(S_t = 0|S_{t-1} = 0) = .92$ and $\mathbb{P}(S_t = 1|S_{t-1} = 1) = .74$.

the closely related autoregression model, although we note that the OLS estimator used to construct $\widehat{\gamma}$ is only asymptotically equivalent to the MLE,

$$y_t = \mu + \delta s_t + \rho y_{t-1} + u_t \qquad s_t \sim Bernoulli\left(\pi\right).$$

For (ii) we employ the model

$$y_t = \mu + \delta s_t + x_t^{\mathrm{T}}\beta + s_t x_t^{\mathrm{T}}\alpha + u_t \qquad s_t \sim Bernoulli\left(\pi\right),$$

which can also inclued several other departures, namely

$$Var\left(u_t|s_t = i\right) = \sigma_i^2$$

and

$$\pi = f\left(x_t\right).$$

This latter generalization is likely to cause difficulties, as larger values of x both increase y directly and indirectly (through increasing the probability of state 1). Note that Q_n implicitly sets $\alpha = 0$, $\sigma_i^2 = \sigma^2$, and treats π as an unknown constant that does not depend on x. In this way, $L_n\left(\widehat{\theta}, \widehat{\delta}, \widehat{\gamma}\right)$ does not correspond to the unconstrained MLE for the full model. If the full model is used to estimate the unconstrained likelihood, then the asymptotic null distribution of the corresponding LR test is unknown. As our test is valid in these cases, we investigate the power of the test under these more complex models.

We should compare our test to the unconstrained QLR test, with critical values from Hall and Stewart, to simple shape tests, of the type listed below, and to the Carrasco, Hu, and Ploberger test. Perhaps the simplest way to understand what power we will have, is to look at the residuals from the OLS regressions (as these are the estimates under the null). The degree of skewness and kurtosis in these residuals will go a long way to determining the power of the test.

To determine the finite-sample performance of the QLR statistic using the subsample critical value, we compare the empirical size and power of the Q_n statistic against two other useful tests. If the errors follow the normal distribution, as they do for our data, then either skewness or kurtosis can reveal evidence of regime switching. The Bera-Jarque statistic is designed to test for skewness and excess kurtosis, and is given by

$$BJ_n = n\left(\frac{s_n^2}{6} + \frac{(k_n - 3)^2}{24}\right),$$

where the sample skewness s_n and kurtosis k_n are formed from the residuals generated by the null estimates $(0, \cdot, \tilde{\gamma})$. Under the null hypothesis, BJ_n has an asymptotic χ^2 distribution with 2 degrees of freedom.

While the Bera-Jarque statistic is widely used to test for regime switching, it is less well known that the $C(\alpha)$ statistic of Neyman and Scott can also be used to test for regime switching. The original $C(\alpha)$ statistic, which considers dispersion of the empirical distribution of the original data, must be modified to account for the zero second derivative of the log-likelihood. Cho and White construct the modified statistic

$$C(\alpha)_n = n \cdot \max \left[\frac{s_n^2}{6}, \min \left[0, \frac{k_n - 3}{24^{1/2}} \right]^2 \right].$$

The limit distribution of this $C(\alpha)$ statistic is $\max \left[Z_1^2, \min \left[0, Z_2 \right]^2 \right]$, where Z_1 and Z_2 are independent standard normal random variables.

5   Remarks

We establish the limiting distribution of Q_n under a sequence of local alternatives.

One could also determine the power to detect Markov switching. To determine the power of the test to detect regimes with Markov switching, we also consider specifications in which the latent regime is governed by

$$P\left(R_t = 1 | R_{t-1} = 1 \right) = \lambda_{22},$$

while $P\left(R_t = 0 | R_{t-1} = 0 \right) = \lambda_{11}$.

6   Appendix A: Proof of Results

PROOF OF LEMMA 1: To write (2) as a function of $\widetilde{v}_t$, by Taylor expansion

$$
\begin{aligned}
\sum Z_\delta(v_t) &= \sum Z_\delta(\tilde{v}_t) + \sum (v_t - \tilde{v}_t)\left[(1 + \delta Z_\delta(\tilde{v}_t)\right] + O\left(\sum(v_t - \tilde{v}_t)^2\right) \\
&= \sum Z_\delta(\tilde{v}_t) + \frac{1}{\sigma}\mathbf{1}^{\mathrm{T}}X(b - \beta) + \delta\left(\frac{s}{\sigma} - 1\right)\sum \tilde{v}_t Z_\delta(\tilde{v}_t) + \\
&\quad + \delta\left[\sum x_t Z_\delta(\tilde{v}_t)\right](b - \beta)/\sigma + O\left(\sum(v_t - \tilde{v}_t)^2\right),
\end{aligned}
$$

and

$$
\sum Z_\delta(v_t)^2 = \sum Z_\delta(\tilde{v}_t)^2 + O\left(\sum(v_t - \tilde{v}_t)Z_\delta(\tilde{v}_t)\right). \tag{5}
$$

Therefore, (2) is

$$
\begin{aligned}
q_n(\beta, \sigma, \theta, \delta) &= n\log\left[\frac{s^2}{\sigma^2}\right] + n\left(1 - \frac{s^2}{\sigma^2}\right) - \frac{1}{\sigma^2}(b - \beta)^{\mathrm{T}}X^{\mathrm{T}}X(b - \beta) + \\
&\quad + 2\theta\sum Z_\delta(\tilde{v}_t) + \frac{2\theta}{\sigma}\mathbf{1}^{\mathrm{T}}X(b - \beta) + 2\delta\theta\left(\frac{s}{\sigma} - 1\right)\sum \tilde{v}_t Z_\delta(\tilde{v}_t) - \theta^2\sum_{t=1}^{n} Z_\delta(\tilde{v}_t)^2 + \\
&\quad + 2\delta\theta\left[\sum x_t Z_\delta(\tilde{v}_t)\right](b - \beta)/\sigma + C_n
\end{aligned}
$$

where the error term in the approximation is on the order of $n\theta^3 + \theta\sum(v_t - \tilde{v}_t)^2 + \theta^2\sum(v_t - \tilde{v}_t)Z_\delta(\tilde{v}_t)$ from (3) and (4) and (5).

The properties of the OLS estimator, together with a Taylor expansion around $s = \sigma$, yields

$$
\max_{\beta,\sigma} q_n(\beta, \sigma, \theta, \delta) = n\theta^2 + \frac{n\delta^2\theta^2}{2}\left[\frac{1}{n}\sum \tilde{v}_t Z_\delta(\tilde{v}_t)\right]^2 + 2\theta\sum Z_\delta(\tilde{v}_t) - \theta^2\sum_{t=1}^{n} Z_\delta(\tilde{v}_t)^2 + C_n.
$$

This implies that under Assumption 1

$$
\sqrt{n}\hat{\theta} = \frac{\frac{1}{\sqrt{n}}\sum Z_\delta(\tilde{v}_t)}{\frac{1}{n}\sum_{i=1}^{n} Z_\delta(\tilde{v}_t)^2 - 1 - \frac{\delta^2}{2}} + o_{\mathbb{P}}(1)
$$

and

$$
\max_{\beta,\sigma,\theta} q_n(\beta, \sigma, \theta, \delta) = \frac{\left[\frac{\delta}{\sqrt{n}}\sum Z_\delta(\tilde{v}_t)\right]^2}{\exp(\delta^2) - 1 - \delta^2 - \frac{\delta^4}{2}} + o_P(1)
$$

as long as we can argue that the error term $C_n \xrightarrow{\mathbb{P}} 0$.

12

This error term includes the following approximations:

$$\delta\theta^2 \mathbf{1}^{\mathrm{T}} \vec{Z}_\delta = n\delta\theta^2 \left[\frac{1}{n}\sum_{t=1}^{n} Z_\delta(\tilde{v}_t)\right] \xrightarrow{\mathbb{P}} 0,$$

$$\theta^2\delta^2 \vec{Z}_\delta^{\mathrm{T}}\left(X\left[X^{\mathrm{T}}X\right]^{-1}X^{\mathrm{T}}\right)\vec{Z}_\delta \xrightarrow{\mathbb{P}} 0,$$

$$
\begin{aligned}
\theta\sum(v_t - \tilde{v}_t)^2 &= n\theta\left(\frac{s}{\sigma}-1\right)^2 + \theta\left(\frac{b-\beta}{\sigma}\right)^{\mathrm{T}} X^{\mathrm{T}}X\left(\frac{b-\beta}{\sigma}\right)^{\mathrm{T}} \\
&= \frac{n\theta^3\delta^2}{4}\left[\frac{1}{n}\sum\tilde{v}_t Z_\delta(\tilde{v}_t)\right]^2 + \theta^3\left(1+\delta\vec{Z}_\delta\right)^{\mathrm{T}} X\left[X^{\mathrm{T}}X\right]^{-1}X^{\mathrm{T}}\left(1+\delta\vec{Z}_\delta\right) \xrightarrow{\mathbb{P}} 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\theta^2\sum(v_t - \tilde{v}_t)Z_\delta(\tilde{v}_t) &= \theta^2\left(\frac{s}{\sigma}-1\right)\sum_{t=1}^{n}\tilde{v}_t Z_\delta(\tilde{v}_t) + \theta^2\sum_{t=1}^{n} Z_\delta(\tilde{v}_t)x_t^{\mathrm{T}}\left(\frac{b-\beta}{\sigma}\right) \\
&= \frac{n\delta\theta^3}{2}\left[\frac{1}{n}\sum\tilde{v}_t Z_\delta(\tilde{v}_t)\right]^2 + \theta^3\vec{Z}_\delta^{\mathrm{T}} X\left[X^{\mathrm{T}}X\right]^{-1}X^{\mathrm{T}}\left(1+\delta\vec{Z}_\delta\right) \xrightarrow{\mathbb{P}} 0,
\end{aligned}
$$

where we use $\vec{Z}_\delta$ as the column vector of the $Z_\delta(\tilde{v}_t)$ terms so that $\sum_{t=1}^{n} x_t Z_\delta(\tilde{v}_t) = \vec{Z}_\delta^{\mathrm{T}} X$. We also used these law of large number limits in the denominator,

$$\frac{1}{n}\sum\tilde{v}_t Z_\delta(\tilde{v}_t) \xrightarrow{\mathbb{P}} 1$$

$$\frac{1}{n}\sum_{t=1}^{n} Z_\delta(\tilde{v}_t)^2 \xrightarrow{\mathbb{P}} \frac{1}{\delta^2}\left(e^{\delta^2}-1\right).$$

$$Q.E.D.$$

PROOF OF LEMMA 2: Let $\hat{e}_t = \frac{y_t - \hat{\mu} - x_t^{\mathrm{T}}\hat{\beta}}{\hat{\sigma}}$ be the standardized residuals constructed from estimates of the full model with two regimes, so $\hat{\mu} = \bar{x} - \hat{\theta}$. Observe first that

$$L_n\left(\hat{\theta},\delta,\hat{\gamma}\right) = -\frac{n}{2}\ln\hat{\sigma}^2 - \frac{1}{2}\sum_{t=1}^{n}\hat{e}_t^2 + \frac{1}{2}\left[\frac{\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_\delta\left(\hat{e}_t\right)\right]^2}{\frac{1}{n}\sum_{t=1}^{n} Z_\delta^2\left(\hat{e}_t\right)}\right] + o_{\mathbb{P}}\left(1\right).$$

A Taylor series approximation yields

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_\delta\left(\hat{e}_t\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n} Z_\delta\left(e_t\right) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\hat{e}_t - e_t\right)\left[\delta Z_\delta\left(e_t\right)+1\right] + O\left(\left[\hat{e}_t - e_t\right]^2\right),$$

13

where $\frac{d}{dx} Z_\delta(x) = \exp\left(x\delta - \delta^2/2\right) = \delta Z_\delta(x) + 1$. Because our analysis is local to $\theta = 0$, the quantity $\hat\theta$ is in a local neighborhood of 0, implying $\sum_{t=1}^n Z_\delta\left(\frac{y_t - \mu - x_t^{\mathrm{T}}\beta}{\sigma}\right)$ lies in a local neighborhood of 0 and so the term $(\hat{e}_t - e_t) \cdot \delta Z_\delta(e_t)$ is negligible

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(\hat{e}_t) = \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(e_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^n (\hat{e}_t - e_t) + o_{\mathbb{P}}(1)$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(e_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{e}_t + o_{\mathbb{P}}(1),$$

because $\sum_{t=1}^n e_t = 0$.

In similar fashion

$$\frac{1}{n} \sum_{t=1}^n Z_\delta^2(\hat{e}_t) = \frac{1}{n} \sum_{t=1}^n Z_\delta^2(e_t) + \frac{2}{n} \sum_{t=1}^n (\hat{e}_t - e_t)\left[\delta Z_\delta^2(e_t) + Z_\delta(e_t)\right] + O\left([\hat{e}_t - e_t]^2\right)$$

$$= \frac{1}{n} \sum_{t=1}^n Z_\delta^2(e_t) + \frac{2}{n} \sum_{t=1}^n (\hat{e}_t - e_t) Z_\delta(e_t) + o_{\mathbb{P}}(1).$$

To show that $\frac{1}{n} \sum_{t=1}^n (\hat{e}_t - e_t) Z_\delta(e_t)$ is asymptotically negligible, by the Cauchy-Schwarz inequality

$$\left[\frac{1}{n} \sum_{t=1}^n (\hat{e}_t - e_t) Z_\delta(e_t)\right]^2 \le \left[\frac{1}{n} \sum_{t=1}^n (\hat{e}_t - e_t)^2\right]\left[\frac{1}{n} \sum_{t=1}^n Z_\delta^2(e_t)\right].$$

Further, there exists a constant vector c such that $(\hat{e}_t - e_t) \le c^{\mathrm{T}} v$ for all t, where $v^{\mathrm{T}} = \left(|\hat\mu - \tilde\mu|, \left|\hat\beta - \tilde\beta\right|, |\hat\sigma - \tilde\sigma|\right)$. Because $v \underset{\mathbb{P}}{\to} 0$ under both H_0 and $H_{1,n}$, $\left[\frac{1}{n} \sum_{t=1}^n (\hat{e}_t - e_t)^2\right] \underset{\mathbb{P}}{\to} 0$ and

$$\frac{1}{n} \sum_{t=1}^n Z_\delta^2(\hat{e}_t) = \frac{1}{n} \sum_{t=1}^n Z_\delta^2(e_t) + o_{\mathbb{P}}(1).$$

Therefore

$$L_n\left(\hat\theta, \delta, \hat\gamma\right) = -\frac{n}{2} \ln \hat\sigma^2 - \frac{1}{2} \sum_{t=1}^n \hat{e}_t^2 + \frac{1}{2}\left[\frac{\left[\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta(e_t) + \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{e}_t\right]^2}{\frac{1}{n} \sum_{t=1}^n Z_\delta^2(e_t)}\right] + o_{\mathbb{P}}(1).$$

To proceed, and replace the terms involving $\hat{e}_t$ from the likelihood, we employ the first-order conditions for $\hat{\gamma}$, which yield $\sum \hat{e}_t = \left[n^{-1} \sum Z_\delta^2 (e_t) - 1 \right]^{-1} \sum Z_\delta (e_t)$ and $\sum \hat{e}_t^2 = n + n^{-1} \left(\sum \hat{e}_t \right)^2$. Hence

$$L_n \left(\hat{\theta}, \delta, \hat{\gamma} \right) = -\frac{n}{2} \ln \hat{\sigma}^2 - \frac{n}{2} + \frac{1}{2} \left[\frac{\left[\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_\delta (e_t) \right]^2}{\frac{1}{n} \sum_{t=1}^n Z_\delta^2 (e_t) - 1} \right] + o_\mathbb{P} (1) .$$

$$Q.E.D.$$

We thus must understand the behavior of $Z_\delta \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right)$ to understand the behavior of $R_n (\delta)$. To do so, we use the expansion

$$e^{x\delta - \frac{1}{2}\delta^2} = \sum_{j=0}^\infty H_j (x) \frac{\delta^j}{j!},$$

where $H_j (\cdot)$ is the jth Hermite polynomial (defined in Appendix B). Because $H_0 (x) = 1$,

$$Z_\delta \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) = \sum_{j=1}^\infty H_j \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) \frac{\delta^{j-1}}{j!}.$$

By construction, for the standardized argument $\frac{y_t - \hat{\mu}}{\hat{\sigma}}$, $H_1 \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) = H_2 \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) = 0$. In consequence

$$Z_\delta \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) = \sum_{j=3}^\infty H_j \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) \frac{\delta^{j-1}}{j!}.$$

Because

$$\frac{1}{n} \sum_{j=1}^\infty \frac{\delta^{j-1}}{j!} H_j \left(\frac{y_t - \mu}{\sigma} \right) = \frac{1}{n} \sum_{j=1}^J \frac{\delta^{j-1}}{j!} H_j \left(\frac{y_t - \mu}{\sigma} \right) + O_\mathbb{P} (1) ,$$

we then have

$$R_{n,J} (\delta) = \left[\frac{\sum_{j=1}^J \left(n^{-1/2} \sum_{t=1}^n H_j \left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) \right) \frac{\delta^{j-1}}{j!}}{\left(\frac{1}{n\delta^2} \sum_{t=1}^n \left[\exp \left(\left(\frac{y_t - \hat{\mu}}{\hat{\sigma}} \right) \delta - \frac{\delta^2}{2} \right) - 1 \right]^2 \right)^{1/2}} \right]^2 + O_\mathbb{P} (1) ,$$

under both H_0 and $H_{1,n}$.

15

PROOF OF THEOREM 2A: Under the null hypothesis $\frac{Y_t - \mu}{\sigma} \sim N(0,1)$, so the properties of the moment generating function imply:

$$\mathbb{E}\left[Z_\delta\left(\frac{Y_t - \mu}{\sigma}\right)\right] = 0,$$

$$Var\left(Z_\delta\left(\frac{Y_t - \mu}{\sigma}\right)\right) = \frac{e^{\delta^2} - 1}{\delta^2},$$

$$Cov\left(Z_{\delta_1}\left(\frac{Y_t - \mu}{\sigma}\right), Z_{\delta_2}\left(\frac{Y_t - \mu}{\sigma}\right)\right) = \frac{e^{\delta_1\delta_2} - 1}{\delta_1\delta_2}.$$

Hence by a pointwise law of large numbers, for each δ:

$$\left(\frac{1}{n\delta^2}\sum_{t=1}^{n}\left[\exp\left(y_t\delta - \frac{\delta^2}{2}\right) - 1\right]^2\right) \xrightarrow[H_0]{\mathbb{P}} \frac{e^{\delta^2} - 1}{\delta^2}.$$

Under the alternative hypothesis $\frac{Y_t - \mu}{\sigma} \sim N(0,1)$ with probability $1 - \frac{h}{\sqrt{n}}$ and $\frac{Y_t - \mu}{\sigma} \sim N(\delta^*, 1)$ with probability $\frac{h}{\sqrt{n}}$, so the properties of the moment generating function imply:

$$\mathbb{E}\left[Z_\delta\left(\frac{Y_t - \mu}{\sigma}\right)\right] = \frac{h}{\sqrt{n}}\frac{\left(e^{\delta^*\delta} - 1\right)}{\delta},$$

$$Var\left(Z_\delta\left(\frac{Y_t - \mu}{\sigma}\right)\right) = \left(1 - \frac{h}{\sqrt{n}}\right)\frac{\left(e^{\delta^2} - 1\right)}{\delta^2} + \frac{h}{\sqrt{n}}\frac{\left(e^{\delta^2 + 2\delta^*\delta} - 2e^{\delta^*\delta} + 1\right)}{\delta^2}.$$

Because $\frac{h}{\sqrt{n}} \to 0$, a pointwise law of large numbers implies, for each δ:

$$\left(\frac{1}{n\delta^2}\sum_{t=1}^{n}\left[\exp\left(y_t\delta - \frac{\delta^2}{2}\right) - 1\right]^2\right) \xrightarrow[H_1]{\mathbb{P}} \frac{e^{\delta^2} - 1}{\delta^2}.$$

PROOF OF THEOREM 3A:

Part i) The results in Appendix B on Hermite polynomials imply that under the null hypothesis $H_j(Y_t)$ is a random variable with mean zero and variance $j!$. Hence, given j, a pointwise central limit theorem applies and for each δ:

$$n^{-1/2}\sum_{t=1}^{n}H_j(Y_t) \underset{H_0}{\rightsquigarrow} \mathcal{N}(0, j!).$$

Part *ii*) The results in Appendix B on shifted Hermite polynomials imply that under the alternative hypothesis $H_j(Y_t)$ is a random variable with mean $\pi_n \delta_*^j$ and variance $j! \left(1 + \pi_n \sum_{m=1}^j \begin{pmatrix} j \\ m \end{pmatrix} \frac{\delta_*^{2m}}{m!} - \pi_n^2 \frac{\delta_*^{2j-2}}{j!} \right)$. This implies that for $\pi_n = \frac{h}{\sqrt{n}}$:

$$ n^{-1/2} \sum_{t=1}^n \left[H_j(Y_t) - h\delta_*^j \right] \rightsquigarrow \mathcal{N}(0, j!), $$

which in turn implies

$$ n^{-1/2} \sum_{t=1}^n H_j(Y_t) \underset{H_1}{\rightsquigarrow} \mathcal{N}\left(h\delta_*^j, j! \right). $$

<div align="right">Q.E.D.</div>

PROOF OF THEOREM 4A:

Part *i*) From the results in Appendix B for Hermite polynomials $\left\{ \frac{H_j(Y_t)}{\sqrt{j!}} \right\}_{j \geq 1}$ is a sequence of uncorrelated random variables, so Theorem 3A part *i* together with a multivariate central limit theorem yields

$$ \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{H_j(Y_t)}{\sqrt{j!}} \right\}_{j=1}^J \underset{H_0}{\rightsquigarrow} \left\{ \xi_j \right\}_{j=1}^J, $$

where $\left\{ \xi_j \right\}_{j=1}^J$ is a sequence of independent $\mathcal{N}(0,1)$ random variables. By Theorem 2A and Slutsky's Lemma we have

$$ \frac{\sum_{j=1}^J \frac{\delta^{j-1}}{\sqrt{j!}} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{H_j(Y_t)}{\sqrt{j!}}}{\left(\frac{1}{n\delta^2} \sum_{t=1}^n Z_\delta^2(Y_t) \right)^{\frac{1}{2}}} \underset{H_0}{\rightsquigarrow} \sum_{j=1}^J \left(\frac{\delta^j}{\sqrt{j!}\sqrt{e^{\delta^2}-1}} \right) \xi_j $$

pointwise for each value of δ in Δ.

Part *ii*) From the results in Appendix B for shifted Hermite polynomials $\left\{ \frac{H_j(Y_t)}{\sqrt{j!}} \right\}_{j \geq 1}$ is a sequence of correlated random variables, so Theorem 3A part *ii* together with a multivariate central limit theorem yields

$$ \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{H_j(Y_t)}{\sqrt{j!}} \right\}_{j=1}^J \underset{H_{1,n}}{\rightsquigarrow} \left\{ \xi_j \right\}_{j=1}^J, $$

<div align="center">17</div>

where $\{\xi_j\}_{j=1}^J$ is a sequence of independent $\mathcal{N}\left(h\delta_*^j, j!\right)$ random variables. By Theorem 2A and Slutsky's Lemma we have

$$\frac{\sum_{j=1}^J \frac{\delta^{j-1}}{\sqrt{j!}} \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{H_j(Y_t)}{\sqrt{j!}}}{\left(\frac{1}{n\delta^2}\sum_{t=1}^n Z_\delta^2(Y_t)\right)^{\frac{1}{2}}} \underset{H_{1,n}}{\rightsquigarrow} \sum_{j=1}^J \left(\frac{\delta^j}{\sqrt{j!}\sqrt{e^{\delta^2}-1}}\right)\xi_j$$

pointwise for each value of δ in Δ.

<div align="right">Q.E.D.</div>

PROOF OF THEOREM 5A:

Part i) From Theorem 4A we have

$$R_{n,J}(\delta) \underset{H_o}{\rightsquigarrow} \mathcal{G}_J(\delta),$$

where $\mathcal{G}_J(\delta) := \sum_{j=1}^J \left(\frac{\delta^j}{\sqrt{j!}\sqrt{e^{\delta^2}-1}}\right)\xi_j$. Further, if $J \to \infty$ and $\frac{J}{n} \to 0$, we have

$$\mathbb{E}\left[\mathcal{G}(\delta) - \mathcal{G}_J(\delta)\right]^2 \underset{\mathbb{P}}{\to} 0.$$

Finally, observe that θ must be between 0 and δ, so if $\sum_{t=1}^n Z_\delta(Y_t)$ differs in sign from δ then $\widehat{\theta} = 0$. Because the sign of $\sum_{t=1}^n Z_\delta(Y_t)$ corresponds to the sign of $\mathcal{G}(\delta)$, we have

$$R_n(\delta) \underset{H_0}{\rightsquigarrow} \begin{cases} \left[\max\left(0, \mathcal{G}(\delta)\right)\right]^2 & \text{if } \delta > 0 \\ 0 & \text{if } \delta = 0 \\ \left[\min\left(0, \mathcal{G}(\delta)\right)\right]^2 & \text{if } \delta < 0 \end{cases}.$$

From the moment calculations for $Z_\delta(Y_t)$ we have

$$Corr\left(\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{\delta_1}(Y_t)}{\left(\frac{1}{n}\sum_{t=1}^n Z_{\delta_1}^2(Y_t)\right)^{\frac{1}{2}}}, \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{\delta_2}(Y_t)}{\left(\frac{1}{n}\sum_{t=1}^n Z_{\delta_2}^2(Y_t)\right)^{\frac{1}{2}}}\right) = \frac{\mathbb{E}\left(Z_{\delta_1}(Y_t) Z_{\delta_2}(Y_t)\right)}{\left(\mathbb{E}\left[Z_{\delta_1}^2(Y_t)\right]\mathbb{E}\left[Z_{\delta_2}^2(Y_t)\right]\right)^{\frac{1}{2}}}$$

$$= \frac{e^{\delta_1\delta_2}-1}{\left(e^{\delta_1^2}-1\right)^{\frac{1}{2}}\left(e^{\delta_2^2}-1\right)^{\frac{1}{2}}} \frac{\sqrt{\delta_1^2\delta_2^2}}{\delta_1\delta_2},$$

with $\frac{\sqrt{\delta_1^2\delta_2^2}}{\delta_1\delta_2} = sign(\delta_1\delta_2)$. An implication of the multivariate central limit theorem is

$$Cov\left(\frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{\delta_1}(Y_t)}{\left(\frac{1}{n}\sum_{t=1}^n Z_{\delta_1}^2(Y_t)\right)^{\frac{1}{2}}}, \frac{\frac{1}{\sqrt{n}}\sum_{t=1}^n Z_{\delta_2}(Y_t)}{\left(\frac{1}{n}\sum_{t=1}^n Z_{\delta_2}^2(Y_t)\right)^{\frac{1}{2}}}\right) \underset{\mathbb{P}}{\to} Cov\left(\mathcal{G}(\delta_1), \mathcal{G}(\delta_2)\right).$$

Part $ii)$ The argument is identical to part i with the exception of the logic for

$$\mathbb{E}\left[\mathcal{G}\left(\delta\right)-\mathcal{G}_J\left(\delta\right)\right]^2 \underset{\mathbb{P}}{\to} 0.$$

Because the Hermite polynomials are correlated under the alternative hypothesis, and because there are n^2-n such terms, the convergence requires $J=o\left(\sqrt{n}\right)$.

$$Q.E.D.$$

PRESENTATION OF LEMMA A1

The key component of the limit distribution is the correlation structure of the Gaussian process. Let δ_1 and δ_2 be two distinct values, then as Carter and Steigerwald (2013) detail

$$Cov\left[\mathcal{G}\left(\delta_1\right),\mathcal{G}\left(\delta_2\right)\right]=\frac{e^{\delta_1\delta_2}-1-\delta_1\delta_2-\frac{(\delta_1\delta_2)^2}{2}}{\left(e^{\delta_1^2}-1-\delta_1^2-\frac{\delta_1^4}{2}\right)^{\frac{1}{2}}\left(e^{\delta_2^2}-1-\delta_2^2-\frac{\delta_2^4}{2}\right)^{\frac{1}{2}}}.$$

To verify that the likelihood ratio diverges asymptotically we establish that the covariance between two widely separated elements of the process converges to zero.

Lemma A1: *For any fixed value δ_1,*

$$\lim_{\delta_2\to\infty} Cov\left[\mathcal{G}\left(\delta_1\right),\mathcal{G}\left(\delta_2\right)\right]=0.$$

PROOF: If we multiply $Cov\left[\mathcal{G}\left(\delta_1\right),\mathcal{G}\left(\delta_2\right)\right]$ by $e^{-\delta_2^2/2}/e^{-\delta_2^2/2}$, then for the numerator we have

$$\lim_{\delta_2\to\infty}\left(\frac{e^{\delta_1\delta_2}}{\sqrt{e^{\delta_2^2}}}-\frac{1}{\sqrt{e^{\delta_2^2}}}-\frac{\delta_1\delta_2}{\sqrt{e^{\delta_2^2}}}-\frac{(\delta_1\delta_2)^2}{2\sqrt{e^{\delta_2^2}}}\right)=0,$$

and for the denominator we have

$$\lim_{\delta_2\to\infty}\left(e^{\delta_1^2}-1-\delta_1^2-\frac{\delta_1^4}{2}\right)^{\frac{1}{2}}\left(1-\frac{1}{e^{\delta_2^2}}-\frac{\delta_2^2}{e^{\delta_2^2}}-\frac{\delta_2^4}{2e^{\delta_2^2}}\right)^{\frac{1}{2}}=\left(e^{\delta_1^2}-1-\delta_1^2-\frac{\delta_1^4}{2}\right)^{\frac{1}{2}},$$

which follow from the basic properties of the exponential function that $\lim_{x\to\infty} x^k/\sqrt{e^{x^2}}=0$ and $\lim_{x\to\infty} e^x/\sqrt{e^{x^2}}=0$ for any k.

$$Q.E.D.$$

Thus for any δ_1 it is possible to select δ_2 to make the correlation between these elements of the process arbitrarily small. Also, for any k it is possible

to select a set of values $\{\delta_1, \ldots, \delta_k\}$ to make all the pair wise correlations arbitrarily small. This is the essence of the argument in Hartigan and implies that $\mathcal{G}(\cdot)$ contains infinitely many asymptotically independent components. In consequence, for any given $\varepsilon \in (0, 1)$ and positive number M, we can select $\{\delta_i\}_{i=1}^{k(\varepsilon, M)}$ such that $\mathbb{P}\left(\sup_{\delta \in \{\delta_i\}_{i=1}^{k(\varepsilon, M)}} |\mathcal{G}(\delta)| > M\right) > 1 - \varepsilon$. Because M can take an arbitrarily large value, it follows that the supremum of $|\mathcal{G}(\delta)|$ over the real line diverges to infinity.

PROOF OF THEOREM 6A:

Parts i and ii) Because $\dfrac{\delta^j}{\sqrt{j!\left(e^{\delta^2} - 1\right)}}$ is infinitely differentiable it is uniformly continuous in δ, so the convergence in Theorem 5A is uniform over Δ.

PROOF OF RESULT 3.2: The full asymptotic theory requires that we analyze the asymptotic behavior of Q_n in neighborhoods of the null space $\theta = 0$ corresponding to both $\pi = 0$ and $\delta = 0$. If we fix δ, then a local neighborhood of $\theta = 0$ corresponds to a local neighborhood of $\pi = 0$. If we let δ shrink to zero at a certain rate, then a local neighborhood of $\theta = 0$ corresponds to π fixed in a local neighborhood of $\delta = 0$.

To begin we consider the limiting behavior of the likelihood ratio for a fixed value of δ. With $\sigma^2 = 1$ the log-likelihood, ignoring the constant in the Gaussian density, is

$$L_\delta(\pi, \mu) = -\frac{1}{2} \sum_{t=1}^n (y_t - \mu)^2 + \sum_{t=1}^n \log\left[1 + \theta Z_\delta(y_t - \mu)\right],$$

where $Z_\delta(y_t - \mu) := \frac{1}{\delta}\left(\exp\left[(y_t - \mu)\delta - \frac{1}{2}\delta^2\right] - 1\right)$. Because

$$\exp\left[(y_t - \mu)\delta - \frac{1}{2}\delta^2\right] - 1 = e^{(\bar{y} - \mu)\delta}\delta Z_\delta(y_t - \bar{y}) + \left(e^{(\bar{y} - \mu)\delta} - 1\right),$$

we have

$$\begin{aligned} L_\delta(\theta, \mu) &= -\frac{1}{2} \sum_{t=1}^n (y_t - \bar{y})^2 - \frac{n}{2}(\bar{y} - \mu)^2 \\ &\quad + \sum_{t=1}^n \log\left[1 + \theta e^{(\bar{y} - \mu)\delta} Z_\delta(y_t - \bar{y}) + \theta\left(\frac{e^{(\bar{y} - \mu)\delta} - 1}{\delta}\right)\right]. \end{aligned}$$

To explore the behavior of the log-likelihood as a function of δ we need the maximizing values of μ and θ. As $\mathbb{E}Y_t = \mu + \theta$, we approximate the

20

maximum of the log-likelihood by replacing μ with $\bar{y} - \theta$ so that

$$
\begin{aligned}
\max_{\theta,\mu} L_\delta\left(\theta,\mu\right) &= \max_\theta L_\delta\left(\theta,\hat{\mu}\right) \approx \max_\theta -\frac{1}{2}\sum_{t=1}^n\left(y_t-\bar{y}\right)^2 - \frac{n}{2}\theta^2 \\
&\quad + \sum_{t=1}^n \log\left[1+\theta e^{\theta\delta}Z_\delta\left(y_t-\bar{y}\right)+\theta\left(\frac{e^{\theta\delta}-1}{\delta}\right)\right].
\end{aligned}
$$

We write the likelihood ratio for a fixed value of δ as

$$
R_n\left(\theta,\delta\right)=2\sum_{t=1}^n\left(\log\left[1+\theta Z_{t,\delta,\bar{y}}+\theta\left(\delta Z_{t,\delta,\bar{y}}+1\right)\left(\frac{e^{\theta\delta}-1}{\delta}\right)\right]-\frac{\theta^2}{4}\right).
$$

7   Appendix B: Hermite Polynomials

HERMITE POLYNOMIALS

We define the Hermite polynomial $H_j\left(x\right)$ in terms of the generating function

$$
e^{x\delta-\delta^2/2}=\sum_{j=0}^\infty \frac{H_j\left(x\right)}{j!}\delta^j. \tag{6}
$$

Specifically, the first 6 Hermite polynomials are

$$
\begin{aligned}
H_0\left(x\right) &= 1 \\
H_1\left(x\right) &= x \\
H_2\left(x\right) &= x^2-1 \\
H_3\left(x\right) &= x^3-3x \\
H_4\left(x\right) &= x^4-6x^2+3 \\
H_5\left(x\right) &= x^5-10x^3+15x \\
H_6\left(x\right) &= x^6-15x^4+45x^2-15.
\end{aligned}
$$

The Hermite polynomials can be used to form an orthonormal series. That is, for a standard normal V,

$$
\mathbb{E}\left[H_j\left(V\right)H_k\left(V\right)\right]=\left\{\begin{array}{ll} j! & \text{for } j=k \\ 0 & \text{for } j\neq k \end{array}\right. . \tag{7}
$$

To prove this, we follow the arguments described in Lebedev (1965, pages 60-76) and take a product of the generating series

$$
e^{V\delta-\delta^2/2}e^{V\gamma-\gamma^2/2}=\sum_{j=0}^\infty\sum_{k=0}^\infty\frac{H_j\left(V\right)H_k\left(V\right)}{j!k!}\delta^j\gamma^k.
$$

21

From the properties of the moment generating function of a normal random variable, the left side of this equation has expectation

$$\mathbb{E}e^{V(\delta+\gamma)-\delta^2/2-\gamma^2/2} = e^{\delta\gamma}.$$

From the power series expansion of an exponential function we have

$$e^{\delta\gamma} = \sum_{j=0}^{\infty} \frac{\delta^j \gamma^j}{j!}.$$

Thus

$$\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{\mathbb{E}\left[H_j\left(V\right) H_k\left(V\right)\right]}{j!k!} \delta^j \gamma^k = \sum_{j=0}^{\infty} \frac{\delta^j \gamma^j}{j!}.$$

Because this is true for any δ and γ, the coefficients in the two expansions must be the same. This implies (7), so $H_j\left(V\right)\left(j!\right)^{-1/2}$ for $j \geq 1$, is an uncorrelated set of random variables with mean 0 and variance 1.

SHIFTED HERMITE POLYNOMIALS
Let $W = V + \delta$, with $V \sim \mathcal{N}\left(0,1\right)$. Then

$$H_j\left(W\right) = \sum_{m=0}^{j} \left(\begin{array}{c} j \\ m \end{array}\right) H_m\left(V\right) \delta^{j-m}.$$

It follows from the fact that $\mathbb{E}H_m\left(V\right) = 0$ for $m > 0$ that $\mathbb{E}H_j\left(V\right) = \delta^j$. Furthermore,

$$\begin{aligned}
\mathbb{E}H_m\left(V\right)^2 &= \sum_{m=0}^{j} \left(\begin{array}{c} j \\ m \end{array}\right)^2 \mathbb{E}H_m\left(V\right)^2 \delta^{2(j-m)} \\
&= \sum_{m=0}^{j} \left(\begin{array}{c} j \\ m \end{array}\right)^2 m! \delta^{2(j-m)},
\end{aligned}$$

which implies

$$\begin{aligned}
Var\left(H_m\left(V\right)\right) &= \sum_{m=1}^{j} \left(\begin{array}{c} j \\ m \end{array}\right)^2 m! \delta^{2(j-m)} \\
&= j! \sum_{q=0}^{j-1} \left(\begin{array}{c} j \\ q \end{array}\right)^2 \frac{\delta^{2q}}{q!}.
\end{aligned}$$

22

Further, for $j < k$,

$$
\begin{aligned}
Cov\left(H_j\left(V\right), H_k\left(V\right)\right) &= \sum_{m=0}^{j}\sum_{q=0}^{k}\binom{j}{m}\binom{k}{q}\mu^{j-m+k-q}Cov\left(H_m\left(V\right), H_q\left(V\right)\right) \\
&= \sum_{m=1}^{j}\binom{j}{m}\binom{k}{m}\mu^{j+k-2m}m! \\
&= k!\left(\frac{\mu^{k-j}}{(k-j)!} + j\frac{\mu^{k-j-2}}{(k-j-1)!} + \cdots + j\frac{\mu^{k+j-2}}{(k-1)!}\right).
\end{aligned}
$$

References

[1] Andrews, D. and X. Cheng, "Estimation and Inference with Weak, Semi-Strong, and Strong Identification," *Econometrica*, 2012, **80**, 2153-2211.

[2] Bostwick, V. and D. Steigerwald, "Obtaining Critical Values for Test of Markov Regime Switching," *Economics Department Discussion Papers, UC Santa Barbara*, 2012.

[3] Carter, A. and D. Steigerwald, "Testing for Regime Switching: A Comment," *Econometrica*, 2012, **80**, 1809-1812.

[4] Carter, A. and D. Steigerwald, "Markov Regime-Switching Tests: Asymptotic Critical Values," *Journal of Econometric Methods*, 2013, **2**, 25-34.

[5] Chen H. and J. Chen, "The Likelihood Ratio Test for Homogeneity in Finite Mixture Models," *Canadian Journal of Statistics*, 2001, **29**, 201-215.

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

[7] Hamilton, J.,"Calling Recessions in Real Time," *International Journal of Forecasting*, 2011, **27**, 1006-1026.

[8] Hamilton, J., "Macroeconomic Regimes and Regime Shifts," *UC San Diego Economics Staff Papers*, 2014.

[9] Hansen, B., "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis", *Econometrica*, 1996, **64**, 413-430.

[10] Hartigan, J., "A Failure of Likelihood Asymptotics for Normal Mixtures," in *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume 2, ed. by L. LeCam and R. Olshen, Belmont: Wadsworth Press, 1985, 807-810.

[11] Kasahara, H., T. Okimoto and K. Shimotsu,"Modified Quasi-Likelihood Ratio Test for Regime Switching," *Japanese Economic Review* 2014, **65**, 25-41.

[12] Lebedev, N., *Special Functions and Their Applications*, 1965, Prentice-Hall.

[13] Liu, X. and Y. Shao, "Asymptotics for the Likelihood Ratio Test in a Two-Component Normal Mixture Model," *Journal of Statistical Planning and Inference* 2004, **123**, 61-81.