

Testing for Regime Switching: A Comment

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

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

May 5, 2011

JEL Classification: C13

Key Words: consistent, Markov regime switching, quasi-maximum likelihood

1 Introduction

In Cho and White (2007) “TESTING FOR REGIME SWITCHING” the authors obtain the asymptotic null distribution of a quasi-likelihood ratio (QLR) statistic. The statistic is designed to test the null hypothesis of one regime against the alternative of Markov switching between two regimes. Likelihood ratio statistics are used because the test involves nuisance parameters that are not identified under the null hypothesis, together with other nonstandard features. Cho and White focus on a quasi-likelihood, which ignores certain serial correlation properties but allows for a tractable factorization of the likelihood. While the majority of their paper focuses on asymptotic behavior under the null hypothesis, Theorem 1(b) states that the quasi-maximum likelihood estimator (QMLE) is consistent under the alternative hypothesis. Consistency of the QMLE requires that the expected quasi-log-likelihood attain a global maximum at the population parameter values. This requirement holds for some Markov regime-switching processes but, as we show below, not for an autoregressive process as analyzed in Cho and White.

The quasi-likelihood approximates the Markov likelihood by replacing the probabilities of the state variable that indicates regimes. Specifically, the probabilities from the distribution of state variables conditional on past values of the observed data (the conditional state probabilities) are replaced with the

*Corresponding author: doug@econ.ucsb.edu We thank Jin Cho and Hal White for helpful comments.

unconditional probabilities from the stationary distribution of state variables.¹ In so doing the quasi-likelihood ignores serial correlation in the state variables. Ignoring this serial correlation can lead to inconsistency if the conditional state probabilities depend upon the regressors that enter the state-specific conditional densities.

To understand the source of inconsistency, we use the classic structure that Wald (1949) proposed to demonstrate consistency of the MLE. The structure was adapted by Levine (1983) to demonstrate a general property of consistency for a QMLE, where the quasi-log-likelihood is constructed from conditional density functions. As we show below, in applying the logic of Levine to Markov regime-switching processes, a key requirement is that the conditional state probabilities be independent of the regressors that enter the state-specific conditional densities. For a process in which the conditional densities depend only on exogenous regressors, the conditional state probabilities are independent of the regressors and a QMLE is consistent. For the autoregressive process analyzed in Cho and White (Section 3, p. 1697), however, the conditional state probabilities depend upon regressors that enter the state-specific conditional densities and a QMLE is inconsistent. Lack of consistency of a QMLE extends generally to autoregressive processes as lagged values of the dependent variable, which are regressors in the conditional densities, contain information about lagged values of the state variable, which in turn contain information about the current value of the state variable.

We organize the results as follows. The key assumptions from Cho and White together with the quasi-likelihood are contained in Section 2. In Section 3 we present the Markov-regime switching autoregressive process analyzed by Cho and White and show that the gradient of the quasi-log-likelihood does not equal zero when evaluated at the population parameter values. In Section 4 we establish a condition that ensures that the expected quasi-log-likelihood attains a global maximum (under the alternative) at the population parameter values. To understand the distinction between our sufficient condition and the condition contained in Cho and White we first relate the autoregressive process to each condition and then refer to the proof of Theorem 1(b) in Cho and White to determine the impact of each condition. We then discuss the implications of these results for the consistency of a QLR test statistic.

2 Markov Regime-Switching Processes

Rather than presenting all the assumptions from Cho and White, we focus only on the assumptions that pertain to establishing a global maximum of the quasi-log-likelihood at the population parameter values. The first assumption defines the class of Markov regime-switching process, which have strictly stationary

¹Other quasi-likelihood approximations for Markov regime-switching processes have been shown to lead to inconsistent estimators. Campbell (2002) analyzes the case in which the underlying innovation density is misspecified. Kim, Piger and Startz (2008) discuss the case in which the endogeneity of regime switching is ignored.

random variables.

ASSUMPTION 1:

(i) The observable random variables $\{X_t \in \mathbb{R}^d\}_{t=1}^n$, $d \in \mathbb{N}$, are generated as a sequence of strictly stationary β -mixing random variables such that for some $c > 0$ and $\rho \in [0, 1)$ the β -mixing coefficient, δ_τ , is at most $c\rho^\tau$.

(ii) The sequence of unobserved state variables that indicate regimes, $\{S_t \in \{1, 2\}\}_{t=1}^n$, is generated as a first-order Markov process such that $\mathbb{P}(S_t = 2 | S_{t-1} = 1) = p_{12}^*$ and $\mathbb{P}(S_t = 1 | S_{t-1} = 2) = p_{21}^*$ with $p_{ij}^* \in [0, 1]$ ($i, j = 1, 2$).

(iii) The given $\{X_t\}$ is a Markov regime-switching process. That is, for some $\theta^* := (\theta_0^*, \theta_1^*, \theta_2^*) \in \mathbb{R}^{r_0+2}$,

$$X_t | \mathcal{F}_{t-1} \sim \begin{cases} F(\cdot | X^{t-1}; \theta_0^*, \theta_1^*) & \text{if } S_t = 1 \\ F(\cdot | X^{t-1}; \theta_0^*, \theta_2^*) & \text{if } S_t = 2 \end{cases},$$

where $\mathcal{F}_{t-1} := \sigma(X^{t-1}, S^t)$ is the smallest σ -algebra generated by $(X^{t-1}, S^t) := (X'_{t-1}, \dots, X'_1, S_t, \dots, S_1)$; $r_0 \in \mathbb{N}$; and the conditional cumulative distribution function of $X_t | \mathcal{F}$, $F(\cdot | X^{t-1}; \theta_0^*, \theta_j^*)$ has a probability density function $f(\cdot | X^{t-1}; \theta_0^*, \theta_j^*)$ ($j = 1, 2$). Further, for $(p_{12}^*, p_{21}^*) \in (0, 1] \times (0, 1] \setminus \{(1, 1)\}$, θ^* is unique in \mathbb{R}^{r_0+2} .

The point $p_{12}^* = p_{21}^* = 1$ is excluded from the parameter space to rule out a deterministically periodic process for $\{S_t\}$, which would imply that $\{X_t\}$ is not strictly stationary. Throughout the following discussion, \mathbb{E} and \mathbb{P} refer to the distribution of the process described in Assumption 1.

To ensure the likelihood is well defined, it is assumed that the data generating process specifies an \mathcal{F}_{t-1} -measurable probability density function.

ASSUMPTION 2:

(i) A model for $f(\cdot | X^{t-1}; \theta_0^*, \theta_j^*)$ is $\{f(\cdot | X^{t-1}; \theta^j) : \theta^j := (\theta_0, \theta_j) \in \tilde{\Theta}\}$, where $\tilde{\Theta} := \Theta_0 \times \Theta_* \in \mathbb{R}^{r_0+1}$. Further, for each $\theta^j \in \tilde{\Theta}$, $f(\cdot | X^{t-1}; \theta^j)$ is a measurable probability density function with cumulative distribution function $F(\cdot | X^{t-1}; \theta^j)$ ($j = 1, 2$).

The relevant hypotheses, for test of one regime against the alternative of two regimes, are

$$\begin{aligned} H_0 &: p_{12}^* = 0 ; p_{21}^* = 0 ; \text{ or } \theta_1^* = \theta_2^*; \\ H_1 &: (p_{12}^*, p_{21}^*) \in (0, 1] \times (0, 1] \setminus \{(1, 1)\} \text{ and } \theta_1^* \neq \theta_2^*. \end{aligned}$$

Because the log-likelihood cannot be reduced to a sum of individual log-likelihoods, if $p_{12}^* = 0$ (or $p_{21}^* = 0$) then the population variance of the associated first derivative grows geometrically under the null. To avoid this difficulty, Cho and White focus on the quasi-log-likelihood for a mixture model. The quasi-log-likelihood replaces the (likelihood) Markov conditional density of $X_t | \sigma(X^{t-1})$, which is given by

$$\mathbb{P}(S_t = 1 | \sigma(X^{t-1})) f(X_t | X^{t-1}; \theta_0^*, \theta_1^*) + \mathbb{P}(S_t = 2 | \sigma(X^{t-1})) f(X_t | X^{t-1}; \theta_0^*, \theta_2^*), \quad (1)$$

with the mixture conditional density,

$$(1 - \pi^*) f(X_t | X^{t-1}; \theta_0^*, \theta_1^*) + \pi^* f(X_t | X^{t-1}; \theta_0^*, \theta_2^*). \quad (2)$$

The mixture model (2) captures the serial correlation in X_t , through $f(X_t | X^{t-1}; \cdot)$, but ignores the serial correlation in $\{S_t\}$ as the random variable $\zeta_t := \mathbb{P}(S_t = 2 | \sigma(X^{t-1}))$ is replaced with $\pi^* = \mathbb{E}[\mathbb{P}(S_t = 2 | \sigma(X^{t-1}))] = p_{12}^* / (p_{12}^* + p_{21}^*)$. Note that because

$$\zeta_t = (1 - p_{21}^*) \mathbb{P}(S_{t-1} = 2 | \sigma(X^{t-1})) + p_{12}^* \mathbb{P}(S_{t-1} = 1 | \sigma(X^{t-1})),$$

if $p_{12}^* + p_{21}^* = 1$ then $\zeta_t = \pi^* = p_{12}^*$ and the quasi-log-likelihood is identical to the log-likelihood.

The resultant quasi-log-likelihood is defined for each $(\pi, \theta) \in [0, 1] \times \Theta$ with $\Theta := \Theta_0 \times \Theta_* \times \Theta_*$, as

$$L_n^*(\pi, \theta) := \sum_{t=1}^n l_t(\pi, \theta),$$

where $l_t(\pi, \theta) := \log((1 - \pi) f(X_t | X^{t-1}; \theta^1) + \pi f(X_t | X^{t-1}; \theta^2))$.

To ensure that expectations are well defined, the following regularity condition is assumed.

ASSUMPTION 3:

For all $(\pi, \theta) \in [0, 1] \times \Theta$, the quantity $\mathbb{E}[l_t(\pi, \theta)]$ exists and is finite.

3 QMLE Inconsistency: An Example

The Markov-regime switching autoregressive process analyzed by Cho and White (Section 3, p. 1697) is

$$X_t = \theta_* \cdot 1_{\{S_t=1\}} - \theta_* \cdot 1_{\{S_t=2\}} + .5X_{t-1} + u_t,$$

where $u_t \sim i.i.d.N(0, 1)$ and the transition probabilities satisfy

$$\mathbb{P}(S_t = 1 | S_{t-1} = 1) = \mathbb{P}(S_t = 2 | S_{t-1} = 2),$$

so that $p_{12}^* = p_{21}^*$. The quasi-log-likelihood employed by Cho and White for this process is constructed from the mixture model (p. 1697, line 14)

$$(1 - \pi) \cdot N(\theta_1 + \alpha X_{t-1}, \sigma^2) + \pi N(\theta_2 + \alpha X_{t-1}, \sigma^2). \quad (3)$$

To isolate the source of inconsistency in the QMLE from (3), we set the variance to 1 and let $\theta_1 = \mu$ and $\theta_2 - \theta_1 = \gamma$. The conditional density functions that enter (3) are

$$\begin{aligned} N(\mu + \alpha X_{t-1}, 1) &= f(X_t | X_{t-1}; \theta^1) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(X_t - \alpha X_{t-1} - \mu)^2\right] \\ N(\mu + \gamma + \alpha X_{t-1}, 1) &= f(X_t | X_{t-1}; \theta^2) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(X_t - \alpha X_{t-1} - \mu - \gamma)^2\right]. \end{aligned}$$

A necessary condition for consistency is that $n^{-1}\mathbb{E}[\sum_{t=1}^n l_t(\pi, \alpha, \mu, \gamma)]$ be maximized at the population parameter values. Because the process is stationary

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E}[l_t(\pi, \alpha, \mu, \gamma)] = \mathbb{E}[l_t(\pi, \alpha, \mu, \gamma)] := M(\pi, \alpha, \mu, \gamma).$$

We then have

$$M(\pi, \alpha, \mu, \gamma) = \mathbb{E} \log [\pi \lambda(X_t, X_{t-1}) + (1 - \pi)] + \mathbb{E} \log f(X_t | X_{t-1}; \theta^1),$$

where

$$\lambda(X_t, X_{t-1}, \theta^1, \theta^2) = \frac{f(X_t | X_{t-1}; \theta^2)}{f(X_t | X_{t-1}; \theta^1)} = \exp \left[\gamma (X_t - \alpha X_{t-1} - \mu) - \frac{\gamma^2}{2} \right].$$

Thus

$$\begin{aligned} M(\pi, \alpha, \mu, \gamma) &= \mathbb{E} \log \left[\pi \exp \left[\gamma (X_t - \alpha X_{t-1} - \mu) - \frac{\gamma^2}{2} \right] + (1 - \pi) \right] \\ &\quad - \frac{1}{2} \left[\log 2\pi + \mathbb{E} (X_t - \alpha X_{t-1} - \mu)^2 \right]. \end{aligned}$$

The key is to calculate the first derivative of $M(\pi, \alpha, \mu, \gamma)$ with respect to the autoregressive coefficient α evaluated at the population values of the parameters $(\pi^*, \alpha^*, \mu^*, \gamma^*)$. To begin, let $Z_t = X_t - \alpha^* X_{t-1} - \mu^*$. The derivative for α is

$$\frac{\partial}{\partial \alpha} M(\pi, \alpha, \mu, \gamma) \Big|_{(\pi^*, \alpha^*, \mu^*, \gamma^*)} = \mathbb{E} \left(\frac{-\pi^* \gamma^* X_{t-1} \exp \left[\gamma^* Z_t - \frac{(\gamma^*)^2}{2} \right]}{\pi^* \exp \left[\gamma^* Z_t - \frac{(\gamma^*)^2}{2} \right] + (1 - \pi^*)} \right) + \mathbb{E} (X_{t-1} Z_t). \quad (4)$$

To calculate these expectations we must account for the correlation between Z_t and X_{t-1} . The distribution of Z_t conditional on X_{t-1} depends on this previous observation through S_{t-1} . The conditional density is

$$\begin{aligned} f(z | X_{t-1}) &= \phi(z - \gamma^*) \mathbb{P}(S_t = 2 | X_{t-1}) + \phi(z) \mathbb{P}(S_t = 1 | X_{t-1}) \\ &= \left[\zeta_t \exp \left[\gamma^* Z_t - \frac{(\gamma^*)^2}{2} \right] + (1 - \zeta_t) \right] \phi(z) \end{aligned}$$

where $\phi(z)$ is the density of a standard Gaussian random variable.

The first expectation in (4) is

$$\begin{aligned} &\mathbb{E} \left(\frac{-\pi^* \gamma^* X_{t-1} \exp \left[\gamma^* Z_t - \frac{(\gamma^*)^2}{2} \right]}{\pi^* \exp \left[\gamma^* Z_t - \frac{(\gamma^*)^2}{2} \right] + (1 - \pi^*)} \right) \\ &= -\pi^* \gamma^* \mathbb{E} \left[X_{t-1} \int \exp \left[\gamma^* z - \frac{(\gamma^*)^2}{2} \right] \left(\frac{\zeta_t \exp \left[\gamma^* z - \frac{(\gamma^*)^2}{2} \right] + (1 - \zeta_t)}{\pi^* \exp \left[\gamma^* z - \frac{(\gamma^*)^2}{2} \right] + (1 - \pi^*)} \right) \phi(z) dz \right]. \end{aligned}$$

Because $\exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] \phi(z) = \phi(z - \gamma^*)$,

$$\begin{aligned} & \int \exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] \left(\frac{\zeta_t \exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] + (1 - \zeta_t)}{\pi^* \exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*)} \right) \phi(z) dz \\ &= \int \left(\frac{\zeta_t \exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] + (1 - \zeta_t)}{\pi^* \exp\left[\gamma^*z - \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*)} \right) \phi(z - \gamma^*) dz \\ &= \int \left(\frac{\zeta_t \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \zeta_t)}{\pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*)} \right) \phi(v) dv, \end{aligned}$$

and it follows that this integral is

$$\begin{aligned} & \int \left(\frac{\zeta_t \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \zeta_t)}{\pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*)} \right) \phi(v) dv \\ &= \frac{\zeta_t}{\pi^*} + \int \left[\frac{\zeta_t \pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \zeta_t) \pi^* - \left(\zeta_t \pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + \zeta_t (1 - \pi^*) \right)}{\pi^* \left[\pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*) \right]} \right] \phi(v) dv \\ &= \frac{\zeta_t}{\pi^*} + \frac{\pi^* - \zeta_t}{\pi^*} \int \left[\pi^* \exp\left[\gamma^*v + \frac{(\gamma^*)^2}{2}\right] + (1 - \pi^*) \right]^{-1} \phi(v) dv. \end{aligned}$$

Therefore, substituting this expression back into the above expectation

$$\frac{\partial}{\partial \alpha} M(\pi, \alpha, \mu, \gamma) \Big|_{(\pi^*, \alpha^*, \mu^*, \gamma^*)} = -\gamma^* \mathbb{E}(X_{t-1} \zeta_t) + \gamma^* \mathbb{E}[X_{t-1} (\zeta_t - \pi^*)] C_{\pi^*, \gamma^*} + \mathbb{E}(X_{t-1} Z_t),$$

where C_{π^*, γ^*} is the expectation

$$\mathbb{E} \left[\pi^* \exp \left[\gamma^* \xi + \frac{(\gamma^*)^2}{2} \right] + (1 - \pi^*) \right]^{-1}$$

for ξ a standard Gaussian random variable. Clearly, C_{π^*, γ^*} does not depend on X_{t-1} or S_t and is a bounded positive quantity: $0 < C_{\pi^*, \gamma^*} \leq (1 - \pi^*)^{-1}$.

Furthermore, we can use that $\mathbb{E}(Z_t | X_{t-1}) = \gamma^* \mathbb{P}(S_t = 2 | X_{t-1}) = \gamma^* \zeta_t$ to simplify the expression

$$\frac{\partial}{\partial \alpha} M(\pi, \alpha, \mu, \gamma) \Big|_{(\pi^*, \alpha^*, \mu^*, \gamma^*)} = \gamma^* C_{\pi^*, \gamma^*} \mathbb{E}[X_{t-1} (\zeta_t - \pi^*)].$$

The last factor in this expression is

$$\mathbb{E}[X_{t-1} (\zeta_t - \pi^*)] = \mathbb{E}(X_{t-1} \zeta_t) - \mathbb{E}X_{t-1} \mathbb{E}S_t = \text{Cov}(X_{t-1}, S_t).$$

The last equality follows from $\mathbb{E}(X_{t-1}S_t) = \mathbb{E}(X_{t-1}\mathbb{E}(S_t|X_{t-1}))$ and $\mathbb{E}(S_t|X_{t-1}) = \zeta_t$.

An Expression for the Covariance

To obtain an expression for the covariance of X_{t-1} and S_t , we first use the recursive expression

$$X_t = \mu + \alpha X_{t-1} + \gamma S_t + \xi_t = \sum_{k=0}^{\infty} \alpha^k (\mu + \gamma S_{t-k} + \xi_{t-k}),$$

where $\xi_t \sim i.i.d.N(0,1)$. This implies that the covariance is

$$Cov(X_{t-1}, S_t) = \gamma \sum_{k=0}^{\infty} \alpha^k Cov(S_{t-1-k}, S_t).$$

The covariance of the binary state variables is

$$Cov(S_{t-k}, S_t) = \mathbb{P}(S_t = 2|S_{t-k} = 2) \pi - \pi^2,$$

where $\pi = \mathbb{P}(S_t = 2) = \mathbb{P}(S_{t-k} = 2)$ is the stationary probability in the Markov chain. It can be shown that the conditional probability is

$$\mathbb{P}(S_t = 2|S_{t-k} = 2) = \pi + (1 - \pi) \left(\frac{\pi - p_{12}}{\pi} \right)^k.$$

Thus

$$Cov(S_{t-k}, S_t) = \pi(1 - \pi) \left(\frac{\pi - p_{12}}{\pi} \right)^k,$$

and

$$\begin{aligned} Cov(X_{t-1}, S_t) &= \gamma \pi (1 - \pi) \left(\frac{\pi - p_{12}}{\pi} \right) \sum_{k=0}^{\infty} \alpha^k \left(\frac{\pi - p_{12}}{\pi} \right)^k \\ &= \gamma (1 - \pi) (\pi - p_{12}) \left(\frac{\pi}{\pi - \alpha(\pi - p_{12})} \right). \end{aligned}$$

Therefore, the expression for the partial derivative of the M function with respect to α becomes

$$\frac{\partial}{\partial \alpha} M(\pi, \alpha, \mu, \gamma) = \gamma^2 C_{\pi, \gamma} (\pi - p_{12}) \left(\frac{\pi(1 - \pi)}{\pi - \alpha(\pi - p_{12})} \right).$$

Therefore, if the Markov regime process includes a dependence between subsequent time points, the gradient along α is not equal to 0 at the population parameter values and the expected value of the quasi-log-likelihood is maximized away from the population parameter values.

Of course, this derivative vanishes under the null hypothesis where $\pi = 0$, $1 - \pi = 0$ or $\gamma = 0$, as under the null hypothesis there is effectively only one regime. This derivative also vanishes for $\pi = p_{12}$ because, if $\pi = p_{12}$ then $\mathbb{P}(S_t = j) = \mathbb{P}(S_t = j|S_{t-1})$ and the Markov regime process reduces to independent draws from the stationary distribution. In this case L_n^* is the population log-likelihood rather than the quasi-log-likelihood.

4 QMLE Consistency: A Sufficient Condition

To define a class of processes for which a QMLE is consistent, it is helpful to distinguish the elements of X that correspond to endogenous variables. Let a partition of X be given by $X = (Y', W')'$, where Y denotes the sub-vector of endogenous variables and W denotes the sub-vector of exogenous variables. Let $X^{t-1} := (Y^{t-1}, W^t) = (Y'_{t-1}, \dots, Y'_1, W'_t, \dots, W'_1)$. Exogeneity implies that $\mathbb{P}(S_t = 2 | \sigma(X^{t-1}))$ is independent of W^t . To obtain a sufficient condition we augment Assumption 2 to remove the correlation between the conditional state probabilities, $\mathbb{P}(S_t = 2 | \sigma(X^{t-1}))$, and the regressors that enter the conditional densities.

Let

$$\lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^2) = \frac{f(Y_t | Y^{t-1}, W^t; \theta^2)}{f(Y_t | Y^{t-1}, W^t; \theta^{1*})}.$$

ASSUMPTION 2:

(ii) *The ratio $\lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^2)$ is a random variable that is independent of Y^{t-1} for every value of θ^2 .*

Assumption 2(ii) is parallel to the assumption used to define a state-space model in the statistical literature.² Because Assumption 2(ii) implies that the state-specific conditional density of Y_t is independent of Y_s for $s < t$, exogeneity of W_t implies that $f(Y_t | X^{t-1}, W_t; \theta^i) = f(Y_t | W_t; \theta^i)$. While the assumption does not include the autoregressive model analyzed by Cho and White, it does include the second model they analyze, namely a simultaneous equations model of the type in Griffen et al. (2009).

We begin with the classic structure adapted by Levine (1983) to establish that the expected quasi-log-likelihood attain a global maximum at the population parameter values. We employ the structure to prove that, under the sufficient condition Assumption 2(ii), the unconditional population mean of the quasi-log-likelihood attains a global maximum at (π^*, θ^*) . The object of analysis is $\mathbb{E}[l_t(\pi, \theta) - l_t(\pi^*, \theta^*)]$, where \mathbb{E} denotes expectation with respect to the Markov conditional density (1).

THEOREM 1:

Under Assumptions 1, 2(i)-(ii) and 3,

$$\mathbb{E}[l_t(\pi, \theta) - l_t(\pi^*, \theta^*)] \leq 0.$$

PROOF:

We have

$$\begin{aligned} \mathbb{E}[l_t(\pi, \theta) - l_t(\pi^*, \theta^*)] &= \mathbb{E} \left[\log \left(\frac{\pi \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) \right] \\ &\leq \log \mathbb{E} \left[\left(\frac{\pi \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda(Y_t, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) \right], \end{aligned}$$

²See, for example, Durbin and Koopman (2001).

where the last line follows by Jensen's inequality.

The expected value of the ratio is

$$\begin{aligned}
& \mathbb{E} \left[\frac{\pi \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right] \\
= & \mathbb{E} \left\{ \mathbb{E} \left[\left(\frac{\pi \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda (Y_t, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) \middle| \sigma (X^{t-1}) \right] \right\} \\
= & \mathbb{E} \left\{ \int \left(\frac{\pi \lambda (y, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (y, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) \right. \\
& \quad \left. \bullet (\zeta_t \lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \zeta_t)) f (y | X^{t-1}; \theta_1^*) dy \right\} \\
= & \mathbb{E} \left\{ \int (\pi \lambda (y, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (y, X^{t-1}, \theta^{1*}, \theta^1)) f (y | X^{t-1}; \theta_1^*) dy \right\} \\
& + \mathbb{E} \left\{ \delta_t \int \left(\frac{\pi \lambda (y, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (y, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) (\lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) - 1) f (y | X^{t-1}; \theta_1^*) dy \right\},
\end{aligned}$$

where $\delta_t = \zeta_t - \pi^*$ is Y^{t-1} -measurable and so can be moved outside the conditional expectation. The first expectation is 1 because this is the expectation of a density. For the second term, let

$$g(\lambda(X^{t-1})) = \left(\frac{\pi \lambda (y, X^{t-1}, \theta^{1*}, \theta^2) + (1 - \pi) \lambda (y, X^{t-1}, \theta^{1*}, \theta^1)}{\pi^* \lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) + (1 - \pi^*)} \right) (\lambda (y, X^{t-1}, \theta^{1*}, \theta^{2*}) - 1).$$

Under Assumption 2(ii), $g(\lambda(X^{t-1}))$ is independent of Y^{t-1} , so the second term becomes $\mathbb{E}[\delta_t \cdot g(\lambda(W^t))]$. Because δ_t is independent of W^t , which follows from the exogeneity of W^t , the second term is

$$\mathbb{E}[\delta_t \cdot g(\lambda(X^{t-1}))] = \mathbb{E}_{Y^{t-1}}[\delta_t] \mathbb{E}_{W^t}[g(\lambda(W^t))].$$

The quantity δ_t has expectation zero because $\mathbb{E}_{Y^{t-1}}[\delta_t] = \mathbb{E}[\delta_t]$ and stationarity implies $\mathbb{E}\zeta_t = \pi^*$. Therefore

$$\mathbb{E}[l_t(\pi, \theta) - l_t(\pi^*, \theta^*)] \leq \log(1) = 0.$$

■

Given Theorem 1, establishing that the QMLE for the mixture model, $(\hat{\pi}_n, \hat{\theta}_n)$, is consistent follows from standard application of a uniform law of large numbers as referenced in Cho and White (Proof of Theorem 1(b) p. 1704).

At this point, it is natural to ask how the sufficient condition for consistency of a QMLE that we introduce, Assumption 2(ii), relates to the condition

that Cho and White refer to in their proof of Theorem 1(b) on consistency of a QMLE. To answer, we use the autoregressive process analyzed in Cho and White and presented in (3). Assumption 2(ii), which restricts the class of processes allowed by Assumptions 1 and 2(i), explicitly rules out dependence of the state-specific conditional densities on past values of the endogenous variables, and so rules out autoregressive processes. Cho and White consider all processes allowed under Assumptions 1 and 2(i). Given that the autoregressive process that enters (3) is a leading example from their paper, it is perhaps not surprising that this process appears to be allowed under Assumptions 1 and 2(i). To see this, note that under Assumption 1 the density of X_t given the entire past of X and the current and past values of the state variable is

$$X_t | \mathcal{F}_{t-1} = 1_{\{S_t=1\}} f(X_t | X^{t-1}; \theta^{1*}) + 1_{\{S_t=2\}} f(X_t | X^{t-1}; \theta^{2*}). \quad (5)$$

The density of X_t given only the entire past of X , which enters the likelihood, is obtained by replacing $1_{\{S_t=i\}}$ with $\mathbb{P}(S_t = i | X^{t-1})$, so that the mixture model in (3) follows from (5) by replacing $1_{\{S_t=1\}}$ with $1 - \pi$ and $1_{\{S_t=2\}}$ with π . In consequence $f(X_t | X^{t-1}; \theta^{1*}) = N(\theta_1 + \alpha X_{t-1}, \sigma^2)$ and $f(X_t | X^{t-1}; \theta^{2*}) = N(\theta_2 + \alpha X_{t-1}, \sigma^2)$.

If the autoregressive process that enters (3) is allowed under Assumptions 1 and 2(i), how can the proof of Theorem 1(b) in Cho and White be reconciled with the result in Section 3? It appears that a key step in the proof of Theorem 1(b) requires that, given $S_t = 2$, the marginal density of X_t equals the density of X_t conditional on X^{t-1} . In detail, the marginal density of X_t given $S_t = 2$ is $\int f(x_1, \dots, x_t | S_t = 2) dx^{t-1}$. As Cho and White note (p. 1704, line 8 from the bottom of the page)

$$f(x_1, \dots, x_t | S_t = 2) = \frac{1}{\pi^*} \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) [\Pi_{\tau=2}^{t-1} \mathbf{P}^* \mathbf{F}_\tau(\theta^*)] \mathbf{P}^* \mathbf{F}_t(\theta^*) [1 \ 0]',$$

which implies that the marginal density of X_t given $S_t = 2$ is

$$\int f(x_1, \dots, x_t | S_t = 2) dx^{t-1} = \frac{1}{\pi^*} \int \{ \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) [\Pi_{\tau=2}^{t-1} \mathbf{P}^* \mathbf{F}_\tau(\theta^*)] \mathbf{P}^* \mathbf{F}_t(\theta^*) [1 \ 0]' \} dx^{t-1}.$$

Because \mathbf{F}_τ is a function of $x^{\tau-1}$, the expression on the right side is

$$\begin{aligned} & \frac{1}{\pi^*} \int \{ \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) [\Pi_{\tau=2}^{t-1} \mathbf{P}^* \mathbf{F}_\tau(\theta^*)] \mathbf{P}^* \mathbf{F}_t(\theta^*) [1 \ 0]' \} dx^{t-1} \\ &= \frac{1}{\pi^*} \int \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) \int \mathbf{P}^* \mathbf{F}_2(\theta^*) \cdots \int \mathbf{P}^* \mathbf{F}_{t-1}(\theta^*) \mathbf{P}^* \mathbf{F}_t(\theta^*) dx_1 dx_2 \cdots dx_{t-1} [1 \ 0]'. \end{aligned}$$

If \mathbf{F}_τ is not a function of $x^{\tau-1}$, then

$$\begin{aligned} & \frac{1}{\pi^*} \int \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) \int \mathbf{P}^* \mathbf{F}_2(\theta^*) \cdots \int \mathbf{P}^* \mathbf{F}_{t-1}(\theta^*) \mathbf{P}^* \mathbf{F}_t(\theta^*) dx_1 dx_2 \cdots dx_{t-1} [1 \ 0]' \\ &= \frac{1}{\pi^*} \int \boldsymbol{\pi}^* \mathbf{F}_1(\theta^*) dx_1 \int \mathbf{P}^* \mathbf{F}_2(\theta^*) dx_2 \cdots \int \mathbf{P}^* \mathbf{F}_{t-1}(\theta^*) dx_{t-1} \mathbf{P}^* \mathbf{F}_t(\theta^*) [1 \ 0]'. \end{aligned}$$

In their proof of Theorem 1(b), Cho and White state (p. 1704, line 5 from the bottom)

$$\begin{aligned} & \int f(x_1, \dots, x_t | S_t = 2) dx^{t-1} \\ &= \frac{1}{\pi^*} \int \boldsymbol{\pi}^{*\prime} \mathbf{F}_1(\theta^*) dx_1 \int \mathbf{P}^* \mathbf{F}_2(\theta^*) dx_2 \cdots \int \mathbf{P}^* \mathbf{F}_{t-1}(\theta^*) dx_{t-1} \mathbf{P}^* \mathbf{F}_t(\theta^*) [1 \ 0]' , \end{aligned}$$

which holds only if, given $S_t = 2$, the distribution of X_t does not depend on X_s for $s < t$. Because Assumptions 1 and 2(i) do not imply that, given the value of S_t , the distribution of X_t not depend on X_s for $s < t$, proof of Theorem 1(b) requires an additional condition of the form of Assumption 2(ii).

5 Remarks

The inconsistency of a QMLE for Markov regime-switching processes with autoregressive components extends to processes with moving-average components. Inconsistency of a QMLE, however, does not imply that a test based on the quasi-likelihood ratio is inconsistent. Consistency of a QLR test requires only that $\mathbb{E} l_t(\pi, \theta)$ be maximized at some point outside the null hypothesis space. For the autoregressive process in Section 3, the gradient of the M function is zero in every coordinate except α , which indicates that M is maximized away from the null hypothesis and that the class of models for which the QLR test is consistent is larger than the class of models for which the QMLE is consistent. A definitive treatment of consistency could be based on bounding the value of the likelihood under the null hypothesis and then demonstrating that under the alternative there is always a point in the alternative space for which the value of the likelihood exceeds the bound. Even for a consistent test, the power may be affected by the consistency properties of the QMLE under the alternative.

References

- [1] Campbell, S., “Specification Testing and Semiparametric Estimation of Regime Switching Models: An Examination of the US Short Term Interest Rate”, *Economics Department Working Papers, Brown University*, 2002.
- [2] Cho, J. and H. White, “Testing for Regime-Switching”, *Econometrica*, 2007, **75**, 1671-1720.
- [3] Durbin, J. and S. Koopman, *Time Series Analysis by State Space Methods*, Oxford University: 2001.
- [4] Griffin, J., C. Kolstad and D. Steigerwald, “Estimating Market Power with Multiple Product Markets”, *Economics Department Working Papers, UC Santa Barbara*, 2009.
- [5] Kim, C., J. Piger and R. Startz, “Estimation of Markov Regime-Switching Regression Models with Endogenous Switching”, *Journal of Econometrics*, 2008, **143**, 263-273.
- [6] Levine, D., “A Remark on Serial Correlation in Maximum Likelihood”, *Journal of Econometrics* 1983, **23**, 337-342.
- [7] Wald, A., “Note on the Consistency of the Maximum Likelihood Estimate”, *Annals of Mathematical Statistics*, 1949, **20**, 595-601.