TESTING FOR REGIME SWITCHING

JIN SEO CHO
School of Economics and Finance
Victoria University of Wellington
PO Box 600, Wellington, 6001, New Zealand
Email: JinSeo.Cho@vuw.ac.nz

AND

HALBERT WHITE
Department of Economics
University of California San Diego
9500 Gilman Dr., La Jolla, CA, 92093-0508, U.S.A.
Email: hwhite@weber.ucsd.edu

First version: Feb. 17, 2001. This version: Apr. 7, 2007.

Abstract We analyze use of a quasi-likelihood ratio (QLR) statistic for a mixture model to test the null hypothesis of one regime versus the alternative of two regimes in a Markov regime-switching context. This test exploits mixture properties implied by the regime-switching process but ignores certain implied serial correlation properties. When formulated in the natural way, the setting is non-standard, involving nuisance parameters on the boundary of the parameter space, nuisance parameters identified only under the alternative, or approximations using derivatives higher than the second order. We exploit recent advances by Andrews (2001) and contribute to the literature by extending the scope of mixture models, obtaining asymptotic null distributions different from those in the literature. We further provide critical values for popular models or bounds for tail probabilities useful in constructing conservative critical values for regime-switching tests. We compare the size and power of our statistics to other useful tests for regime switching via Monte Carlo and find relatively good performance. We apply our methods to re-examine the classic cartel study of Porter (1983) and reaffirm Porter's findings.

Key Words Markov Regime Switching, Mixture Model, Likelihood-Ratio Statistic, Null Distribution, Bounds for Critical Values, Cartel Stability.
Subject Class Primary C12, Secondary L13.

Acknowledgements The co-editor, Joel Horowitz, and three anonymous referees provided very helpful comments for which we are most grateful. We also have benefited from discussions with Swarnjit Arora, Roger Bowden, Robert Davies, Wouter den Haan, Steven Durlauf, Graham Elliott, Patrick Fitzsimmons, Clive Granger, James Hamilton, Chirok Han, Bruce Hansen, Tae-Hwan Kim, Bruce Lehmann, Mark Machina, Angelo Melino, Masao Ogaki, Peter Phillips, Ehsan Soofi, Kamil Tahmiscioglu, Allan Timmermann, Peter Thomson, and seminar participants at UCSD, Wisconsin-Milwaukee, SRA, the University of Auckland, ESAM03, VUW, and NZESG03. We are indebted to Robert Porter for provision of his data.

1. Introduction

Models of regime-switching behavior play an important role in analyzing economic data. For example, in industrial organization, Porter (1983) in a classic paper uses a two regime model to investigate cartel behavior. In macroeconomics, Hamilton (1989) in another classic paper uses a two regime model to investigate the properties of postwar U.S. real GNP.

Conducting inference about such processes is often a main goal. It is critical that such inferences be properly drawn, as these can be used to argue the innocence or guilt of firms accused of antitrust violations or to inform key economic policy decisions. Nevertheless, as Hamilton (1996), among others, has pointed out, conducting proper inference in regime-switching models is particularly challenging. As we discuss in detail later, this challenge arises due to the fact that when formulated in the natural way, testing the null hypothesis that there is a single regime (versus the alternative of, say, two regimes) can involve a nuisance parameter identified only under the alternative, as well as a parameter on the boundary of the parameter space. Standard likelihood ratio (LR) tests (and related Lagrange multiplier or Wald tests) cannot be conducted in the usual manner.

A main goal of this paper is therefore to develop straightforward methods that researchers can use to draw large sample inferences, testing the null of one regime versus the alternative of two regimes in a regime-switching model. Recent significant advances by Andrews (1999, 2001) play an important role in attaining this goal. A further goal of this paper is to revisit the work of Porter (1983). This serves the dual purpose of illustrating our methods in a classical setting and, as it turns out, affirming Porter's original inferences.

In the prior literature, attempts to test the number of regimes have proceeded by addressing certain aspects of the problem. For example, Hansen (1992) considers this problem using Markov regime-switching models, and obtains a lower bound for the limiting distribution of a standardized LR statistic. As will be clear later, however, the null parameter space can be partitioned into two mutually exclusive subsets: one with the boundary parameter problem and one without the boundary parameter problem. Hansen's bound considers the behavior of the LR statistic on only one of these two subsets. Garcia (1998) reviews Hansen's problem. As we see below, however, both subsets are indeed relevant, and due to the boundary parameter problem, standard arguments cannot apply to the LR statistic.

As we discuss in Section 2, applying the LR statistic to testing one vs. two Markov regimes is challenging, because the log-likelihood for the two-regime alternative does not factor in the usual way. This leads to geometric growth of the population variance of the log-likelihood first derivative under the null, ruling out application of standard central limit results. Moreover, the power of such a test turns out to be weaker than in the standard case. Instead, we proceed by applying mixture models, ignoring certain time series dependence properties implied by the regime-switching process. This yields a quasi-log-likelihood that *does* factor in the usual way and whose analysis is much more tractable. The resulting quasi-likelihood ratio (QLR) test is thus sensitive to the mixture aspect of the regime-switching process, delivering a test with appealing power under the alternative.

Mixture models are widely used for identically and independently distributed (i.i.d.) data. Testing the number of mixture components also has problems similar to those encountered in testing Markov regime switching, namely the boundary parameter problem and nuisance parameters present only under the alternative. Much of the literature attempts to avoid these problems or must confront associated difficulties in attempting to test the number of components. For example, Chesher (1984) and Lancaster (1984) test for unobserved heterogeneity by testing the hypothesis of correct model specification, an indirect method of testing the mixture hypothesis. Neyman and Scott's (1966) C(α) statistic tests the mixture hypothesis, motivated by the properties of the dispersion of the dependent variable of interest under the null and alternative hypotheses relevant here, as well reviewed by Lindsay (1995). On the other hand, Hartigan (1985), Ghosh and Sen (1985), Liu and Shao (2003), and references therein consider the LR statistic for testing the number of components of a mixture model, and show that it converges weakly to a functional of a continuous Gaussian process on a compact parameter space. As these authors show, compactness plays a particularly important role in determining the null distribution of the LR statistic. This is true here, too, and we devote particular attention to the crucial role played by the parameter space. As we show, mixture models can give rise to variety of interesting behaviors. In particular, the model considered by Porter (1983) doesn't have a continuous Gaussian process as the limit of the LR statistic.

We study the mixture model in the Markov regime-switching context, and contribute to the literature in several ways. First, we contribute not only by providing a way to exploit the associated QLR statistic in such a way that the previously encountered difficulties can be avoided, but by doing so in a context that explicitly allows the observable random variables to exhibit time-series dependence. To this end, we show how the mixture model estimators behave when the data are generated by a β-mixing process that is a Markov regime-switching process under the alternative. This yields the asymptotic null distribution for the QLR statistic on a compact parameter space. Next, we extend the mixture literature by examining models whose null weak limits are functionals of discontinuous Gaussian processes, as well as those models whose limits are continuous Gaussian processes. We carefully examine these, providing examples and comparing critical values obtained with and without account taken of the boundary parameter problem. As we show, if boundary conditions are ignored, then critical values can be too conservative. Next, we demonstrate our methods by using them to revisit Porter's (1983) empirical study. Finally, we provide approximate null distributions for popular models and a method for obtaining conservative bounds when approximations are otherwise hard to obtain.

This paper is organized as follows. In Section 2, we assume that given data follow a Markov regime-switching process, and we show that when a mixture model is applied to these data, we can obtain an associated QLR statistic that can be used to test the number of regimes. Further, we discuss how to obtain critical values or their conservative approximations. Section 3 provides results of Monte Carlo experiments in which we compare the size and the power of our statistics with others in the literature. We revisit Porter's (1983) analysis in Section 4. Mathematical proofs are collected in the Appendix.

Before proceeding, we introduce some useful mathematical notation. We let "$\Rightarrow$" denote "weakly converges to," and $\|\cdot\|$ and $\|\cdot\|_\infty$ are the Euclidean and the uniform norms respectively.

2. Markov Regime-Switching Processes and Mixture Models

We consider a specific framework designed to facilitate analysis of key aspects of the problem of interest, using the following data generating process (DGP).

A1: (*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,* β_τ, *is at most* $c\rho^\tau$.
(*ii*) *The sequence of unobserved regime indicators,* $\{S_t \in \{1, 2\}\}_{t=1}^n$, *is generated as a first-order Markov process such that* $P(S_t = j|S_{t-1} = i) = p^*_{ij}$ with $p^*_{ii} \in [0, 1]$ $(i, j = 1, 2)$.
(*iii*) *The given* $\{X_t\}$ *is a* Markov regime-switching process (hidden Markov 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* $(X^{t-1}, S^t) := (X'_{t-1}, \cdots X'_1, S_t, \cdots, S_1)$; $r_0 \in \mathbb{N}$; *and the conditional cumulative distribution function* (*CDF*) *of* $X_t|\mathcal{F}_{t-1}$, $F(\,\cdot\,|X^{t-1}; \theta^*_0, \theta^*_j)$, *has a probability density function* (*PDF*) $f(\,\cdot\,|X^{t-1}; \theta^*_0, \theta^*_j)$ $(j = 1, 2)$. *Further, for* $(p^*_{11}, p^*_{22}) \in [0, 1) \times [0, 1) \setminus \{(0, 0)\}$, θ^* *is unique in* $\mathbb{R}^{r_0+2}$.

The β-mixing condition is suitable for the Markov regime-switching process, as discussed by Davydov (1973), Doukhan (1994), and Vidyasagar (2003). As well reviewed by Ephraim and Merhav (2002), the popularity of this DGP extends far beyond economics. In economics, Porter (1983) examines the cartel stability problem, assuming $p^*_{11} = p^*_{21}$, as we discuss in Section 4. Hamilton (1989) considers the case in which $X_t|\mathcal{F}_{t-1}$ is a function of $S_{t-m}, \cdots, S_t$ $(m \in \mathbb{N} \cup \{0\})$ and X^{t-1}, so that the unobserved two state process, $\{S_t\}$, induces a DGP for $X_t|\mathcal{F}_{t-1}$ with 2^{m+1} unobserved states. In this paper, we restrict our attention to the case $m = 0$ and focus strictly on testing for regime switching. Also, note that we cannot assume that $p^*_{11} = p^*_{22} = 0$, because if so $\{S_t\}$ becomes deterministically periodic, implying that $\{X_t\}$ is unconditionally heterogeneous, thus violating the stationarity assumption.

Many models for this DGP have been proposed. We consider the following model.

A2: (*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* $\tilde{\Theta} := \Theta_0 \times \Theta_* \in \mathbb{R}^{r_0+1}$; and Θ_0 *and* Θ_* *are convex and compact sets in* $\mathbb{R}^{r_0}$ *and* $\mathbb{R}$ *respectively. Further, for each* $\theta^j \in \tilde{\Theta}$, $f(\,\cdot\,|X^{t-1}; \theta^j)$ *is a measurable PDF with CDF* $F(\,\cdot\,|X^{t-1}; \theta^j)$ $(j = 1, 2)$.
(*ii*) *For every* $x \in \mathbb{R}^d$, $f(x|X^{t-1}; \cdot\,) \in \mathcal{C}^{(2)}(\tilde{\Theta})$ (*the set of twice continuously differentiable functions on* $\tilde{\Theta}$) *almost surely.* (*a.s.*)

For notational simplicity, we will abbreviate $F(X_t|X^{t-1};\theta^j)$ and $f(X_t|X^{t-1};\theta^j)$ as $F_t(\theta^j)$ and $f_t(\theta^j)$ respectively ($j=1,2$). Also, unless confusion will otherwise result, we omit the function argument placeholder, so that as an example, $f_t(\,\cdot\,)$ is also denoted as f_t.

Suppose that the researcher wishes to test whether there is only a single regime. Formally, relevant hypotheses are: for an unknown θ_*,

$$H_0\text{: } p_{11}^*=1 \text{ and } \theta_1^*=\theta_*;\; p_{22}^*=1 \text{ and } \theta_2^*=\theta_*;\; \text{or } \theta_1^*=\theta_2^*=\theta_*;$$

$$H_1\text{: } (p_{11}^*,p_{22}^*)\in[0,1)\times[0,1)\setminus\{(0,0)\} \text{ and } \theta_1^*\neq\theta_2^*,$$

where θ_* is defined in the following A3.

A3: (θ_0^*,θ_*) *maximizes* $n^{-1}E[\sum_{t=1}^n\tilde{\ell}_t]$ *uniquely in the interior of* $\tilde{\Theta}$, *where for each* θ^1, $\tilde{\ell}_t(\theta^1):=\log(f_t(\theta^1))$.

Given the Assumptions A1 and A2(i), the log-likelihood function can be represented as

$$L_n(p_{11},p_{22},\theta):=\log\left(\boldsymbol{\pi}'\left[\prod_{t=1}^n\mathbf{P}\mathbf{F}_t(\theta)\right]\boldsymbol{\iota}\right),$$

where for each t and $\theta\in\Theta:=\Theta_0\times\Theta_*\times\Theta_*$, $\mathbf{F}_t(\theta)$ is a 2×2 diagonal matrix with j-th diagonal element $f_t(\theta^j)$; $\mathbf{P}:=[p_{ij}]$ parameterizes the transition matrix of S_t ($i,j=1,2$); $\boldsymbol{\pi}:=[\pi,\;1-\pi]'$ with $\pi:=(1-p_{22})/(2-p_{11}-p_{22})$; and $\boldsymbol{\iota}$ is a 2×1 vector of ones. Because the log-likelihood function cannot be represented as a sum of individual log-likelihood functions, the LR statistic turns out to behave in unappealing ways. Specifically, if $p_{11}^*=1$ (or $p_{22}^*=1$), then the associated null first-order derivative of the log-likelihood function has a population variance growing geometrically as the sample size increases, and the standard central limit theorem cannot be applied. Further, as implied in Section 2.3, the power of the LR statistic is weaker than in the standard $n^{1/2}$ case when $\theta_1^*=\theta_2^*=\theta_*$. Because of these difficulties, we take a different approach.

2.1. A Quasi-LR Statistic

We avoid these difficulties by focusing instead on the *quasi*-likelihood function for a mixture model. As we show, this permits us to estimate key aspects of the Markov regime-switching process without sacrificing much. Thus, consider the mixture model quasi-log-likelihood function defined as follows: for each $(\pi,\theta)\in[0,1]\times\Theta$, let $L_n^*(\pi,\theta):=\sum_{t=1}^n\ell_t(\pi,\theta)$, where $\ell_t(\pi,\theta):=\log(\pi f_t(\theta^1)+(1-\pi)f_t(\theta^2))$.

This model captures the mixture aspect of the conditional PDF of $X_t|\sigma(X^{t-1})$ and the unconditional PDF of S_t under the alternative. More precisely, the conditional PDF of $X_t|\sigma(X^{t-1})$ is $f_t(\theta_0^*,\theta_1^*)P(S_t=1|\sigma(X^{t-1}))+f_t(\theta_0^*,\theta_2^*)P(S_t=2|\sigma(X^{t-1}))$, and the mixture weights, $P(S_t=1|\sigma(X^{t-1}))$ and $P(S_t=2|\sigma(X^{t-1}))$, are random variables with unconditional means, π^* and $1-\pi^*$, respectively, where $\pi^*:=(1-p_{22}^*)/(2-p_{11}^*-p_{22}^*)$. We replace these with an unknown parameter and estimate by maximizing the quasi-log-likelihood function. This specification ignores the serial correlation in $\{S_t\}$, whereas the serial correlation of X_t is captured by f_t. Thus, we work with a model that ignores the serial correlation in the unobserved state; Theorem 1 below shows this does not matter for testing the number of regimes.

In particular, our QLR test is potentially powerful against any regime-switching process, even if S_t is not a Markov process. Rather than testing for the specific serial correlation implied by regime switching, we test for the mixture properties of $X_t|\sigma(X^{t-1})$ generated by S_t. Here we leave aside testing for this serial correlation. (See Carrasco, Hu, and Ploberger (2004), who recently propose a test statistic of this sort.)

Despite the extensive analysis of mixture models for the case of i.i.d. data (see Hartigan (1985), Ghosh and Sen (1985) Liu and Shao (2003) and references therein), mixture models have not often been used for testing for regime switching. Our analysis thus not only contributes to the mixture model literature by showing its utility for testing the number of regimes, but also contributes to both the Markov switching and mixture model literature by demonstrating the utility of mixture models in testing the number of Markov regimes in a β-mixing context. We also extend the scope of the mixture models by considering other popular mixtures yielding regime-switching test statistics whose null limiting distributions are different from those in the literature. We further contribute by providing formulae and algorithms that can be used to calculate critical values and/or upper bounds for our test statistics that are useful in applications.

The almost sure limits of the estimators maximizing L_n^* can be represented in terms of the coefficients of the DGP both under the null and the alternative. For this, we assume the following regularity conditions.

A4: *For all* $(\pi, \theta) \in [0,1] \times \Theta$, $n^{-1}E[\sum_{t=1}^n \ell_t(\pi, \theta)]$ *exists and is finite.*

A5: (i) *There exists a sequence of positive, strictly stationary, and ergodic random variables,* $\{M_t\}$, *such that* (a) $E[M_t] < \Delta < \infty$; (b) $\sup_{(\pi,\theta)\in[0,1]\times\Theta} \|\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)\|_\infty \leq M_t$.

These assumptions are mild and enable us to apply the strong uniform law of large numbers (SULLN) to the mixture model quasi-log-likelihood.

THEOREM 1: (a) *Given A1, A2(i, ii), A3, A4, A5(i), and* H_0, $(\hat{\pi}_n^q, \hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q) \to \{(\pi, \theta_0^*, \theta_1, \theta_2) \in [0,1] \times \Theta : \pi = 1 \text{ and } \theta_1 = \theta_*;\ \text{or } \theta_1 = \theta_2 = \theta_*;\ \text{or } \pi = 0 \text{ and } \theta_2 = \theta_*\}$ *a.s., and* $(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) \to (\theta_0^*, \theta_*)$ *a.s., where* $(\hat{\pi}_n^q, \hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q)$ *is the ("unrestricted") quasi-MLE (QMLE) of the mixture model, and* $(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)$ *is the ("restricted") QMLE imposing* H_0.[1]
(b) *Given A1, A2(i, ii), A3, A4, A5(i), and* 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.*

Under the null, the QMLE is the MLE, and it converges to a set in Theorem 1(a). The conclusion of Theorem 1(b) is crucial to the goal of this paper. As pointed out by Levine (1983), a correct model specification for the conditional mean is important for consistent estimation of the conditional mean, but correct specification of DGP dynamics is not necessary. Theorem 1 assumes that $X_t|\mathcal{F}_{t-1}$ is correctly specified, and ignores the dynamics induced by $\{S_t\}$. Levine's (1983) point applies to the current context, and from this, it follows that $(\hat{\theta}_{0,n}^q, \hat{\theta}_{1,n}^q, \hat{\theta}_{2,n}^q) \to (\theta_0^*, \theta_1^*, \theta_2^*)$ a.s. We show additionally that the estimator for the parameter π replacing the random weights, $P(S_t = 1|\sigma(X^{t-1}))$, is consistent for the unconditional mean of the random weights. That is, $\hat{\pi}_n^q$ converges to π^* a.s. Thus, the mixture model is correctly specified

[1] The superscripts 'q' and 'n' are used to denote 'quasi-MLE' and 'null-imposing MLE' respectively.

for both $X_t|\sigma(X^{t-1})$ and the unconditional mean of $\{S_t\}$, but misspecified in terms of the dynamics of $\{S_t\}$.

Regime-switching tests can be based on the limits of the estimators. Note that $\pi^* = 1$ if and only if $p^*_{11} = 1$ (and $\pi^* = 0$ if and only if $p^*_{22} = 1$), so that there are two regimes if and only if $\pi^* \in (0,1)$. We exploit this fact and test these hypotheses using the QLR statistic defined as the log-likelihood ratio computed from the QMLEs under the null and the alternative. This modifies our prior hypotheses as follows: for an unknown θ_*,

$$H'_0 : \pi^* = 1 \text{ and } \theta^*_1 = \theta_*; \theta^*_1 = \theta^*_2 = \theta_*; \text{ or } \pi^* = 0 \text{ and } \theta^*_2 = \theta_*; \text{versus } H'_1 : \pi^* \in (0,1) \text{ and } \theta^*_1 \neq \theta^*_2.$$

The null H'_0 can be further partitioned: for an unknown θ_*, $H'_{01} : \pi^* = 1$ and $\theta^*_1 = \theta_*$; $H'_{02} : \theta^*_1 = \theta^*_2 = \theta_*$; or $H'_{03} : \pi^* = 0$ and $\theta^*_2 = \theta_*$. In this context, several standard assumptions are violated, summarized as follows. First, if $\pi^* = 1$ (resp. $= 0$), then θ^*_2 (resp. θ^*_1) is not identified, so that the Davies problem (1977, 1987) occurs: a "nuisance" parameter is present only under the alternative. At the same time, π^* is on the boundary of [0, 1], which also violates the standard condition yielding the chi-square limiting distribution for the QLR statistic. Second, if $\theta^*_1 = \theta^*_2$, then π^* is not identified. Hence, the nuisance parameter problem again occurs, but the boundary parameter problem does not appear. The results of Andrews (2001) thus play a key role in analyzing H'_{01} and H'_{03}, but not H'_{02}. To analyze H'_{02}, it turns out that the standard second-order derivative-based approximation to the QLR statistic has to be improved to approximations using higher-order derivatives. We resolve these challenges under each hypothesis and combine the ensuing results to derive the null limiting distribution of the QLR statistic.

2.2. Null Distribution of the QLR Statistic under H'_{01} and H'_{03}

We first examine the QLR behavior under H'_{01} and H'_{03} with suitable regularity conditions.

A5: (*ii*) *There exists a sequence of positive, strictly stationary, and ergodic random variables*, $\{M_t\}$, *such that* (*a*) *for some* $\delta > 0$, $E[M_t^{1+\delta}] < \Delta < \infty$; (*b*) $\sup_{(\pi,\theta)\in[0,1]\times\Theta} \|\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)'\|_\infty \leq M_t$; *and* (*c*) $\sup_{(\pi,\theta)\in[0,1]\times\Theta} \|\nabla^2_{(\pi,\theta)}\ell_t(\pi,\theta)\|_\infty \leq M_t$.

A6: (*i*) *For each* $(\pi^*, \theta^*_0, \theta^*_1, \theta^*_2)$ *with* $\theta^*_1 \neq \theta^*_2$, $\lambda_{\min}(B(\pi^*, \theta^*_0, \theta^*_1, \theta^*_2)) \geq 0$ *such that* (*a*) *if* $\lambda_{\min}(B(\pi^*, \theta^*_0, \theta^*_1, \theta^*_2)) > 0$, *then* $\lambda_{\max}(B(\pi^*, \theta^*_0, \theta^*_1, \theta^*_2)) < \infty$; *or* (*b*) *if* $\lambda_{\min}(B(\pi^*, \theta^*_0, \theta^*_1, \theta^*_2)) = 0$, *then* $\pi^* = 1$ *or* 0, *and for each* $\theta_2 \neq \theta_*$ *and* $\theta'_2 \neq \theta_*$, $\lambda_{\min}(C^{(\theta)}(\theta_2, \theta'_2)) > 0$ *and* $\lambda_{\max}(C^{(\theta)}(\theta_2, \theta'_2)) < \infty$, *where for each* (π, θ), $B(\pi,\theta) = E[\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)']$; *for each* (θ_2, θ'_2),

$$C^{(\theta)}(\theta_2, \theta'_2) := \begin{bmatrix} C^{(\theta)}_{11}(\theta_2, \theta'_2) & C^{(\theta)}_{12}(\theta'_2) \\ C^{(\theta)}_{21}(\theta_2) & C^{(\theta)}_{22} \end{bmatrix} := \begin{bmatrix} E[r_t(\theta_2)r_t(\theta'_2)] - 1 & -E[r_t(\theta'_2)r^{(1)}_t(\theta_*)] \\ -E[r_t(\theta_2)r^{(1)}_t(\theta_*)] & E[r^{(1)}_t(\theta_*)r^{(1)}_t(\theta_*)'] \end{bmatrix},$$

$r_t(\theta_2) := f_t(\theta^*_0, \theta_2)/f_t(\theta^*_0, \theta_*)$, *and* $r^{(1)}_t(\theta_2) := \nabla_{\theta^1} f_t(\theta^*_0, \theta_2)/f_t(\theta^*_0, \theta_*)$; *and* $\lambda_{\max}(\cdot)$ *and* $\lambda_{\min}(\cdot)$ *denote the maximum and the minimum eigenvalues of a given matrix, respectively.*

These assumptions enable us to apply the CLT on the set of unidentified parameters. In particular, we impose A6(i) to approximate the quasi-log-likelihood functions by quadratic functions. More precisely, although $n^{-1}E[\sum \ell_t]$ is not uniquely maximized under H'_{01} and H'_{03} and therefore cannot be usefully approximated by a quadratic function, A6(i) nevertheless ensures that for each $\theta_2(\neq \theta_*)$, $n^{-1}E[\sum \ell_t(1,\,\cdot\,,\,\cdot\,,\theta_2)]$ (or, for each $\theta_1(\neq \theta_*)$, $n^{-1}E[\sum \ell_t(0,\,\cdot\,,\theta_1,\,\cdot\,)]$) can be locally approximated by quadratic functions under H'_{01} (or H'_{03}), providing the necessary degree of identification through $C^{(\theta)}(\theta_2,\theta_2)$. By A6(i), the null model is identified for each θ_2. The QLR scores have the following properties.

LEMMA 1: (*a*) *For each element in* $\{(1,\theta_0^*,\theta_*,\theta_2)\in[0,1]\times\Theta:\theta_2\in\Theta_*\setminus\{\theta_*\}\}$, *define*

$$S_{1,n}(\theta_2):=\left[n^{-1}\sum_{t=1}^{n}\nabla_{(\pi,\,\theta^1)}\ell_t(\pi,\theta)\nabla_{(\pi,\,\theta^1)}\ell_t(\pi,\theta)'\right]^{-1}\left[n^{-1/2}\sum_{t=1}^{n}\nabla_{(\pi,\,\theta^1)}\ell_t(\pi,\theta)\right].$$

Given A1, A2(i, ii), A3, A4, A5(ii), A6(i), and H'_{01}, $S_{1,n}\Rightarrow\mathcal{S}_1$ *over* $\Theta_*(\epsilon):=\{\theta_2\in\Theta_*:|\theta_2-\theta_*|>\epsilon\}$ *for each* $\epsilon>0$ *such that for each* $\theta_2\in\Theta_*(\epsilon)$, $\mathcal{S}_1(\theta_2)\sim N(0,C^{(\theta)}(\theta_2,\theta_2)^{-1})$, *and for each* θ_2, $\theta_2'\in\Theta_*(\epsilon)$, $E[\mathcal{S}_1(\theta_2)\mathcal{S}_1(\theta_2')]=C^{(\theta)}(\theta_2,\theta_2)^{-1}C^{(\theta)}(\theta_2,\theta_2')C^{(\theta)}(\theta_2',\theta_2')^{-1}$.
(*b*) *Given the same assumptions as in Lemma 1(a),* $\mathcal{G}:\Theta_*(\epsilon)\mapsto\mathbb{R}$ *is differentiable in the mean, where for each* θ_2,θ_2', $\mathcal{G}(\theta_2):=\Omega^{(\theta)}(\theta_2,\theta_2)^{1/2}\mathcal{S}_1^{[1:1]}(\theta_2)$ *and* $\Omega^{(\theta)}(\theta_2,\theta_2'):=C_{11}^{(\theta)}(\theta_2,\theta_2')-C_{12}^{(\theta)}(\theta_2)[C_{22}^{(\theta)}]^{-1}C_{21}^{(\theta)}(\theta_2')$. *Further,* $A^{[i:j]}$ *is a sub-vector containing the i-th through j-th elements of the vector A.*[2]

We omit explicit analysis for H'_{03}, as the same score as $S_{1,n}$ is obtained by symmetry.

The proof of Lemma 1 involves considering the joint behavior of a continuum of random variables. For each θ_2, we can use $S_{1,n}(\theta_2)$ to approximate the quasi-log-likelihood function by a quadratic function. Let the associated quasi-log-likelihood function be defined as $QLR_{1,n}(\theta_2):=2(L_n^*(\hat{\pi}_n^q(\theta_2),\hat{\theta}_{0,n}^q(\theta_2),\hat{\theta}_{1,n}^q(\theta_2),\theta_2)-L_n^*(1,\theta_0^*,\theta_*,\theta_2))$, where $(\hat{\pi}_n^q(\theta_2),\hat{\theta}_{0,n}^q(\theta_2),\hat{\theta}_{1,n}^q(\theta_2)):=\arg\max_{(\pi,\theta^1)\in[0,1]\times\tilde{\Theta}}L_n^*(\pi,\theta)$. Then, for given θ_2, $QLR_{1,n}(\theta_2)=S_{1,n}(\theta_2)'C^{(\theta)}(\theta_2,\theta_2)S_{1,n}(\theta_2)+o_p(1)$ under H'_{01}, if we ignore the boundary parameter for the moment. Theorem 1(a) shows that $\hat{\theta}_{2,n}^q$ does not converge to any particular value in $\Theta_*(\epsilon)$. Thus, the limit of $S_{1,n}$ needs to be derived instead. For this, the finite dimensional distributions of $S_{1,n}$ are first shown to converge weakly to those of $\mathcal{S}_1$, and we show further that this distribution is tight. The desired results of Lemma 1(a) then follow by theorem 7.1 of Billingsley (1999), and this yields the specified Gaussian process as the limiting process of $S_{1,n}$. Tightness is proved by relying on Doukhan, Massart and Rio (1995) and Hansen (1996, 2004), who provide sufficient conditions for tightness in the β-mixing context. Next, we show that the covariance function of $\mathcal{G}$ has a generalized second-order derivative. Differentiability in the mean follows by Grenander (1981, theorem 1, ch. 2-2). Later, this yields a conservative rejection region.

There are several interesting aspects to Lemma 1. First, because the model is correctly specified under the null, each score is a martingale difference sequence, and the information matrix equality holds. This ensures that we can represent the covariance of $\mathcal{S}_1$ by $C^{(\theta)}(\theta,\theta')$. Note also that if $\theta_2=\theta_*$, then $S_{1,n}(\theta_2)$ isn't defined, as $\nabla_\pi L_n(\pi,\theta^*)\equiv 0$, so that $[-n^{-1}\nabla^2_{(\pi,\,\theta^1)}L_n^*(\pi,\theta)]^{-1}$ isn't necessarily defined uniformly

[2] We call a stochastic process, $\{\mathcal{G}:\Theta\mapsto\mathbb{R}^k,k\in\mathbb{N}\}\subset L^2(\Theta)$, *differentiable in the mean on* Θ, if there is a stochastic process, $\{\mathcal{G}':\Theta\mapsto\mathbb{R}^k\}\subset L^2(\Theta)$, such that for all $\theta\in\Theta$, $\lim_{\|h\|\to 0}E[[(\mathcal{G}(\theta+h)-\mathcal{G}(\theta))/\|h\|-\mathcal{G}'(\theta)]^2]=0$.

on $\{(1, \theta_0^*, \theta_*, \theta_2) \in [0,1] \times \Theta : \theta_2 \in \Theta_* \setminus \{\theta_*\}\}$ or in n. It could even be negative definite near θ_* for some n, so that the usual approximation using the Hessian may behave quite badly. To prevent this, we replace the conventional score with $S_{1,n}$, exploiting the information matrix equality. Second, for each θ_2, $\mathcal{S}_1(\theta_2)' C^{(\theta)}(\theta_2, \theta_2) \mathcal{S}_1(\theta_2)$ can be decomposed into $\mathcal{G}_1(\theta_2)^2$ and other terms. As shown below, $\mathcal{G}_1(\theta_2)^2$ forms the weak limit of the QLR statistic under H'_{01}, but the boundary parameter condition needs to be adjusted. Third, the distribution of $\mathcal{G}$ may be stationary or non-stationary. That is, $E[\mathcal{G}(\theta_2)\mathcal{G}(\theta_2 + \tau)]$ can be a function of τ only or of both τ and θ_2. This property depends on the DGP as well as the model. Finally, as θ_2 tends to θ_*, $\Omega^{(\theta)}(\theta_2, \theta_2)$ tends to zero, because $C_{11}^{\theta}(\theta, \theta)$ and $C_{12}^{\theta}(\theta)$ converge to zero. This raises a question about the existence of plim $_{\theta_2 \to \theta_*} \mathcal{G}(\theta_2)$. We investigate this after the QLR statistic is examined under H'_{02}. Also, note that for given θ_2 and θ'_2, $\Omega^{(\theta)}(\theta_2, \theta'_2)$ is the asymptotic covariance between $n^{-1/2} \sum (1 - f_t(\hat{\theta}_{0,n}^n, \theta_2)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n))$ and $n^{-1/2} \sum (1 - f_t(\hat{\theta}_{0,n}^n, \theta'_2)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n))$.

The asymptotic distribution of the QLR statistic under H'_{01} can be derived by using Lemma 1.

THEOREM 2: *Given A1, A2(i, ii), A3, A4, A5(ii), A6(i), and* H'_{01}, *for each* $\epsilon > 0$, $QLR_n(\epsilon) := \max_{\theta_2 \in \Theta_*(\epsilon)} (QLR_{1,n}(\theta_2) - QLR_{2,n}) \Rightarrow \mathcal{H}(\epsilon) := \sup_{\theta_2 \in \Theta_*(\epsilon)} (\min[0, \mathcal{G}(\theta_2)])^2$, *where* $QLR_{2,n} := 2(L_n^*(1, \hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n, \theta_2) - L_n^*(1, \theta_0^*, \theta_*, \theta_2))$, *and* θ_2 *in* $L_n^*(1, \hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n, \theta_2)$ *is a placeholder whose value is irrelevant.*

An advantage of the QLR statistic under H'_{01} is that its weak convergence limit exists under mild conditions. Theorem 2 extends Ghosh and Sen's (1985) result for i.i.d. data to the β-mixing time series context. To interpret the QLR statistic, we note that the first piece, $\max_{\theta_2 \in \Theta_*(\epsilon)} QLR_{1,n}(\theta_2)$, tests a joint hypothesis: there is a single regime and $X_t | \mathcal{F}_{t-1} \sim F(\cdot | X^{t-1}; (\theta_0^*, \theta_*))$, whereas $QLR_{2,n}$ tests the single hypothesis $X_t | \mathcal{F}_{t-1} \sim F(\cdot | X^{t-1}; (\theta_0^*, \theta_*))$. Thus, the QLR statistic tests only the number of regimes, as desired.

The boundary parameter condition involves only the negative part of the score under H'_{01}. The boundary parameter associated with $S_{1,n}$ splits $\mathcal{G}$ into positive and negative pieces and discards the positive piece, resulting in the appearance of the "min" operator in the conclusion of Theorem 2. Technical considerations relevant to the boundary parameter problem trace from Chernoff (1954), Self and Liang (1987) and Andrews (1999). As these results resolve the boundary parameter problem only for identified models, we can apply their conclusions only to $QLR_{1,n}(\theta_2)$ for given θ_2. Andrews (2001) provides further relevant theory for unidentified models with boundary parameter problems. We utilize his advances to obtain Theorem 2. In particular, from the given approximations and the boundary parameter condition, $\max_{\theta_2 \in \Theta_*(\epsilon)} QLR_{1,n}(\theta_2) \Rightarrow \mathcal{H}(\epsilon) + Z'Z$ under H'_{01}, where $Z'Z$ is identically the probability limit of $QLR_{2,n}$, following the chi-square distribution. Thus, $QLR_n(\epsilon)$ weakly converges to $\mathcal{H}(\epsilon)$ under H'_{01}. We emphasize that the convergence limit of $\max_{\theta_2 \in \Theta_*(\epsilon)} QLR_{1,n}(\theta_2)$ separates into two pieces, $\mathcal{H}(\epsilon)$ and $Z'Z$ (depending on whether the boundary parameter problem arises or not), and the probability limit of $QLR_{2,n}$, which is the identical random variable $Z'Z$. As pointed out by one of the referees, the estimation error for parameters not on the boundary has the same limit as the estimation error obtainable when the boundary parameters are known in advance.

The nature of the parameter space Θ_* fundamentally affects the probability law of $\mathcal{H}(\epsilon)$. As pointed out by Hartigan (1985) and Lindsay (1995), if we allow Θ_* to be unbounded, then even the existence of $\mathcal{H}(\epsilon)$ is in question. Theorem 3 formally underscores the importance of assuming compact Θ_*.

THEOREM 3: *Given A1, A2(i, ii), A3, A4, A5(ii), A6(i), and H'_{01}, for all $\epsilon > 0$, $P(\sup_{\theta_2 \in \Theta_*(\epsilon)} |\mathcal{G}(\theta_2)| < \infty) = 1$.*

We prove Theorem 3 using the fact in Lifshits (1995) that a Gaussian process is bounded with probability one if and only if it has a finite oscillation a.s. when the given parameter space, $\Theta_*(\epsilon)$, can be covered by a finite number of open balls with radius measured by the semi-metric $E[(\mathcal{G}(\theta_2) - \mathcal{G}(\theta'_2))^2]$ for $\theta_2 \neq \theta'_2 \in \Theta_*(\epsilon)$. If Θ_* is unbounded and the covariance between $\mathcal{G}(\theta_2)$ and $\mathcal{G}(\theta'_2)$ converges to 0 as $|\theta_2 - \theta'_2|$ gets large, then for given $\eta \in (0,1)$ and $K \in \mathbb{R}^+$, we can choose a set of parameters, say $\{\theta_{2i}\}_{i=1}^{n(\eta,K)}$, such that $P(\sup_{\theta_2 \in \{\theta_{2i}\}_{i=1}^{n(\eta,K)}} |\mathcal{G}(\theta_2)| > K) > 1 - \eta$. This implies that by letting K grow we can ensure that $\mathcal{H}(\epsilon)$ eventually diverges to infinity in probability. We avoid this by requiring Θ_* to be bounded. In terms of Hartigan (1985) and Lindsay (1995), therefore, our focus here should be understood as investigating how to exploit the QLR statistic for a given compact parameter space. Our boundedness requirement is further underscored by the warnings raised by Azaïs, Gassiat, and Mercadier (2006) against using an unbounded parameter space. The Monte Carlo experiments for mixtures of exponentials in Mosler and Seidel (2001) show the lack of convergence in distribution in this context.

Another interesting aspect of Theorem 3 is that the QLR statistic has a model-dependent null distribution. As mentioned following Lemma 1, the null distribution of $\mathcal{G}$ depends on both the DGP and the model. This situation has been recognized in the goodness-of-fit test literature by Darling (1955) and Durbin (1973). Their insights apply here. Further, the parameter space, Θ_*, is another source of model dependence for our QLR statistic. For example, if there are two parameter spaces, say $\Theta_*^{(1)}$ and $\Theta_*^{(2)}$, such that $\Theta_*^{(1)} \subset \Theta_*^{(2)}$, then $P(\sup_{\theta_2 \in \Theta_*^{(1)}(\epsilon)} (\min[0, \mathcal{G}(\theta_2)])^2 > K) \leq P(\sup_{\theta_2 \in \Theta_*^{(2)}(\epsilon)} (\min[0, \mathcal{G}(\theta_2)])^2 > K)$ and $P(\sup_{\theta_2 \in \Theta_*^{(1)}(\epsilon)} (\min[0, \mathcal{G}(\theta_2)])^2 = 0) \geq P(\sup_{\theta_2 \in \Theta_*^{(2)}(\epsilon)} (\min[0, \mathcal{G}(\theta_2)])^2 = 0)$. Thus, for different parameter spaces, different critical values will apply, and the point mass given to 0, the effect due to the boundary parameter problem, can eventually disappear. We investigate this in the model exercises of Section 2.5.

2.3. Null Distribution of the QLR Statistic under H'_{02}

2.3.1. *Non-zero Second-Order Derivative Case.* To examine the QLR statistic under H'_{02}, we first introduce relevant notation for our analysis. For given $(\pi, \theta_2) \in (0,1) \times \Theta_*$, let $(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) := \arg\max_{\theta^1 \in \tilde{\Theta}} L^*_n(\pi, \theta)$, which satisfies the first-order conditions (FOCs), so that for each (π, θ_2),

$$\nabla_{\theta_0} L^*_n(\pi, \tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2), \theta_2) = \sum \frac{\pi f_t^{(1,0)}(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t^{(1,0)}(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)}{\pi f_t(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)} \equiv 0,$$

$$\nabla_{\theta_1} L^*_n(\pi, \tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2), \theta_2) = \sum \frac{(1-\pi) f_t^{(0,1)}(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2))}{\pi f_t(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)} \equiv 0,$$

where $f_t^{(i,j)} := \nabla^i_{\theta_0}\nabla^j_{\theta_1} f_t$. Note that $(\tilde{\theta}^q_{0,n}, \tilde{\theta}^q_{1,n})$ should have been represented as a function of π, too, but we omit this, as π will be taken as given under H'_{02}. Each component in the expressions above can be appropriately exploited for our further analysis. For each (π, θ_2), we thus simplify by letting

$$h_t(\theta_2) := f_t^{(0,1)}(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2), \quad k_t(\theta_2) := f_t^{(0,1)}(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)),$$

$$m_t(\pi, \theta_2) := \pi f_t^{(1,0)}(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t^{(1,0)}(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2),$$

$$g_t(\pi, \theta_2) := 1/(\pi f_t(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)),$$

and $\tilde{L}_n(\pi, \theta_2) := L^*_n(\pi, \tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2), \theta_2)$. Then we can write:

$$\tilde{M}^{(1)}_n(\pi, \theta_2) := \nabla_{\theta_0} L_n(\pi, \tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2), \theta_2) = \sum m_t(\pi, \theta_2) g_t(\pi, \theta_2) \equiv 0,$$

$$\tilde{K}^{(1)}_n(\pi, \theta_2) := \nabla_{\theta_1} L_n(\pi, \tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2), \theta_2) = (1-\pi) \sum k_t(\theta_2) g_t(\pi, \theta_2) \equiv 0.$$

In the expressions immediately above, the superscript (1) denotes the first-order derivative with respect to θ_2. Later, we use the superscript (2) to denote the second-order derivative with respect to θ_2, etc. Thus,

$$\tilde{L}^{(1)}_n(\pi, \theta_2) := \nabla_{\theta_2} \tilde{L}_n(\pi, \theta_2) = \sum \frac{(1-\pi) f_t^{(0,1)}(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)}{\pi f_t(\tilde{\theta}^q_{0,n}(\theta_2), \tilde{\theta}^q_{1,n}(\theta_2)) + (1-\pi) f_t(\tilde{\theta}^q_{0,n}(\theta_2), \theta_2)},$$

implying that $\tilde{L}^{(1)}_n(\pi, \theta_2) = (1-\pi)\sum h_t(\theta_2) g_t(\pi, \theta_2)$. We further let $\hat{r}_t^{(i,j)} := f_t^{(i,j)}(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}) / f_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n})$ and $\hat{r}_t^{(1)} := \nabla_{\theta^1} f_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}) / f_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n})$; and where no confusion arises, for a given function of (π, θ_2), say q_t, $q_t(\pi, \hat{\theta}^n_{1,n})$ is abbreviated as $\hat{q}_t$. For example, $g_t(\pi, \hat{\theta}^n_{1,n})$ and $h_t(\hat{\theta}^n_{1,n})$ are denoted as $\hat{g}_t$ and $\hat{h}_t$.

The QLR statistic can be represented using this notation under H'_{02}. Note that for each π, we have $QLR_n(\pi) := 2(L^*_n(\pi, \hat{\theta}^q_{0,n}(\pi), \hat{\theta}^q_{1,n}(\pi), \hat{\theta}^q_{2,n}(\pi)) - L^*_n(1, \hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}, \theta_2))$, where $(\hat{\theta}^q_{0,n}(\pi), \hat{\theta}^q_{1,n}(\pi), \hat{\theta}^q_{2,n}(\pi)) := \max_\theta L^*_n(\pi, \theta)$; this is identical to $2(\tilde{L}_n(\pi, \hat{\theta}^q_{2,n}(\pi)) - \tilde{L}_n(\pi, \hat{\theta}^n_{1,n}))$, because $(\tilde{\theta}^q_{0,n}(\hat{\theta}^q_{2,n}(\pi)), \tilde{\theta}^q_{1,n}(\hat{\theta}^q_{2,n}(\pi)),$ $\hat{\theta}^q_{2,n}(\pi))$ and $(\tilde{\theta}^q_{0,n}(\hat{\theta}^n_{1,n}), \tilde{\theta}^q_{1,n}(\hat{\theta}^n_{1,n}))$ satisfy the FOCs under the alternative and the null model assumptions respectively, so that $\hat{\theta}^q_{0,n}(\pi) = \tilde{\theta}^q_{0,n}(\hat{\theta}^q_{2,n}(\pi))$, $\hat{\theta}^q_{1,n}(\pi) = \tilde{\theta}^q_{1,n}(\hat{\theta}^q_{2,n}(\pi))$, $\hat{\theta}^n_{0,n} = \tilde{\theta}^q_{0,n}(\hat{\theta}^n_{1,n})$ and $\hat{\theta}^n_{1,n} = \tilde{\theta}^q_{1,n}(\hat{\theta}^n_{1,n})$.

In addition to the identification problem, the QLR statistic exhibits another nonstandard property under H'_{02}. By the definitions of h_t and k_t, $\hat{h}_t := h_t(\hat{\theta}^n_{1,n}) = \hat{k}_t := k_t(\hat{\theta}^n_{1,n})$, implying that

$$\tilde{L}^{(1)}_n(\pi, \hat{\theta}^n_{1,n}) = (1-\pi)\sum \hat{h}_t \hat{g}_t = (1-\pi) \sum \hat{k}_t \hat{g}_t = \tilde{K}^{(1)}_n(\pi, \hat{\theta}^n_{1,n}) = 0.$$

That is, the first-order derivative is identically zero under H'_{02}. The score given by the first-order derivative cannot approximate the null distribution. Neymann and Scott (1966) and Lee and Chesher (1986) consider similar problems in the context of the C(α) statistic and recommend approximating the log-likelihood function using higher-order derivatives. Lindsay (1995) elaborates and shows this approach extends beyond the models of Neymann and Scott (1966). We follow these insights. A little algebra gives that for each (π, θ_2),

$$\tilde{L}^{(2)}_n(\pi, \theta_2) = (1-\pi)\sum\{h_t^{(1)}(\theta_2) g_t(\pi, \theta_2) + h_t(\theta_2) g_t^{(1)}(\pi, \theta_2)\},$$

$$\tilde{M}^{(2)}_n(\pi, \theta_2) = \sum\{m_t^{(1)}(\pi, \theta_2) g_t(\pi, \theta_2) + m_t(\pi, \theta_2) g_t^{(1)}(\pi, \theta_2)\} = 0,$$

$$\tilde{K}_n^{(2)}(\pi, \theta_2) = (1-\pi)\sum\{k_t^{(1)}(\theta_2)g_t(\pi,\theta_2) + k_t(\theta_2)g_t^{(1)}(\pi,\theta_2)\} = 0.$$

The last two equations hold because $\tilde{M}_n^{(1)}(\pi, \theta_2)$ and $\tilde{K}_n^{(1)}(\pi, \theta_2)$ are identically zero. Thus,

$$\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = \tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) - \tilde{K}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t.$$

Note that this involves first computing $\hat{\tilde{\theta}}_i^{(1)} := \tilde{\theta}_i^{(1)}(\hat{\theta}_{1,n}^n)$ $(i = 0, 1)$, and if these are plugged back into $\tilde{K}_n^{(2)}$ and $\tilde{M}_n^{(2)}$, then another set of identities is obtained. We can also compute $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)$ in the same way. Once again, $\hat{\tilde{\theta}}_i^{(2)}$ $(i = 0, 1)$ appears; we iterate this process to obtain $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$. Then for each π,

$$\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t + 2(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(1)}\},$$

$$\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(1)} + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)}\}.$$

The asymptotic behaviors of $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)$ to $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ are determined by each element on the right-hand side (RHS). We now collect the regularity conditions that enable us to apply the law of large numbers (LLN) and the central limit theorem (CLT) to each element.

A2: *(iii) For every* $x \in \mathbb{R}^d$, $f(x|X^{t-1}; \cdot) \in \mathcal{C}^{(4)}(\tilde{\Theta})$ *a.s.*

A5: *(iii) There exists a sequence of positive, strictly stationary, and ergodic random variables,* $\{M_t\}$, *such that (a) for some* $\delta > 0$, $E[M_t^{1+\delta}] < \Delta < \infty$; *(b)* $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1} f_t(\theta^1)/f_t(\theta^1)|^4 \leq M_t$; *(c)* $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2} f_t(\theta^1)$ $/f_t(\theta^1)|^2 \leq M_t$; *(d)* $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\theta^1)/f_t(\theta^1)|^2 \leq M_t$; *(e)* $\sup_{\theta^1 \in \tilde{\Theta}} |\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4} f_t(\theta^1)/f_t(\theta^1)| \leq M_t$, *where* $i_1, \cdots, i_4 \in \{\theta_{01}, \theta_{02}, \cdots, \theta_{0r_0}, \theta_1\}$.

A6: *(ii) For each* $(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$ *with* $\theta_1^* = \theta_2^*$ *and* $\pi^* \in (0, 1)$, $\lambda_{\min}(C^{(2)}) \geq 0$ *such that if* $\lambda_{\min}(C^{(2)}) > 0$, *then* $\lambda_{\max}(C^{(2)}) < \infty$, *where for* $k = 2, 3, \cdots$, $r_t^{(k)}(\theta_*) := (r_t^{(0,k)}(\theta_*), r_t^{(1)}(\theta_*)')$ *and*

$$C^{(k)} := E[r_t^{(k)}(\theta_*) r_t^{(k)}(\theta_*)'] := \begin{bmatrix} C_{11}^{(k)} & C_{12}^{(k)} \\ C_{21}^{(k)} & C_{22}^{(\theta)} \end{bmatrix} := \begin{bmatrix} E[r_t^{(0,k)}(\theta_*)^2] & E[r_t^{(0,k)}(\theta_*) r_t^{(1)}(\theta_*)]' \\ E[r_t^{(0,k)}(\theta_*) r_t^{(1)}(\theta_*)] & C_{22}^{(\theta)} \end{bmatrix}$$

whenever they exist.

We impose A6(ii) in order to approximate the quasi-log-likelihood function by a quartic function. Specifically, A6(ii) provides a condition relating the second-order and first-order derivatives under the alternative and the null respectively. Using these, we can obtain the following asymptotic properties.

LEMMA 2: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(ii), and* H'_{02}, *for each* π,
(a) $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = \frac{(1-\pi)}{\pi}\sum \hat{r}_t^{(0,2)} + o_p(n^{1/2})$;
(b) $n^{-1/2}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) \Rightarrow (\frac{1-\pi}{\pi})G_0^{(2)}$, *where* $G_0^{(2)} \sim N(0, \Omega^{(2)})$ *and* $\Omega^{(2)} := C_{11}^{(2)} - C_{12}^{(2)\prime}[C_{22}^{(\theta)}]^{-1}C_{21}^{(2)}$;
(c) $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(d) $n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = -3(\frac{1-\pi}{\pi})^2\Omega^{(2)} + o_p(1)$.

Lemma 2 is proved in the Appendix. The fact that $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$ implies that the QLR statistic can be non-degenerate even with the zero first-order derivative. Also, it is of interest that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$ can estimate the asymptotic variance of $n^{-1/2}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)$, as a result of the information matrix equality.

The asymptotic null distribution of the QLR statistic under H_{02}' can now be derived using Lemma 2. Note that for each π and $\bar{\theta}_2$ between θ_2 and $\hat{\theta}_{1,n}^n$,

$$\tilde{L}_n(\pi, \theta_2) = \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n) + \frac{1}{2!}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^2 + \frac{1}{3!}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^3 + \frac{1}{4!}\tilde{L}_n^{(4)}(\pi, \bar{\theta}_2)(\theta_2 - \hat{\theta}_{1,n}^n)^4$$

by the mean value theorem. Lemma 2 and theorem 3.9 of Billingsley (1999) imply that

$$\left(n^{-1/2}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n), n^{-3/4}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n), n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)\right) \Rightarrow \left(\left[\frac{1-\pi}{\pi}\right] G_0^{(2)}, 0, -3\left[\frac{1-\pi}{\pi}\right]^2 \Omega^{(2)}\right)$$

for each π. Thus, given the differentiability and the moment conditions, for each π,

$$\sup_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \sup_{\xi}\left[\frac{1-\pi}{\pi}\right] G_0^{(2)}\xi^2 - \frac{1}{4}\left[\frac{1-\pi}{\pi}\right]^2 \Omega^{(2)}\xi^4,$$

where ξ captures the asymptotic behavior of $n^{1/4}(\theta_2 - \hat{\theta}_{1,n}^n)$. From this, we obtain the following.

THEOREM 4: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(ii), and H_{02}', for each $\pi \in (0, 1)$,*
(a) $\max_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \max[0, G_0]^2$, *where* $G_0 \sim N(0, 1)$;
(b) for $\epsilon \in (0, 1/2)$, $QLR_n(\epsilon) := \max_{\pi\in[\epsilon,1-\epsilon]} \max_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \max[0, G_0]^2$.

There are a number of interesting aspects to Theorem 4. First, the proof of Theorem 4 is not too different from the standard argument, but it involves a sign constraint. Note that the QLR statistic is approximated mainly by the second and fourth-order derivatives, and ξ is raised to even powers, which cannot be less than zero. This gives rise to the square of the half normal distribution as the asymptotic null distribution, even without the boundary parameter condition. Second, the nuisance parameter, π, is present only in scaling $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)$ and $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$, so that the QLR statistic turns out to be nuisance parameter free by the information matrix equality. This also implies that the asymptotic null distribution of $2(\tilde{L}_n(\cdot, \theta_2) - \tilde{L}_n(\cdot, \hat{\theta}_{1,n}^n))$ is automatically tight, leading to Theorem 4(b). Third, $n^{-1/2}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)$ doesn't have to obey asymptotic normality for Theorem 4. It can be degenerate. What is required for the result is that $n^{-3/4}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = o_p(1)$. Fourth, the convergence rate of the estimator is different from the standard $n^{1/2}$ case. In this case, the convergence rate is $n^{1/4}$, which is the same as for the C(α) statistic of Neymann and Scott (1966) and Lindsay (1995). Thus, under H_{02}', the QLR test can have power comparable to that of the C(α) test asymptotically. Also, in turn, the "general quadratic approximation" of Lin and Shao (2003) cannot be applied to the likelihood ratio under H_{02}'. Goffinet, Loisel and Laurent (1992) report the same results as Lemma 4(a) in the case of a mixture of normals with unknown means but known common variance. Our analysis provides a general theory that nests theirs as a special case. Fifth, the given limiting random variable, G_0, is the probability limit of $\mathcal{G}(\theta_2)$ as θ_2 tends to θ_*. This feature will be explained in further detail below. Sixth, the literature often reports the tendency of mixtures of exponential family distributions to yield more stable simulation results than other distributions. This is mainly because they are infinitely differentiable, so that the fourth-order differentiation condition holds automatically. Simulation results can be unstable if the model is differentiable only up to the second-order. Finally, many mixture models can be analyzed by the fourth-order approximation, although there are many other popular mixtures that cannot be approximated using a fourth-order Taylor expansion.

2.3.2. *Zero Second-Order Derivative Case.* We often observe mixture models to have zero second-order derivatives under the null, because the second-order derivative turns out to be a linear function of the first-order derivatives: for each θ^1 and for some non-zero $(\alpha'\ \beta)' \in \mathbb{R}^{r_0+1}$,

$$f_t^{(0,2)}(\theta^1) = \alpha' f_t^{(1,0)}(\theta^1) + \beta f_t^{(0,1)}(\theta^1),$$

so that $\sum \hat{r}_t^{(0,2)} = 0$. The empirical example of Porter (1983) belongs to this case, so the quartic approximation of the previous section has to be improved. As will be clear later, the required approximation order is of the eighth order.[3] By further elaborating the prior derivatives, we have for $i = 3, \cdots, 8$,

$$\tilde{L}_n^{(i)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum \left\{ \left(\hat{h}_t^{(i-1)} - \hat{k}_t^{(i-1)}\right)\hat{g}_t + \sum_{j=1}^{i-2} \begin{pmatrix} i-1 \\ j \end{pmatrix} \left(\hat{h}_t^{(i-j-1)} - \hat{k}_t^{(i-j-1)}\right)\hat{g}_t^{(j)} \right\}.$$

As before, each component in the RHS contributes to the asymptotic null distribution of the QLR statistic. We provide suitable regularity conditions as follows.

A2: (*iv*) *For every* $x \in \mathbb{R}^d$, $f(x|X^{t-1};\cdot) \in \mathcal{C}^{(8)}(\tilde{\Theta})$ *a.s.*

A5: (*iv*) *There exists a sequence of positive, strictly stationary, and ergodic random variables*, $\{M_t\}$, *such that for some* $\delta > 0$, $E[M_t^{1+\delta}] < \Delta < \infty$; $\sup_{\theta^1\in\tilde{\Theta}} |\nabla_{i_1}\cdots\nabla_{i_k} f_t(\theta^1)/f_t(\theta^1)|^4 \le M_t$; $\sup_{\theta^1\in\tilde{\Theta}} |\nabla_{i_1}\cdots\nabla_{i_\ell} f_t(\theta^1)/f_t(\theta^1)|^2 \le M_t$; $\sup_{\theta^1\in\tilde{\Theta}} |\nabla_{\theta_1}^8 f_t(\theta^1)/f_t(\theta^1)| \le M_t$; $\sup_{\theta^1\in\tilde{\Theta}} |\nabla_{j_1}\nabla_{\theta_1}^7 f_t(\theta^1)/f_t(\theta^1)| \le M_t$; $\sup_{\theta^1\in\tilde{\Theta}} |\nabla_{j_1}\nabla_{j_2}\nabla_{\theta_1}^6 f_t(\theta^1)/f_t(\theta^1)| \le M_t$, *where* $k = 1,2,3,4$; $\ell = 5,6,7$; $i_1,\cdots,i_7 \in \{\theta_{01},\theta_{02},\cdots,\theta_{0r_0},\theta_1\}$; *and* $j_1, j_2 \in \{\theta_{01},\theta_{02},\cdots,\theta_{0r_0}\}$.

Assumptions A5(iii and iv) are not the most efficient possible moment conditions. For expositional purposes, we provide simple assumptions that are stronger than is strictly necessary. Note that the second, third, and fourth-order derivative conditions in A5(iv) are strengthened compared to A5(iii), and also that the highest moment-order condition is of fourth order even if the eighth-order derivative is involved. This contrasts sharply with the previous case. Recall that in the prior case, $(\hat{\hat{\theta}}_{0,n}^{(3)}, \hat{\hat{\theta}}_{1,n}^{(3)})$ must first be computed to obtain $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)$. Here we don't need to compute $(\hat{\hat{\theta}}_{0,n}^{(7)}, \hat{\hat{\theta}}_{1,n}^{(7)})$ to compute $\tilde{L}_n^{(8)}(\pi, \hat{\theta}_{1,n}^n)$, only $(\hat{\hat{\theta}}_{0,n}^{(6)}, \hat{\hat{\theta}}_{1,n}^{(6)})$. This is mainly because $\hat{\hat{\theta}}_{1,n}^{(7)}$ appears as a coefficient of $\sum \hat{r}_t^{(0,2)}$ (which is zero), and $\hat{\hat{\theta}}_{0,n}^{(7)}$ is not required to compute $\tilde{L}_n^{(8)}(\pi, \hat{\theta}_{1,n}^n)$. Consequently, our regularity conditions do not require finite eighth-order moments even if the eighth-order derivatives are involved.

Assumption A5(iv) and the next assumption enable us to apply the CLT.

[3] We are indebted to Robert Davies for guidance with this aspect of the problem.

A6: *(iii) For each* $(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$ *with* $\theta_1^* = \theta_2^*$ *and* $\pi^* \in (0,1)$, $\lambda_{\min}(C^{(2)}) \geq 0$ *such that if* $\lambda_{\min}(C^{(2)}) = 0$, *then* $\lambda_{\min}(C^{(s)}) > 0$ *and* $\lambda_{\max}(C^{(s)}) < \infty$, *where*

$$C^{(s)} := \begin{bmatrix} C_{11}^{(s)} & C_{12}^{(s)} \\ C_{21}^{(s)} & C^{(3)} \end{bmatrix} := \begin{bmatrix} E[s_t(\theta_*)^2] & E[s_t(\theta_*) r_t^{(3)}(\theta_*)'] \\ E[s_t(\theta_*) r_t^{(3)}(\theta_*)] & C^{(3)} \end{bmatrix},$$

and $s_t(\theta_*) := r_t^{(0,4)}(\theta_*) - 6\beta r_t^{(0,3)}(\theta_*) - 6\alpha' r_t^{(1,2)}(\theta_*) + 6\alpha' r_t^{(1,1)}(\theta_*)\beta + 3\alpha' r_t^{(2,0)}(\theta_*)\alpha$.

We partition $C_{12}^{(s)}$ into $[C_3^{(s)}, C_1^{(s)\prime}] := [E[s_t(\theta_*) r_t^{(3,0)}(\theta_*)], E[s_t(\theta_*) r_t^{(1)}(\theta_*)]']$ for future reference. As given below, $\sum \hat{s}_t$ is asymptotically equivalent to the fourth-order derivative, affecting the asymptotic distribution of the QLR statistic. Thus we require both $C^{(s)}$ and $C^{(3)}$ to be positive definite. We now have

LEMMA 3: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H_{02}', *for each* π,
(a) $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = -\frac{(1-\pi)(1-2\pi)}{\pi^2} \sum \hat{r}_t^{(0,3)}$;
(b) $n^{-1/2}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) \Rightarrow -\frac{(1-\pi)(1-2\pi)}{\pi^2} G_0^{(3)}$, *where* $G_0^{(3)} \sim N(0, \Omega^{(3)})$ *and* $\Omega^{(3)} := C_{11}^{(3)} - C_{12}^{(3)}[C_{22}^{(\theta)}]^{-1} C_{21}^{(3)}$;
(c) $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(d) $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$.

Lemma 3 holds for any $\pi \in (0,1)$. Nevertheless, care is needed. If $\pi = 1/2$, $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = 0$, so that the third-order derivative also turns out to be zero, implying that we need to differentiate one more time when $\pi = 1/2$. Goffinet, Loisel and Laurent (1992) observe the same phenomenon when considering the mixture of normals with unknown different means and an unknown common variance. Nevertheless, their analysis is incorrect, as they approximate the log-likelihood function only to the fourth order when $\pi = 1/2$. We suppose first that $\pi \neq 1/2$. In this case, the asymptotic variance of $n^{-1/2}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)$ can be estimated by the sixth-order derivative.

LEMMA 4: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii) and* H_{02}', *if* $\pi \in \{x \in (0,1) : x \neq 1/2\}$, *then*
(a) $n^{-1}\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = -10(\frac{(1-\pi)(1-2\pi)}{\pi^2})^2 \Omega^{(3)} + o_p(1)$;
(b) $\max_{\theta_2} 2(L_n(\pi, \theta_2) - L_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow G_0^2$.

The proof of Lemma 4 is straightforward. Using Lemmas 3 and 4, we can approximate the likelihood function as before. For each π and $\bar{\theta}_2$ between θ_2 and $\hat{\theta}_{1,n}^n$, the mean value theorem gives

$$\begin{aligned} \tilde{L}_n(\pi, \theta_2) = \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n) &+ \frac{1}{3!}\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^3 + \frac{1}{4!}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^4 \\ &+ \frac{1}{5!}\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n)(\theta_2 - \hat{\theta}_{1,n}^n)^5 + \frac{1}{6!}\tilde{L}_n^{(6)}(\pi, \bar{\theta}_2)(\theta_2 - \hat{\theta}_{1,n}^n)^6. \end{aligned}$$

Applying theorem 3.9 of Billingsley (1999) and Lemma 6 ensures that for each π,

$$\max_{\theta_2} 2(L_n(\pi, \theta_2) - L_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \sup_{\xi} -\frac{1}{3}\left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right] G_0^{(3)}\xi^3 - \frac{20}{6!}\left[\frac{(1-\pi)(1-2\pi)}{\pi^2}\right]^2 \Omega^{(3)}\xi^6.$$

Solving for the RHS gives us $[\Omega^{(3)}]^{-1}[G^{(3)}]^2$, which has the same distribution as G_0^2. This explains why the standard chi-square distribution is obtained here as the limiting distribution of the QLR statistic. Note that the information matrix equality holds here, and that the limiting distribution is nuisance parameter free.

If $\pi = 1/2$, we examine derivatives up to eighth order. We have the following large sample properties.

Lemma 5: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H'_{02}, *if* $\pi = 1/2$, *then*
(a) $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = \sum \hat{s}_t + o_p(n^{1/2})$;
(b) $n^{-1/2}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) \Rightarrow G_0^{(s)}$, *where* $G_0^{(s)} \sim N(0, \Omega^{(s)})$ *and* $\Omega^{(s)} := C_{11}^{(s)} - C_1^{(s)\prime}[C_{22}^{(\theta)}]^{-1}C_1^{(s)}$;
(c) $\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(d) $\tilde{L}_n^{(7)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;
(e) $n^{-1}\tilde{L}_n^{(8)}(\pi, \hat{\theta}_{1,n}^n) = -35\Omega^{(s)} + o_p(1)$;
(f) $\max_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \max[0, G_*]^2$, *where* $G_* \sim N(0, 1)$.

Note the sharp difference between the behavior of $\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n)$ here and that in Lemma 4(a). The $O_p(n)$ term in Lemma 4(a) turns out to have a zero coefficient given $\pi = 1/2$. The other aspects are the same as before. Applying theorem 3.9 of Billingsley (1999) and Lemmas 5(a to e) leads to Lemma 5(f) by the same argument as before. The sign condition applies here also, so that the square of the half-normal distribution is obtained as the limiting distribution of $\max_{\theta_2} 2(\tilde{L}_n(1/2, \theta_2) - \tilde{L}_n(1/2, \hat{\theta}_{1,n}^n))$. The asymptotic variance of the fourth-order derivative can be estimated by the eighth-order derivative, and the information matrix equality follows from this.

The asymptotic distribution of the QLR statistic under H'_{02} follows as a corollary of Lemmas 4(b) and 5(f). By the Cramér-Wold device, theorem 3.9 of Billingsley (1999), and Lemmas 3 to 5, we have

$$\max_{(\pi,\theta_2)} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \max[G_0^2, \max[0, G_*]^2],$$

where $\mathrm{cov}(G_0, G_*) = \Omega^{(3,s)}/[\Omega^{(3)}\Omega^{(s)}]^{1/2}$ and $\Omega^{(3,s)} := C_3^{(s)} - C_{12}^{(s)}[C_{22}^{(\theta)}]^{-1}C_1^{(3)}$. In applying theorem 3.9 of Billingsley (1999), we exploit the fact that $C^{(s)}$ is positive definite to show the existence of the RHS.

Theorem 5: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H'_{02}, *for each* $\epsilon > 0$, $QLR_n(\epsilon) \Rightarrow \max[G_0^2, \max[0, G_*]^2]$.

Several items are noteworthy. First, there are only two limits, G_0 and G_* under H'_{02}, implying that the tightness trivially follows for the same reason as in Theorem 4. Second, as explained below, the limiting random variable, G_0, can be inferred from $\mathcal{G}$ by moving θ_2 to θ_*, but G_* cannot. Third, the rate of convergence of the QLR statistic is $n^{1/8}$ or $n^{1/6}$ depending on whether $\pi = 1/2$ or not, so that the power of the QLR statistic under H'_{02} is much weaker than the standard case where the rate of convergence is $n^{1/2}$, and also weaker than the non-zero second-order derivative case. This property is also expected even for the standard LR statistic under the same hypothesis, because the mixture model is a special case of the HMM specification. Finally, the analysis of Neyman and Scott's (1966) C(α) statistic requires combining $\sum \hat{r}_t^{(0,3)}$ and $\sum \hat{s}_t$. For our later Monte Carlo simulations, we use the C(α) statistic asymptotically equivalent to the LR statistic under H'_{02}.

2.4. Null Distribution of the QLR Statistic under H'_0

Given the null distribution of the QLR statistic under H'_{01} or H'_{02}, we can test regime switching by restricting our attention to a particular null hypothesis. As an example, we can compute $QLR_{1,n}(\epsilon)$ or $QLR_{2,n}(\epsilon)$ for a given ϵ and apply the distribution appropriate to each test statistic. Indeed, the LR statistic of Hartigan (1985) and Ghosh and Sen (1985) focuses on H'_{01}, and the C(α) test in Neyman and Scott (1966) focuses on H'_{02}. It is, however, of general interest to obtain the asymptotic null distribution of the QLR statistic when the null parameter space is unrestricted. We derive this asymptotic null distribution, and implement Monte Carlo simulations below to compare it with other statistics.

The null asymptotic distribution can be given as the distribution of the maximum value of the random variables given under each hypothesis. We elaborate further, however, because these random variables are not independent: there exists a regular relationship between them. This requires a further regularity condition.

A6: *(iv) For each* $(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$, $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) \geq 0$ *such that (a) if* $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) > 0$, *then* $\lambda_{\max}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) < \infty$; *or (b) if* $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) = 0$, *then for each* $\theta_2 \neq \theta_*$ *and* $\theta'_2 \neq \theta_*$, $\lambda_{\min}(C^{(u)}(\theta_2, \theta'_2)) > 0$ *and* $\lambda_{\max}(C^{(u)}(\theta_2, \theta'_2)) < \infty$, *where for each* (θ_2, θ'_2), $r_t^{(u)}(\theta_2) := [1 - r_t(\theta_2), r_t^{(1)}(\theta_2)']'$, *and*

$$C^{(u)}(\theta_2, \theta'_2) := \begin{bmatrix} E[r_t^{(0,2)}(\theta_*)^2] & E[r_t^{(0,2)}(\theta_*) r_t^{(u)}(\theta'_2)'] \\ E[r_t^{(u)}(\theta_2) r_t^{(0,2)}(\theta_*)] & C^{(\theta)}(\theta_2, \theta'_2) \end{bmatrix}.$$

(v) For each $(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)$, $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) \geq 0$ *such that (a) if* $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) > 0$, *then* $\lambda_{\max}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) < \infty$; *or (b) if* $\lambda_{\min}(B(\pi^*, \theta_0^*, \theta_1^*, \theta_2^*)) = 0$, *then for each* $\theta_2 \neq \theta_*$ *and* $\theta'_2 \neq \theta_*$, $\lambda_{\min}(C^{(v)}(\theta_2, \theta'_2)) > 0$ *and* $\lambda_{\max}(C^{(v)}(\theta_2, \theta'_2)) < \infty$, *where for each* (θ_2, θ'_2),

$$C^{(v)}(\theta_2, \theta'_2) := \begin{bmatrix} E[s_t(\theta_*)^2] & E[s_t(\theta_*) r_t^{(0,3)}(\theta_*)] & E[s_t(\theta_*) r_t^{(u)}(\theta'_2)] \\ E[r_t^{(0,3)}(\theta_*) s_t(\theta_*)] & E[r_t^{(0,3)}(\theta_*)^2] & E[r_t^{(0,3)}(\theta_*) r_t^{(u)}(\theta'_2)'] \\ E[r_t^{(u)}(\theta_2) s_t(\theta_*)] & E[r_t^{(u)}(\theta_2) r_t^{(0,3)}(\theta_*)] & C^{(\theta)}(\theta_2, \theta'_2) \end{bmatrix}.$$

A6(iv and v) ensure that the asymptotic joint distribution of the random variables obtained under H'_{01} and H'_{02} is well-defined and properly behaved.

Lemma 6: *(a) Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(iv), and* H'_0, $\text{plim}_{\,\theta_2 \to \theta_*} \mathcal{G}(\theta_2) = -G_0$, *where* G_0 *is given in Theorem 4(a).*
(b) Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(v), and H'_0, $\text{plim}_{\,\theta_2 \uparrow \theta_*} \mathcal{G}(\theta_2) = G_0$ *and* $\text{plim}_{\,\theta_2 \downarrow \theta_*} \mathcal{G}(\theta_2) = -G_0$, *where* G_0 *is given in Lemma 4(b).*

The null parameter restrictions enforced through ϵ in Theorems 2, 4, and 5 are eliminated here to let θ_2 approach θ_*. We prove Lemma 6 by approximating the sample scores of $\mathcal{G}$ around θ_*. Lemma 6(a) implies that the Gaussian process, $\mathcal{G}$, is continuous at θ_* in probability (that is, with probability approaching one), when the second-order derivative isn't zero. Otherwise, $\mathcal{G}$ is discontinuous at θ_* in probability. Thus, the limiting distribution of the QLR statistic cannot be represented by a functional of a continuous Gaussian process. Even $\mathcal{G}(\theta_*)$ is not identified appropriately.

We make several remarks. First, if the second-order derivative isn't zero, then $\mathcal{G}$ contains all the information on the QLR statistic under H_0'. Second, however, if the second-order derivative is zero, we cannot neglect the event that $\max[0, G_*]^2 \geq \sup_{\Theta_*\backslash\{\theta_*\}} \min[0, \mathcal{G}(\theta_2)]^2$, but must consider this separately. Third, Lemma 6(b) implies that $\sup_{\Theta_*\backslash\{\theta_*\}} \min[0, \mathcal{G}(\theta_2)] \geq G_0^2$, so that the G_0^2 term in Theorem 5 can be ignored in considering the asymptotic distribution under H_0'. Thus, we have the following.

Theorem 6: (*a*) *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(iv), and* H_0', $QLR_n \Rightarrow \sup_{\Theta_*} \min[0, \mathcal{G}(\theta_2)]^2$. (*b*) *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(v), and* H_0', $QLR_n \Rightarrow \max[\max[0, G_*]^2, \sup_{\Theta_*\backslash\{\theta_*\}} \min[0, \mathcal{G}(\theta_2)]^2]$.

By the definition of "sup," $\sup_{\Theta_*\backslash\{\theta_*\}} \min[0, \mathcal{G}(\theta_2)]^2 = \sup_{\Theta_*} \min[0, \mathcal{G}(\theta_2)]^2$. We nevertheless maintain this notation to indicate that $\mathcal{G}$ is discontinuous at θ_*.

Theorem 6 extends the literature in several ways. Bickel and Chernoff (1993) and Lindsay (1995) corroborate the claim in Hartigan (1985) that the (Q)LR statistics are not bounded in probability unless the parameter space is bounded. Chernoff and Lander (1995) consider mixtures of binomials and derive the limiting distribution of the LR statistics. Chen and Chen (2000) generalize the analysis of Chernoff and Lander (1995) by considering general mixture models. But their generalization is restricted in the sense that their regularity conditions do not allow for the presence of other nuisance parameters. Under each regime, the conditional PDF has to have a single parameter. Dacunha-Castelle and Gassiat (1999) allow for the presence of other nuisance parameters and derive the asymptotic null distribution using the so-called "locally conic parameterizations." Nevertheless, their locally conic parametrization does not consider the case in which second-order derivatives are zero, so that their analysis cannot accommodate the models considered by Theorem 6(b). In this respect, Theorem 6(b) extends the scope of mixture models by showing that the asymptotic null distributions of (Q)LR statistics can be different from those previously obtained. Specifically, the relevant Gaussian process can be discontinuous. We revisit the popular empirical model considered by Porter (1983) below. His model requires Theorem 6(b) for its analysis.

2.5. Model Exercises

We consider two popular model specifications to explore the behaviors identified above. Suppose that $\{X_t\}$ is generated by the AR(1) process $X_t = 0.5X_{t-1} + u_t$, $u_t \sim$ *i.i.d.* $N(0, 1)$; and consider the following mixtures of normals as alternatives: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, 1) + (1-\pi) \cdot N(\theta_2 + \theta_0 X_{t-1}, 1)$, where $\pi \in [0, 1]$, $\theta_1, \theta_2 \in \Theta_*$, and Θ_* is taken in turn to be $\Theta_* := [-1.0, 1.0]$, $[-2.0, 2.0]$, $[-3.0, 3.0]$, $[-4.0, 4.0]$, $[-5.0, 5.0]$, $[0.0, 1.0]$, $[0.0, 2.0]$, $[0.0, 3.0]$, $[0.0, 4.0]$ or $[0.0, 5.0]$. The other parameters are not restricted. These choices for Θ_* are considered to illustrate how different parameter spaces or the boundary parameter can affect the QLR statistic. The last five choices for Θ_* are motivated by the consideration that many econometric models have structures similar to those considered by Porter (1983). The limiting behavior of the QLR statistics for these models is given as $\sup_{\theta_2\in\Theta_*}(\min[0, \bar{\mathcal{G}}(\theta_2)])^2$ by Theorem 6(a), where for each $\theta_2 \in \Theta_*$, $\bar{\mathcal{G}}(\theta_2) \sim N(0,1)$, and for each $\theta_2, \theta_2' \in \Theta_*$, $E[\bar{\mathcal{G}}(\theta_2)\bar{\mathcal{G}}(\theta_2')] = (\exp(\theta_2\theta_2') - 1 - \theta_2\theta_2')/[(\exp(\theta_2^2) - 1 - \theta_2^2)^{1/2}(\exp(\theta_2'^2) - 1 - \theta_2'^2)^{1/2}]$. Further, consider a Gaussian process defined as $\tilde{\mathcal{G}}(\theta_2) := \sum_{\ell=2}^{\infty} \theta_2^\ell Y_\ell/[\ell!(\exp(\theta_2^2) - 1 - \theta_2^2)]^{1/2}$ for each θ_2 and $\{Y_\ell\}_{\ell=1}^{\infty} \sim$ *i.i.d.* $N(0,1)$; then $\bar{\mathcal{G}}$ has the same covariance structure as $\tilde{\mathcal{G}}$, and this implies that their distributions are identical. Thus, simulating $\tilde{\mathcal{G}}$ yields the critical values of the QLR statistic. We approximate $\tilde{\mathcal{G}}(\theta_2)$ using $\sum_{\ell=2}^{150} \theta_2^\ell Y_\ell/[\ell!(\exp(\theta_2^2) - 1 - \theta_2^2)]^{1/2}$ and compute the maximum by grid search with a grid distance of 0.01. Table 1 contains results for the various parameter spaces at the 5% level, using 10,000 replications. As explained above, the critical value gets larger, and the point mass at zero gets smaller, as the size of Θ_* increases. Also, we compute the critical values when the boundary parameter problem is neglected. That is, we compute K' such that $P(\sup_{\theta_2\in\Theta_*} \tilde{\mathcal{G}}(\theta_2)^2 > K') = 0.05$ by simulation. This is intended to examine how serious the boundary parameter problem can be. Table 1 shows that

Table 1. Critical Values for Various Parameter Spaces (5% nominal level)

DGP: $X_t = 0.5X_{t-1} + u_t$ and $u_t \sim$ *i.i.d.* $N(0,1)$

Model: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, 1) + (1-\pi)\cdot N(\theta_2 + \theta_0 X_{t-1}, 1)$

Θ_*	$[-1.0, 1.0]$	$[-2.0, 2.0]$	$[-3.0, 3.0]$	$[-4.0, 4.0]$	$[-5.0, 5.0]$
Critical values	4.01	4.92	5.67	6.26	6.76
Point mass at zero (in percent)	31.66	16.37	7.61	3.12	1.48
Critical values w/o boundary condition	5.33	6.28	7.06	7.62	8.19
Critical values given by Corollary 1	4.05	5.40	6.06	6.58	7.00
Θ_*	$[0.0, 1.0]$	$[0.0, 2.0]$	$[0.0, 3.0]$	$[0.0, 4.0]$	$[0.0, 5.0]$
Critical values	3.49	4.09	4.76	5.17	5.65
Point mass at zero (in percent)	40.44	30.12	19.64	12.73	9.00
Critical values w/o boundary condition	4.63	5.54	6.09	6.53	6.99
Critical values given by Corollary 1	2.53	4.05	4.73	5.23	5.85

Model: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1-\pi)\cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$

Θ_*	$[-1.0, 1.0]$	$[-2.0, 2.0]$	$[-3.0, 3.0]$	$[-4.0, 4.0]$	$[-5.0, 5.0]$
Critical values	5.01	5.61	6.35	6.54	7.06
Point mass at zero (in percent)	0.00	0.00	0.00	0.00	0.00
Critical values w/o boundary condition	5.30	6.23	7.16	7.59	8.15
Critical values given by Corollary 1	5.02	5.62	6.18	6.67	7.07
Θ_*	$[0.0, 1.0]$	$[0.0, 2.0]$	$[0.0, 3.0]$	$[0.0, 4.0]$	$[0.0, 5.0]$
Critical values	4.21	4.51	5.28	5.51	5.84
Point mass at zero (in percent)	17.05	10.93	8.06	5.44	3.17
Critical values w/o boundary condition	4.98	5.42	6.28	6.75	7.04
Critical values given by Corollary 1	3.94	4.65	5.11	5.53	5.88

there is a tendency for the impact of the boundary parameter to increase as the size of the parameter space increases. That is, the distance between the critical values and K' increases as Θ_* becomes larger.

As another example, we consider the case in which the variance is unknown and must be estimated, taking the same AR(1) process and the same Θ_* as above. That is, the model for $X_t|\mathcal{F}_{t-1}$ is $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1-\pi)\cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$. Accommodating the unknown variance modifies the limiting distributions of the QLR statistics: now $QLR_n \Rightarrow \max[(\max[0, G_*])^2, \sup_{\theta_2 \in \Theta_* \setminus \{\theta_*\}} (\min[0, \dot{\mathcal{G}}(\theta_2)])^2]$ by Theorem 6(b), where for each $\theta_2 \in \Theta_*$, $\dot{\mathcal{G}}(\theta_2)$ and $G_* \sim N(0,1)$. The processes $\dot{\mathcal{G}}$ and G_* are the weak limits derived under H'_{01} and H'_{02} respectively. In particular, G_* must be derived using an eighth-order Taylor expansion. Note that G_*, plim $_{\theta_2 \downarrow 0}\dot{\mathcal{G}}(\theta_2)$, and plim $_{\theta_2 \uparrow 0}\dot{\mathcal{G}}(\theta_2)$ are different. Further, for each θ_2, $E[\dot{\mathcal{G}}(\theta_2)G_*] = \theta_2^4/(\exp(\theta_2^2) - 1 - \theta_2^2 - \theta_2^4/2)^{1/2}$ and for each θ_2, θ_2', $E[\dot{\mathcal{G}}(\theta_2)\dot{\mathcal{G}}(\theta_2')] = (\exp(\theta_2\theta_2') - 1 - \theta_2\theta_2' - \theta_2^2{\theta_2'}^2/2)/[(\exp(\theta_2^2) - 1 - \theta_2^2 - \theta_2^4/2)^{1/2}(\exp({\theta_2'}^2) - 1 - {\theta_2'}^2 - {\theta_2'}^4/2)^{1/2}]$. The given Gaussian process $\dot{\mathcal{G}}$ has the same distribution as $\ddot{\mathcal{G}}$, which is defined as $\sum_{\ell=3}^{\infty} \theta_2^{\ell} Y_\ell/[\ell!(\exp(\theta_2^2) - 1 - \theta_2^2 - \theta_2^4/2)]^{1/2}$ for each θ_2, and if $G_* = Y_4$ then $E[\dot{\mathcal{G}}(\theta_2)G_*] = E[\ddot{\mathcal{G}}(\theta_2)G_*]$ for each θ_2. Thus, we generate critical values by simulating $\max[(\max[Y_4, 0])^2, \sup_{\theta_2 \in \Theta_* \setminus \{\theta_*\}} (\min[0, \ddot{\mathcal{G}}(\theta_2)])^2]$. Table 1 also contains the critical values for these models with the same parameter spaces as before, and we observe similar behavior. An interesting feature of this case is that the point mass at zero disappears when θ_* is an interior point of Θ_*. This is mainly because $\ddot{\mathcal{G}}$ is an odd function at zero with probability one, as implied by Lemma 6(b).

2.6. Conservative Approximation of the Null Distribution

The properties of $\mathcal{G}$ are highly model-dependent and generally will not coincide with those of the above examples. The exact behavior of $\mathcal{G}$ can be difficult to pin down, as pointed out by Davies (1977, 1987), mainly because of the need to simultaneously consider a continuum of Gaussian random variables whose covariance structures are not necessarily representable by a Markovian Gaussian probability law. For the same reason, simulation methods based on a finite number of parameter elements can provide too rough a lower bound for the QLR statistic, despite extensive simulation.

Our next result overcomes this difficulty by providing a large deviation inequality that gives a relatively sharp lower bound on the tail distribution of the test statistic. These tail lower bounds are designed to yield quick reference critical values rather than to define a precise critical region. We recommend referring to other more precise critical values (when available) upon accepting the null hypothesis using our conservative critical values.

First we consider the case in which the covariance of $\mathcal{G}$ is uniformly greater than that of another stationary Gaussian process, say $\mathcal{B}^s$. Then the desired tail lower bound is obtained from $\mathcal{B}^s$.

Theorem 7: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), and A6(iv) (or A1, A2(i, iv), A3, A4, A5(ii, iv), and A6(v)), if for all* $\theta_2, \theta_2' \in \{\theta_2 \in \Theta_* : |\theta_2 - \theta_2'| \leq \delta, \exists\, \delta > 0\}$ *we have* $E[\mathcal{B}^s(\theta_2)\mathcal{B}^s(\theta_2')] \leq E[\mathcal{G}(\theta_2)\mathcal{G}(\theta_2')]$, *where* $\mathcal{B}^s$ *is a Gaussian process with mean* 0 *and* $\text{cov}(\mathcal{B}^s(\theta_2), \mathcal{B}^s(\theta_2')) = 1 - |\theta_2 - \theta_2'|^\gamma(1 + o(1))$ *for some* $\gamma \in \mathbb{R}$, *then under* H'_{01},

$$P\left(\sup_{\theta_2\in\Theta_*\backslash\{\theta_*\}}(\min[0,\mathcal{G}(\theta_2)])^2>u^2\right)\le T_1(u,\Theta_*):=H_\gamma\cdot\lambda(\Theta_*)\cdot u^{2/\gamma}\cdot(1-\Phi(u))(1+o(1)),$$

as $u\to\infty$, *where* $\Phi(\cdot)$ *is the standard normal CDF,* $H_\gamma := \lim_{\overline{\theta}_2\to\infty} H_\gamma([0,\overline{\theta}_2])/\overline{\theta}_2$, $H_\gamma([0,\overline{\theta}_2]) := E[\exp(\max_{\theta_2\in[0,\overline{\theta}_2]}\mathcal{B}^F(\theta_2))]$, λ *is the Lebesgue measure of the argument set, and* $\mathcal{B}^F$ *is a fractional Gaussian process with mean* $-|\theta_2|^\gamma$, *and* $\mathrm{cov}(\mathcal{B}^F(\theta_2),\mathcal{B}^F(\theta_2'))=|\theta_2|^\gamma+|\theta_2'|^\gamma-|\theta_2-\theta_2'|^\gamma$ *on* Θ_*.

A leading case occurs when $\gamma=2$, in which case H_2 is the Pickand constant, $1/\sqrt{\pi}$. If a stationary Gaussian process $\mathcal{B}^s$ defined on an arbitrary set Θ_i is such that $\mathrm{cov}(\mathcal{B}^s(\theta_2),\mathcal{B}^s(\theta_2'))=1-|\theta_2-\theta_2'|^\gamma(1+o(1))$, then it has a well-known tail distribution for its extremum on the same set, which is given as $T_1(u,\Theta_i)$. In other words, $|P(\sup_{\theta_2\in\Theta_i}\mathcal{B}^s(\theta_2)>u)-T_1(u,\Theta_i)|\to 0$, as $u\to\infty$ (see Piterbarg (1996)). Further, if the covariance function of $\mathcal{B}^s$ is uniformly bounded from above on Θ_i by that of $\mathcal{G}$, then the distribution of the extremum for $\mathcal{G}$ is uniformly bounded by $T_1(\cdot,\Theta_i)$ from above. That is, the Slepian inequality (cf. Dudley (1999)) can link the distributions of the extremes for $\mathcal{G}$ and $\mathcal{B}^s$ in such a way that an upper bound for the critical value of the test can be found for small values of the level of the test on Θ_i, because $P(\sup_{\theta_2\in\Theta_i}\mathcal{G}(\theta_2)\ge u)=P(\sup_{\theta_2\in\Theta_i}(\min[0,\mathcal{G}(\theta_2)])^2\ge u^2)$ for $u>0$. We partition the relevant parameter space of Θ_* into small pieces, Θ_i, and gather these upper bounds using the Bonferroni inequality to deliver the conservative tail critical values of the statistic on Θ_* in Theorem 7. Tighter bounds can be obtained using improvements to Bonferroni, but we content ourselves with this straightforward approach. It also follows from Theorem 7 that the closer the covariance function of $\mathcal{B}^s$ is to that of $\mathcal{G}$, the sharper the lower bound is. When $\mathcal{G}$ is stationary, then $\mathcal{B}^s$ directly gives the tail critical value.

Another lower bound for the test statistic is available based on the number of up-crossings for a given level of $\mathcal{G}$ (cf. Rice (1944, 1945)), which is also the source of Davies's (1977) lower bound. Heuristically, as the state level of $\mathcal{G}$ increases, the number of up-crossings decreases under suitable conditions. Thus, investigating the number of up-crossings can reveal information on the extremum of $\mathcal{G}$. Using the expected number of up-crossings, we obtain another tail critical bound for the test as follows.

THEOREM 8: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(iv) (or A1, A2(i, iv), A3, A4, A5(ii, iv), A6(v)) and* H_{01}',

$$P\left(\sup_{\theta_2\in\Theta_*\backslash\{\theta_*\}}(\min[0,\mathcal{G}(\theta_2)])^2>u^2\right)\le T_2(u,\Theta_*):=E[N_u^{\mathcal{G}}[\underline{\theta},\theta_*)]+E[N_u^{\mathcal{G}}(\theta_*,\overline{\theta}]]+2(1-\Phi(u)),$$

as $u\to\infty$, *where* $\underline{\theta}:=\min\{\Theta_*\}$, $\overline{\theta}:=\max\{\Theta_*\}$ *and* $N_u^{\mathcal{G}}[\Theta]:=\#\{\theta_2\in\Theta:\mathcal{G}(\theta_2)=u,\exists\delta>0,\mathcal{G}(\theta_2')>u,\forall\theta_2'\in[\theta_2-\delta,\theta_2]\}$.

The expected number of up-crossings in Theorem 8 is given in Cramér and Leadbetter (1967, pp. 288–289) as $E[N_u^{\mathcal{G}}[\Theta]]=\int_\Theta\int_0^\infty y p_{\theta_2}(u,y)dyd\theta_2$, where $p_{\theta_2}(x,y)$ is the probability density function of $(\mathcal{G}(\theta_2),\mathcal{G}'(\theta_2))$, and $\mathcal{G}'(\theta_2)$ is a (mean square) derivative of $\mathcal{G}(\theta_2)$ with respect to θ_2, whose existence is proved in Lemma

1. Davies (1977) elaborates this to obtain that $E[N_u^{\mathcal{G}}[\Theta]] = (2\pi)^{-1/2}\phi(u)\int_\Theta \gamma(\theta_2)d\theta_2$, where $\gamma(\theta_2) := \text{var}\{\mathcal{G}'(\theta_2)\}$ and ϕ is the PDF of a standard normal.

We can combine these lower bounds. The lower bounds provided by Theorem 7 and 8 generally differ. In such cases, using the higher lower bound delivers preferred performance. Further, Theorems 7 and 8 are derived under H'_{01}. If $\mathcal{G}$ is continuous at θ_*, then the lower bounds in Theorems 7 and 8 are the lower bound under H'_0 by Lemma 6(a). Otherwise, this needs to be modified; a modified bound can be obtained using the Bonferoni inequality.

Corollary 1: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(iv) (or A1, A2(i, iv), A3, A4, A5(ii, iv), A6(v)) and* H'_0, $\lim_{n\to\infty} P(QLR_n > u^2) \leq \min[T_1(u,\Theta_*), T_2(u,\Theta_*)] + (1-\Phi(u))\cdot \mathbf{1}_{\{C\}}$, *as* $u \to \infty$, *where* $\mathbf{1}_{\{C\}}$ *is* 0 *or* 1 *whether* $\mathcal{G}$ *is continuous or not.*

Corollary 1 follows because $\lim_{n\to\infty} P(QLR_n > u^2) \leq P(\sup_{\theta\in\Theta_*}\min[0,\mathcal{G}(\theta)]^2 > u^2) + P(\max[0,G_*]^2 > u^2)$ and $P(\max[0,G_*]^2 > u^2) = 1-\Phi(u)$ for $u > 0$.

Table 1 provides the bounds of Corollary 1 for the models in Section 2.5.[4] These bounds are very close to the critical values when the size of parameter space is moderate. If the parameter space is relatively small, then conservative bounds are ensured by choosing a small significance level.

3. Monte Carlo Simulation

We compare the size and the power of the (Q)LR statistic with other statistics in this section. For the size comparison, we use simulation environments corresponding to the examples of Section 2.5. That is, X_t is generated as an AR(1) process such that $X_t = 0.5X_{t-1} + u_t$, where $u_t \sim i.i.d.\ N(0,1)$; the model for $X_t|\mathcal{F}_{t-1}$ is a mixture of normals with unknown variance, $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1-\pi)\cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$, where $\theta_1, \theta_2 \in [-2.0, 2.0]$; and the restrictions for the other parameters are as before. For our power comparisons, we consider the DGP with $X_t = \theta_* \cdot \mathbf{1}_{\{S_t=1\}} - \theta_* \cdot \mathbf{1}_{\{S_t=2\}} + 0.5X_{t-1} + u_t$ with $P(S_t = 1|S_{t-1} = 1) = P(S_t = 2|S_{t-1} = 2) = \pi^*$, where $\pi^* \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and $\theta_* \in \{0.2, 0.4, 0.6, 0.8, 1.0\}$. Note that these DGPs are Markov regime-switching processes, so that the QLR statistic ignores the serial correlation of $\{S_t\}$. We also consider another group of DGPs, in which $\{S_t\}$ is identically and independently distributed, so that $P(S_t = 1|S_{t-1} = 1) = 1 - P(S_t = 2|S_{t-1} = 2) = \pi^*$, and the other conditions are the same as the first group of DGPs. In this latter case, the QLR statistic is the LR statistic.

There are several statistics in the literature that can be used for the same purpose as the (Q)LR statistic. The Bera and Jarque (BJ) statistic tests the normality assumption by jointly testing for skewness and kurtosis. It can thus be used to test the number of regimes when the residuals follow the normal distribution, as in our case. Because of its computational simplicity, the BJ statistic is widely used in applications. Its formula is as follows:

$$\text{BJ}_n := n\left(\frac{S_n^2}{6} + \frac{(K_n-3)^2}{24}\right),$$

[4]The program code for this bound is provided at the following URL: http://www.vuw.ac.nz/staff/js-cho/mixtures.html or ***.

Table 2. Test Statistics (levels in percent)

Number of replications: 3,000

DGP: $X_t = 0.5X_{t-1} + u_t$ and $u_t \sim$ *i.i.d.* $N(0,1)$

Model for $X_t|\mathcal{F}_{t-1}$: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1-\pi) \cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$, $\theta_1, \theta_2 \in [-2, 2]$

Sample Size	Nominal Levels (%)	10.00	7.50	5.00	2.50
	BJ_n	6.06	4.83	3.66	2.56
50	$\text{C}(\alpha)_n$	6.03	4.40	3.06	1.40
	$(Q)LR_n$	13.53	10.56	7.90	4.20
	BJ_n	7.06	5.96	4.73	3.40
100	$\text{C}(\alpha)_n$	7.60	5.46	3.86	1.90
	$(Q)LR_n$	11.60	9.16	6.73	3.20
	BJ_n	8.23	6.53	5.23	3.76
200	$\text{C}(\alpha)_n$	8.16	5.96	3.96	2.46
	$(Q)LR_n$	10.00	7.50	5.30	2.43
	BJ_n	8.73	6.66	4.50	2.63
500	$\text{C}(\alpha)_n$	8.33	5.73	3.73	1.80
	$(Q)LR_n$	10.20	7.53	5.43	2.33
Standard Errors		0.0054	0.0048	0.0039	0.0028

where S_n and K_n stand for the sample skewness and the sample kurtosis respectively of the prediction residuals of the null (restricted) model. This is known to have the χ^2_2 distribution under the null, asymptotically. Neyman and Scott's C(α) statistic can be also used to test regime switching. Essentially, it tests the number of regimes by considering the dispersion of the empirical distribution; for more details, see Lindsay (1995). The C(α) statistic hasn't been formally examined in the literature for our model specifications, so its limiting distribution must be obtained using third and fourth-order derivatives, following the analysis in Section 2.3.2. We define C(α) as

$$\text{C}(\alpha)_n := n \max\left[\frac{S_n^2}{6}, \min\left[0, \frac{K_n - 3}{\sqrt{24}}\right]^2\right].$$

The C(α) statistic weakly converges to $\max[Z_1^2, \min[0, Z_2]^2]$, where Z_1 and Z_2 are independent standard normal random variables. Interestingly, the C(α) statistic uses S_n and K_n just as BJ does.

Table 2 contains results for the size comparison computed for 3,000 replications. All of the statistics have good size. Nevertheless, the QLR statistic seems to be more stable than the others. Our power comparisons are reported in Tables 3 and 4. Table 3 contains the results for serially correlated $\{S_t\}$ with $P(S_t = 1) = 0.5$. It compares the powers of test statistics at the 5% level. Given that $P(S_t = 1) = 0.5$, the QLR statistic is most powerful for most cases. The presence of serial correlation in the Markov regime-switching process enhances the appeal of the QLR statistic. Table 4 reports the case for i.i.d $\{S_t\}$ with $P(S_t = 1) = \pi^*$ for the same level of test. Given the symmetry of the mixtures, we do not consider the

Table 3. Power of Test Statistics (in percent, 5% nominal level)

Number of observations: 500, Number of replications: 3,000

DGP: $X_t = -\theta_* \cdot \mathbf{1}_{\{S_t=1\}} + \theta_* \cdot \mathbf{1}_{\{S_t=2\}} + 0.5X_{t-1} + u_t$

$P(S_t = 1|S_{t-1} = 1) = P(S_t = 2|S_{t-1} = 2) = \pi^*$ and $u_t \sim$ $i.i.d.$ $N(0, 1)$

Model for $X_t|\mathcal{F}_{t-1}$: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1 - \pi) \cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$, $\theta_1, \theta_2 \in [-2, 2]$

	θ_*	0.20	0.40	0.60	0.80	1.00
	BJ_n	4.83	3.33	2.46	3.60	7.10
$\pi^* = 0.1$	$\mathrm{C}(\alpha)_n$	3.63	3.40	3.90	6.33†	12..80†
	QLR_n	5.76†	6.63†	7.06†	7.13	7.30
	BJ_n	4.93†	3.73	2.60	7.90	30.93
$\pi^* = 0.3$	$\mathrm{C}(\alpha)_n$	4.26	3.23	4.63	12.23	42.93
	QLR_n	5.60	6.06†	9.50†	23.93†	62.16†
	BJ_n	4.70	3.66	3.20	10.46	51.73
$\pi^* = 0.5$	$\mathrm{C}(\alpha)_n$	3.70	3.63	4.86	16.06	64.26
	QLR_n	6.03†	6.83†	11.33†	35.20†	85.10†
	BJ_n	5.00	3.33	3.26	6.20	24.73
$\pi^* = 0.7$	$\mathrm{C}(\alpha)_n$	3.66	3.53	4.70	10.93	35.36
	QLR_n	6.16†	6.33†	9.46†	26.80†	68.83†
	BJ_n	5.10†	3.93	4.20	4.40	4.83
$\pi^* = 0.9$	$\mathrm{C}(\alpha)_n$	3.93	3.73	4.70	4.20	4.76
	QLR_n	5.53	6.53†	8.13†	14.83†	29.26†

† indicates the most powerful test statistic when size distortion-adjusted critical values are applied.

cases in which π^* is greater than 0.5. Here, the C(α) and BJ statistics are more powerful than the LR statistic when θ_* close to zero and π^* is close to zero. Otherwise, the LR statistic is most powerful.

We do not report the results of other experiments due to space constraints, but our general experience is that the overall performance of the (Q)LR statistic is better than that of the other statistics. In particular, its performance is appealing when θ_* is large and $P(S_t = 1)$ is close to zero or one.

4. Empirical Application

In this section, we apply our results to the classic study by Porter (1983), who investigates cartel behavior using the prices for railroad freight shipment of grain between Chicago and the Atlantic seaboard between 1880 and 1886. During this period, there was no Sherman act (1890) to deter firms from colluding. Porter's main interest was to see if prices were consistent with the theory of cartel pricing.

Porter (1983) exploits a regime-switching model. He begins with the following demand function:

$$\log(Q_d) = c_1 + \alpha \log(P) + \beta' X_d + U_1,$$

Table 4. Power of Test Statistics (in percent, 5% nominal level)

Number of observations: 500, Number of replications: 3,000

DGP: $X_t = -\theta_* \cdot \mathbf{1}_{\{S_t=1\}} + \theta_* \cdot \mathbf{1}_{\{S_t=2\}} + 0.5X_{t-1} + u_t$

$P(S_t = 1|S_{t-1} = 1) = 1 - P(S_t = 2|S_{t-1} = 2) = \pi^*$ and $u_t \sim$ $i.i.d.$ $N(0, 1)$

Model for $X_t|\mathcal{F}_{t-1}$: $\pi \cdot N(\theta_1 + \theta_0 X_{t-1}, \sigma^2) + (1 - \pi) \cdot N(\theta_2 + \theta_0 X_{t-1}, \sigma^2)$, $\theta_1, \theta_2 \in [-2, 2]$

	θ_*	0.20	0.40	0.60	0.80	1.00
	BJ_n	4.26	6.36†	13.33	41.33	84.73
$\pi^* = 0.1$	$\text{C}(\alpha)_n$	4.00†	4.70	13.46†	43.16†	85.93†
	LR_n	5.93	5.76	13.80	43.76	86.73
	BJ_n	4.76†	4.60	8.93	26.70	73.96
$\pi^* = 0.3$	$\text{C}(\alpha)_n$	3.80	4.33	11.13	30.50	71.06
	LR_n	4.93	7.53†	15.33†	49.73†	93.56†
	BJ_n	4.83	4.00	3.36	11.40	52.20
$\pi^* = 0.5$	$\text{C}(\alpha)_n$	4.00	4.23	5.50	18.20	64.70
	LR_n	6.40†	6.86†	10.60†	35.50†	84.86†

† indicates the most powerful test statistic when size distortion-adjusted critical values are applied.

where $Q = Q_d$ is the total quantity of grain shipped by rail measured in tonnage, P is the price level of shipping measured in dollars per hundred pounds, c_1 and α are unknown parameters to be estimated, X_d contains Porter's demand shifters, β is the vector of corresponding parameters, and U_1 is a stochastic disturbance. Next, Porter specifies a supply function of the form:

$$\log(P) = \begin{cases} c_2 + \gamma \log(Q_s) + \delta' X_s + U_2, & \text{w.p. } 1 - \pi, \\ c_2 + c_3 + \gamma \log(Q_s) + \delta' X_s + U_2, & \text{w.p. } \pi, \end{cases}$$

where π is an unknown probability, c_2, c_3, and γ are unknown parameters, X_s contains Porter's supply shifters, δ contains the corresponding parameters, and U_2 is a stochastic disturbance such that

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix} \right].$$

Green and Porter (1984) motivate the use of the two-regime supply function using the fact that a cartel is not a stable organization, as each firm has an incentive to cheat on the agreement to increase its profits. Consequently, the cartel needs a mechanism to prevent deviations. In particular, a two-regime process is posited: if the market price is observed to be less than a reservation level for whatever reason, then all of the firms cut their price for a given period, leading to a competitive equilibrium. After the given period, the firms return to the cooperative equilibrium. If this strategy is adopted, then in fact no firm has any incentive to deviate from the agreement. Nevertheless, prices and output will follow a two-regime process as a result of demand shocks. See Green and Porter (1984) for further details.

Table 5. Maximum Likelihood Estimation[a]

	Porter (1983)		Ellison (1994)	
Coefficients	Demand Curve	Supply Curve	Demand Curve	Supply Curve
c_1	0.909		7.677	
	(0.149)		(1.882)	
α	–0.800		–1.802	
	(0.091)		(1.287)	
c_2		–2.416		–4.764
		(0.710)		(1.863)
c_3		0.545		0.637
		(0.032)		(0.104)
γ		0.090		0.306
		(0.068)		(0.178)

a: standard errors in parentheses.

Table 5 contains the estimation results given by Porter (1983) for the two-regime model (the alternative). The corresponding hypotheses are $H_0^\star : c_3 = 0$ versus $H_1^\star : 0 < c_3 < \log(\frac{\alpha}{1+\alpha})$. Unfortunately, the upper bound for c_3, $\log(\frac{\alpha}{1+\alpha})$, is not defined for the estimated parameters, as the model specification requires elastic demand, i.e., $|\alpha| > 1$, but in fact we estimate $\alpha = -0.8$. This difficulty is avoided by Ellison (1994), who specified the same model but with serially correlated errors:[5] $U_{1t} = \rho U_{1t-1} + V_t$, $|\rho| < 1$, and

$$\begin{pmatrix} V \\ U_2 \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11}^2 & \sigma_{12} \\ \sigma_{21} & \sigma_{22}^2 \end{pmatrix}\right].$$

Thus, the model of Ellison (1994) is as follows:

$$\log(Q_{d,t}) = c_1(1-\rho) + \rho\log(Q_{d,t-1}) + \alpha\log(P_t) - \rho\alpha\log(P_{t-1}) + \beta' X_{d,t} - \rho\beta' X_{d,t-1} + V_t,$$

$$\log(P_t) = \begin{cases} c_2 + \gamma\log(Q_{s,t}) + \delta' X_{s,t} + U_{2,t}, & \text{w.p. } 1-\pi, \\ c_2 + c_3 + \gamma\log(Q_{s,t}) + \delta' X_{s,t} + U_{2,t}, & \text{w.p. } \pi. \end{cases}$$

Ellison reports that the estimated demand elasticity is around -1.802, so that the railroad industry indeed faced an elastic demand function between 1880 and 1886. Ellison estimates various other Markov regime-switching models and obtains similar results. As shown above, however, specifying a mixture for the QLR statistic is enough for a test of regime switching. With this specification, the limiting distribution of the QLR statistic under $H_0^\star$ is obtained using Theorem 6(b) as follows:

$$QLR_n \Rightarrow \mathcal{K} := \max\left[\max[0, Y_4], \sup_{\varpi\in[0,\,\xi]}\left(\min\left[0, \frac{\sum_{m=3}^{\infty}\frac{\varpi^m}{\sqrt{m!}}Y_m}{(\exp(\varpi^2) - 1 - \varpi^2 - 0.5\varpi^4)^{1/2}}\right]\right)^2\right],$$

[5]Cosslett and Lee (1985) also specify this model with serial correlation for the same data set.

where $Y_m \sim \; i.i.d.\, N(0,1)$, $\varpi := \frac{\sigma_{11}c_3}{\sqrt{2}(\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^2)^{1/2}}$ and $\xi := \log(\frac{\alpha}{1+\alpha})\frac{\sigma_{11}}{\sqrt{2}(\sigma_{11}^2\sigma_{22}^2-\sigma_{12}^2)^{1/2}}$. Note that this is similar to one of the models considered in Section 2.5; the limiting distribution of the QLR statistic can be computed by simulating $\mathcal{K}$. For this, we set ξ to 5.00 to accommodate the estimation error for the upper bound of ϖ around 3.625 estimated using the null model. Table 1 reports the critical value at the 5% level. The corresponding critical value at the .1% level is 14.128. The computed QLR statistic value is 152.27, quite far from these critical values. We thus reject the null hypothesis as Porter did and affirm Porter's original inference. As illustrated in this example, it is essential to specify an appropriate parameter space for ϖ to ensure a valid testing procedure. Otherwise, even with a statistic value as large as 152.27, the null hypothesis could be accepted for a moderately large parameter space.

5. Conclusion

We consider Markov regime-switching processes, and investigate the limiting distributions of statistics for testing the null hypothesis of one regime versus the alternative of two regimes. We obtain a useful test by specifying the multiple regime model as a mixture of the two PDFs of a Markov regime-switching process. This specification ignores the serial correlation of the unobserved data switching process. Despite this neglect, the quasi-maximum likelihood estimator is consistent and yields reliable testing procedures. Under the null, one has a non-standard situation, in which one has an identification problem and a boundary parameter problem or a log-likelihood function that cannot be approximated using second-order derivatives. These conditions imply that the quasi-LR statistic based on the mixture model follows a non-standard distribution.

We derive the asymptotic null distributions of the quasi-LR statistics for various mixture models not previously examined in the mixture literature, and provide critical values of the QLR statistics for various popular mixture models. Also, we suggest formulae to compute the tail upper critical values when the exact critical values are difficult to compute. Although these methods do not produce precise critical values, they may prove useful in practice. Further, we consider the size and power of several test statistics used to test the number of regimes. Overall, the QLR statistic performs well as to size and power, and outperforms the other standard statistics for a number of cases considered in our Monte Carlo simulations. Nevertheless, the QLR statistic is not plausibly optimal; better performing statistics for the specific hidden Markov type serial correlation may be found following the approach in Andrews and Ploberger (1994).

As an empirical application, we re-examine the cartel stability problem investigated by Porter (1983). By comparing the model-specific critical value with the empirical QLR statistic value, we corroborate Porter's findings, as we reject the null hypothesis that there was a single equilibrium in the market for railroad shipments between Chicago and the Atlantic seaboard between 1880 and 1886.

6. Appendix

First, we provide supplementary results used to prove the main claims.

Lemma A1: *(a) Given A1, A2(i, iii), A3, and A5(i),* $\sup_{(\pi,\theta)} |n^{-1}\sum \ell_t(\pi,\theta) - E[\ell_t(\pi,\theta)]| \to 0$ *a.s.*

(b) *Given A1, A2(i, ii), A3, and A5(i),* $\sup_{(\pi,\theta)} \|n^{-1}\sum \nabla_{(\pi,\theta)}\ell_t(\pi,\theta) - E[\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)]\|_\infty \to 0$ *a.s.*

(c) *Given A1, A2(i, ii), A3, and A5(ii),* $\sup_{(\pi,\theta)} \|n^{-1}\sum \nabla^2_{(\pi,\theta)}\ell_t(\pi,\theta) - E[\nabla^2_{(\pi,\theta)}\ell_t(\pi,\theta)]\|_\infty \to 0$ *a.s.*

Proof of Lemma A1: (a) First, note that ℓ_t is differentiable a.s. by Lemma A2(b) below, and for some positive, stationary, and ergodic random variable, M_t, $\|\nabla_{(\pi,\theta)}\ell_t(\pi,\theta)\|_\infty < M_t$ by A5(i), so that for each (π,θ) and $(\tilde{\pi},\tilde{\theta})$, $|\ell_t(\pi,\theta) - \ell_t(\tilde{\pi},\tilde{\theta})| \leq M_t\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\|$. This also implies that $|n^{-1}\sum \ell_t(\pi,\theta) - n^{-1}\sum \ell_t(\tilde{\pi},\tilde{\theta})| \leq n^{-1}\sum M_t\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\|$. Further, we can apply the ergodic theorem to $\{n^{-1}\sum M_t\}$, so that for any $\omega \in F$, $P(F) = 1$, and $\varepsilon > 0$, there is an $n^*(\omega,\varepsilon)$ such that if $n \geq n^*(\omega,\varepsilon)$, then $|n^{-1}\sum M_t - E[M_t]| \leq \varepsilon$, and this implies that $n^{-1}\sum M_t \leq E[M_t] + \varepsilon$. For the same ε, we may let $\delta := \varepsilon/(E[M_t] + \varepsilon)$; then $n^{-1}\sum M_t\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\| \leq \varepsilon$, whenever $\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\| \leq \delta$, because $n^{-1}\sum M_t\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\| \leq n^{-1}\sum M_t\delta = n^{-1}\sum M_t\varepsilon/(\varepsilon + E[M_t]) \leq \varepsilon$. That is, for any $\omega \in F$, $P(F) = 1$ and $\varepsilon > 0$, there is $n^*(\omega,\varepsilon)$ and δ such that if $n \geq n^*(\omega,\varepsilon)$ and $\|(\pi,\theta')' - (\tilde{\pi},\tilde{\theta}')'\| \leq \delta$, then $|n^{-1}\sum \ell_t(\pi,\theta) - n^{-1}\sum \ell_t(\tilde{\pi},\tilde{\theta})| < \varepsilon$, which means that $\{n^{-1}\sum \ell_t\}^\infty_{n^*(\omega,\varepsilon)}$ is equicontinuous. Thus, with probability one, $n^{-1}\sum \ell_t$ converges to $E[\ell_t]$ uniformly on $[0,1]\times\Theta$ by Rudin (1976, p.168).

(b *and* c) We can apply Ranga Rao (1962) to each case. To save space, for each $(\pi,\theta) \in [0,1]\times\Theta$, let $q_t := \nabla_{(\pi,\theta)}\ell_t(\pi,\theta)$ (of (b)) or $\nabla^2_{(\pi,\theta)}\ell_t(\pi,\theta)$ (of (c)) respectively. Then, $\sup_{\pi,\theta}|q_t(\pi,\theta)| \leq M_t$ by A5(i or ii), and as M_t has a finite first moment, $E[q_t]$ is continuous on $[0,1]\times\Theta$, and $\sup_{(\pi,\theta)}\|n^{-1}\sum q_t(\pi,\theta) - E[q_t(\pi,\theta)]\|_\infty \to 0$ a.s., as $n\to\infty$ by Ranga Rao (1962). □

Lemma A2: (a) *Suppose that* $(\tilde{\pi},\tilde{\theta}_0,\tilde{\theta}_1,\tilde{\theta}_2)$ *is a boundary element with* $\tilde{\pi} = 0$ *and* $(\tilde{\theta}_0,\tilde{\theta}_1,\tilde{\theta}_2)$ *an interior element of* Θ *with* $\tilde{\theta}_1 \neq \tilde{\theta}_2$. *Then, under A2(i) and A3, for each* $\theta_1 \in \Theta_*$, *the domain of* $\ell_t(\,\cdot\,,\cdot\,,\theta_1,\cdot\,)$ *includes a set* $P^0_1 \times \Theta^0_0 \times \Theta^0_{2+}$, *where for some* $\varepsilon > 0$, $P^0_1 := [0,\varepsilon]$, *and* Θ^0_0 *and* Θ^0_{2+} *are open cubes centered at* $\tilde{\theta}_0$ *and* $\tilde{\theta}_2$ *respectively.*

(b) *Under A2(ii),* $\ell_t \in \mathcal{C}^{(2)}([0,1]\times\Theta)$ *a.s.*

(c) *Under A1, A2(i, ii), A3, A4, A5(ii), and A6(i),* $\sqrt{n}[(\hat{\theta}^{n\prime}_{n,0},\hat{\theta}^n_{n,1}) - (\theta^{*\prime}_0,\theta_*)]' \overset{\text{A}}{\sim} N[0,[C^{(\theta)}_{22}]^{-1}]$.

Proof of Lemma A2: (a) Since $(\tilde{\theta}_0,\tilde{\theta}_2)$ is an interior element of $\tilde{\Theta}$, there are open balls, $B(\tilde{\theta}_0,\varepsilon_0) \subset \Theta_0$ and $B(\tilde{\theta}_2,\varepsilon_1) \subset \Theta_*$. Now, let $\varepsilon := \min[\varepsilon_0/\sqrt{2},\varepsilon_1/\sqrt{2}]$. Then the desired result follows by letting $\Theta^0_0 := C(\tilde{\theta}_0,\varepsilon)$ and $\Theta^0_{2+} := C(\tilde{\theta}_2,\varepsilon)$, where $C^{(\theta)}(\theta,\varepsilon)$ is an open cube centered at θ with length 2ε.

(b) By the definition of f_t, its composition with the log-function must be in $\mathcal{C}^{(2)}([0,1]\times\Theta)$ a.s.

(c) By the mean value theorem, the interiority of (θ^*_0,θ_*), and the FOC for $(\hat{\theta}^n_{0,n},\hat{\theta}^n_{1,n})$, $[(\hat{\theta}^{n\prime}_{n,0},\hat{\theta}^n_{n,1}) - (\theta^{*\prime}_0,\theta_*)]' = [\nabla^2_{\theta^1}L^*_n(1,\theta^*_0,\theta_*\theta_2)]^{-1}[\nabla_{\theta^1}L^*_n(1,\theta^*_0,\theta_*,\theta_2)] + o_p(1)$. The desired result follows by the LLN, the CLT, the information matrix equality, and the definition of $C^{(\theta)}_{22}$. □

Remarks: 1. Assumption 2^{2*} of Andrews (2001) for the mixture is satisfied by Lemma A2.

2. In considering the open cubes of Lemma A2(a), we consider the locations of $(\pi^*,\theta^*_0,\theta^*_1,\theta^*_2)$ generated by the null hypotheses, so that Lemma A2(a) treats the null parameter space generated by H'_{01}.

Proof of Theorem 1: (a) The given claim follows directly by the SULLN of Lemma A1(a).

(b) To show the result, we prove that the unconditional population means of the FOCs are zeros at (π^*, θ^*). Then this implies that $(\hat{\pi}_n^q, \hat{\theta}_n^q) \to (\pi^*, \theta^*)$ a.s. by Lemma A1(a) (cf. White (1994, theorem 4.6)).

For this, we first prove that $E[1/(\pi^* f_t(\theta_0^*, \theta_1^*) + (1-\pi^*) f_t(\theta_0^*, \theta_2^*))] = 1$, which is the common element in the FOCs. For each (π, θ), $E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))] = \sum_{k=1}^{2} E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))|S_t = k]P(S_t = k)$, $P(S_t = 1) = \pi^*$ and $f(X_1, \cdots, X_t|S_t = 1) = \pi^{*-1}\boldsymbol{\pi}^{*\prime}\mathbf{F}_1(\theta^*)[\prod_{\tau=1}^{t-1} \mathbf{P}^*\mathbf{F}_\tau(\theta^*)]\mathbf{P}^*\mathbf{F}_t(\theta^*)[1\ 0]'$, where $\mathbf{P}^*$ is $\mathbf{P}$ evaluated at p_{11}^* and p_{22}^*. Note that $E[[\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2)]^{-1}|S_t = 1] = \int f(x_1, \cdots, x_t|S_t = 1)/[\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2)]dx^t$, and also $\int f(x_1, \cdots, x_t|S_t = 1)dx^{t-1} = \pi^{*-1}\boldsymbol{\pi}^{*\prime} \int \mathbf{F}_1(\theta^*)dx_1[\prod_{\tau=2}^{t-1} \mathbf{P}^* \int \mathbf{F}_\tau(\theta^*)dx_\tau]\mathbf{P}^*\mathbf{F}_t(\theta^*)[1, 0]' = \pi^{*-1}\boldsymbol{\pi}^{*\prime}I[\prod_{\tau=2}^{t-1} \mathbf{P}^* I]\mathbf{P}^*\mathbf{F}_t(\theta^*)[1, 0]' = f_t(\theta_0^*, \theta_1^*)$, implying that $E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))|S_t = 1] = \int f_t(\theta_0^*, \theta_1^*)/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))dx_t$. In the same way, it follows that $E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))|S_t = 2] = \int f_t(\theta_0^*, \theta_2^*)/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))dx_t$, so that $E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))] = \int (\pi^* f_t(\theta_0^*, \theta_1^*) + (1-\pi^*) f_t(\theta_0^*, \theta_2^*))/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))dx_t$. This shows that $E[1/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))] = 1$ when $(\pi, \theta) = (\pi^*, \theta^*)$.

Second, we consider the expected value of each FOC. For θ_0, $E[\nabla_{\theta_0} \ell_t(\pi, \theta)] = E[(\pi \nabla_{\theta_0} f_t(\theta^1) + (1-\pi)\nabla_{\theta_0} f_t(\theta^2))/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))] = \int (\pi \nabla_{\theta_0} f_t(\theta^1) + (1-\pi)\nabla_{\theta_0} f_t(\theta^2))(\pi^* f_t(\theta_0^*, \theta_1^*) + (1-\pi^*) f_t(\theta_0^*, \theta_2^*))/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))dx_t$, where the last equality follows by the same reasoning as above. Thus, if $(\pi, \theta) = (\pi^*, \theta^*)$, then $E[\nabla_{\theta_0} \ell_t(\pi, \theta)] = \pi^* \int \nabla_{\theta_0} f_t(\theta_0^*, \theta_1^*)dx_t + (1-\pi^*) \int \nabla_{\theta_0} f_t(\theta_0^*, \theta_2^*)dx_t = \pi^* \nabla_{\theta_0} \int f_t(\theta_0^*, \theta_1^*)dx_t + (1-\pi^*)\nabla_{\theta_0} \int f_t(\theta_0^*, \theta_2^*)dx_t = 0$, where the second last equality holds by the Lebesgue dominated convergence theorem (LDCT) and A5(ii); and the last equality follows by that $\int f_t(\theta^1)dx_t = \int f_t(\theta^2)dx_t = 1$ for each θ^1, θ^2. Next, for θ_1, if $(\pi, \theta) = (\pi^*, \theta^*)$, $E[\nabla_{\theta_1} \ell_t(\pi, \theta)] = \pi^* \int \nabla_{\theta_1} f_t(\theta_0^*, \theta_1^*)dx_t = \pi^* \nabla_{\theta_1} \int f_t(\theta_0^*, \theta_1^*)dx_t = 0$ by the same argument as above. Similarly, $E[\nabla_{\theta_2} \ell_t(\pi, \theta)] = 0$, if $(\pi, \theta) = (\pi^*, \theta^*)$. Finally, $E[\nabla_\pi \ell_t(\pi, \theta)] = \int (f_t(\theta^1) - f_t(\theta^2))(\pi^* f_t(\theta_0^*, \theta_1^*) + (1-\pi^*) f_t(\theta_0^*, \theta_2^*))/(\pi f_t(\theta^1) + (1-\pi) f_t(\theta^2))dx_t$. Thus, if $(\pi, \theta) = (\pi^*, \theta^*)$, then $E[\nabla_\pi \ell_t(\pi, \theta)] = \int (f_t(\theta^1) - f_t(\theta^2))dx_t = 1 - 1 = 0$, implying the desired result by Lemma A1. □

Proof of Lemma 1: (a) We show Lemma 1(a) in two steps. First, we show that $\{n^{-1} \sum \nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \cdot)\nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \cdot)'\}$ satisfies the SULLN. Second, we prove that $\{n^{-1/2} \sum \nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \cdot)\}$ weakly converges to $\mathcal{S}_1$. Then we can apply theorem 7.1 of Billingsley (1999) to obtain the desired result.

First, from the assumption that $\{X_t\}$ is a geometric β-mixing process, for each $\theta_2 \in \Theta_*$, $\{\nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \theta_2)\nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \theta_2)'\}$ is a strictly stationary β-mixing process with a mixing coefficient less than $c\rho^\tau$ and a finite $1+\delta$ moment by A1 and A5(ii). Thus, for each $\theta_2 \in \Theta_*$, $n^{-1} \sum \nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \theta_2)\nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \theta_2)' \to C^{(\theta)}(\theta_2, \theta_2)$ a.s. by the ergodic theorem. Further, this holds uniformly on $\Theta_*(\epsilon)$ by Lemma A1. Second, for each $\theta_2 \in \Theta_*(\epsilon)$, $\{n^{-1/2}\nabla_{(\pi,\theta^1)} L_n^*(1, \theta_0^*, \theta_*, \theta_2)\}$ obeys the CLT given that $\{\nabla_{(\pi,\theta^1)} \ell_t(1, \theta_0^*, \theta_*, \theta_2)\}$ has a β-mixing coefficient less than $c\rho^\tau$ and a finite $2+\delta$ moment by A1 and A5(ii) respectively (Doukhan, Massart, and Rio (DMR) (1995, theorem 1)). Thus, for each $\theta_2 \in \Theta_*(\epsilon)$, $n^{-1/2}\nabla_{(\pi,\theta^1)} L_n^*(1, \theta_0^*, \theta_*, \theta_2) \overset{\text{A}}{\sim} N(0, C^{(\theta)}(\theta_2, \theta_2))$, where $\text{var}(n^{-1/2}\nabla_{(\pi,\theta^1)} L_n^*(1, \theta_0^*, \theta_*, \theta_2)) = C^{(\theta)}(\theta_2, \theta_2)$. Further, applying the Cramér-Wold device establishes the finite dimensional distributional convergence. Next, we apply DMR (1995, theorem 1) to prove the tightness of $\{n^{-1/2} \sum \nabla_{(\pi,\theta^1)} L_n^*(1, \theta_0^*, \theta_*, \cdot) : \Theta_*(\epsilon) \mapsto \mathbb{R}^{2+r_0}\}$. First of all, from the definition of $\nabla_{(\pi,\theta^1)} L_n^*(1, \theta_0^*, \theta_*, \cdot)$, $\nabla_{\theta^1} L_n^*(1, \theta_0^*, \theta_*, \cdot) = \sum r_t(\theta_*)$, which is not a function of θ_2; thus, we can ignore it in proving tightness, and pay attention to only $\{n^{-1/2}\nabla_\pi L_n^*(1, \theta_0^*, \theta_*, \cdot) = n^{-1/2} \sum (1 - r_t(\cdot))\}$.

We proceed in three steps. (i) Given A5(ii), let $\phi(x) := x^{1+\delta/2}$ $(x > 0)$; then for any $s \in (0,\infty)$, $E[\sup_{\theta_2\in\Theta_*(\epsilon)} \phi(s(1-r_t(\theta_2))^2] = s^{1+\delta/2}E[\sup_{\theta_2\in\Theta_*(\epsilon)}(1-r_t(\theta_2))^{2+\delta}] < s^{1+\delta/2}\Delta < \infty$. Further, if we let $\phi^*(y) := \sup_{x\geq 0}(xy - \phi(x))$, then $\phi^*(y) = \delta^* y^{1+2/\delta}$, where $\delta^* := \frac{2}{2+\delta} - (\frac{2}{2+\delta})^{(2+\delta)/\delta}$. From this and the geometric β-mixing condition, it follows that $\delta^* \int_0^1 (\beta^{-1}(u))^{1+2/\delta} du < \delta^*(|\log(\rho)|)^{-(1+2/\delta)}\{|\log(c)| \times (|\log(c)|)^{m_*} + \Gamma(m_*)\} < \infty$ by some algebra, where $\beta^{-1}(u) := \inf\{t : \beta(t) \leq u\}$, Γ is the gamma function and $m_* := \min\{2i : i \in \mathbb{N}, i \geq 1 + 2/\delta\}$. This implies that $\int_0^1 \beta^{-1}(u)Q^*(u)^2 du < \infty$ by lemma 2 of DMR (1995), where Q^* is the quantile function of $\sup_{\theta_2\in\Theta_*(\epsilon)} |1 - r_t(\theta_2)|$. (ii) It is trivial to show that $\sum_{k=1}^{\infty} \beta_k \leq \sum_{k=1}^{\infty} c\rho^k = c\rho/(1-\rho) < \infty$. (iii) Finally, from A5(ii) and given that $f_t \in \mathcal{C}^{(2)}(\Theta_*)$ a.s. in A2(ii), Lipschitz continuity holds a.s., thus $E[|r_t(\theta_2) - r_t(\theta_2')|] \leq E[\sup_{\theta_2\in\Theta_*(\epsilon)} |\nabla_{\theta_2} r_t(\theta_2)|^{2+\delta}]|\theta_t - \theta_2'|^{2+\delta}$, implying that Ossiander's $L^{2+\delta}$ entropy is finite by Andrews (1994, theorem 5). From these facts and theorem 1 of DMR (1995), $\{n^{-1/2}\sum(1 - r_t(\,\cdot\,)) : \Theta_*(\epsilon) \mapsto \mathbb{R}\}$ must have a tight distribution, and so the convergence limit of $S_{1,n}$ must be a Gaussian process with covariance structure the same as that of $\mathcal{S}_1$.

(b) To show this, we prove that the covariance function of $\mathcal{S}_1$ has a generalized second-order derivative. That is, for each $\theta_2 \in \Theta_*(\epsilon)$, $\nabla_{\theta_2}\nabla_{\theta_2'} K(\theta_2, \theta_2') := \lim_{h,h'\to 0}(hh')^{-1}(K(\theta_2 + h\boldsymbol{\iota}_r, \theta_2' + h'\boldsymbol{\iota}_r) - K(\theta_2 + h\boldsymbol{\iota}_r, \theta_2') - K(\theta_2, \theta_2' + h'\boldsymbol{\iota}_r) + K(\theta_2, \theta_2'))$ exists at $\theta_2 = \theta_2'$, where $\boldsymbol{\iota}_r$ is an $r \times 1$ vector of ones, and $K(\theta_2, \theta_2') = C^{(\theta)}(\theta_2, \theta_2)^{-1}C^{(\theta)}(\theta_2, \theta_2')C^{(\theta)}(\theta_2', \theta_2')^{-1}$ for each $\theta_2, \theta_2' \in \Theta_*(\epsilon)$. The claim then follows as a corollary of Grenander (1981, theorem 1 of Chapter 2-2). Note that for each $\theta_2 \in \Theta_*(\epsilon)$, $C^{(\theta)}(\theta_2, \theta_2)$ is positive definite by A6(i); thus, if $C^{(\theta)}(\theta_2, \theta_2')$ has a generalized second-order derivative, then $K(\theta_2, \theta_2')$ must have a generalized second-order derivative. Further, $C_{22}^{(2)}$ is not a function of θ_2. Thus, we can restrict our attention only to the generalized second-order derivative of $C_{11}^{(\theta)}(\theta_2, \theta_2')$, and prove its existence. $\nabla_{\theta_2}\nabla_{\theta_2'} C_{11}^{(\theta)}(\theta_2, \theta_2') = \lim_{h,h'\to 0} E[(hh')^{-1}(r_t(\theta_2 + h\boldsymbol{\iota}_r, \theta_2' + h'\boldsymbol{\iota}_r) - r_t(\theta_2 + h\boldsymbol{\iota}_r, \theta_2') - r_t(\theta_2, \theta_2' + h'\boldsymbol{\iota}_r) + r_t(\theta_2, \theta_2'))] = E[\lim_{h,h'\to 0}(hh')^{-1}(r_t(\theta_2 + h\boldsymbol{\iota}_r, \theta_2' + h'\boldsymbol{\iota}_r) - r_t(\theta_2 + h\boldsymbol{\iota}_r, \theta_2') - r_t(\theta_2, \theta_2' + h'\boldsymbol{\iota}_r) + r_t(\theta_2, \theta_2'))] = E[\lim_{h\to 0} h^{-1}(\nabla_{\theta_2'} r_t(\theta_2 + h\iota_r') - \nabla_{\theta_2'} r_t(\theta_2, \theta_2'))] = E[\nabla_{\theta_2}\nabla_{\theta_2'} r_t(\theta_2, \theta_2')]$, where the second equality follows by the LDCT, A5(ii) and the second-order differentiability of A2(ii). The final term is well defined by the moment condition in A5(ii). This completes the proof. □

Remarks: 1. Hansen (1996, 2004) considers various other econometric models applying DMR (1995) in a time series context.

2. In Lemma 1(a), we estimate $E[\nabla^2_{(\pi,\theta^1)}\ell_t(1, \theta_0^*, \theta_*, \cdot)]$ using $n^{-1}\sum \nabla_{(\pi,\theta^1)}\ell_t(1, \theta_0^*, \theta_*, \cdot) \cdot \nabla_{(\pi,\theta^1)}\ell_t(1, \theta_0^*, \theta_*, \cdot)'$ to ensure a positive definite estimator.

Proof of Theorem 2: Note that $QLR_n(\epsilon) = \sup_{\theta_2\in\Theta_*(\epsilon)} QLR_{1,n}(\theta_2) - QLR_{2,n}$. We examine the limiting behavior of each element of the RHS.

First, note that $QLR_{2,n} := 2(L_n^*(1, \hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n, \theta_2)) - L_n^*(1, \theta_0^*, \theta_*, \theta_2))$ and $L_n^*(1, \theta_0, \theta_1, \theta_2) = \sum \tilde{\ell}_t(\theta^1)$. Thus, by the mean value theorem, the interiority of (θ_0^*, θ_*), and the FOC for $(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)$, $QLR_{2,n} = [\nabla_{\theta^1} L_n^*(1, \theta_0^*, \theta_*, \theta_2)]'[\nabla^2_{\theta^1} L_n^*(1, \theta_0^*, \theta_*\theta_2)]^{-1}[\nabla_{\theta^1} L_n^*(1, \theta_0^*, \theta_*, \theta_2)] + o_p(1)$.

Second, to prove the remaining claim, we verify the conditions (assumptions 2-5, 7 and 8) given in theorem 2(b) of Andrews (2001). Now Lemma A2(a–b) verifies assumption 2; Lemma 1 and its proof verify assumption 3; Theorem 1, Lemma A2(a–b) and Lemma 1 verify assumption 4; assumption 5 is satisfied given that 1 is a boundary element of $[0,1]$, thus $\{\pi - 1 : \pi \in [0,1]\}$ is locally equal to $\mathbb{R}^-$; and assumption 7 and 8 trivially hold by A2(i). In particular, assumption 7(a) does not apply to our problem, and assumptions 7(b) and 8 hold by the assumption on $\tilde{\Theta}$ given in A2(i). From these, under H_{01}', $\sup_{\theta_2 \in \Theta_*(\epsilon)} QLR_{1,n}(\theta_2) = \sup_{\theta_2 \in \Theta_*(\epsilon)} \min[0, \Omega^{(\theta)}(\theta_2,\theta_2)^{1/2} S_{1,n}^{[1:1]}(\theta_2)]^2 + [n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]'[n^{-1}\sum \nabla_{\theta^1}\ell_t(\pi, \theta)\nabla_{\theta^1}\ell_t(\pi,\theta)']^{-1}[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]_{(\pi,\theta^1)=(1,\theta_0^*,\,\theta_*)} + o_p(1)$ using theorem 2(b) of Andrews (2001). Next, note that for any $\theta_2 \in \Theta_*(\epsilon)$, $[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]'[n^{-1}\sum \nabla_{\theta^1}\ell_t(\pi,\theta)\nabla_{\theta^1}\ell_t(\pi,\theta)']^{-1}[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]$ evaluated at $(1,\theta_0^*,\theta_*,\theta_2)$ is identical to $[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]'[n^{-1}\sum \nabla_{\theta^1}\ell_t(\pi,\theta)\nabla_{\theta^1}\ell_t(\pi,\theta)']^{-1}[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\ \theta)]$ at $(\pi,\theta_0^*,\theta_*,\theta_*)$ for any $\pi \in [0,1]$. Therefore, it follows that $[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]'[n^{-1}\sum \nabla_{\theta^1}\ell_t(\pi,\theta)\nabla_{\theta^1}\ell_t(\pi,\theta)']^{-1}[n^{-1/2}\nabla_{\theta^1} L_n^*(\pi,\theta)]_{(\pi,\theta^1)=(1,\theta_0^*,\,\theta_*)} = QLR_{2,n} + o_p(1)$ by the information matrix equality.

Thus, by these two facts we have that $QLR_n(\epsilon) = \sup_{\theta_2 \in \Theta_*(\epsilon)} \min[0, \Omega^{(\theta)}(\theta_2,\theta_2)^{1/2} S_{1,n}^{[1:1]}(\theta_2)]^2 + o_p(1) \Rightarrow \sup_{\theta_2 \in \Theta_*(\epsilon)} \min[0, \mathcal{G}(\theta_2)]^2$ under H_{01}', where the weak convergence follows by Lemma 1(a) and the continuous mapping theorem. □

Remark: It's straightforward that $QLR_{2,n} \Rightarrow Z'Z$ by the standard argument, where $Z \sim N[0, I_{r_0+1}]$.

Proof of Theorem 3: To show the result, we use the following definition and fact.

Definition: A semi-metric space, $(\Theta_*, d(\theta_2,\theta_2'))$, is *precompact* if for all $\varepsilon > 0$, $N(\Theta_*, d;\varepsilon) < \infty$, where $N(\Theta_*, d;\varepsilon)$ is the smallest number of open balls with radius ε measured by d.

Fact (*Lifshits, 1995, p. 65*): *Let* $\{\mathcal{G} : \Theta_* \mapsto \mathbb{R}\}$ *be a set of separable Gaussian random functions and suppose that* $(\Theta_*, d(\theta_2,\theta_2'))$ *is precompact for some* d. *Then* $P(\sup_{\theta_2\in\Theta_*} |\mathcal{G}(\theta_2)| < \infty) = 1$ *if and only if* $\alpha(\mathcal{G}(\theta_2),\Theta_*) < \infty$ *a.s., where* $\alpha(\mathcal{G}(\theta_2),\Theta_*) := \lim_{\delta\to 0}\sup_{\theta_2,\theta_2'\in\Theta_*, d(\theta_2,\theta_2')<\delta} |\mathcal{G}(\theta_2) - \mathcal{G}(\theta_2')|$.

By Lemma 1(b), $\mathcal{G}$ is continuous in probability on $\Theta_*^1 := [\underline{\theta}, \theta_*)$ and $\Theta_*^2 := (\theta_*, \overline{\theta}]$ so that $\alpha(\mathcal{G},\Theta_*)$ equals 0 in probability. Further, Ossiander's $L^{2+\delta}$ entropy is finite as shown in the proof of Lemma 1(a). That is, $\int_0^1 [\log(N_{[\,]}(\mathcal{F}, \|\cdot\|_{2+\delta};\varepsilon))]^{1/2} d\varepsilon < \infty$, where $N_{[\,]}(\mathcal{F}, \|\cdot\|_{2+\delta};\varepsilon)$ is the bracketing number of functions on Θ_*^i with respect to $\|\cdot\|_{2+\delta}$ $(i = 1,2)$. From the relationship that for any $\varepsilon > 0$, $N(\Theta_*^i, d_{2+\delta};\varepsilon) \le N_{[\,]}(\mathcal{F}, \|\cdot\|_{2+\delta};\varepsilon)$, we have $\int_0^1 [\log(N(\Theta_*^i, d_{2+\delta};\varepsilon))]^{1/2} d\varepsilon < \infty$, where $d_{2+\delta}$ is the metric corresponding to $\|\cdot\|_{2+\delta}$. Thus, if for some $\varepsilon_1 > 0$, $N(\Theta_*^i, d_{2+\delta};\varepsilon_1) = \infty$, then for any $\varepsilon < \varepsilon_1$, $N(\Theta_*^i, d_{2+\delta};\varepsilon) = \infty$, implying that $\int_0^1 [\log(N(\Theta_*^i, d_{2+\delta};\varepsilon))]^{1/2} d\varepsilon = \infty$. This is a contradiction to the fact that $\int_0^1 [\log(N(\Theta_*^i, d_{2+\delta};\varepsilon))]^{1/2} d\varepsilon < \infty$. Therefore, for any $\varepsilon > 0$, $N(\Theta_*^i, d_{2+\delta};\varepsilon) < \infty$ $(i = 1,2)$. That is, $(\Theta_*, d_{2+\delta})$ is precompact. Hence, the desired result follows by the above fact. □

We use the following supplementary lemmas to show the given main claims.

Lemma B1: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(ii), and* H_{02}',
(a) $\sum \nabla_{i_1}\nabla_{i_2} f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n)/f_t(\hat{\theta}_{0,n}^n, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$;

(*b*) $\sum \nabla_{i_1}\nabla_{i_2}\nabla_{i_3} f_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n})/f_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}) = O_p(n^{1/2})$, *where* $i_1, i_2, i_3 \in \{\theta_{01}, \cdots, \theta_{0r}, \theta_1\}$.

Proof of Lemma B1: (*a and b*) For notational simplicity, let $q_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n})$ be the functions in (a) and (b). By the mean value theorem, $\sum q_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}) = \sum q_t(\theta^*_0, \theta_*) + \sum \nabla_{\theta^1} q_t(\bar{\theta}_0, \bar{\theta}_1)[(\hat{\theta}^{n\prime}_{n,0}, \hat{\theta}^n_{n,1}) - (\theta^{*\prime}_0, \theta_*)]'$ for some $(\bar{\theta}_0, \bar{\theta}_1)$. Given A5(iii), we can apply the CLT in DMR (1995) to $n^{-1/2}\sum q_t(\theta^*_0, \theta_*)$, because $E[q_t(\theta^*_0, \theta_*)] = 0$. Also, $\sum \nabla_{\theta^1} q_t(\bar{\theta}_0, \bar{\theta}_1) = O_p(n)$ by the ergodic theorem and A5(iii). Lemma A2(c) completes the proof. □

Lemma B2: *Given A1, A2(i, iii), A3, A4, A5(ii, iii), A6(ii), and* H'_{02}, *for each* π,
(*a*) $(\hat{\tilde{\theta}}^{(1)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(1)}_{1,n} + (1-\pi))' = O_p(n^{-1/2})$;
(*b*) $(\hat{\tilde{\theta}}^{(2)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(2)}_{1,n})' = -(\frac{1-\pi}{\pi})[C^{(\theta)}_{22}]^{-1}C^{(2)}_{21} + o_p(1)$;
(*c*) $(\hat{\tilde{\theta}}^{(3)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(3)}_{1,n})' = O_p(1)$;
(*d*) *further, for a sequence of random variables,* $\{\hat{q}_t\}$ *say, if* $\sum \hat{q}_t \hat{r}^{(1)}_t = O_p(n)$, *then* $\sum \hat{q}_t \hat{f}_t \hat{g}^{(1)}_t = O_p(n^{1/2})$;
(*e*) $n^{-1/2}\sum \hat{r}^{(0,2)}_t \overset{\text{A}}{\sim} N(0, \Omega^{(2)})$.

Proof of Lemma B2: (*a to c*) First, for notational simplicity, for each π let

$$\bar{R}^c_n(\pi) := \begin{bmatrix} -\bar{R}^{(2,0)}_n + \bar{R}^{(1,0)(1,0)}_n & -\bar{R}^{(1,1)}_n + \bar{R}^{(1,0)(0,1)'}_n \\ -\bar{R}^{(1,1)}_n + \bar{R}^{(1,0)(0,1)}_n & -\pi^{-1}\bar{R}^{(0,2)}_n + \bar{R}^{(0,1)(0,1)}_n \end{bmatrix},$$

which converges to $C^{(\theta)}_{22}$ a.s. by the SULLN, A5(iii), and Lemma A2(c), where $\bar{R}^{(i,j)}_n := n^{-1}\sum \hat{r}^{(i,j)}_t$ and $\bar{R}^{(i,j)(k,l)}_n := n^{-1}\sum \hat{r}^{(i,j)}_t \hat{r}^{(k,l)\prime}_t$. Given this, we can solve for $(\hat{\tilde{\theta}}^{(1)\prime}_{0,n}, \hat{\tilde{\theta}}^{(1)}_{1,n})$ from $\tilde{M}^{(2)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}^{(2)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$:

$$\bar{R}^c_n(\pi)\begin{bmatrix} \hat{\tilde{\theta}}^{(1)}_{0,n} \\ \pi\hat{\tilde{\theta}}^{(1)}_{1,n} \end{bmatrix} = \begin{bmatrix} (1-\pi)(\bar{R}^{(1,1)}_n - \bar{R}^{(1,0)(0,1)}_n) \\ -(1-\pi)\bar{R}^{(0,1)(0,1)}_n \end{bmatrix}.$$

From this, Lemma B2(a) follows by Lemma B1.

In a similar way, we can derive $(\hat{\tilde{\theta}}^{(2)\prime}_{0,n}, \hat{\tilde{\theta}}^{(2)}_{1,n})$ from $\tilde{M}^{(3)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}^{(3)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$. Then

$$\bar{R}^c_n(\pi)\begin{bmatrix} \hat{\tilde{\theta}}^{(2)}_{0,n} \\ \pi\hat{\tilde{\theta}}^{(2)}_{1,n} \end{bmatrix} = \begin{bmatrix} (\pi(\hat{\tilde{\theta}}^{(1)}_{1,n})^2 + 1 - \pi)(\bar{R}^{(1,2)}_n - \bar{R}^{(1,0)(0,2)}_n) \\ -(3\pi(\hat{\tilde{\theta}}^{(1)}_{1,n})^2 + 2(1-\pi)\hat{\tilde{\theta}}^{(1)}_{1,n} + 1 - \pi)\bar{R}^{(0,1)(0,2)}_n + (\hat{\tilde{\theta}}^{(1)}_{1,n})^2\bar{R}^{(0,3)}_n \end{bmatrix}$$
$$+ O_p(\hat{\tilde{\theta}}^{(1)}_{0,n}) + O_p(\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + (1-\pi)),$$

where $O_p(\hat{\tilde{\theta}}^{(1)}_{0,n})$ and $O_p(\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + (1-\pi))$ are $O_p(1)$ terms given in A5(iii), whose coefficients are $\hat{\tilde{\theta}}^{(1)}_{0,n}$ or $\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + (1-\pi)$. We have simplified our presentation for brevity. The RHS converges to $-(\frac{1-\pi}{\pi})C^{(2)}_{21}$ in probability by Lemma B2(a), implying Lemma B2(b).

Finally, $(\hat{\tilde{\theta}}^{(3)\prime}_{0,n}, \hat{\tilde{\theta}}^{(3)}_{1,n})$ can be derived from $\tilde{M}^{(4)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}^{(4)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ in the same way. Then,

$$\bar{R}^c_n(\pi)\begin{bmatrix} \hat{\tilde{\theta}}^{(3)}_{0,n} \\ \pi\hat{\tilde{\theta}}^{(3)}_{1,n} \end{bmatrix} = \begin{bmatrix} (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})(-\bar{R}^{(1,3)}_n + \bar{R}^{(1,0)(0,3)}_n) \\ -(\frac{1-\pi}{\pi})^3\bar{R}^{(0,4)}_n + 3(\frac{1-\pi}{\pi})^2\bar{R}^{(0,2)(0,2)}_n + (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})\bar{R}^{(0,1)(0,3)}_n \end{bmatrix}$$
$$+ O_p(\hat{\tilde{\theta}}^{(1)}_{0,n}) + O_p(\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + (1-\pi)) + O_p(1),$$

where $O_p(1)$ is a finite collection of $O_p(1)$ terms given by A5(iii) and Lemmas B2(a and b). Applying the SULLN leads to the desired result.

(*d*) Note that $\hat{f}_t\hat{g}_t^{(1)} = \hat{r}_t^{(1,0)}\hat{\hat{\theta}}_{0,n}^{(1)} + \hat{r}_t^{(0,1)}(\pi\hat{\hat{\theta}}_{1,n}^{(1)} + (1-\pi))$, so that the conclusion follows from Lemma B2(a).

(*e*) By the mean value theorem, $n^{-1/2}\sum\hat{r}_t^{(0,2)} = n^{-1/2}\sum r_t^{(0,2)}(\theta_0^*, \theta_*) + n^{-1/2}\sum\nabla_{\theta^1} r_t^{(0,2)}(\bar{\theta}_0, \bar{\theta}_1)[(\hat{\theta}_{n,0}^{n\prime}, \hat{\theta}_{n,1}^n) - (\theta_0^{*\prime}, \theta_*)]'$ for some $(\bar{\theta}_0, \bar{\theta}_1)$. Given this, $n^{-1}\sum\nabla_{\theta^1} r_t^{(0,2)}(\bar{\theta}_0, \bar{\theta}_1)$ converges to $-C_{12}^{(2)}$ a.s. by A5(iii), Lemma A1(a), and the SULLN. Lemma A2(c), the CLT in DMR (1995), and A6(ii) complete the proof. □

Proof of Lemma 2: (*a*) The result follows by Lemma B2(a) and the fact that $\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)(1-\hat{\hat{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}$.

(*b*) We can apply Lemmas B2(a, e) to $n^{-1}\tilde{L}_n^{(2)}(\pi, \hat{\theta}_{1,n}^n)$.

(*c*) Note that $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\{\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t + 2\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(1)}\}$. First, $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t = 2\hat{\hat{\theta}}_{0,n}^{(1)\prime}(1-\hat{\hat{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(1,2)} + (1-(\hat{\hat{\theta}}_{1,n}^{(1)})^2)\sum\hat{r}_t^{(0,3)} - \hat{\hat{\theta}}_{1,n}^{(2)}\sum\hat{r}_t^{(0,2)}$ by the definition of $\hat{h}_t$ and $\hat{k}_t$, so that $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t = O_p(n^{1/2})$ by Lemmas B1 and B2(a, b). Second, $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(1)} = (1-\hat{\hat{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}\hat{g}_t^{(1)}$, implying that $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(1)} = O_p(n^{1/2})$ by Lemma B2(d). Thus, $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$.

(*d*) Note that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(1)} + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)}\}$. We examine each element on the RHS. First, note that $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t = (1-(\hat{\hat{\theta}}_{1,n}^{(1)})^3)\sum\hat{r}_t^{(0,4)} + 3(1-\hat{\hat{\theta}}_{1,n}^{(1)})\hat{\hat{\theta}}_{0,n}^{(1)\prime}\sum\hat{r}_t^{(2,2)}\hat{\hat{\theta}}_{0,n}^{(1)} + 3(1-(\hat{\hat{\theta}}_{1,n}^{(1)})^2)\hat{\hat{\theta}}_{0,n}^{(1)\prime}\sum\hat{r}_t^{(1,3)} - 3\hat{\hat{\theta}}_{1,n}^{(2)}\hat{\hat{\theta}}_{1,n}^{(1)}\sum\hat{r}_t^{(0,3)} + 3(\hat{\hat{\theta}}_{0,n}^{(1)\prime}\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(2)\prime}(\hat{\hat{\theta}}_{1,n}^{(1)} - 1))\sum\hat{r}_t^{(1,2)} - \hat{\hat{\theta}}_{1,n}^{(3)}\sum\hat{r}_t^{(0,2)}$. Given this, it easily follows that $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t = o_p(n)$ by A5(iii) and Lemmas B2(a to c). Second, $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(1)} = 2\hat{\hat{\theta}}_{0,n}^{(1)\prime}(1-\hat{\hat{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(1,2)}\hat{g}_t^{(1)} + (1-(\hat{\hat{\theta}}_{1,n}^{(1)})^2)\sum\hat{r}_t^{(0,3)}\hat{g}_t^{(1)} - \hat{\hat{\theta}}_{1,n}^{(2)}\sum\hat{r}_t^{(0,2)}\hat{g}_t^{(1)}$ after some algebra. Thus, $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(1)} = O_p(n^{1/2})$ by Lemma B2(d). Third, note that $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)} = -(1-\hat{\hat{\theta}}_{1,n}^{(1)})\sum\hat{r}_t^{(0,2)}\{(\pi(\hat{\hat{\theta}}_{1,n}^{(1)})^2 + 1 - \pi)\hat{r}_t^{(0,2)} + \hat{\hat{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,0)} + \pi\hat{\hat{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,1)}\} + O_p(n\hat{\hat{\theta}}_{0,n}^{(1)}) + O_p(n(\pi\hat{\hat{\theta}}_{1,n}^{(1)} + 1 - \pi))$, so that $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)} = -(1-\pi)\pi^{-2}\sum(\hat{r}_t^{(0,2)})^2 - \hat{\hat{\theta}}_{1,n}^{(2)}\sum\hat{r}_t^{(0,2)}\hat{r}_t^{(0,1)} - \pi^{-1}\hat{\hat{\theta}}_{0,n}^{(2)\prime}\sum\hat{r}_t^{(0,2)}\hat{r}_t^{(1,0)} + o_p(n)$ by Lemma B2(a), implying that $n^{-1}\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)} = -(1-\pi)\pi^{-2}\Omega^{(2)} + o_p(1)$ by Lemma B2(b) and the definition of $\Omega^{(2)}$. Finally, $n^{-1}\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = -3(\frac{1-\pi}{\pi})^2\Omega^{(2)} + o_p(1)$ by combing all these, as claimed. □

Proof of Theorem 4: (*a*) By Lemma 2 and A5(ii, iii), for each π,

$$\sup_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \sup_{\xi}\left(\frac{1-\pi}{\pi}\right)[\Omega^{(2)}]^{1/2}G_0\xi^2 - \frac{1}{4}\left(\frac{1-\pi}{\pi}\right)^2\Omega^{(2)}\xi^4.$$

As ξ^2 cannot be less than zero, the optimal solution for ξ has to depend on the value of G_0. Suppose that $G_0 \geq 0$. Then the maximum is attained when $\xi^2 = 2(\frac{1-\pi}{\pi})[\Omega^{(2)}]^{-1/2}G_0$, so that $\max_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow G_0^2$. If $G_0 < 0$, the maximum is attained when $\xi = 0$, so that $\max_{\theta_2} 2(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow 0$. Thus, $\max_{\theta_2} 2\sum(\tilde{L}_n(\pi, \theta_2) - \tilde{L}_n(\pi, \hat{\theta}_{1,n}^n)) \Rightarrow \max[0, G_0]^2$, as claimed.

(*b*) The conclusion follows from the proof of Theorem 4(a), and the fact that the limiting distribution does not depend on π, as this coefficient function of $[\Omega^{(2)}]^{1/2}G_0$ and $\Omega^{(2)}$ vanishes in the FOC. □

Lemma C1: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H_{02}',

(*a*) $\sum\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4}\hat{f}_t/\hat{f}_t = O_p(n^{1/2})$;

(*b*) $\sum\nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4}\nabla_{i_5}\hat{f}_t/\hat{f}_t = O_p(n^{1/2})$;

(c) $\sum \nabla_{i_1}\nabla_{i_2}\nabla_{i_3}\nabla_{i_4}\nabla_{i_5}\nabla_{i_6}\hat{f}_t/\hat{f}_t = O_p(n^{1/2})$;

(d) $\sum \nabla_{j_1}\nabla^6_{\theta_1}\hat{f}_t/\hat{f}_t = O_p(n^{1/2})$ and $\sum \nabla^7_{\theta_1}\hat{f}_t/\hat{f}_t = O_p(n^{1/2})$, *where* $i_1, \cdots, i_6 \in \{\theta_{01}, \cdots, \theta_{0r}, \theta_1\}$ *and* $j_1 \in \{\theta_{01}, \cdots, \theta_{0r}\}$.

Proofs of Lemma C1: (*a to d*) Let $q_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n})$ be the functions of interests. By the mean value theorem, $\sum q_t(\hat{\theta}^n_{0,n}, \hat{\theta}^n_{1,n}) = \sum q_t(\theta^*_0, \theta_*) + \sum \nabla_{\theta^1} q_t(\hat{\theta}^m_{n,0}, \hat{\theta}^m_{n,1})[(\hat{\theta}^{n\prime}_{n,0}, \hat{\theta}^n_{n,1}) - (\theta^{*\prime}_0, \theta_*)]'$ for some $(\hat{\theta}^m_{n,0}, \hat{\theta}^m_{n,1})$. Given A5(ii, iv), we can apply the CLT in DMR (1995) to $n^{-1/2}\sum q_t(\theta^*_0, \theta_*)$, because $E[q_t(\theta^*_0, \theta_*)] = 0$. Also, $\sum \nabla_{\theta^1} q_t(\hat{\theta}^m_{n,0}, \hat{\theta}^m_{n,1}) = O_p(n)$ by the SULLN and A5(iv). This completes the proof. □

Lemma C2: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H'_{02}, *for each* π,

(a) $\hat{\tilde{\theta}}^{(1)}_{0,n} = 0$ *and* $(\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + 1 - \pi) = 0$;

(b) $(\hat{\tilde{\theta}}^{(2)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(2)}_{1,n})' = -(\frac{1-\pi}{\pi})(\alpha', \beta)' + O_p(n^{-1/2})$;

(c) $(\hat{\tilde{\theta}}^{(4)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(4)}_{1,n})' = O_p(1)$;

(d) $\hat{g}^{(1)}_t = 0$ *and* $\hat{m}^{(1)}_t = 0$;

(e) *for a sequence of random variables,* $\{\hat{q}_t\}$ *say, if* $\sum \hat{q}_t\hat{r}^{(1)}_t = O_p(n)$, *then* $\sum \hat{q}_t\hat{f}_t\hat{g}^{(2)}_t = O_p(n^{1/2})$;

(f) $\sum \hat{m}_t\hat{g}^{(3)}_t = O_p(n^{1/2})$ *and* $\sum \hat{k}_t\hat{g}^{(3)}_t = O_p(n^{1/2})$;

(g) $\sum \hat{m}_t\hat{g}^{(4)}_t = O_p(n^{1/2})$ *and* $\sum \hat{k}_t\hat{g}^{(4)}_t = O_p(n^{1/2})$.

(h) $n^{-1/2}\sum \hat{r}^{(0,3)}_t \overset{\text{A}}{\sim} N(0, \Omega^{(3)})$.

Proof of Lemma C2: (*a and b*) Let $\hat{r}^{(0,2)}_t = \alpha'\hat{r}^{(1,0)}_t + \beta\hat{r}^{(0,1)}_t$, and iterate the proof of Lemma B2(a and b respectively).

(c) If we rearrange $\tilde{M}^{(5)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}^{(5)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$, $\bar{R}^c_n(\pi)(\hat{\tilde{\theta}}^{(4)\prime}_{0,n}, \pi\hat{\tilde{\theta}}^{(4)}_{1,n})'$ equals

$$\begin{bmatrix} \frac{(1-\pi)(1-3\pi+3\pi^2)}{\pi^3}(\bar{R}^{(1,4)}_n - \bar{R}^{(1,0)(0,4)}_n) \\ (\frac{1-\pi}{\pi})^4\bar{R}^{(0,5)}_n + \frac{2(1-\pi)^2(7\pi-5)}{\pi^3}\bar{R}^{(0,3)(0,2)}_n - \frac{(1-\pi)(1-3\pi+3\pi^2)}{\pi^3}\bar{R}^{(0,1)(0,4)}_n \end{bmatrix} + 6\begin{bmatrix} -\left(\frac{1-\pi}{\pi}\right)^2\bar{R}^{(0,2)(1,2)}_n \\ \left(\frac{1-\pi}{\pi}\right)^2\bar{R}^{(0,1)(0,2)(0,2)}_n \end{bmatrix} + O_p(1),$$

using Lemma C2(a), where $\bar{R}^{(k,l)(i,j)(m,q)}_n := n^{-1}\sum \hat{r}^{(k,l)}_t\hat{r}^{(i,j)}_t\hat{r}^{(m,q)}_t$ and the remainder is the collection of $O_p(1)$ terms given in A5(iv). The given terms are those having the highest moment or the highest-order derivatives. Given this, it's straightforward to obtain the desired result by A5(ii, iv) and the SULLN.

(d) This is obvious from the facts that $\hat{g}^{(1)}_t = (\pi - 1)(\hat{h}_t - \hat{k}_t)(\hat{g}_t)^2$, $\hat{m}^{(1)}_t = (\pi\hat{\tilde{\theta}}^{(1)}_{1,n} + 1 - \pi)\hat{f}^{(0,2)}_t$, $\hat{h}_t = \hat{k}_t$ and Lemma C2(a).

(e) From $\hat{r}^{(0,2)}_t = \alpha'\hat{r}^{(1,0)}_t + \beta\hat{r}^{(0,1)}_t$ and Lemma C2(a), $\hat{f}_t\hat{g}^{(2)}_t = -(\frac{1-\pi}{\pi}\alpha + \hat{\tilde{\theta}}^{(2)}_{0,n})'\hat{r}^{(1,0)}_t - (\frac{1-\pi}{\pi}\beta + \pi\hat{\tilde{\theta}}^{(2)}_{1,n})\hat{r}^{(0,1)}_t$. The given result follows by Lemma C2(b).

(f) From Lemma C2(d), $\tilde{M}^{(4)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}^{(4)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$, $\sum \hat{m}_t\hat{g}^{(3)}_t = -\sum \hat{m}^{(3)}_t\hat{g}_t$ and $\sum \hat{k}_t\hat{g}^{(3)}_t = -\sum \hat{k}^{(3)}_t\hat{g}_t - 3\sum \hat{k}^{(1)}_t\hat{g}^{(2)}_t$ are obtained. Given these, $\sum \hat{m}^{(3)}_t\hat{g}_t$, $\sum \hat{k}^{(3)}_t\hat{g}_t$, and $\sum \hat{k}^{(1)}_t\hat{g}^{(2)}_t$ are $O_p(n^{1/2})$ by Lemmas B2(c) and C2(a, b and e), implying the given conclusion.

(g) Lemma C2(d), $\tilde{M}^{(5)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$, and $\tilde{K}^{(5)}_n(\pi, \hat{\theta}^n_{1,n}) = 0$ imply that $\sum \hat{m}_t\hat{g}^{(4)}_t = -\sum\{\hat{m}^{(4)}_t\hat{g}_t + 6\hat{m}^{(2)}_t\hat{g}^{(2)}_t\}$ and $\sum \hat{k}_t\hat{g}^{(4)}_t = -\sum\{\hat{k}^{(4)}_t\hat{g}_t + 6\hat{k}^{(2)}_t\hat{g}^{(2)}_t + 4\hat{k}^{(1)}_t\hat{g}^{(3)}_t\}$. Further, $\sum \hat{m}^{(4)}_t\hat{g}_t$, $\sum \hat{m}^{(2)}_t\hat{g}^{(2)}_t$, $\sum \hat{k}^{(4)}_t\hat{g}_t$ and $\sum \hat{m}^{(2)}_t\hat{g}^{(2)}_t$ are $O_p(n^{1/2})$ by Lemmas B2(c) and C2(a to c and e). Finally, $\sum \hat{k}^{(1)}_t\hat{g}^{(3)}_t = \hat{\tilde{\theta}}^{(1)}_{1,n}\sum(\alpha'\hat{f}^{(1,0)}_t + \beta\hat{f}^{(0,1)}_t)\hat{g}^{(3)}_t = \hat{\tilde{\theta}}^{(1)}_{1,n}\sum(\alpha'\hat{m}_t\hat{g}^{(3)}_t + \beta\hat{k}_t\hat{g}^{(3)}_t)$. Thus, $\sum \hat{k}^{(1)}_t\hat{g}^{(3)}_t = O_p(n^{1/2})$ by Lemma C2(f).

(h) We iterate the proof of Lemma B2(e). □

Proof of Lemma 3: (a) By Lemma C2(d), $\tilde{L}_n^{(3)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t = (1-\pi)\{2\hat{\hat{\theta}}_{0,n}^{(1)\prime}(1 - \hat{\hat{\theta}}_{1,n}^{(1)})\sum \hat{r}_t^{(1,2)} + (1-(\hat{\hat{\theta}}_{1,n}^{(1)})^2)\sum \hat{r}_t^{(0,3)}\}$, where the last equality holds by the fact that $\sum \hat{r}_t^{(0,2)} = 0$. The given result follows by Lemma C2(a).

(b) Apply Lemma C2(h) to $\sum \hat{r}_t^{(0,3)}$.

(c) Note that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)}\}$ by applying Lemma C2(d). We already showed that $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t = O_p(n^{1/2})$ in the proof of Lemma 2(d). Further, $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)} = O_p(n^{1/2})$ by Lemma C2(e). Thus, $\tilde{L}_n^{(4)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$.

(d) Note that $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t + 6(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(2)} + 4(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(3)}\}$ by Lemma C2(d). We examine each element in the RHS. First, some algebra gives that $\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t = -\pi^{-4}(1-2\pi)(1-2\pi+2\pi^2)\sum \hat{r}_t^{(0,5)} - 6\pi^{-2}(1-\pi)^2\hat{\hat{\theta}}_{1,n}^{(2)}\sum \hat{r}_t^{(0,4)} + 12\hat{\hat{\theta}}_{0,n}^{(2)}\pi^{-1}(1-0.5\pi^{-1})\sum \hat{r}_t^{(1,3)} - (3(\hat{\hat{\theta}}_{1,n}^{(2)})^2 + 4\hat{\hat{\theta}}_{1,n}^{(3)}(\pi^{-1} - 1))\sum \hat{r}_t^{(0,3)} - 2\pi^{-1}(3\pi\hat{\hat{\theta}}_{0,n}^{(2)}\hat{\hat{\theta}}_{1,n}^{(2)} - 2\hat{\hat{\theta}}_{0,n}^{(3)})'\sum \hat{r}_t^{(1,2)}$. Thus, $\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t = O_p(n^{1/2})$ by Lemmas B1(b) and C2(a, b and c). Second, it's trivial that $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(2)} + \sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(3)} = O_p(n^{1/2})$ by Lemma C2(e and f). Therefore, $\tilde{L}_n^{(5)}(\pi, \hat{\theta}_{1,n}^n) = O_p(n^{1/2})$, as desired. □

Proof of Lemma 4: (a) Note that $\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = (1-\pi)\sum\{(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t + 10(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} + 10(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} + 5(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)}\}$ by Lemma C2(d). We examine each element in the RHS. First, using Lemma C2(a), $\sum(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t = (1 - 5\pi(1-\pi)(1-\pi-\pi^2))\pi^{-5}\sum \hat{r}_t^{(0,6)} + O_p(n^{1/2})$, where the remainder is the collection of $O_p(n^{1/2})$ terms in Lemma C1(a to d) multiplied by the $O_p(1)$ terms in Lemmas B2(c) and C2(a, b and d). Thus, $\sum(\hat{h}_t^{(5)} - \hat{k}_t^{(5)})\hat{g}_t = O_p(n^{1/2})$ by Lemma C1(c). Second, $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(2)} = O_p(n^{1/2})$ by Lemma C2(e). Third, $\sum(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(4)} = \pi^{-1}\sum(\alpha'\hat{m}_t + \beta\hat{k}_t)\hat{g}_t^{(4)} = O_p(n^{1/2})$ by Lemma C2(g). Finally, $\sum(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(3)} = (-1+2\pi)/(\pi^2)\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} - \hat{\hat{\theta}}_{1,n}^{(2)}\sum \hat{f}_t^{(0,2)}\hat{g}_t^{(3)} = (-1+2\pi)/(\pi^2)\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} + o_p(n)$, where the last equality follows from the fact that $\sum \hat{f}_t^{(0,2)}\hat{g}_t^{(3)} = \sum(\alpha'\hat{m}_t + \beta\hat{k}_t)\hat{g}_t^{(3)} = O_p(n^{1/2})$ by Lemma C2(f). Thus, we pay attention only to $n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)}$. Note that $n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} = (3(1-\pi)\alpha\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(3)})'\bar{R}_n^{(1,0)(0,3)} + (3(1-\pi)\beta\hat{\hat{\theta}}_{1,n}^{(2)} - \pi\hat{\hat{\theta}}_{1,n}^{(3)})\bar{R}_n^{(0,1)(0,3)} + (1-\pi)(1-2\pi)/(\pi^2)\bar{R}_n^{(0,3)(0,3)}$ from the fact that $\hat{f}_t\hat{g}_t^{(3)} = (3(1-\pi)\alpha\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(3)})'\hat{r}_t^{(1,0)} + (3(1-\pi)\beta\hat{\hat{\theta}}_{1,n}^{(2)} - \pi\hat{\hat{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,1)} + (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})\hat{r}_t^{(0,3)}$. Further, note that $\sum \hat{k}_t\hat{g}_t^{(3)} = o_p(n)$ and $\sum \hat{m}_t\hat{g}_t^{(3)} = o_p(n)$ by Lemma C2(f), so that $(3(1-\pi)\alpha\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(3)})'\bar{R}_n^{(1,0)(1,0)} + (3(1-\pi)\beta\hat{\hat{\theta}}_{1,n}^{(2)} - \pi\hat{\hat{\theta}}_{1,n}^{(3)})\bar{R}_n^{(0,1)(1,0)} + (1-\pi)(1-2\pi)/(\pi^2)\bar{R}_n^{(0,3)(1,0)} = o_p(1)$ and $(3(1-\pi)\alpha\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(3)})'\bar{R}_n^{(1,0)(0,1)} + (3(1-\pi)\beta\hat{\hat{\theta}}_{1,n}^{(2)} - \pi\hat{\hat{\theta}}_{1,n}^{(3)})\bar{R}_n^{(0,1)(0,1)} + (1-\pi)(1-2\pi)/(\pi^2)\bar{R}_n^{(0,3)(0,1)} = o_p(1)$. Using these two equations, we can solve for $(3(1-\pi)\alpha\hat{\hat{\theta}}_{1,n}^{(2)} - \hat{\hat{\theta}}_{0,n}^{(3)})$ and $(3(1-\pi)\beta\hat{\hat{\theta}}_{1,n}^{(2)} - \pi\hat{\hat{\theta}}_{1,n}^{(3)})$, and plug these back into $n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)}$. Then, $n^{-1}\sum \hat{f}_t^{(0,3)}\hat{g}_t^{(3)} = -(1-\pi)(\frac{(1-2\pi)}{\pi^2})^2(\bar{R}_n^{(0,3)(0,3)} - \bar{R}_n^{(0,3)(1)}\bar{R}_n^{(1)(1)^{-1}}\bar{R}_n^{(1)(0,3)}) + o_p(1)$. Thus, $n^{-1}\tilde{L}_n^{(6)}(\pi, \hat{\theta}_{1,n}^n) = -10(\frac{(1-\pi)(1-2\pi)}{\pi^2})^2\Omega^{(3)} + o_p(1)$ by the SUULN, A5(iv), and the definition of $\Omega^{(3)}$.

(b) We can iterate the proof of Theorem 4(a) using $\tilde{L}_n^{(3)}(\pi, \theta_2)$ and $\tilde{L}_n^{(6)}(\pi, \theta_2)$ instead of $\tilde{L}_n^{(2)}(\pi, \theta_2)$ and $\tilde{L}_n^{(4)}(\pi, \theta_2)$. The absence of the sign condition leads to the desired conclusion. □

Lemma C3: *Given A1, A2(i, iv), A3, A4, A5(ii, iv), A6(iii), and* H_{02}', *if* $\pi = 1/2$, *then*

(a) $(\hat{\hat{\theta}}_{0,n}^{(3)\prime}, \pi\hat{\hat{\theta}}_{1,n}^{(3)})' = 1.5(\alpha', \beta)' + O_p(n^{-1/2})$;

(b) $(\hat{\hat{\theta}}^{(5)\prime}_{0,n}, \pi\hat{\hat{\theta}}^{(5)}_{1,n})' = O_p(1)$;

(c) $(\hat{\hat{\theta}}^{(6)\prime}_{0,n}, \pi\hat{\hat{\theta}}^{(6)}_{1,n})' = O_p(1)$;

(d) *for a sequence of random variables,* $\{\hat{q}_t\}$ *say, if* $\sum \hat{q}_t \hat{r}_t^{(1)} = O_p(n)$*, then* $\sum \hat{q}_t \hat{f}_t \hat{g}_t^{(3)} = O_p(n^{1/2})$;

(e) $\sum \hat{m}_t \hat{g}_t^{(5)} = O_p(n^{1/2})$ *and* $\sum \hat{k}_t \hat{g}_t^{(5)} = O_p(n^{1/2})$;

(f) $15\sum \hat{k}_t^{(2)} \hat{g}_t^{(4)} + \sum \hat{k}_t \hat{g}_t^{(6)} = O_p(n^{1/2})$ *and* $15\sum \hat{m}_t^{(2)} \hat{g}_t^{(4)} + \sum \hat{m}_t \hat{g}_t^{(6)} = O_p(n^{1/2})$.

Proof of Lemma C3: $(a$ *and* $d)$ By Lemma C2(a), it follows that $\hat{f}_t \hat{g}_t^{(3)} = (3(1-\pi)\alpha - \hat{\hat{\theta}}^{(3)}_{0,n})' \hat{r}_t^{(1,0)} + (3(1-\pi)\beta - \pi\hat{\hat{\theta}}^{(3)}_{1,n})\hat{r}_t^{(0,1)} + (\frac{1-\pi}{\pi})(\frac{1-2\pi}{\pi})\hat{r}_t^{(0,3)}$; and the last term vanishes if $\pi = 1/2$. Further, applying Lemma C2(f) leads to $\sum \hat{m}_t \hat{g}_t^{(3)} = O_p(n^{1/2})$ and $\sum \hat{k}_t \hat{g}_t^{(3)} = O_p(n^{1/2})$, so that $(1.5\beta - 0.5\hat{\hat{\theta}}^{(3)}_{1,n})\sum(\hat{r}_t^{(0,1)})^2 = O_p(n^{1/2})$ and $(1.5\alpha - \hat{\hat{\theta}}^{(3)}_{0,n})' \sum \hat{r}_t^{(1,0)} \hat{r}_t^{(1,0)\prime} = O_p(n^{1/2})$. Given these, $\sum(\hat{r}_t^{(0,1)})^2 = O_p(n)$ and $\sum \hat{r}_t^{(1,0)} \hat{r}_t^{(1,0)\prime} = O_p(n)$ by A5(iv), implying that $(\hat{\hat{\theta}}^{(3)\prime}_{0,n},\ \pi\hat{\hat{\theta}}^{(3)}_{1,n})' = 1.5(\alpha',\ \beta)' + O_p(n^{-1/2})$. Lemma C3(d) follows.

$(b$ *and* $c)$ Rearranging $\tilde{M}_n^{(6)}(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}_n^{(6)}(\pi, \hat{\theta}^n_{1,n}) = 0$; and $\tilde{M}_n^{(7)}(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{K}_n^{(7)}(\pi, \hat{\theta}^n_{1,n}) = 0$ lead to

$$\bar{R}_n^c(\pi)\begin{bmatrix}\hat{\hat{\theta}}^{(5)}_{0,n}\\ \pi\hat{\hat{\theta}}^{(5)}_{1,n}\end{bmatrix} = -\begin{bmatrix}0\\ \bar{R}_n^{(0,6)} - 15\bar{R}_n^{(0,2)(0,4)} + 30\bar{R}_n^{(0,2)(0,2)(0,2)}\end{bmatrix} + O_p(1),$$

and to $\bar{R}_n^c(\pi)[\hat{\hat{\theta}}^{(6)\prime}_{0,n}, \pi\hat{\hat{\theta}}^{(6)}_{1,n}]'$ equaling

$$\begin{bmatrix}30\bar{R}_n^{(1,0)(0,2)(0,4)} + 90\bar{R}_n^{(1,2)(0,2)(0,2)} - 90\bar{R}_n^{(1,0)(0,2)(0,2)(0,2)} + \bar{R}_n^{(1,6)} - \bar{R}_n^{(1,0)(0,6)} - 15\bar{R}_n^{(1,4)(0,2)} - 15\bar{R}_n^{(1,2)(0,4)}\\ 30\bar{R}_n^{(0,1)(0,2)(0,4)} + 90\bar{R}_n^{(0,2)(0,2)(0,3)} - 90\bar{R}_n^{(0,1)(0,2)(0,2)(0,2)} + \bar{R}_n^{(1,4)(0,2)} - 15\bar{R}_n^{(1,2)(0,4)}\end{bmatrix}$$

plus an $O_p(1)$ term, where $\bar{R}_n^{(k,l)(i,j)(m,q)(x,y)} := n^{-1}\sum \hat{r}_t^{(k,l)}\hat{r}_t^{(i,j)}\hat{r}_t^{(m,q)}\hat{r}_t^{(x,y)}$, and the remainders are the collections of $O_p(1)$ terms in A5(iv) multiplied by other $O_p(1)$ terms in Lemmas C2(a, b, d) and C3(a). It's now straightforward to obtain the given result by applying the SULLN and A5(iv).

(e) Given Lemma C3(d), this is identical to the proof of Lemma C2(f).

(f) From the facts that $\tilde{K}_n^{(7)}(\pi, \hat{\theta}^n_{1,n}) = 0$ and $\tilde{M}_n^{(7)}(\pi, \hat{\theta}^n_{1,n}) = 0$, it follows that $\sum\{15\hat{h}_t^{(2)}\hat{g}_t^{(4)} + \hat{h}_t\hat{g}_t^{(6)}\} = -\sum\{\hat{h}_t^{(6)}\hat{g}_t + 15\hat{h}_t^{(4)}\hat{g}_t^{(2)} + 20\hat{h}_t^{(3)}\hat{g}_t^{(3)} + 6\hat{h}_t^{(1)}\hat{g}_t^{(5)}\}$ and $\sum\{15\hat{m}_t^{(2)}\hat{g}_t^{(4)} + \hat{m}_t\hat{g}_t^{(6)}\} = -\sum\{\hat{m}_t^{(6)}\hat{g}_t + 15\hat{m}_t^{(4)}\hat{g}_t^{(2)} + 20\hat{m}_t^{(3)}\hat{g}_t^{(3)}\}$. Expanding all the elements on each RHS shows that each RHS is a sum of $O_p(n^{1/2})$ terms, verified by Lemmas B2(c and d), C2(a, b and d) and C3(a to e). □

Proof of Lemma 5: (a) Note that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}^n_{1,n}) = (1-\pi)\sum\{(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t + 3(\hat{h}_t^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(2)}\}$ by Lemma C2(d). Some algebra reveals that $\tilde{L}_n^{(4)}(\pi, \hat{\theta}^n_{1,n}) = \sum \hat{r}_t^{(0,4)} + 3(0.5\hat{\hat{\theta}}^{(2)}_{1,n} - \beta)\sum \hat{r}_t^{(0,3)} + 3(\hat{\hat{\theta}}^{(2)}_{0,n} - \alpha)'\sum \hat{r}_t^{(1,2)} - 3(0.5\alpha\hat{\hat{\theta}}^{(2)}_{1,n} + \beta\hat{\hat{\theta}}^{(2)}_{0,n})'\sum \hat{r}_t^{(1,1)} - 3\hat{\hat{\theta}}^{(2)\prime}_{0,n}\sum \hat{r}_t^{(2,0)}\alpha$ if $\pi = 1/2$. Thus, the desired result follows by Lemma C2(b) and the definition of $\hat{s}_t$.

(b) We iterate the proof of Lemma B2(e).

(c) Let $\pi = 1/2$ in the proof of Lemma 4(a).

(d) Note that $\tilde{L}_n^{(7)}(\pi, \hat{\theta}^n_{1,n}) = (1-\pi)\sum\{(\hat{h}_t^{(6)} - \hat{k}_t^{(6)})\hat{g}_t + 15(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t^{(2)} + 20(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(3)} + 15(\hat{h}_t^{(2)} - \hat{k}_t^{(2)})\hat{g}_t^{(4)} + 6(\hat{h}^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(5)}\}$ by Lemma C2(d). First, $\sum(\hat{h}_t^{(6)} - \hat{k}_t^{(6)})\hat{g}_t$, $\sum(\hat{h}_t^{(4)} - \hat{k}_t^{(4)})\hat{g}_t^{(2)}$ and $\sum(\hat{h}_t^{(3)} - \hat{k}_t^{(3)})\hat{g}_t^{(3)}$ are $O_p(n^{1/2})$ by Lemmas C2(a to c and e) and C3(a to d). Also, $\sum(\hat{h}^{(1)} - \hat{k}_t^{(1)})\hat{g}_t^{(5)} = O_p(n^{1/2})$ by Lemma C3(e), as $(\hat{h}^{(1)} - \hat{k}_t^{(1)}) = \pi^{-1}(\alpha'\hat{m}_t + \beta\hat{k}_t)$. Further, if $\pi = 1/2$, $(\hat{h}_t^{(2)} - \hat{k}_t^{(2)}) = -\hat{\hat{\theta}}^{(2)}_{1,n}\hat{f}_t^{(0,2)} =$

$-\hat{\tilde{\theta}}_{1,n}^{(2)}\pi^{-1}(\alpha'\hat{m}_t+\beta\hat{k}_t)$, so that $\sum(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t^{(4)}=O_p(n^{1/2})$ by Lemma C2(g). All the elements in the RHS are $O_p(n^{1/2})$, leading to the given claim.

(*e*) Note that $\tilde{L}_n^{(8)}(\pi,\hat{\theta}_{1,n}^n)=(1-\pi)\sum\{(\hat{h}_t^{(7)}-\hat{k}_t^{(7)})\hat{g}_t+21(\hat{h}_t^{(5)}-\hat{k}_t^{(5)})\hat{g}_t^{(2)}+35(\hat{h}_t^{(4)}-\hat{k}_t^{(4)})\hat{g}_t^{(3)}+35(\hat{h}_t^{(3)}-\hat{k}_t^{(3)})\hat{g}_t^{(4)}+21(\hat{h}^{(2)}-\hat{k}_t^{(2)})\hat{g}_t^{(5)}+7(\hat{h}^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(6)}\}$ by Lemma C2(d). Given this, $\sum(\hat{h}_t^{(7)}-\hat{k}_t^{(7)})\hat{g}_t=o_p(n)$ and $\sum(\hat{h}_t^{(5)}-\hat{k}_t^{(5)})\hat{g}_t^{(2)}$, $\sum(\hat{h}_t^{(4)}-\hat{k}_t^{(4)})\hat{g}_t^{(3)}$ and $\sum(\hat{h}_t^{(2)}-\hat{k}_t^{(2)})\hat{g}_t^{(5)}$ are $O_p(n^{1/2})$ by the same reasoning as in the proof of Lemma 5(c). Further, $\sum(\hat{h}^{(1)}-\hat{k}_t^{(1)})\hat{g}_t^{(6)}=\sum 2(\alpha'\hat{m}_t+\beta\hat{k}_t)\hat{g}_t^{(6)}=-30\sum(\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)})\hat{g}_t^{(4)}+o_p(n)$ by Lemma C3(f), implying that $n^{-1}\tilde{L}_n^{(8)}(\pi,\hat{\theta}_{1,n}^n)=35n^{-1}\sum\{(0.5[\hat{h}_t^{(3)}-\hat{k}_t^{(3)}]-3[\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)}])\hat{g}_t^{(4)}\}+o_p(1)$. By some algebra, $(0.5[\hat{h}_t^{(3)}-\hat{k}_t^{(3)}]-3[\alpha'\hat{m}_t^{(2)}+\beta\hat{k}_t^{(2)}])/\hat{f}_t=\hat{r}_t^{(0,4)}+(1.5\hat{\tilde{\theta}}_{1,n}^{(2)}-3\beta)\hat{r}_t^{(0,3)}+3(\hat{\tilde{\theta}}_{0,n}^{(2)}-\alpha)'\hat{r}_t^{(1,2)}-3(\beta\hat{\tilde{\theta}}_{0,n}^{(2)}+0.5\alpha\hat{\tilde{\theta}}_{1,n}^{(2)})'\hat{r}_t^{(1,1)}+(6\beta^2-0.5\hat{\tilde{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}-3\alpha'\hat{r}_t^{(2,0)}\hat{\tilde{\theta}}_{0,n}^{(2)}$ and $\hat{g}_t^{(4)}\hat{f}_t=-\hat{r}_t^{(0,4)}-3\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,3)}-6\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,2)}-3\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,1)}-3\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(2,0)}\hat{\tilde{\theta}}_{0,n}^{(2)}-(1.5(\hat{\tilde{\theta}}_{1,n}^{(2)})^2-2\hat{\tilde{\theta}}_{1,n}^{(3)})\hat{r}_t^{(0,2)}-\hat{\tilde{\theta}}_{0,n}^{(4)\prime}\hat{r}_t^{(1,0)}-0.5\hat{\tilde{\theta}}_{1,n}^{(4)}\hat{r}_t^{(0,1)}+1.5(\hat{\tilde{\theta}}_{1,n}^{(2)}\hat{r}_t^{(0,1)}+2\hat{r}_t^{(0,2)}+2\hat{\tilde{\theta}}_{0,n}^{(2)\prime}\hat{r}_t^{(1,0)})^2$. By the definition of $\hat{s}_t$, it follows that $n^{-1}\tilde{L}_n^{(8)}(\pi,\hat{\theta}_{1,n}^n)=-35(n^{-1}\sum\hat{s}_t^2-\bar{\upsilon}_n'\bar{R}_n^{(1)(1)}\bar{\upsilon}_n)+o_p(1)$, where $\bar{\upsilon}_n':=(\hat{\tilde{\theta}}_{0,n}^{(4)\prime},0.5\hat{\tilde{\theta}}_{1,n}^{(4)})+(1.5(\hat{\tilde{\theta}}_{1,n}^{(2)})^2-2\hat{\tilde{\theta}}_{1,n}^{(3)})(\alpha',\beta)$. Finally, from $\sum\hat{m}_t\hat{g}_t^{(4)}=o_p(n)$ and $\sum\hat{k}_t\hat{g}_t^{(4)}=o_p(n)$ (which are given in Lemma C2(g)) we have $\bar{\upsilon}_n=-[\bar{R}_n^{(1)(1)}]^{-1}n^{-1}\sum\hat{s}_t\hat{r}_t^{(1)}+o_p(1)$. The desired result follows by plugging this into $n^{-1}\tilde{L}_n^{(8)}(\pi,\hat{\theta}_{1,n}^n)=-35(n^{-1}\sum\hat{s}_t^2-\bar{\upsilon}_n'\bar{R}_n^{(1)(1)}\bar{\upsilon}_n)+o_p(1)$. This completes the proof.

(*f*) We can iterate the proof of Theorem 4(a) by approximating the QLR statistic using $\tilde{L}_n^{(4)}(\pi,\theta_2)$ and $\tilde{L}_n^{(8)}(\pi,\theta_2)$ instead of $\tilde{L}_n^{(2)}(\pi,\theta_2)$ and $\tilde{L}_n^{(4)}(\pi,\theta_2)$. □

Proof of Lemma 6: (*a*) To show this, we approximate $n^{-1/2}\sum\nabla_\pi\ell_t(1,\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n,\,\cdot\,)=n^{-1/2}\sum(1-f_t(\hat{\theta}_{0,n}^n,\,\cdot\,)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n))$ around $\hat{\theta}_{1,n}^n$, which forms the main argument of Lemma 1. It then follows that $n^{-1/2}\sum(1-f_t(\hat{\theta}_{0,n}^0,\theta_2)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n))=-\frac{1}{2n^{1/2}}\sum\hat{r}_t^{(0,2)}(\theta_2-\theta_*)^2+o_p(|\theta_2-\theta_*|^2)$, because $\sum\hat{r}_t^{(0,1)}\equiv 0$ by the FOC and $(\hat{\theta}_{1,n}^n-\theta_*)=O_p(n^{-1/2})$. Further, the asymptotic variance is $\frac{1}{4}\Omega^{(2)}(\theta_2-\theta_*)^4+o(|\theta_2-\theta_*|^4)$, implying that the standardized score is $-n^{-1/2}\sum\hat{r}_t^{(0,2)}/[\Omega^{(2)}]^{1/2}+o_p(|\theta_2-\theta_*|)$. The negative value of the first component is asymptotically identical to the score used to derive Theorem 4. Hence, $\mathcal{G}(\theta_2)=-G_0+o_p(|\theta_2-\theta_*|)$, as desired.

(*b*) As $\sum\hat{r}_t^{(0,2)}=0$, we approximate the function of interest using the next order derivative, so that $n^{-1/2}\sum(1-f_t(\hat{\theta}_{0,n}^0,\theta_2)/f_t(\hat{\theta}_{0,n}^n,\hat{\theta}_{1,n}^n))=-\frac{1}{3!n^{1/2}}\sum\hat{r}_t^{(0,3)}(\theta_2-\theta_*)^3+o_p(|\theta_2-\theta_*|^3)$. This has asymptotic variance $\frac{1}{36}\Omega^{(3)}(\theta_2-\theta_*)^6+o(|\theta_2-\theta_*|^6)$, and its standardized score is $-n^{-1/2}\sum\hat{r}_t^{(0,3)}(\theta_2-\theta_*)^3/[\Omega^{(3)}(\theta_2-\theta_*)^6]^{1/2}+o_p(|\theta_2-\theta_*|)$. Note that this equals $-n^{-1/2}\sum\hat{r}_t^{(0,3)}/[\Omega^{(3)}]^{1/2}+o_p(|\theta_2-\theta_*|)$ if $\theta_2\geq\theta_*$ or $n^{-1/2}\sum\hat{r}_t^{(0,3)}/[\Omega(0,3)]^{1/2}+o_p(|\theta_2-\theta_*|)$ otherwise. As before, $n^{-1/2}\sum\hat{r}_t^{(0,3)}/[\Omega^{(3)}]^{1/2}$ is asymptotically equivalent to the score used for Lemma 4(b), but its sign depends whether θ_2 approaches θ_* from above or below. This completes the proof. □

Proof of Theorem 6 (*a and b*): By the continuous mapping theorem, $QLR_n\Rightarrow\max[\max[0,G_0]^2,\sup_{\Theta_*}\min[0,\mathcal{G}(\theta_2)]^2]$ if (a) is considered; and $QLR_n\Rightarrow\max[\sup_{\Theta_*\setminus\{\theta_*\}}\min[0,\mathcal{G}(\theta_2)]^2,G_0^2,\max[0,G_*]^2]$ if (b) is considered. By Lemma 6, $\max[0,G_0]^2\leq\sup_{\Theta_*\setminus\{\theta_*\}}\min[0,\mathcal{G}(\theta_2)]^2$ in (a); and $G_0^2\leq\sup_{\Theta_*\setminus\{\theta_*\}}\min[0,\mathcal{G}(\theta_2)]^2$ in (b). The conclusions follow. □

Proof of Theorem 7: We use the following facts to prove the result.[6]

Fact 1: *If* $\mathcal{B}^s(\cdot)$ *is a stationary stochastic process on* $[\underline{\theta}, \overline{\theta}]$ *such that* $E[\mathcal{B}^s(\theta_2)\mathcal{B}^s(\theta_2')] = 1 - |\theta_2 - \theta_2'|^\gamma(1 + o(1))$, *then* $P(\sup_{\theta_2 \in [\underline{\theta}, \overline{\theta}]} \mathcal{B}^s(\theta_2) > u) = H_\gamma(\overline{\theta} - \underline{\theta})u^{2/\gamma}(1 - \Phi(u))(1 + o(1))$ *as* $u \to \infty$, *where* $H_\gamma := \lim_{\overline{\theta} \to \infty} H_\gamma(\overline{\theta})/\overline{\theta}$ *and* $H_\gamma(\overline{\theta}) := E[\exp(\max_{\theta_2 \in [0, \overline{\theta}]} \mathcal{B}^F(\theta_2))]$.

Fact 2 (*Slepian inequality*): *Let* $\mathcal{U}(\cdot)$ *and* $\mathcal{V}(\cdot)$ *be separable Gaussian stochastic processes on* Θ_*. *If* $E[\mathcal{U}(\theta_2)^2] = E[\mathcal{V}(\theta_2)^2]$, $E[\mathcal{U}(\theta_2)] = E[\mathcal{V}(\theta_2)]$ *and* $E[\mathcal{U}(\theta_2)\mathcal{U}(\theta_2')] \leq E[\mathcal{V}(\theta_2)\mathcal{V}(\theta_2')]$ *for all* $\theta_2, \theta_2' \in \Theta_*$, *then for all* $x \in \mathbb{R}$, $P(\sup_{\Theta_*} \mathcal{U}(\theta_2) < x) \leq P(\sup_{\Theta_*} \mathcal{V}(\theta_2) < x)$.

Given our assumptions, it easily follows that $E[\mathcal{G}(\theta_2)\mathcal{G}(\theta_2')] \geq 1 - |\theta_2 - \theta_2'|^\gamma(1 + o(1))$, as $|\theta_2 - \theta_2'| \to 0$, and from this, $P(\sup_{[\underline{\theta}, \overline{\theta}]} \mathcal{G}(\theta_2) \geq u) \leq \sum_{i=1}^{u^*} P(\sup_{\theta_2 \in \Theta_i} \mathcal{G}(\theta_2) \geq u) \leq \sum_{i=1}^{u^*} P(\sup_{\theta_2 \in \Theta_i} \mathcal{B}^s(\theta_2) \geq u) = H_\gamma(\sum_{i=1}^{u^*}(\theta_2^i - \theta_2^{i-1}))u^{2/\gamma}(1 - \Phi(u))(1 + o(1)) = H_\gamma(\overline{\theta} - \underline{\theta})u^{2/\gamma}(1 - \Phi(u))(1 + o(1))$, as $u \to \infty$, where u^* is the number of sets, Θ_i, partitioning $[\underline{\theta} = \theta_2^0, \theta_*) \cup (\theta_*, \overline{\theta} = \theta_2^{u^*}]$ such that $E[\mathcal{B}^s(\theta_2)\mathcal{B}^s(\theta_2')] = 1 - |\theta_2 - \theta_2'|^\gamma(1 + o(1))$ for all θ_2, θ_2' in Θ_i whose closure is $[\theta_2^{i-1}, \theta_2^i]$. From the fact that $\underline{\theta}$ and $\overline{\theta}$ are bounded, u^* must be finite. The first and the second inequalities are from the Bonferoni and the Slepian inequalities respectively, and the first equality follows from Fact 1. Finally, for any $u > 0$, $P(\sup_{\Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u) = P(\sup_{\Theta_* \backslash \{\theta_*\}} \max[0, \mathcal{G}(\theta_2)] \geq u) = P(\sup_{\Theta_* \backslash \{\theta_*\}}(\max[0, \mathcal{G}(\theta_2)])^2 \geq u^2) = P(\sup_{\Theta_* \backslash \{\theta_*\}}(\min[0, \mathcal{G}(\theta_2)])^2 \geq u^2)$, where the last equality follows by the symmetry of the Gaussian process distribution, as desired. □

Proof of Theorem 8: Note that $P(\sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u) = P(|G_0| \geq u) + P(|G_0| < u,\ \sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u)$. We can decompose the second element on the RHS as follows: $P(|G_0| < u, \sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u) = P(|G_0| < u, N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) \geq 1) = P(N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) = 1) + P(N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) \geq 2) - P(|G_0| \geq u, N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) \geq 1)$. First, note that $P(N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) = 1) = E[N_u^{\mathcal{G}}((\theta_*, \overline{\theta}_2]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*))] - \sum_{k=2}^\infty k \cdot p_k(u)$, where $p_k(u) = P(N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]\ ) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) = k)$. Next, it follows trivially that $P(N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) \geq 2) - P(|G_0| > u, N_u^{\mathcal{G}}((\theta_*, \overline{\theta}]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*)) \geq 1) < \sum_{k=2}^\infty k \cdot p_k(u)$. Thus, combining these gives $P(\sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u) = P(|G_0| \geq u) + P(|G_0| < u,\ \sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u) \leq E[N_u^{\mathcal{G}}((\theta_*, \overline{\theta}_2]) + N_u^{\mathcal{G}}([\underline{\theta}, \theta_*))]$. Finally, $P(\sup_{\theta_2 \in \Theta_* \backslash \{\theta_*\}}(\min\ [0, \mathcal{G}(\theta_2)])^2 \geq u^2) = P(\sup_{\Theta_* \backslash \{\theta_*\}} \mathcal{G}(\theta_2) \geq u)$, as in the proof of Theorem 7, and $P(|G_0| > u) = 2(1 - \Phi(u))$, yielding the desired result. □

References

[1] Andrews, D. (1994), "Empirical Process Methods in Econometrics," in R. Engle and D. McFadden, eds. *Handbook of Econometrics*, 4, pp. 2247–2294.

[2] Andrews, D. (1999), "Estimation When a Parameter is on a Boundary," *Econometrica*, 67, 543–563.

[3] Andrews, D. (2001), "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

[6]These are proved in Piterbarg (1996, p. 16) and Dudley (1999, p. 31) respectively.

[4] ANDREWS, D. AND PLOBERGER, W. (1994), "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.

[5] AZAÏS, J., GASSIAT, É., AND MERCADIER, C. (2006), "Asymptotic Distribution and Power of the Likelihood Ratio Test for Mixtures: Bounded and Unbounded Cases," *Bernoulli*, 12, 775– 799.

[6] BICKEL, P. AND CHERNOFF, H. (1993), "Asymptotic Distribution of the Likelihood Ratio Statistic in a Prototypical Nonregular Problem," in J. Ghosh, S. Mitra, K. Parthasarathy, and B. Prakasa Rao, eds. *Statistics and Probability*: *A Raghu Raj Bahadur Festschrift*, 83–96. New York: Wiley.

[7] BILLINGSLEY, P. (1999), *Convergence of Probability Measures*. New York: Wiley.

[8] BÖHNING, D., DIETZ, E., SCHAUB, R., SCHLATTMAN, P., AND LINDSAY, B. (1994), "The Distribution of the Likelihood Ratio for Mixtures of Densities from the One Parameter Exponential Family," *Annals of the Institute of Statististical Mathematics*, 46, 373–388.

[9] CARRASCO, M., HU, L. AND PLOBERGER, W. (2004), "Optimal Test for Markov Switching," mimeo, University of Rochester.

[10] CHEN. H. AND CHEN, J. (2001), "The Likelihood Ratio Test for Homogeneity in the Finite Mixture Models," *Canadian Journal of Statistics*, 29, 201–216.

[11] CHERNOFF, H. (1954), "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 25, 573–578.

[12] CHERNOFF, H. AND LANDER, E. (1995), "Asymptotic Distribution of the Likelihood Ratio Test That a Mixture of Two Binomials is a Single Binomial," *Journal of Statistical Planning and Inference*, 43, 19–40.

[13] CHESHER, A. (1984), "Testing for Neglected Heterogeneity," *Econometrica*, 22, 865–872.

[14] COSSLETT, S., AND LEE, L.-F. (1985), "Serial Correlation in Discrete Variable Models," *Journal of Econometrics*, 27, 79-97.

[15] CRAMÉR, H. AND LEADBETTER, M. (1967), *Stationary and Related Stochastic Processes*: *Sample Function Properties and Their Application*. New York: Wiley.

[16] DACUNHA-CASTELLE, D. AND GASSIAT, É. (1999), "Testing the Order of a Model Using Locally Conic Parameterization: Population Mixtures and Stationary ARMA Processes," *The Annals of Statistics*, 27, 1178–1209.

[17] DARLING, D. (1955), "The Cramér-Smirnov test in the Parametric Case," *Annals of Mathematical Statistics*, 26, 1–20.

[18] DAVYDOV, Y. (1973), "Mixing Conditions for Markov Chains," *Theory of Probability and Its Applications*, 18, 312–328.

[19] DAVIES, R. (1977), "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

[20] DAVIES, R. (1987), "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

[21] Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics, 85. New York: Springer-Verlag.

[22] Doukhan, P., Massart, P., and Rio, E. (1995), "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré, Probabilites et Statistiques*, 31, 393-427.

[23] Dudley, M. (1999), *Uniform Central Limit Theorems*. New York: Cambridge University Press.

[24] Durbin, J. (1973), "Weak Convergence of the Sample Distribution Function When Parameters Are Estimated," *Annals of Statistics*, 1, 279–290.

[25] Ephraim, Y. and Merhav, N. (2002), "Hidden Markov Processes," *IEEE Transactions on Information Theory*, 48, 1518–1569.

[26] Ellison, G. (1994), "Theories of Cartel Stability and the Joint Executive Committee," *RAND Journal of Economics*, 25, 37–57.

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

[28] Ghosh, J. and Sen, P. (1985), "On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results," in L. Le Cam and R. Olshen, eds. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, 2. Berkeley: University of California Press, pp. 789–806.

[29] Goffinet, B., Loisel, P., and Laurent, B. (1992), "Testing in Normal Mixture Models When the Proportions are Known," *Biometrika*, 79, 842-846.

[30] Grenander, U. (1981), *Abstract Inference*. New York: Wiley.

[31] Green, E. and Porter, R. (1984), "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87–100.

[32] Hamilton, J. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–384.

[33] Hamilton, J. (1996), "Specification Testing in Markov-Switching Time-Series Models," *Journal of Econometrics*, 70, 127–157.

[34] Hartigan, J. (1985), "Failure of Log-Likelihood Ratio Test," in L. Le Cam and R. Olshen, eds. *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, 2. Berkeley: University of California Press, pp. 807–810.

[35] Hansen, B. (1992), "The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

[36] Hansen, B. (1996), "Inference When a Nuisance Parameter is Not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

[37] Hansen, B. (2004), "Interval Forecasts and Parameter Uncertainty," mimeo.

[38] Lancaster, T. (1984), "The Covariance Matrix of the Information Matrix Test," *Econometrica*, 52, 1051–1053.

[39] Lee, L.-F. and Chesher, A. (1986), "Specification Testing When Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 121–149.

[40] Levine, D. (1983), "A Remark on Serial Correlation in Maximum Likelihood," *Journal of Econometrics*, 23, 337–342.

[41] Lifshits, M. (1995), *Gaussian Random Functions*. Boston: Kluwer Academic Publishers.

[42] Lindsay, B. (1995), *Mixture Models: Theory, Geometry, and Applications*. Hayward: Institute for Mathematical Statistics.

[43] Liu, X. and Shao, Y. (2003), "Asymptotics for Likelihood Ratio Tests under Loss of Identifiability," *Annals of Statistics*, 31, 807–832.

[44] Mosler, K. and Seidel, W. (2001), "Testing for Homogeneity in an Exponential Mixture Model," *Australian & New Zealand Journal of Statistics*, 43, 231–247.

[45] McLachlan, G. and Basford, K. (1988), *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker.

[46] Neyman, J. and Scott, E. (1966), "On the Use of C(α) Optimal Tests of Composite Hypothesis," *Bulletin of the Institute of International Statististics*, 41, 477–497.

[47] Piterbarg, V. (1996), *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Translations of Mathematical Monographs, 148. Providence: American Mathematical Society.

[48] Porter, R. (1983), "A Study of Cartel Stability: The Joint Executive Committee, 1880–1886," *Bell Journal of Economics*, 14, 301–314.

[49] Ranga Rao, R. (1962), "Relations Between Weak and Uniform Convergence of Measures with Applications," *Annals of Mathematical Statistics*, 33, 659-680.

[50] Rice, S. (1944, 1945), "Duration of Fades in Radio Transmission," *Bell System Technical Journal*, 37, 581–635.

[51] Rudin, W. (1976), *Principles of Mathematical Analysis*. New York: McGraw-Hill.

[52] Self, S. and Liang, K. (1987), "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests under Nonstandard Conditions," *Journal of American Statistical Association*, 82, 605–610.

[53] Stinchcombe, M. and White, H. (1998), "Consistent Specification Testing with Nuisance Parameters Present Only under the Alternative," *Econometric Theory*, 14, 295–324.

[54] Vidyasagar, M. (2003), *Learning and Generalization: With Applications to Neural Networks*. London: Springer Verlag.

[55] White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

[56] White, H. (1994), *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

[57] White, H. (2001), *Asymptotic Theory for Econometricians*. Orlando: Academic Press.