Testing for Regime Switching: Rejoinder

JIN SEO CHO
School of Economics, Yonsei University

HALBERT WHITE
Department of Economics, UC San Diego

May 21, 2011

Abstract

We revisit Theorem 1 of Cho and White (2007, "CW") in light of Carter and Steigerwald (2010) and give a set of sufficient conditions for CW's quasi-likelihood ratio (QLR) statistic to yield a consistent test.

Key Words: regime switching; quasi-likelihood ratio test; consistent test

JEL Classification: C12, C22, C45, C52.

Acknowledgements: The authors thank Doug Steigerwald and Andrew Carter for helpful discussions.

As Carter and Steigerwald (2010) correctly point out, Theorem $1(b)$ of Cho and White (2007,"CW"), which was intended to provide conditions ensuring the consistency of the quasi-maximum likelihood estimator (QMLE) for the parameters of a regime-switching process, requires further assumptions for its validity.

Since the null distribution of CW's quasi-likelihood ratio (QLR) test statistic is unaffected by this omission and since the QLR test generally will have power even when the QMLE is not consistent for the true regime-switching parameters, this deficiency does not have disastrous consequences for CW's central goal, testing for regime switching, Nevertheless, it is important to accurately delineate conditions under which the QLR test is consistent. Carter and Steigerwald's further condition, that[1] $X_t|\mathcal{F}_{t-1} \sim F(\cdot; \theta_0^*, \theta_k^*)$ when $S_t = k$ $(k \in \{1, 2\})$ ensures the consistency of the QMLE for the regime-switching parameters. This then implies the consistency of the QLR test.

There are, however, other conditions ensuring the consistency of the QLR test; consistency of the QMLE for the true regime-switching parameters is sufficient but not necessary for this. Consistency of the test occurs whenever the *pseudo*-true parameters for the (misspecified) regime switching process differ. Accordingly, we state a modified version of CW's theorem 1 that provides several such sufficient conditions, including, for completeness, those of Carter and Steigerwald (2010) (Theorem 1 (b')).

---

[1]Notation and references to assumptions are as in Cho and White (2007), unless otherwise specified.

Theorem 1 (b') *Given Assumptions A.1, A.2(i,ii), A.3, A.4, A.5(i), suppose* $X_t|\mathcal{F}_{t-1} \sim F(\cdot;\theta_0^*,\theta_k^*)$ *when* $S_t = k$, $k \in \{1,2\}$. *Under* H_1, $(\hat{\pi}_n^q, \hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q) \to (\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$ *a.s.*

(c) *Given Assumptions A.1, A.2(i,ii), A.3, A.4, A.5(ii), suppose* $\{S_t\}$ *is an independent sequence. Under* H_1, $(\hat{\pi}_n^q, \hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q) \to (\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$ *a.s.*

(d) *Given Assumptions A.1, A.2(i,ii), A.3, A.4, A.5(ii), suppose that for any* $(\theta_0,\theta_1) \in \tilde{\Theta}$, *there is* $\theta_\dagger \in \Theta_*$ *such that* $E[f(X_t|X^{t-1};\theta_0,\theta_\dagger)/f(X_t|X^{t-1};\theta_0,\theta_1)] > 1$. *Then* $(\hat{\pi}_n^q, \hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q) \to (\pi^\dagger, \theta_0^\dagger, \theta_1^\dagger, \theta_2^\dagger)$ *a.s., where* $\pi^\dagger \in (0,1)$ *and* $\theta_1^\dagger \neq \theta_2^\dagger$.

Theorem 1(c) is elementary, so we just sketch its proof: As $\{S_t\}$ is an independent process, the conditional PDF of $X_t|\mathcal{F}_{t-1}$ is $\pi^* f(\cdot|X^{t-1};\theta^{1*}) + (1-\pi^*)f(\cdot|X^{t-1};\theta^{2*})$, so that $E[\ell_t(\cdot,\cdot)]$ is maximized at $(\pi^*,\theta_0^*,\theta_1^*,\theta_2^*)$. Lemma A1$(a)$ of CW then gives Theorem 1(c).

Theorem 1(d) provides other sufficient conditions for the consistency of the QLR test. We prove Theorem 1(d) by showing that the first-order condition for maximization does not hold on the parameter space given by CW's H_0'. Moreover, the key condition in Theorem 1(d) holds for many popular regime-switching models. We now illustrate this.

First, suppose $X_t = \theta_1^*\mathbf{1}_{\{S_t=1\}} + \theta_2^*\mathbf{1}_{\{S_t=2\}} + u_t$, with $u_t \sim$ IID $N(0,1)$. Then $X_t|\mathcal{F}_{t-1}$ has the same conditional distribution as $X_t|\sigma(S_t)$, so $E[f(X_t;\theta_\dagger)/f(X_t;\theta_1)] = \pi^*\exp[(\theta_1-\theta_\dagger)(\theta_1-\theta_1^*)] + (1-\pi^*)\exp[(\theta_1-\theta_\dagger)(\theta_1-\theta_2^*)]$. We view this as a function of $(\theta_\dagger,\theta_1)$ for given $(\pi^*,\theta_1^*,\theta_2^*)$, say $g(\theta_\dagger,\theta_1)$, where we suppress $(\pi^*,\theta_1^*,\theta_2^*)$ for simplicity. We now show that this example satisfies the additional requirement in Theorem 1(d) by contradiction. For this, we suppose that for some θ_1, $g(\cdot,\theta_1) \leq 1$, so that $\frac{\partial}{\partial\theta_\dagger}g(\theta_\dagger,\theta_1)|_{\theta_\dagger=\theta_1} = 0$ and $\frac{\partial^2}{\partial\theta_\dagger^2}g(\theta_\dagger,\theta_1)|_{\theta_\dagger=\theta_1} \leq 0$, because for each θ_1, $g(\theta_1,\theta_1) \equiv 1$ by the definition of function g. That is, $g(\cdot,\theta_1)$ must be at least weakly concave at θ_1. Nevertheless, $\frac{\partial^2}{\partial\theta_\dagger^2}g(\theta_\dagger,\theta_1)|_{\theta_\dagger=\theta_1} = \pi^*(\theta_1-\theta_1^*)^2 + (1-\pi^*)(\theta_1-\theta_2^*)^2 > 0$, so it is a convex function of $\theta_\dagger$ at θ_1. This is a contradiction caused by the supposition that for some θ_1, $g(\cdot,\theta_1) \leq 1$, so that for each θ_1, there is $\theta_\dagger$ such that $g(\theta_\dagger,\theta_1) > 1$.

Second, suppose $X_t = \theta_1^*\mathbf{1}_{\{S_t=1\}} + \theta_2^*\mathbf{1}_{\{S_2=2\}} + \theta_0^*X_{t-1} + u_t$, with $u_t \sim$ IID $N(0,1)$. This differs from the first example due to the included lag. Then $E[f(X_t|X_{t-1};\theta_0,\theta_\dagger)/\ f(X_t|X_{t-1};\theta_0,\theta_1)] = E[P(S_t = 1|X^{t-1})\exp[(\theta_\dagger-\theta_1)\{(\theta_1^*-\theta_1)+(\theta_0^*-\theta_0)X_{t-1}\}] + P(S_t=2|X^{t-1})\exp\left[(\theta_\dagger-\theta_1)\{(\theta_2^*-\theta_1)+(\theta_0^*-\theta_0)X_{t-1}\}\right]$. If we view this as a function of $(\theta_\dagger,\theta_1,\theta_0)$ for given $(\pi^*,\theta_0^*,\theta_1^*,\theta_2^*)$, say $g(\theta_\dagger,\theta_1,\theta_0)$, then for given (θ_0,θ_1), $g(\theta_1,\theta_1,\theta_0) \equiv 1$ by the definition of function g. We also note that $\frac{\partial^2}{\partial\theta_\dagger^2}g(\theta_\dagger,\theta_1,\theta_0)|_{\theta_\dagger=\theta_1} = E[P(S_t=1|X^{t-1})\{(\theta_1^*-\theta_1)+(\theta_0^*-\theta_0)X_{t-1}\}^2 + P(S_t=2|X^{t-1})\{(\theta_2^*-\theta_1)+(\theta_0^*-\theta_0)X_{t-1}\}^2] > 0$. Thus, $g(\cdot,\theta_1,\theta_0)$ is convex at θ_1, so the additionally required condition in Theorem 1(d) holds.

Third, suppose $X_t = \theta_1^*\mathbf{1}_{\{S_t=1\}} + \theta_2^*\mathbf{1}_{\{S_2=2\}} + u_t$, with $u_t \sim$ IID $N(0,\sigma^{*2})$. This differs from the first example, as the variance of u_t is no longer fixed at 1. If we let $g(\theta_\dagger,\theta_1,\sigma_0^2) := E[f(X_t;\sigma_0^2,\theta_\dagger)/f(X_t;$

$\sigma_0^2, \theta_1)]$, then $\frac{\partial^2}{\partial \theta_\dagger^2} g(\theta_\dagger, \theta_1, \sigma_0^2)|_{\theta_\dagger = \theta_1} = [\pi^*(\theta_1 - \theta_1^*)^2 + (1-\pi^*)(\theta_1 - \theta_2^*)^2 + (\sigma^{*4} - \sigma_0^2)]/\sigma_0^4$. Note that $\pi^*(\theta_1 - \theta_1^*)^2 + (1-\pi^*)(\theta_1 - \theta_2^*)^2$ attains its minimum $\pi^*(1-\pi^*)(\theta_1^* - \theta_2^*)^2$ when $\theta_1 = \pi^*\theta_1^* + (1-\pi^*)\theta_2^*$. From this, if $\sigma_0^2 < \pi^*(1-\pi^*)(\theta_1^* - \theta_2^*)^2 + \sigma^{*4}$, then for every (θ_1, σ_0^2), $\frac{\partial^2}{\partial \theta_\dagger^2} g(\theta_\dagger, \theta_1, \sigma_0^2)|_{\theta_\dagger = \theta_1} > 0$, implying that $g(\cdot, \theta_1, \sigma_0^2)$ is convex at θ_1, so the condition in Theorem 1(*d*) is satisfied.

Finally, let $X_t = \theta_1^* \mathbf{1}_{\{S_t = 1\}} + \theta_2^* \mathbf{1}_{\{S_2 = 2\}} + \theta_0^* X_{t-1} + u_t$, with $u_t \sim$ IID $N(0, \sigma^{*2})$. We have a lagged dependent variable and the variance of u_t is not 1. If we denote $E[f(X_t; \sigma_0^2, \theta_0, \theta_\dagger)/f(X_t; \sigma_0^2, \theta_0, \theta_1)]$ by $g(\theta_\dagger, \theta_1, \theta_0, \sigma_0^2)$, then $\frac{\partial^2}{\partial \theta_\dagger^2} g(\theta_\dagger, \theta_1, \theta_0, \sigma_0^2)|_{\theta_\dagger = \theta_1} = \frac{\sigma_*^2 - \sigma_0^2 + h(\mu, \delta)}{\sigma_0^4}$, where $h(\mu, \delta) := E[P(S_t = 1|X^{t-1})(\mu + \delta X_{t-1})^2 + P(S_t = 2|X^{t-1})(\mu + \gamma^* + \delta X_{t-1})^2]$, $\mu := \theta_1 - \theta_1^*$, $\delta := \theta_0 - \theta_0^*$, and $\gamma^* := \theta_1^* - \theta_2^*$. By definition of $h(\cdot, \cdot)$, it is uniformly positive, and its minimum is attained when $(\mu, \delta) = (\hat{\mu}, \hat{\delta})$, where $\hat{\mu} := \gamma^*[\kappa^* E[X_t] - (1-\pi^*)E[X_t^2]]/\text{var}[X_t]$, $\hat{\delta} := \gamma^*[(1-\pi^*)E[X_t] - \kappa^*]/\text{var}[X_t]$, and $\kappa^* := E[\mathbf{1}_{\{S_t = 2\}} X_{t-1}]$. From this, if $\sigma_0^2 < h(\hat{\mu}, \hat{\delta}) + \sigma^{*4}$, then for every $(\theta_1, \theta_0, \sigma_0^2)$, $\frac{\partial^2}{\partial \theta_\dagger^2} g(\theta_\dagger, \theta_1, \theta_0, \sigma_0^2)|_{\theta_\dagger = \theta_1} > 0$, implying that $g(\cdot, \theta_1, \theta_0, \sigma_0^2)$ is convex at θ_1, and the required condition in Theorem 1(*d*) holds.

Appendix

Proof of Theorem 1(d) We show that the first-order condition does not hold on the null parameter space given in CW (p. 1667). We consider the null parameter space constrained by H_{01}: for some θ_*, $\{(\pi, \theta) \in [0,1] \times \Theta : \pi = 1, \theta_1 = \theta_*\}$ and evaluate the first derivatives of $E[\ell_t(\cdot, \cdot)]$ on this space.

For $E[\ell_t(\cdot, \cdot)]$ to be maximized on the the null parameter space, there must be $(\theta_{0*}, \theta_*) \in \tilde{\Theta}$ such that for every $\theta_2 \in \Theta_*$,

$$\frac{\partial}{\partial \pi} E[\ell_t(\pi, \theta)]_{H_{01}} = E\left[1 - \frac{f_t(\theta_{0*}, \theta_2)}{f_t(\theta_{0*}, \theta_*)}\right] \geq 0; \quad \text{and} \tag{1}$$

$$\nabla_{\theta_1} E[\ell_t(\pi, \theta)]_{H_{01}} = E\left[\frac{\nabla_{\theta_1} f_t(\theta_{0*}, \theta_*)}{f_t(\theta_{0*}, \theta_*)}\right] = 0. \tag{2}$$

Here, the inequality (1) is used to accommodate the fact that $\pi = 1$ is on the boundary of $[0, 1]$. Also, we do not consider $\nabla_{\theta_2} E[\ell_t(\pi, \theta)]_{H_{01}}$, as it is identical to zero. Condition (2) is implied by condition (1) since $E[f_t(\theta_{0*}, \cdot)/f_t(\theta_{0*}, \theta_*)]$ has a maximum value 1 by condition (1), implying that $E[\nabla_{\theta_1} f_t(\theta_{0*}, \theta_1)/f_t(\theta_{0*}, \theta_*)] = 0$, provided that $\theta_1 = \theta_*$. Thus, condition (2) follows if condition (1) holds. Nevertheless, from the condition given in the theorem, this cannot hold because there is $\theta_\dagger \in \Theta_*$ such that $E[f_t(\theta_{0*}, \theta_\dagger)/f_t(\theta_{0*}, \theta_*)] > 1$. This implies that $E[\ell_t(\cdot, \cdot)]$ is not maximized on the null parameter space. *Q.E.D.*

References

Carter, A. and Steigerwald, D. (2010), "Testing for Regime Switching: A Comment," *Discussion Paper*, UC Santa Barbara Department of Economics.

Cho, J.S. and White, H. (2007), "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.