COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 208281 Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1058

Admissibility of the Likelihood Ratio Test When a Nuisance Parameter is Present Only Under the Alternative

by

Donald W.K. Andrews

and

Werner Ploberger

July 1993

ADMISSIBILITY OF THE LIKELIHOOD RATIO TEST WHEN A NUISANCE PARAMETER IS PRESENT ONLY UNDER THE ALTERNATIVE

Donald W. K. Andrews
Cowles Foundation for Research in Economics
Yale University

Werner Ploberger
Technische Universität Wien

July 1993

ABSTRACT

This paper establishes the asymptotic admissibility of the likelihood ratio (LR) test for a general class of testing problems in which a nuisance parameter is present only under the alternative hypothesis. The paper also establishes the finite sample admissibility of the LR test for testing problems of this sort that arise in Gaussian linear regression models with known variance.

1. INTRODUCTION

This paper considers hypothesis tests when a nuisance parameter is present only under the alternative hypothesis. Such tests are non-standard and the classical likelihood ratio (LR) test does not possess its usual chi-square asymptotic null distribution in this context. It also does not possess its usual asymptotic optimality properties (of the sort considered by Wald (1943)).

Davies (1977, 1987) first provided a general asymptotic analysis of the testing problems considered here. He established the asymptotic null distribution of the LR test under a set of high-level assumptions. He also provided approximations to the asymptotic critical values of the LR test.

Andrews and Ploberger (1992) (denoted AP) developed a class of tests, called exponential LR tests, that exhibit explicit asymptotic optimality properties in terms of weighted average power when a nuisance parameter is present only under the alternative. The weight functions they consider are particular multivariate normal densities. The class of tests that are optimal with respect to these weight functions does not include the LR test. These results, Davies' adoption of the LR test, and the omnibus use of the LR test make the question of the asymptotic admissibility of the LR test one of considerable interest (to some at least). It is this question that is addressed in the present paper.

We show that the LR test and two asymptotically equivalent tests, viz., the sup Wald and sup Lagrange multiplier (LM) tests, are asymptotically admissible. In fact, we show that these tests are best tests, in a certain sense, against alternatives that are sufficiently distant from the null hypothesis. We establish these results first under a set of high-level assumptions. Then, we provide primitive sufficient conditions for a number of examples. The examples considered include tests of (i) changepoints in nonlinear dynamic models, (ii) cross-sectional constancy in nonlinear models, (iii) threshold effects in autoregressive models, (iv) variable relevance in nonlinear models, such as Box-Cox transformed regressor models, and (v) functional form in nonlinear models. Two examples that are covered by the high-level results, but for which primitive conditions are not provided, are tests of (i) white noise versus first-order autoregressive-moving average structure (ARMA (1, 1)) and (ii) white noise versus first-order generalized autoregressive conditional heteroskedasticity (GARCH (1, 1)).

Next, we consider finite sample admissibility of the LR test for the Gaussian linear regression model with known variance. We show that minor modifications to the proof of the asymptotic admissibility result yield finite sample admissibility. The types of hypotheses covered by this result include tests of (i) single and multiple changepoints, (ii) variable relevance for Box-Cox transformed regressors, and (iii) cross-sectional constancy, among others. The admissibility result for a single changepoint in the case of an iid univariate Gaussian location model replicates a recent result of Chang and Hartigan (1993). (Rather unbelievably, Chang and Hartigan's work was done independently and almost contemporaneously -- their paper was finished six months earlier -- fifty yards down Hillhouse Avenue.)

The remainder of this paper is organized as follows. Section 2 presents the main asymptotic admissibility result under a set of high-level assumptions. Section 3 presents examples and provides primitive sufficient conditions for the high-level assumptions. Section 4 states the finite sample admissibility results for tests concerning a Gaussian linear regression model. Section 5 gives proofs of the results stated in earlier sections.

2. ASYMPTOTIC ADMISSIBILITY

This section introduces notation and assumptions and states the asymptotic admissibility result of the paper. The notation and assumptions are very similar to those of AP. The problem considered is that of testing whether a subvector $\beta \in R^p$ of a parameter $\theta \in \Theta \subset R^s$ equals zero when the likelihood function depends on an additional parameter $\pi \in \Pi$ under the alternative.

2.1. Notation and Definitions

Let $(\Omega, \mathcal{F}, P)$ denote a probability space on which all of the random elements introduced below are defined. Let $\underset{\sim}{Y}_T$ denote the data matrix when the sample size is T for $T = 1, 2, \ldots$. Consider a parametric family $\{f_T(\underset{\sim}{y}_T, \theta, \pi) : \theta \in \Theta, \pi \in \Pi\}$ of densities of $\underset{\sim}{Y}_T$ with respect to some σ-finite measure μ_T, where $\Theta \subset R^s$ and Π is some metric space (usually a subset of Euclidean space). The likelihood function of the data is given by $f_T(\theta, \pi) = f_T(\underline{Y}_T, \theta, \pi)$.

The parameter θ is taken to be of the form $\theta = (\beta', \delta')'$, where $\beta \in R^p$, $\delta \in R^q$, and $s = p+q$. For example, in one-time changepoint problems, the parameter $\pi \in (0, 1)$ indicates the point of change as a fraction of the sample size, δ is of the form $(\delta_1', \delta_2')'$, δ_1 is a pre-change parameter vector, $\delta_1 + \beta$ is a post-change parameter vector, and δ_2 is a parameter vector that is constant across regimes.

The null hypothesis of interest is

(2.1) $\quad \mathrm{H}_0 : \beta = \underset{\sim}{0}$.

In the changepoint problem, this is the hypothesis of no change. The alternative hypothesis is

(2.2) $\quad \mathrm{H}_1 : \beta \neq \underset{\sim}{0}$ and the likelihood function depends on the parameter π .

We let θ_0 denote the true value of θ under the null hypothesis. Under the null hypothesis, the likelihood function $f_T(\theta_0, \pi)$ does not depend on the parameter π and is denoted $f_T(\theta_0)$. Let $\ell_T(\theta, \pi) = \log f_T(\theta, \pi)$. Let $D\ell_T(\theta, \pi)$ denote the s-vector of partial derivatives of $\ell_T(\theta, \pi)$ with respect to θ. Let $D^2\ell_T(\theta, \pi)$ denote the $s \times s$ matrix of second partial derivatives of $\ell_T(\theta, \pi)$ with respect to θ. (Note that $D\ell_T(\theta_0, \pi)$ and $D^2\ell_T(\theta_0, \pi)$ depend on π in general even though $f_T(\theta_0, \pi)$ and $\ell_T(\theta_0, \pi)$ do not.)

We consider the case where the appropriate norming factors for $D\ell_T(\theta, \pi)$ and $D^2\ell_T(\theta, \pi)$ (so that each is $O_p(1)$ but not $o_p(1)$) are non-random diagonal $s \times s$ matrices B_T^{-1} and $B_T^{-1} \times B_T^{-1}$, respectively, where $[B_T^{-1}]_{jj} \to 0$ as $T \to \infty$ $\forall j \leq s$. For non-trending data, the matrix B_T is just $\sqrt{T}\, I_s$. For data with deterministic time trends, B_T is more complicated, see AP. The local alternatives to H_0 that we consider are of the form $f_T(\theta_0 + B_T^{-1}h, \pi)$ for $h \in R^s$ and $\pi \in \Pi$.

All limits below are taken "as $T \to \infty$" unless stated otherwise. We say that a statement holds "under θ_0" (i.e., under the null hypothesis) if it holds when the true density of $\underline{Y}_T$ is $f_T(\theta_0)$ for $T = 1, 2, \ldots$. Let $\lambda_{\min}(A)$ denote the smallest eigenvalue of a matrix A. Let $|\cdot|$ denote the Euclidean norm. Let wp $\to 1$ abbreviate "with probability that goes to 1 as $T \to \infty$."

2.2. Assumptions

The likelihood function/parametric model is assumed to satisfy:

ASSUMPTION 1: (a) $f_T(\theta, \pi)$ *does not depend on* π *for all* θ *in the null hypothesis.*

(b) θ_0 *is an interior point of* Θ.

(c) $f_T(\theta, \pi)$ *is twice continuously partially differentiable in* θ *for all* $\theta \in \Theta_0$ *and* $\pi \in \Pi$ *with probability one under* Θ_0, *where* Θ_0 *is some neighborhood of* θ_0.

(d) $-B_T^{-1} D^2 \ell_T(\theta, \pi) B_T^{-1} \xrightarrow{p} \mathcal{I}(\theta, \pi)$ *uniformly over* $\pi \in \Pi$ *and* $\theta \in \Theta_0$ *under* θ_0 *for some non-random* $s \times s$ *matrix function* $\mathcal{I}(\theta, \pi)$ *and some sequence of non-random diagonal* $s \times s$ *matrices* $\{B_T : T \geq 1\}$ *that satisfies* $[B_T]_{jj} \to \infty$ *as* $T \to \infty$ $\forall j \leq s$.

(e) $\mathcal{I}(\theta, \pi)$ *is uniformly continuous in* (θ, π) over $\Theta_0 \times \Pi$.

(f) $\mathcal{I}(\theta_0, \pi)$ *is uniformly positive definite over* $\pi \in \Pi$ (*i.e.* $\inf_{\pi \in \Pi} \lambda_{\min}(\mathcal{I}(\theta_0, \pi)) > 0$).

The matrix function $\mathcal{I}(\theta, \pi)$ introduced in Assumption 1 is the asymptotic information matrix for θ for given π, which depends on both θ and π. See AP for comments on Assumption 1. Note that Assumption 1 is a "high level" assumption. Primitive conditions that imply Assumption 1 are provided below.

Let $\hat{\theta}(\pi)$ ($= \hat{\theta}_T(\pi)$) be the *(unrestricted) maximum likelihood (ML) estimator* of θ for fixed $\pi \in \Pi$. That is, $\hat{\theta}(\pi)$ satisfies

$$(2.3) \qquad \ell_T(\hat{\theta}(\pi), \pi) = \sup_{\theta \in \Theta} \ell_T(\theta, \pi) \quad \forall \pi \in \Pi \quad \text{wp} \to 1 \text{ under } \theta_0 .$$

Let $\tilde{\theta}$ be the *restricted maximum likelihood estimator* of θ. That is, $\tilde{\theta}$ satisfies

$$(2.4) \qquad \begin{aligned} &\tilde{\theta} \in \tilde{\Theta} = \{\theta \in \Theta : \theta = (0', \delta')' \text{ for some } \delta \in R^q\} \quad \text{and} \\ &\ell_T(\tilde{\theta}, \pi) = \sup_{\theta \in \tilde{\Theta}} \ell_T(\theta, \pi) \quad \text{wp} \to 1 \text{ under } \theta_0 . \end{aligned}$$

Note that $\tilde{\theta}$ does not depend on π by Assumption 1(a).

We assume that the parametric model is sufficiently regular that the ML estimators $\hat{\theta}(\pi)$ and $\tilde{\theta}$ are consistent for θ_0 under the null hypothesis uniformly over $\pi \in \Pi$.

ASSUMPTION 2: $\sup_{\pi \in \Pi} |\hat{\theta}(\pi) - \theta_0| \xrightarrow{p} 0$ *under* θ_0.

ASSUMPTION 3: $\tilde{\theta} - \theta_0 \xrightarrow{p} 0$ *under* θ_0.

The parameter space Π is assumed to satisfy:

ASSUMPTION 4: Π *is a compact metric space with metric* ρ.

We now specify high-level conditions under which the asymptotic null distribution of the sup LR, Wald, and LM test statistics (defined below) can be determined. Let "$\xrightarrow{d}$" denote convergence in distribution. Let "$\Rightarrow$" denote weak convergence of stochastic processes indexed by $\pi \in \Pi$. Below we consider weak convergence of the process $B_T^{-1} D\ell_T(\theta_0, \pi)$ $(\in R^s)$ indexed by $\pi \in \Pi$ to a process $G(\theta_0, \pi)$. Note that the definition of weak convergence requires the specification of a metric d on the space $\mathcal{Z}$ of R^s-valued functions on Π. We assume d is chosen such that (i) the function

$$(2.5) \qquad G(\cdot) \to \sup_{\pi \in \Pi} (HG(\pi))' \left[H \mathcal{I}^{-1}(\theta_0, \pi) H' \right]^{-1} HG(\pi)$$

is continuous at each function $G \in \mathcal{Z}$ that is continuous on Π, where $H = [I_p \,\vdots\, 0] \in R^{p \times s}$, and (ii) if $g_n \in \mathcal{Z}$ $\forall n \geq 0$, g_0 is continuous on Π, and $d(g_n, g_0) \to 0$ as $n \to \infty$, then

$$(2.6) \qquad \sup_{\pi \in \Pi} |g_n(\pi) - g_0(\pi)| \to 0 \quad \text{as} \quad n \to \infty .$$

These conditions hold, for example, if the uniform metric is used, as in Pollard (1984), or if the Skorohod metric is used in the case when $\Pi \subset [0, 1]$ or $\Pi \subset [0, 1]^r$, as in Billingsley (1968).

We assume that the normalized score function satisfies:

ASSUMPTION 5: $B_T^{-1} D\ell_T(\theta_0, \cdot) \Rightarrow G(\theta_0, \cdot)$ *under* θ_0 *(as processes indexed by* $\pi \in \Pi$*) for some mean zero* R^s*-valued Gaussian stochastic process* $\{G(\theta_0, \pi) : \pi \in \Pi\}$ *that has* $EG(\theta_0, \pi)G(\theta_0, \pi)' = \mathcal{I}(\theta_0, \pi)$ $\forall \pi \in \Pi$ *and has continuous sample paths (as functions of* π *for fixed* θ_0*) with probability one.*

In applications, Assumption 5 is verified by applying a functional CLT. Assumptions 1-3 and 5 above are the same as in AP.

2.3. Specification of Weight Functions

The admissibility result given below is stated in terms of *weighted average power*. That is, we show that for certain weight functions the sup LR, Wald, and LM tests have greater weighted average power than any other *asymptotically distinct* test. To achieve this, a weight function $J(\cdot)$ needs to be specified for the parameter $\pi \in \Pi$. Given π, a weight function $Q_{r,\pi}(\cdot)$ needs to be specified for the perturbation vector h that appears in the local alternative density $f_T(\theta_0 + B_T^{-1}h, \pi)$.

Let $S(\pi, \varepsilon)$ denote the open sphere in Π centered at π with radius $\varepsilon > 0$. Of the weight function $J(\cdot)$, we only assume:

ASSUMPTION 6: *$J(\cdot)$ is a probability measure on Π for which* $\inf_{\pi \in \Pi} J(S(\pi, \varepsilon)) > 0 \quad \forall \varepsilon > 0$.

If Π is separable (and satisfies Assumption 4), then Assumption 6 holds provided the support of J is Π.

The weight functions $\{Q_{r,\pi} : \pi \in \Pi\}$ for $r > 0$ are taken to be ellipses of radius proportional to r. The ellipses are the same as those considered by Wald (1943) for a single fixed π.

ASSUMPTION 7: *$Q_{r,\pi}$ is the distribution of $rA_\pi(A_\pi' \mathcal{I}_\pi A_\pi)^{-1/2}X$, where $X \sim U_p$, U_p is the uniform distribution on the unit sphere in R^p,* $\mathcal{I}_\pi = \mathcal{I}(\theta_0, \pi) = \begin{bmatrix} \mathcal{I}_{1\pi} & \mathcal{I}_{2\pi} \\ \mathcal{I}_{2\pi}' & \mathcal{I}_{3\pi} \end{bmatrix}$, and $A_\pi = \begin{bmatrix} I_p \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}' \end{bmatrix}$.

2.4. Definition of the sup LR, Wald, and LM Test Statistics

For known $\pi \in \Pi$, the standard LR, Wald, and LM test statistics for testing H_0 against H_1 (as defined in (2.1) and (2.2)) are given by

$$LR_T(\pi) = -2(\ell_T(\tilde{\theta}, \pi) - \ell_T(\hat{\theta}(\pi), \pi)) ,$$

$$W_T(\pi) = (HB_T\hat{\theta}(\pi))'\left[H\mathcal{I}_T^{-1}(\hat{\theta}(\pi), \pi)H'\right]^{-1}HB_T\hat{\theta}(\pi) , \quad \text{and} \tag{2.7}$$

$$LM_T(\pi) = \left[B_T^{-1}D\ell_T(\tilde{\theta}, \pi)\right]'\mathcal{I}_T^{-1}(\tilde{\theta}, \pi)B_T^{-1}D\ell_T(\tilde{\theta}, \pi) , \quad \text{where}$$

$$H = [I_p \; \vdots \; 0] \in R^{p \times s} \quad \text{and} \quad \mathcal{I}_T(\theta, \pi) = -B_T^{-1}D^2\ell_T(\theta, \pi)B_T^{-1} .$$

Alternatively, one can define $\mathcal{I}_T(\theta, \pi)$ to be of outer product, rather than Hessian, form.

The sup LR, Wald, and LM test statistics are now defined as

(2.8) $$\sup_{\pi\in\Pi} LR_T(\pi)\ , \quad \sup_{\pi\in\Pi} W_T(\pi)\ , \quad \text{and} \quad \sup_{\pi\in\Pi} LM_T(\pi)\ .$$

Note that the sup LR test statistic is the standard LR test statistic for the case of unknown π.

Let $\{k_{T\alpha} : T \geq 1\}$ be a sequence of critical values (possibly random) such that the sup LR, Wald, or LM tests $\{\xi_T : T \geq 1\}$ have asymptotic significance level α. That is, $\int \xi_T f_T(\theta_0) d\mu_T \to \alpha$ for all θ_0 that satisfy the null hypothesis, where

(2.9) $$\xi_T = 1(\sup_{\pi\in\Pi} LR_T(\pi) > k_{T\alpha})$$

or where ξ_T is defined analogously with $LR_T(\pi)$ replaced by $W_T(\pi)$ or $LM_T(\pi)$.

Under Assumptions 1-5, the asymptotic null distribution of $\sup_{\pi\in\Pi} LR_T(\pi)$, $\sup_{\pi\in\Pi} W_T(\pi)$, and $\sup_{\pi\in\Pi} LM_T(\pi)$ is that of

(2.10) $$\sup_{\pi\in\Pi}(HG(\theta_0, \pi))'(H\mathcal{I}^{-1}(\theta_0, \pi)H')^{-1}HG(\theta_0, \pi)\ .$$

This is proved by an argument analogous to that used to prove Theorem 1 of AP.

2.5. *Asymptotic Admissibility*

Let φ_T denote a test of H_0. That is, φ_T is a [0, 1]-valued function that is determined by $\underline{Y}_T$ (and perhaps some randomization scheme) that rejects H_0 with probability γ when $\varphi_T = \gamma$. The power of φ_T against the local alternative $f_T(\theta_0 + B_T^{-1}h, \pi)$ is denoted $\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi) d\mu_T$.

DEFINITION: A sequence of tests $\{\varphi_T : T \geq 1\}$ is *asymptotically distinct* from the sup LR, Wald, or LM tests $\{\xi_T : T \geq 1\}$ if

(2.11) $$\delta = \lim_{T\to\infty} \int (1 - \varphi_T)\xi_T f_T(\theta_0) d\mu_T > 0\ .$$

Note that $\int (1 - \varphi_T)\xi_T f_T(\theta_0) d\mu_T$ is just the null probability that the test φ_T accepts H_0 and the sup test ξ_T rejects H_0. If two tests are not equal almost surely (under θ_0) and are not nested, then this probability is positive.

The sup LR, Wald, and LM tests are asymptotically equivalent under the null and local alternatives under Assumptions 1-5, see AP. In consequence, if a sequence of tests is asymptotically distinct from any one of the three, it is asymptotically distinct from all three.

The main result of this paper is the following admissibility result.

THEOREM 1: *Suppose Assumptions 1-7 hold and $\{\varphi_T : T \geq 1\}$ is a sequence of tests that is asymptotically distinct from a sequence of asymptotically level α supremum* LR, *Wald, or* LM *tests $\{\xi_T : T \geq 1\}$. Then, there exists an $r_0 < \infty$ such that for all $r \geq r_0$,*

$$\overline{\lim_{T\to\infty}} \int\left[\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi) d\mu_T\right] dQ_{r,\pi}(h) dJ(\pi) < \underline{\lim}_{T\to\infty} \int\left[\int \xi_T f_T(\theta_0 + B_T^{-1}h, \pi) d\mu_T\right] dQ_{r,\pi}(h) dJ(\pi) .$$

(In addition, the $\underline{\lim}_{T\to\infty}$ on the right-hand side equals $\lim_{T\to\infty}$.)

REMARKS: 1. Theorem 1 shows the sup LR, Wald, and LM tests are best tests against alternatives that are sufficiently distant from the null.

2. The weighted average power of a test can exceed that of another test only if its power at some (h, π) exceeds that of the other test. In consequence, Theorem 1 implies that there exist sequences $\{(h_T, \pi_T): T \geq 1\}$ such that ξ_T has higher power asymptotically against $\{f_T(\theta_0 + B_T^{-1}h_T, \pi_T) : T \geq 1\}$ than any other sequence of tests. Thus, ξ_T is asymptotically admissible.

3. The proof of Theorem 1 actually shows that the ratio of the asymptotic (as $T \to \infty$) weighted average type II error of φ_T (with respect to $(Q_{r,\pi}, J)$) over that of ξ_T converges to infinity as $r \to \infty$.

4. Theorem 1 holds for all weight functions J that satisfy Assumption 6. Thus, the optimal performance of the sup tests against distant alternatives holds for a wide variety of weight functions J.

5. Theorem 1 holds for *any* sequence of asymptotically distinct tests $\{\varphi_T : T \geq 1\}$ -- it need not be a sequence of tests of asymptotic significance level α. Thus, for certain alternatives the only way to increase the asymptotic power of a sequence of sup tests is to enlarge its critical regions.

6. Assumption 3 is not required in Theorem 1 for the case of the sup Wald test.

3. EXAMPLES

For a number of examples, this section provides primitive sufficient conditions for Assumptions 1-3 and 5 of Section 2.

3.1. Changepoint Tests

In this subsection we consider changepoint tests. The tests are designed to detect a one-time change in the value of a parameter (but they have power against more general forms of change, e.g., see Andrews (1993b, Thm. 5 and Cor. 2) and Ploberger, Kramer and Kontrus (1989, Cor. 1)). The models we consider are stationary dynamic nonlinear models. A simple example where the results of this section can be applied is in a test for constancy of the intercept in an AR or ARMA model for the growth rate of a macroeconomic variable such as GNP. Tests of this sort have attracted some attention in the literature, e.g., see Perron (1991) and Bai, Lumsdaine, and Stock (1991). Asymptotic critical values for the LR tests considered in this section are tabulated in Andrews (1993b).

The sample of observations is given by

$$(3.1) \qquad \underline{Y}_T = \{(Y_t, X_t) : t \leq T\} .$$

$\{X_t : t \leq T\}$ are weakly exogenous variables (i.e., regressor-like variables, see below). We consider the relatively simple case where the data are strictly stationary, ergodic, and Markov under the null hypothesis. In particular, we suppose that $\{(Y_t, X_t) : t \leq T\}$ is part of a doubly infinite strictly stationary ergodic sequence $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$ and $\{Y_t : t = ..., 0, 1, ...\}$ is m-th order Markov for some integer $m \geq 0$. By definition, $\{Y_t : t = ..., 0, 1, ...\}$ is m-th order Markov if the conditional distribution of Y_t given $\mathcal{F}_{t-1} = \sigma(..., Y_{t-2}, Y_{t-1}; ..., X_{t-1}, X_t)$ equals the conditional distribution of Y_t given $Y_{t,m} = (Y_{t-m}, ..., Y_{t-1})$ and $X_{t,m} = (X_{t-m}, ..., X_t)$ for all t.

Let

$$(3.2) \qquad \{g_t(\delta_1, \delta_2) : \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\} = \{g_t(Y_t | Y_{t,m}, X_{t,m}; \delta_1, \delta_2) : \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\}$$

denote a parametric family of conditional densities (with respect to some measure) of Y_t given $Y_{t,m}$ and $X_{t,m}$ evaluated at the random variables Y_t, $Y_{t,m}$, and $X_{t,m}$, where $\Delta_1 \subset R^p$, $\Delta_2 \subset R^{q-p}$, and $p \leq q$. Let

(3.3) $$h_t = h_t(X_t | Y_1, ..., Y_{t-1}, X_1, ..., X_{t-1})$$

denote the conditional density (with respect to some measure) of X_t given $Y_1, ..., Y_{t-1}, X_1, ..., X_{t-1}$ evaluated at the random variables $Y_1, ..., Y_{t-1}, X_1, ..., X_t$. By the definition of *weak exogeneity*, h_t does not depend on $\delta = (\delta_1', \delta_2')'$.

Let $\Pi \subset (0, 1)$ and let $\pi \in \Pi$. Suppose the parameter vector equals (δ_1, δ_2) for the observations $t = 1, ..., [T\pi]$ and $(\delta_1 + \beta, \delta_2)$ for the observations $t = [T\pi]+1, ..., T$, where $\beta \in B \subset R^p$ and $[\cdot]$ denotes the integer part of $\cdot$. Then, π is the changepoint and $\theta = (\beta', \delta')'$ for $\delta = (\delta_1', \delta_2')'$ contains the pre- and post-change parameter values. Note that the parameter δ_{20} is constant across the whole sample under H_0 and H_1. Of course, no such parameter δ_{20} need appear in the model.

In the present case, the likelihood function is given by

(3.4) $$f_T(\theta, \pi) = \left(\prod_{t=1}^{[T\pi]} g_t(\delta_1, \delta_2)\right)\left(\prod_{t=[T\pi]+1}^{T} g_t(\delta_1 + \beta, \delta_2)\right)\left(\prod_{t=1}^{T} h_t\right).$$

The norming matrix B_T of Section 2 is taken to be $\sqrt{T} I_s$.

The Markov assumption yields the simplification that under the null hypothesis the summands $\log g_t(\delta_1, \delta_2)$ in the log-likelihood function are strictly stationary and ergodic for $t > m$.

The following assumption is sufficient for Assumptions 1-3 and 5 of Section 2. This is proved in Theorem 3 of AP. All expectations E below are taken under θ_0.

ASSUMPTION CP: (a) Π *has closure contained in* $(0, 1)$.

(b) Θ *is compact and* θ_0 *lies in the interior of* Θ.

(c) *Under* θ_0, $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$ *is strictly stationary and ergodic,* $\{Y_t : t = ..., 0, 1, ...\}$ *is* m-th *order Markov, and* $\{X_t : t = ..., 0, 1, ...\}$ *is weakly exogenous.*

(d) $g_t(\delta_1, \delta_2)$ *is continuous in* (δ_1, δ_2) *on* $\Delta_{10} \times \Delta_{20}$ *with probability one under* θ_0 *and twice continuously partially differentiable in* (δ_1, δ_2) *on* $\Delta_{10} \times \Delta_{20}$ *with probability one under* θ_0, *where* Δ_{10} *and* Δ_{20} *are compact neighborhoods of* δ_{10} *and* δ_{20} *respectively.*

(e) $g_t(\delta_1, \delta_2) \neq g_t(\delta_{10}, \delta_{20})$ *with positive probability under* θ_0 $\forall(\delta_1, \delta_2) \in \Delta_1 \times \Delta_2$ *such that* $(\delta_1, \delta_2) \neq (\delta_{10}, \delta_{20})$.

(f) $E \sup_{\delta_1 \in \Delta_1, \delta_2 \in \Delta_2} |\log g_t(\delta_1, \delta_2)| < \infty$, $E \sup_{\delta_1 \in \Delta_{10}, \delta_2 \in \Delta_{20}} \left\| \frac{\partial}{\partial(\delta_1', \delta_2')} \log g_t(\delta_1, \delta_2) \right\| < \infty$,

$E \left\| \frac{\partial}{\partial(\delta_1', \delta_2')} \log g_t(\delta_{10}, \delta_{20}) \right\|^2 < \infty$, *and* $E \sup_{\delta_1 \in \Delta_{10}, \delta_2 \in \Delta_{20}} \left\| \frac{\partial^2}{\partial(\delta_1', \delta_2')'\partial(\delta_1', \delta_2')} \log g_t(\delta_1, \delta_2) \right\| < \infty$.

(g) $\mathcal{J} = -E \frac{\partial^2}{\partial(\delta_1', \delta_2')'\partial(\delta_1', \delta_2')} \log g_t(\delta_{10}, \delta_{20})$ *is positive definite.*

Note that Assumption CP is quite similar to standard assumptions in the literature for the consistency and asymptotic normality of ML estimators in stationary contexts. The results could be extended to some nonstationary contexts. In addition, the example could be extended to cover multiple changepoints.

3.2. *Empirical Process Examples*

This subsection provides primitive sufficient conditions for Assumptions 1-3 and 5 for empirical process examples. Such examples are ones in which the log likelihood can be written as the sum of terms of the form $\log g(W_t, \theta, \pi)$, plus a term that does not depend on (θ, π), where W_t is a random variable and g is a fixed function. We now provide several examples.

EXAMPLE 1 (Cross-sectional Constancy): In this example, the observations are iid and the unknown parameter π partitions the sample space of some observed variable(s) into $m+1$ regions. In one region the model is indexed by the parameter $(\delta_1', \delta_2')'$ and in other regions it is indexed by $(\delta_1' + \beta_j', \delta_2')'$ for $j \leq m$. In this case, $\theta = (\beta', \delta')'$ for $\beta = (\beta_1', ..., \beta_m')'$ and $\delta = (\delta_1', \delta_2')'$. In this model, a test of cross-sectional constancy of the parameters corresponds to a test of the null hypothesis $H_0 : \beta = 0$.

To be concrete, consider the following special case given by a linear regression model with two regions:

$$(3.5) \qquad Y_t = \begin{cases} X_t'\delta_1 + U_t & \text{for } Z_t \leq \pi \\ X_t'(\delta_1 + \beta) + U_t & \text{for } Z_t > \pi \end{cases} \quad \text{for } t = 1, ..., T,$$

where $\{(Y_t, X_t, Z_t, U_t) : t = 1, ..., T\}$ are iid, (X_t, Z_t) and U_t are independent; U_t is an unobserved

$N(0, \delta_2)$ error; Y_t is an observed scalar random variable; X_t is an observed random p-vector with $EX_t'X_t < \infty$; Z_t is an observed scalar random variable that may be an element of X_t; Z_t has bounded density with respect to Lebesgue measure on the intersection of its support and Π; $\inf_{\pi \in \Pi} \lambda_{\min}\left(E\begin{pmatrix} X_t 1(Z_t > \pi) \\ X_t \end{pmatrix}\begin{pmatrix} X_t 1(Z_t > \pi) \\ X_t \end{pmatrix}'\right) > 0$; the parameter $\theta = (\beta', \delta_1', \delta_2)'$ lies in a compact set $\Theta \subset R^{2p+1}$ that excludes δ_2 values ≤ 0; the parameter π lies in a compact set $\Pi \subset R$; and the true parameter θ_0 lies in the interior of Θ under H_0.

EXAMPLE 2 (Threshold Autoregression): This example generalizes Example 1 to time series contexts in which the variable (or vector) Z_t is often given by a lagged value(s) of a dependent variable. In particular, consider the simple threshold autoregressive model defined by (3.5) with $X_t = (1, Y_{t-1})'$, $Z_t = Y_{t-d}$ for some integer $d > 0$, $\{U_t : t = 1, ..., T\}$ are iid, (Y_0, Y_{1-d}) have distributions that correspond to a stationary start-up of the AR model when $\beta = 0$, and Θ and Π are as defined above with $p = 2$ and $|\delta_1| < 1$. (In this case, the assumptions of Example 1 on X_t and Z_t automatically hold.) Models of this sort have been applied in the physical and biological sciences, e.g., see Tong (1990), as well as in economics, e.g., see Potter (1989). Typically, it is of interest with these models to test for the existence of a threshold effect, which corresponds to testing the null $H_0 : \beta = 0$.

EXAMPLE 3 (Variable Relevance): This example considers tests of variable relevance in nonlinear models. For specificity, consider a nonlinear regression model

$$(3.6) \qquad Y_t = g(X_t, \delta_1) + \beta h(Z_t, \pi) + U_t \quad \text{for} \quad t = 1, ..., T ,$$

where $\{(Y_t, X_t, Z_t, U_t) : t = 1, ..., T\}$ are iid; (X_t, Z_t) and U_t are independent; U_t is an unobserved $N(0, \delta_2)$ error; Y_t is an observed scalar random variable; X_t and Z_t are observed random vectors; g and h are known functions; β is a scalar parameter; π is an R^b-valued parameter; $\theta = (\beta, \delta_1', \delta_2)'$ and π lie in compact sets Θ and Π respectively; Θ excludes δ_2 values ≤ 0; the true parameter θ_0 lies in the interior of Θ under H_0; $g(X_t, \delta_1)$ is two times continuously differentiable in δ_1 $\forall \theta \in \Theta_0$ with probability one under θ_0, where Θ_0 is some neighborhood of θ_0; $h(Z_t, \pi)$ is differentiable in π with probability one under θ_0 $\forall \pi \in \Pi$; $E \sup_{\theta \in \Theta} g^2(X_t, \delta_1) < \infty$;

$E \sup_{\pi\in\Pi} h^2(Z_t, \pi)\log^+(|h(Z_t, \pi)|) < \infty$, where $\log^+(x) = \max\{\log(x), 0\}$ for $x \geq 0$; $E \sup_{\theta\in\Theta_0}\left\|\frac{\partial}{\partial\delta_1}g(X_t, \delta_1)\right\|^2 < \infty$; $E \sup_{\theta\in\Theta_0}\left\|\frac{\partial^2}{\partial\delta_1\delta_1'}g(X_t, \delta_1)\right\|^2 < \infty$; $E \sup_{\pi\in\Pi}\left\|\frac{\partial}{\partial\pi}h(Z_t, \pi)\right\|^r < \infty$ for some $r > 2$; $\inf_{\pi\in\Pi} \lambda_{\min}\left(E\begin{pmatrix} h(Z_t, \pi) \\ \frac{\partial}{\partial\delta_1}g(X_t, \delta_{10})\end{pmatrix}\begin{pmatrix} h(Z_t, \pi) \\ \frac{\partial}{\partial\delta_1}g(X_t, \delta_{10})\end{pmatrix}'\right) > 0$; and $E(g(X_t, \delta_1) - g(X_t, \delta_{10}) + \beta h(Z_t, \pi))^2 > 0 \ \forall\theta \in \Theta$ with $\theta \neq \theta_0$.

For example, $h(Z_t, \pi)$ might be of the Box-Cox form $(Z_t^\pi - 1)/\pi$. A test for the relevance of the regressors Z_t is a test of the null hypothesis $H_0 : \beta = 0$. Under H_0, the parameter π is no longer present.

We note that the results given below cover the case where π is infinite dimensional, but the resulting test statistic and critical values may be difficult to compute in this case, so we have focussed on the finite dimensional case above.

EXAMPLE 4 (Functional Form): This example consists of tests of functional form for nonlinear models. The model set-up is the same as in Example 3 except that the variables that are being tested for relevance in Example 3 are variables that are already in the model in the present Example. For example, for the nonlinear regression model (3.6), Z_t is taken to be a sub-vector of X_t. In this case, the nonlinear regression function depends on the same variables under the null and alternative hypotheses, but is of a more complicated form under the alternative. Neural network tests of functional form and some consistent tests of model specification are designed for this testing problem.

We now introduce the requisite definitions and assumptions used for the empirical process examples. For simplicity, we consider stationary random variables. The results could be extended to cover non-identically distributed random variables. The data are given by $\{(Y_t, X_t) : t = 1, ..., T\}$, which are part of a strictly stationary, absolutely regular process $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$, where $\{Y_t\}$ is an m-th order Markov sequence of random variables and $\{X_t\}$ is a sequence of weakly exogenous variables (both as defined in Section 3.1).

By definition, a sequence $\{W_t : t = ..., 0, 1, ...\}$ is *absolutely regular* (β-mixing) if $\beta(s) \to 0$ as

$s \to \infty$, where $\beta(s)$ is defined as follows. For any two σ-fields $\mathcal{A}$ and $\mathcal{B}$, define

$$(3.7) \qquad \beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2}\sup \sum_{(i,j)\in I\times J} |P(A_i \cap B_j) - P(A_i)P(B_j)| \ ,$$

where the supremum is taken over all finite partitions of the sample space $\{A_i : i \in I\}$ and $\{B_j : j \in J\}$ that are $\mathcal{A}$ and $\mathcal{B}$ measurable respectively. Let $\mathcal{F}_t = \sigma(..., W_{t-1}, W_t)$ and $\mathcal{F}^t = \sigma(W_t, W_{t+1}, ...)$, where $\sigma(\cdot)$ denotes a σ-field. Then,

$$(3.8) \qquad \beta(s) = \sup_t \ \beta(\mathcal{F}_t, \mathcal{F}^{t+s}) \ .$$

Absolute regularity is stronger than strong mixing (α-mixing), but weaker than φ-mixing. Examples of absolutely regular processes are given by Davydov (1973), Mokkadem (1986, 1990), and Doukhan (1992). They include, under suitable conditions, finite state space Harris recurrent Markov chains, vector autoregressive moving average processes, bilinear processes, and nonlinear autoregressive processes, among others. In particular, the AR process of Example 2 under H_0 is absolutely regular with $\beta(s) = O(\rho^s)$ for $0 < \rho < 1$.

Let $W_t = (Y'_{t-m}, ..., Y_t', X'_{t-m}, ..., X_t')'$. Let

$$(3.9) \qquad g(W_t, \theta, \pi) = g(Y_t|Y_{t-m}, ..., Y_{t-1}, X_{t-m}, ..., X_t; \theta, \pi)$$

for $\theta \in \Theta$ and $\pi \in \Pi$ denote a parametric family of conditional densities (with respect to some measure) of Y_t given $Y_1, ..., Y_{t-1}, X_1, ..., X_t$ evaluated at the random variables $Y_1, ..., Y_t, X_1, ..., X_t$. By the m-th order Markov property, the above conditional density is a function only of W_t and not of all the data prior to (Y_t, X_t). Let

$$(3.10) \qquad h_t = h(X_t|Y_1, ..., Y_{t-1}, X_1, ..., X_{t-1})$$

denote the conditional density (with respect to some measure) of X_t given $Y_1, ..., Y_{t-1}, X_1, ..., X_{t-1}$ evaluated at the random variables $Y_1, ..., Y_{t-1}, X_1, ..., X_t$. By the assumption of weak exogeneity, h_t does not depend on θ or π.

The parameter space Θ is a subset of R^s. In the primary applications of interest, the parameter space Π also is a subset of Euclidean space. In such cases, $\Theta \times \Pi$ and $\Theta_0 \times \Pi$, where Θ_0 is some neighborhood of the true null parameter θ_0, are metric spaces with the Euclidean metric.

For generality, however, we allow Π to be infinite dimensional in the assumptions below. In particular, we allow for the case where $\Theta \times \Pi$ and $\Theta_0 \times \Pi$ are pseudo-metric spaces with some pseudo-metrics d^* and d_0^* respectively.

The likelihood and log-likelihood functions of the sample are

$$(3.11) \qquad f_T(\theta, \pi) = \prod_{t=1}^{T} g(W_t, \theta, \pi)\cdot\prod_{t=1}^{T} h_t \quad \text{and} \quad \ell_T(\theta, \pi) = \sum_{t=1}^{T} \log g(W_t, \theta, \pi) + \sum_{t=1}^{T} h_t \, .$$

The information matrix for θ given π is defined to be

$$(3.12) \qquad \mathcal{I}(\theta, \pi) = -E\frac{\partial^2}{\partial\theta\partial\theta'}\log g(W_t, \theta, \pi) \, .$$

To obtain the uniform weak laws of large numbers that are needed to verify parts of Assumptions 1-3, we use the concept of L^r-continuity, which we now define. Let $(\mathcal{T}, d)$ be a pseudo-metric space and let $f(W_t, \tau)$ be a vector-valued function of $\tau \in \mathcal{T}$. We say that f is *L^r-continuous at* τ_0 if

$$(3.13) \qquad E \sup_{\tau\in\mathcal{T}:d(\tau,\tau_0)<\delta} \|f(W_t, \tau) - f(W_t, \tau_0)\|^r \to 0 \quad \text{as} \quad \delta \to 0 \, ,$$

where $\|\cdot\|$ is the Euclidean norm. We say that f is *L^r-continuous at* τ_0 *with modulus of continuity* $c(\delta)$ if the left-hand side of (3.13) is $\leq c(\delta)$ $\forall\delta$ small and $c(\delta) \to 0$ as $\delta \to 0$.

Of course, L^r-continuity is implied by almost sure pointwise continuity (viz., $f(W_t, \tau) \to f(W_t, \tau_0)$ as $\tau \to \tau_0$ a.s.) plus a moment condition (viz., $E \sup_{\tau\in\mathcal{T}:d(\tau,\tau_0)<\delta} \|f(W_t, \tau) - f(W_t, \tau_0)\|^r < \infty$ for some $\delta > 0$) by the dominated convergence theorem. We will require the log of the conditional density $g(W_t, \theta, \pi)$ and its second derivative with respect to θ to be L^1-continuous in (θ, π). This holds in each of the examples above.

To obtain the weak convergence property of Assumption 5, we use a bracketing empirical process central limit theorem (CLT) of Doukhan, Massart, and Rio (1992). The latter is a generalization to strictly stationary absolutely regular processes of an empirical process CLT of Ossiander (1987) for iid processes. The empirical process CLT relies on a bracketing cover number condition. We define the cover numbers here. Below we state the aforementioned condition and give primitive sufficient conditions for it.

Let $\mathcal{W}$ include the sample space of W_t. Let $\mathcal{F}$ denote a class of real functions on $\mathcal{W}$. For any $\varepsilon > 0$ and $r \in [2, \infty]$, the *L^r bracketing cover number* $N_r^B(\varepsilon, \mathcal{F})$ of $\mathcal{F}$ is the smallest value of n for which there exist real functions $a_1, ..., a_n$ and $b_1, ..., b_n$ on $\mathcal{W}$ such that for each $f \in \mathcal{F}$ $|f - a_j| \leq b_j$ for some $j \leq n$ and $\max_{j \leq n}(Eb_j^r(W_t))^{1/r} \leq \varepsilon$. By convention, if $r = \infty$, $(Eb_j^r(W_t))^{1/r} = \sup_{w \in \mathcal{W}}|b_j(w)|$. The bracketing cover number $N_r^B(\varepsilon, \mathcal{F})$ of a class of vector-valued functions is equal to the maximum of the element by element bracketing cover numbers of the functions in $\mathcal{F}$.

Throughout this section, we assume $g(\cdot, \theta, \pi)$, $\frac{\partial}{\partial\theta}\log g(\cdot, \theta, \pi)$ and $\frac{\partial^2}{\partial\theta\partial\theta'}\log g(\cdot, \theta, \pi)$ are Borel measurable functions $\forall\theta \in \Theta$ $\forall\pi \in \Pi$, as are their element by element suprema and infima over all balls in $\Theta \times \Pi$ of small radius. Below, C denotes a generic positive finite constant.

The following Assumptions EP1-EP4 are sufficient for Assumptions 1-3 and 5 of Section 2. All expectations E below are taken under θ_0.

ASSUMPTION EP1: (a) *Under θ_0, $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$ is a strictly stationary absolutely regular sequence of random variables with $\sum_{s=1}^{\infty} s^{2/(r-2)}\beta(s) < \infty$ for some constant $r > 2$, $\{Y_t : t = ..., 0, 1, ...\}$ is m-th order Markov, and $\{X_t : t = ..., 0, 1, ...\}$ is weakly exogenous.*

(b) *$g(W_t, \theta, \pi)$ does not depend on π for θ in the null hypothesis.*

(c) *The true parameter θ_0 is in the interior of Θ.*

(d) *$g(W_t, \theta, \pi)$ is twice continuously partially differentiable in θ for all $\theta \in \Theta_0$ and $\pi \in \Pi$ with probability one under θ_0, where Θ_0 is some neighborhood of θ_0.*

(e) $E \sup_{\theta\in\Theta,\pi\in\Pi}|\log g(W_t, \theta, \pi)| < \infty$, *$E \sup_{\pi\in\Pi}\left\|\frac{\partial}{\partial\theta}\log g(W_t, \theta_0, \pi)\right\|^r < \infty$ for r as in part* (a), *and $E \sup_{\theta\in\Theta_0,\pi\in\Pi}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}\log g(W_t, \theta, \pi)\right\| < \infty$. If $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$ is a sequence of independent or m-dependent random variables for some $m < \infty$, then r can be taken to equal 2 here and in Assumptions* EP4 *and* EP4* *below. If $\{(Y_t, X_t) : t = ..., 0, 1, ...\}$ has geometrically declining β-mixing numbers (i.e., $\beta(s) = O(\rho^s)$ for some $0 < \rho < 1$), then $E \sup_{\pi\in\Pi}\left\|\frac{\partial}{\partial\theta}\log g(W_t, \theta_0, \pi)\right\|^r < \infty$ can be replaced by $E \sup_{\pi\in\Pi}\left\|\frac{\partial}{\partial\theta}\log g(W_t, \theta, \pi)\right\|^2 \log^+\left(\left\|\frac{\partial}{\partial\theta}\log g(W_t, \theta_0, \pi)\right\|\right) < \infty$ and r can be taken to be any number > 2 in Assumptions* EP4 *and* EP4* *below.*

(f) $\mathcal{I}(\theta, \pi)$ *is uniformly continuous in* (θ, π) *over* $\Theta_0 \times \Pi$.

(g) $\mathcal{I}(\theta_0, \pi)$ *is uniformly positive definite over* $\pi \in \Pi$.

ASSUMPTION EP2: (a) $\log g(W_t, \theta, \pi)$ *is* L^1*-continuous in* (θ, π) *on* $\Theta \times \Pi$ *under* θ_0 *and* $\Theta \times \Pi$ *is totally bounded under some pseudo-metric* d^*.

(b) $\frac{\partial^2}{\partial\theta\partial\theta'}\log g(W_t, \theta, \pi)$ *is* L^1*-continuous in* (θ, π) *on* $\Theta_0 \times \Pi$ *under* θ_0 *and* $\Theta_0 \times \Pi$ *is totally bounded under some pseudo-metric* d_0^*.

ASSUMPTION EP3: *For all neighborhoods* Θ_0 *of* θ_0,

$\sup_{\pi \in \Pi} \sup_{\theta \in \Theta/\Theta_0}(E \log g(W_t, \theta, \pi) - E \log g(W_t, \theta_0, \pi)) < 0$.

ASSUMPTION EP4: *The* L^r*-bracketing cover numbers* $N_r^B(\varepsilon, \mathcal{M})$ *of the class of functions* $\mathcal{M} = \left\{\frac{\partial}{\partial\theta}\log g(\cdot, \theta_0, \pi) : \pi \in \Pi\right\}$ *satisfy* $\int_0^1\left[\log N_r^B(\varepsilon, \mathcal{M})\right]^{1/2} d\varepsilon < \infty$ *for* r *as in Assumption* EP1(a) *and* (e).

We now discuss Assumptions EP1-EP4. First, we note that they are satisfied in Examples 1-4 above. Assumption EP1 is similar to Assumption 1 except that it does not contain the high-level Assumption 1(d) and it adds temporal dependence and finite moment assumptions instead. In Example 2, Assumption EP1(a) holds with geometrically declining β-mixing numbers by Mokkadem (1986). Assumption EP2 is a fairly standard ML regularity condition, which requires L^1-continuity of the log of the conditional densities and their second derivatives.

Assumption EP3 is an identification condition that is used in the proof of the consistency under H_0 of the ML and restricted ML estimators. If $\Theta \times \Pi$ is compact, then more primitive conditions can be considered, which together with EP2(a) are sufficient for EP3:

ASSUMPTION EP3*: (a) $g(W_t, \theta, \pi) \neq g(W_t, \theta_0, \pi)$ *with positive probability under* θ_0 $\forall\theta \in \Theta$ *with* $\theta \neq \theta_0$ *and* $\forall\pi \in \Pi$.

(b) $\Theta \times \Pi$ *is compact (with respect to the metric* d^* *in Assumption* EP2(a)).

Assumption EP3* holds in Examples 1-4. It can be verified by showing that $E(\log g(W_t, \theta_0, \pi) - \log g(W_t, \theta, \pi)) > 0$ $\forall\theta \in \Theta$ with $\theta \neq \theta_0$ $\forall\pi \in \Pi$. Note that if $g(X_t, \delta_1) = X_t'\delta_1$ in

Examples 3 or 4, then the final condition stated in Example 3 can be omitted, since it is implied by the $\lambda_{\min}(\cdot)$ condition of Example 3.

Assumption EP4 is an L^r-bracketing cover number condition that requires that the class of functions $\mathcal{M}$ is not too complex/large. It is satisfied if the cover numbers $N_r^B(\varepsilon, \mathcal{M})$ do not increase too quickly as $\varepsilon \downarrow 0$. Primitive sufficient conditions for EP4 are given in Andrews (1993a, Thms. 5 and 6). In particular, classes of functions of types II–VI, defined in Andrews (1993a), satisfy EP4. These classes include infinite dimensional classes of smooth functions (of W_t) and finite dimensional classes of functions that satisfy some continuity properties as functions of a Euclidean-valued parameter π. In addition, results are given in Andrews (1993a, Thm. 6) that show that functions in classes of types II–VI can be "mixed and matched" in various ways to yield new classes that satisfy EP4.

For many applications π is a finite dimensional parameter and the following primitive sufficient conditions for EP4 are satisfied.

ASSUMPTION EP4*: *Π is a bounded subset of Euclidean space and* $\frac{\partial}{\partial\theta}\log g(W_t, \theta_0, \pi)$ *is L^r-continuous in π on Π with modulus of continuity $C\delta^{\psi}$ for some positive constants C and ψ, where r is as in Assumption* EP1(a) *and* (e).

In Example 1, the assumption that Z_t has a bounded Lebesgue density is used to verify Assumption EP4* with $\psi = 1$. In Example 2, this assumption on Z_t holds automatically, since Z_t has a normal distribution. In Example 3, the assumptions that $h(Z_t, \pi)$ is differentiable in π and its derivative satisfies a moment condition are used to verify Assumption EP4* with $\psi = 1$.

The results referred to above are summarized as follows:

THEOREM 2: (a) *Assumptions* EP1–EP4 *imply Assumptions* 1–3 *and* 5.

(b) *Assumption* EP2(a) *and* EP3* *imply Assumption* EP3.

(c) *Assumption* EP4* *implies Assumption* EP4.

Note that under Assumptions EP1–EP4, Assumption 5 holds with the Gaussian process $G(\theta_0, \cdot)$ having covariance function given by $EG(\theta_0, \pi_1)G(\theta_0, \pi_2)' = E\frac{\partial}{\partial\theta}\log g(W_t, \theta_0, \pi_1) \times \frac{\partial}{\partial\theta'}\log g(W_t, \theta_0, \pi_2)$ for $\pi_1, \pi_2 \in \Pi$. Continuity (in π) of the sample path of $G(\theta_0, \pi)$ is with respect to the L^r-pseudo-metric, where r is as in Assumptions EP1(a), EP1(e), and EP4 or EP4*. (When the data are independent or m-dependent, continuity is with respect to the L^2-pseudo-metric. When the β-mixing numbers decline geometrically fast, continuity is with respect to the L^r-pseudo-metric for arbitrary $r > 2$.)

4. FINITE SAMPLE ADMISSIBILITY

In this section we show that the LR test is finite sample admissible for a class of testing problems that arise in a Gaussian linear model.

ASSUMPTION 8: *The model is given by*

$$Y_t = X_t(\pi)'\beta + Z_t'\delta + U_t \quad \text{for} \quad t = 1, \ldots, T,$$

where $U_t \sim$ iid $N(0, \sigma^2)$, σ^2 *is known,* $X_t(\pi) \in R^p$, $\beta \in R^p$, $Z_t \in R^q$, $\delta \in R^q$, $\{(X_t(\pi), Z_t) : t = 1, \ldots, T\}$ *are non-random,* $\pi \in \Pi$, $\sum_{t=1}^T \begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}\begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}'$ *is nonsingular for all* $\pi \in \Pi$, *and* $X_t(\pi)$ *is continuous on* Π *for all* $t = 1, \ldots, T$.

Below, the parameter space Π is assumed to satisfy Assumption 4 and the weight functions $J(\cdot)$ and $Q_{r,\pi}(\cdot)$ are assumed to satisfy Assumptions 6 and 7, respectively, with $\mathcal{I}_\pi$ equal to $\sum_{t=1}^T \begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}\begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}'$ in Assumption 7.

The hypotheses of interest are the same as in Section 2 and are specified by (2.1) and (2.2). By varying the definition of $X_t(\pi)$, we obtain hypotheses of different types. For example, if

$$(4.1) \qquad X_t(\pi) = X_t 1(t \le T\pi), \quad Z_t = \begin{pmatrix} X_t \\ X_t^* \end{pmatrix}, \quad \text{and} \quad \Pi \subset \{1/T, 2/T, \ldots, (T-1)/T\},$$

then a test of $H_0 : \beta = 0$ is a test for a single changepoint in a subvector of the regressor vector. This example can be extended straightforwardly to allow for arbitrarily many changepoints.

Another example is a test of relevance of Box-Cox transformed regressors. In this case,

$$(4.2) \qquad X_t(\pi) = \left((X_{1t}^\pi - 1)/\pi, \ldots, (X_{pt}^\pi - 1)/\pi\right)' \quad \text{and} \quad \Pi \subset [0, \infty).$$

Under the null $H_0 : \beta = 0$, the Box-Cox transformed regressors do not belong in the regression model. This example can be extended to allow the Box-Cox parameter to differ across regressors and to allow a more complicated nonlinear transformation than the Box-Cox transformation.

A third example is a test of cross-sectional constancy. In this case,

$$(4.3) \qquad X_t(\pi) = X_t 1(X_{1t} \le \pi) \ , \quad Z_t = \begin{pmatrix} X_t \\ X_t^* \end{pmatrix}, \quad \text{and} \quad \Pi \subset R \ ,$$

where X_{1t} is an element of the regressor X_t. In this example, one is testing for constancy of the regression parameters across two (unknown) regions. The example can be extended to cover multiple regions of more complex form.

For known π, the standard LR, Wald, and LM test statistics for testing H_0 against H_1 are given by (2.7) with $B_T = I_s$, $\hat{\theta}(\pi)$ equal to the unrestricted least squares (LS) estimator of $\theta = (\beta', \delta')'$, $\mathcal{I}_T(\theta, \pi) = \frac{1}{\sigma^2}\sum_{t=1}^{T}\begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}\begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}'$ $\forall \theta \in \Theta = R^{p+q}$, $\tilde{\theta}$ equal to the restricted LS estimator of θ, and $\ell_T(\theta, \pi)$ and $D\ell_T(\theta, \pi)$ equal to the Gaussian regression log likelihood and its vector of derivatives with respect to θ respectively. As is well known, $LR_T(\pi)$, $W_T(\pi)$, and $LM_T(\pi)$ are monotone transformations of each other.

For the case of unknown π, the LR test statistic is

$$(4.4) \qquad \sup_{\pi \in \Pi} LR_T(\pi) \ .$$

By the monotone transform property, the test statistics $\sup_{\pi\in\Pi} W_T(\pi)$ and $\sup_{\pi\in\Pi} LM_T(\pi)$ yield equivalent tests to that based on $\sup_{\pi\in\Pi} LR_T(\pi)$. For convenience in the proof, we focus on the $\sup_{\pi\in\Pi} W_T(\pi)$ version of the LR test.

We say that a test φ_T is *distinct* from the significance level α LR test $\xi_T = 1(\sup_{\pi\in\Pi} W_T(\pi) > k_\alpha)$, where k_α is a positive constant, if

$$(4.5) \qquad \delta = \int(1 - \varphi_T)\xi_T f_T(\theta_0) d\mu_T > 0 \ ,$$

where $f_T(\theta_0)$ is the null Gaussian density and μ_T is Lebesgue measure on R^T. That is, φ_T is distinct from ξ_T if there is positive probability under H_0 that φ_T accepts when ξ_T rejects.

Minor alterations of the proof of asymptotic admissibility of the LR test in Theorem 1 yield finite sample admissibility of the LR test for the Gaussian linear regression model. The following theorem states the result.

THEOREM 3: *Suppose Assumptions* 4 *and* 6-8 *hold with* $\mathcal{I}_\pi = \frac{1}{\sigma^2}\sum_{t=1}^{T}(X_t(\pi)', Z_t')'(X_t(\pi)', Z_t')$ *in Assumption* 7. *Let* φ_T *be a test that is distinct from the level* α LR *test* ξ_T. *Then, there exists an* $r_0 < \infty$ *such that for all* $r \geq r_0$,

$$\int[\int\varphi_T f_T(\theta_0 + h,\ \pi)d\mu_T]dQ_{r,\pi}(h)dJ(\pi) < \int[\int\xi_T f_T(\theta_0 + h,\ \pi)d\mu_T]dQ_{r,\pi}(h)dJ(\pi)\ .$$

REMARK: Remarks 1-5 following Theorem 1 all apply to Theorem 2 (with the references to asymptotics deleted).

5. PROOFS

First we prove Theorem 1. Let $F_T(\pi)$ denote $LR_T(\pi)$, $W_T(\pi)$, or $LM_T(\pi)$. For notational simplicity, let "sup" denote "$\sup_{\pi\in\Pi}$". For $r \geq 0$, let P_r and E_r denote probabilities and expectations with respect to the density $\int f_T(\theta_0 + B_T^{-1}h)dQ_{r,\pi}(h)dJ(\pi)$. The case $r = 0$ corresponds to the null density $f_T(\theta_0)$. The likelihood ratio of P_r to P_0 is denoted

$$(5.1)\qquad LR_{T,r} = \int f_T(\theta_0 + B_T^{-1}h)dQ_{r,\pi}(h)dJ(\pi)/f_T(\theta_0)\ .$$

For $\lambda \geq 0$ and arbitrary $\mu \in R^p$ with $|\mu| = 1$, let

$$(5.2)\qquad \psi_p(\lambda) = \int\exp(\lambda x'\mu)dU_p(x)\ ,$$

where $U_p(\cdot)$ denotes the uniform distribution on the unit sphere in R^p.

Define an approximate standardized ML estimator $\bar{\theta}(\pi)$ and an approximate Wald statistic $\bar{W}_T(\pi)$ by

$$(5.3)\qquad \begin{aligned}\bar{\theta}(\pi) &= \mathcal{I}^{-1}(\theta_0,\ \pi)B_T^{-1}D\ell_T(\theta_0,\ \pi) \quad \text{and}\\ \bar{W}_T(\pi) &= (H\bar{\theta}(\pi))'\left[H\mathcal{I}^{-1}(\theta_0,\ \pi)H'\right]^{-1}H\bar{\theta}(\pi)\ .\end{aligned}$$

The proof of Theorem 1 uses the following lemmas.

LEMMA 1: *Under Assumptions* 1, 2, 5, *and* 7, $\exp(-r^2/2)\int \psi_p(r\bar{W}_T^{1/2}(\pi))dJ(\pi)/LR_{T,r} \xrightarrow{p} 1$ *under* P_0.

LEMMA 2: *For some constants* C_1, C_2, *and* C_3 *in* $(0, \infty)$,

(a) $\psi_p(\lambda) \leq C_1 + C_2 \exp(\lambda) \quad \forall\lambda \geq 0$ *and*

(b) $\psi_p(\lambda) \geq C_3 \exp(\lambda)\lambda^{-(p-1)/2} \quad \forall\lambda \geq 1$.

LEMMA 3: *Under Assumptions* 1-5, $\sup|\bar{W}_T(\pi) - F_T(\pi)| \xrightarrow{p} 0$ *and* $\sup F_T(\pi) \xrightarrow{d} \sup F(\pi) = \sup(HG(\theta_0, \pi))'\left[H\mathcal{I}^{-1}(\theta_0, \pi)H'\right]^{-1}HG(\theta_0, \pi)$ *under* P_0 and $\sup F(\pi)$ *has absolutely continuous distribution.*

LEMMA 4: *Under Assumptions* 1, 2, 5, *and* 7, $\{P_r : T \geq 1\}$ *are contiguous to* P_0 *for all* $r > 0$.

PROOF OF THEOREM 1: For simplicity we consider the case where $k_{T\alpha}$ equals a constant k_α $\forall T \geq 1$. For the case of random $k_{T\alpha}$, we must have $k_{T\alpha} \xrightarrow{p} k_\alpha$ for some constant k_α by Lemma 3 and the corresponding adjustments to the proof are minor.

To prove Theorem 1, it suffices to show that

$$(5.4) \qquad \underline{\lim}_{T\to\infty} E_r(1 - \varphi_T)/\overline{\lim}_{T\to\infty} P_r(\sup F_T(\pi) \leq k_\alpha) \to \infty \quad \text{as} \quad r \to \infty .$$

Below we show that

$$(5.5) \qquad \overline{\lim}_{T\to\infty} P_r(\sup F_T(\pi) \leq k_\alpha) \leq 2\exp(-r^2/2)[C_1 + C_2 \exp(rk_\alpha^{1/2})] \quad \forall r > 0 .$$

We also show that for some $\gamma > 0$ and $0 < C_4 < \infty$

$$(5.6) \qquad \underline{\lim}_{T\to\infty} E_r(1 - \varphi_T) \geq C_4 \exp(-r^2/2)\exp\left(r(k_\alpha + \gamma)^{1/2}\right)\left[r(k_\alpha + \gamma)^{1/2}\right]^{-(p-1)/2}$$

for r sufficiently large. Equation (5.4) follows immediately from (5.5) and (5.6).

We now establish (5.5). Define the event $D_{T,r}^*$ by

$$(5.7) \qquad D_{T,r}^* = \left\{\exp(-r^2/2)\int \psi_p(r\bar{W}_T^{1/2}(\pi))dJ(\pi)/LR_{T,r} \in [1/2, 2]\right\} .$$

By Lemmas 1 and 4, $\lim_{T\to\infty} P_r(D^*_{T,r}) = 1 \ \forall r \geq 0$. In addition, $\sup F_T(\pi) - \sup \bar{W}_T(\pi) \xrightarrow{p} 0$ under P_r by Lemmas 3 and 4.

Using these results, we obtain: $\forall r > 0$,

$$\begin{aligned}
&\overline{\lim_{T\to\infty}}\, P_r(\sup F_T(\pi) \leq k_\alpha) \\
&= \overline{\lim_{T\to\infty}}\, P_r(\sup \bar{W}_T(\pi) \leq k_\alpha,\, D^*_{T,r}) \\
&= \overline{\lim_{T\to\infty}}\, E_0 LR_{T,r}\cdot 1(\sup \bar{W}_T(\pi) \leq k_\alpha,\, D^*_{T,r}) \\
(5.8) \quad &\leq 2\exp(-r^2/2)\overline{\lim_{T\to\infty}}\, E_0\int \psi_p(r\bar{W}_T^{1/2}(\pi))dJ(\pi)\cdot 1(\sup \bar{W}_T(\pi) \leq k_\alpha) \\
&\leq 2\exp(-r^2/2)\overline{\lim_{T\to\infty}}\, E_0\int [C_1 + C_2\exp(r\bar{W}_T^{1/2}(\pi))]dJ(\pi)\cdot 1(\sup \bar{W}_T(\pi) \leq k_\alpha) \\
&\leq 2\exp(-r^2/2)[C_1 + C_2\exp(rk_\alpha^{1/2})] ,
\end{aligned}$$

where the second inequality uses Lemma 2. Note that the first equality of (5.8) actually relies on the results above plus the convergence in distribution, absolute continuity, and contiguity results of Lemmas 3 and 4.

Next, we establish (5.6). The fact that φ_T and ξ_T are asymptotically distinct implies that $\exists\gamma > 0$ such that

$$(5.9) \qquad \underline{\lim_{T\to\infty}}\, E_0(1 - \varphi_T)1(\sup F_T(\pi) > k_\alpha + 2\gamma) \geq \delta/2 ,$$

where δ is as in the definition of asymptotically distinct. This follows because the left-hand sides of (5.9) and (2.11) differ by less than

$$(5.10) \qquad \overline{\lim_{T\to\infty}}\, P_0(\sup F_T(\pi) \in (k_\alpha, k_\alpha + 2\gamma]) = P_0(\sup F(\pi) \in (k_\alpha, k_\alpha + 2\gamma]) \leq \delta/2 ,$$

where the inequality holds for some small $\gamma > 0$ by Lemma 3.

Let K be a compact subset (under the metric d) of the space of continuous R^s-valued functions on Π. For δ as above, K can be chosen such that

$$(5.11) \qquad P_0(\mathcal{I}^{-1}(\theta_0, \cdot)G(\theta_0, \cdot) \in K) \geq 1 - \delta/4$$

using Assumptions 1 and 5. For $\varepsilon > 0$, let $K(\varepsilon) = \{g \in \mathcal{E} : \sup_{\pi \in \Pi} \|g(\pi) - \ell(\pi)\| < \varepsilon$ for some $\ell \in K\}$. Note that $K(\varepsilon)$ is a neighborhood of K in $(\mathcal{E}, d)$ by the condition (2.6) on the metric d. By Assumption 5, $\bar{\theta}(\cdot) \Rightarrow \mathcal{I}^{-1}(\theta_0, \cdot)G(\theta_0, \cdot)$. In consequence,

$$(5.12) \qquad \varliminf_{T \to \infty} P_0(\bar{\theta}(\cdot) \in K(\varepsilon)) \geq 1 - \delta/4 \quad \forall \varepsilon > 0 \ .$$

We claim that given $\gamma > 0$, $\exists \varepsilon_1 > 0$ and $\xi > 0$ such that

$$(5.13) \qquad \sup_{\rho(\pi,\pi') < \xi} |\bar{W}_T(\pi) - \bar{W}_T(\pi')| < \gamma \quad \forall \bar{\theta} \in K(\varepsilon_1) \ .$$

This claim holds because (i) $\mathcal{I}(\theta_0, \pi)$ is uniformly continuous and uniformly positive definite on Π by Assumptions 1(e), 1(f), and 4, (ii) given any $\eta > 0$, $\exists \varepsilon_2 > 0$ and $\xi_1 > 0$ such that

$$(5.14) \qquad \sup_{\rho(\pi,\pi') < \xi_1} \|g(\pi) - g(\pi')\| < \eta \quad \forall g \in K(\varepsilon_2) \ ,$$

which follows from the equicontinuity of K and the definition of $K(\varepsilon_2)$, and (iii) $\forall \varepsilon > 0$, $\sup_{g \in K(\varepsilon)} \sup_{\pi \in \Pi} \|g(\pi)\| < \infty$, which follows from sequential compactness of K, condition (2.6) on the metric d, plus the definition of $K(\varepsilon)$.

For ε_1 as in (5.13), define the event $D_{T,r}$ by

$$(5.15) \qquad D_{T,r} = \{1 - \varphi_T > \delta/8, \ \sup \bar{W}_T(\pi) > k_\alpha + 2\gamma, \ \bar{\theta} \in K(\varepsilon_1), \ D^*_{T,r}\} \ ,$$

Let $\hat{\pi}$ be a random element of Π that satisfies $\sup_{\pi \in S(\hat{\pi},\xi)} \bar{W}_T(\pi) = \sup_{\pi \in \Pi} \bar{W}_T(\pi)$.

We now have

$$(5.16) \qquad \begin{aligned} &E_r(1 - \varphi_T) \geq \frac{\delta}{8} P_r(1 - \varphi_T > \delta/8) \geq \frac{\delta}{8} P_r(D_{T,r}) = \frac{\delta}{8} E_0 LR_{T,r} \cdot 1(D_{T,r}) \\ &\geq \frac{\delta}{16} \exp(-r^2/2) E_0 1(D_{T,r}) \int \psi_p(r \bar{W}_T^{1/2}(\pi)) dJ(\pi) \\ &\geq \frac{\delta}{16} \exp(-r^2/2) E_0 1(D_{T,r}) \int 1(\pi \in S(\hat{\pi}, \xi)) \psi_p(r \bar{W}_T^{1/2}(\pi)) dJ(\pi) \\ &\geq \frac{\delta}{16} \exp(-r^2/2) E_0 1(D_{T,r}) \inf_{\pi \in S(\hat{\pi},\xi)} \psi_p(r \bar{W}_T^{1/2}(\pi)) \inf_{\pi \in \Pi} J(S(\pi, \xi)) \ , \end{aligned}$$

where the third inequality uses Lemma 1.

Note that $D_{T,r}$ has been defined such that for $\omega \in D_{T,r}$ and $\pi \in S(\hat{\pi}, \xi)$, we have $\sup_{\pi \in \Pi} \bar{W}_T(\pi)_\omega > k_\alpha + 2\gamma$, $\bar{\theta}(\cdot)_\omega \in K(\varepsilon_1)$, $|\bar{W}_T(\pi)_\omega - \bar{W}_T(\hat{\pi})_\omega| < \gamma$, and

$$r\bar{W}_T^{1/2}(\pi)_\omega \geq r(\bar{W}_T(\hat{\pi})_\omega - \gamma)^{1/2} \geq r(k_\alpha + \gamma)^{1/2} . \tag{5.17}$$

Let $b = \inf_{\pi \in \Pi} J(S(\pi, \xi))$. By Assumption 6, $b > 0$.

For r such that $r(k_\alpha + \gamma)^{1/2} > 1$, Lemma 2, (5.16), and (5.17) combine to yield

$$\begin{aligned} E_r(1 - \varphi_T) &\geq \frac{b\delta}{16}\exp(-r^2/2)E_0 1(D_{T,r}) \inf_{\pi \in S(\hat{\pi},\xi)} [C_3 \exp(r\bar{W}_T^{1/2}(\pi))(r\bar{W}_T^{1/2}(\pi))^{-(p-1)/2}] \\ &\geq C_3\frac{b\delta}{16}\exp(-r^2/2)P_0(D_{T,r})\exp(r(k_\alpha + \gamma)^{1/2})\left[r(k_\alpha + \gamma)^{1/2}\right]^{-(p-1)/2} . \end{aligned} \tag{5.18}$$

The desired result now follows if

$$\underline{\lim}_{T\to\infty} P_0(D_{T,r}) \geq \delta/8 \quad \forall r \geq 0 . \tag{5.19}$$

We obtain a lower bound on $\underline{\lim}_{T\to\infty} P_0(D_{T,r})$ using (5.9). (5.9) and (5.12) yield

$$\underline{\lim}_{T\to\infty} E_0(1 - \varphi_T)1(\sup F_T(\pi) > k_\alpha + 2\gamma,\ \bar{\theta} \in K(\varepsilon_1)) \geq \frac{\delta}{2} - \overline{\lim}_{T\to\infty} P_0(\bar{\theta} \notin K(\varepsilon_1)) \geq \frac{\delta}{4} . \tag{5.20}$$

This result and Lemma 3 give

$$\begin{aligned} \delta/4 &\leq \underline{\lim}_{T\to\infty} E_0(1 - \varphi_T)1(\sup \bar{W}_T(\pi) > k_\alpha + 2\gamma,\ \bar{\theta} \in K(\varepsilon_1)) \\ &\leq \underline{\lim}_{T\to\infty} P_0(1 - \varphi_T > \delta/8,\ \sup \bar{W}_T(\pi) > k_\alpha + 2\gamma,\ \bar{\theta} \in K(\varepsilon_1)) \\ &\quad + \overline{\lim}_{T\to\infty} E_0(1 - \varphi_T)1(1 - \varphi_T \leq \delta/8) \\ &\leq \underline{\lim}_{T\to\infty} P_0(D_{T,r}) + \delta/8 , \end{aligned} \tag{5.21}$$

where the last inequality uses Lemma 1. □

PROOF OF LEMMA 1: Let $G_T = B_T^{-1}D\ell_T(\theta_0, \pi)$, $\mathcal{I}_T = \mathcal{I}_T(\theta_0, \pi)$, and $\mathcal{I} = \mathcal{I}(\theta_0, \pi)$. Define

$$(5.22) \qquad \omega_T(r) = \sup_{\substack{\pi\in\Pi,\ h:h'\mathcal{I}h\le r^2,\\ 0<\lambda<1}} \|\mathcal{I}_T(\theta_0 + \lambda B_T^{-1}h, \pi) - \mathcal{I}_T(\theta_0, \pi)\| .$$

Then, for h such that $h'\mathcal{I}h \le r^2$, a two-term Taylor expansion yields

$$(5.23) \qquad \begin{aligned} &\ell_T(\theta_0 + B_T^{-1}h, \pi) - \ell_T(\theta_0) = h'G_T - h'\mathcal{I}_T h/2 + R_T , \quad \text{where} \\ &|R_T| \le \omega_T(r)\|h\|^2 \le \omega_T(r)r^2/\inf_{\pi\in\Pi} \lambda_{\min}(\mathcal{I}(\theta_0, \pi)) = C_r\omega_T(r) \end{aligned}$$

for some constant $C_r < \infty$.

By Assumption 1, $\omega_T(r) \xrightarrow{p} 0 \quad \forall r > 0$. Hence, $\forall\pi \in \Pi$ and $\forall h$ with $h'\mathcal{I}_0 h \le r^2$, we have

$$(5.24) \qquad \begin{aligned} K_{1T} &= \exp(-C_r\omega_T(r)) \\ &\le \exp(h'G_T - h'\mathcal{I}_T h/2)/[f_T(\theta_0 + B_T^{-1}h, \pi)/f_T(\theta_0)] \\ &\le \exp(C_r\omega_T(r)) \\ &= K_{2T} , \end{aligned}$$

where $K_{1T} \xrightarrow{p} 1$ and $K_{2T} \xrightarrow{p} 1$.

In turn, this yields

$$(5.25) \qquad K_{1T} \le \int \exp(h'G_T - h'\mathcal{I}_T h/2)dQ_{r,\pi}(h)dJ(\pi)/LR_{T,r} \le K_{2T} .$$

By Assumption 1(d), $\exp(h'\mathcal{I}_T h/2)\exp(h'\mathcal{I}h/2) \xrightarrow{p} 1$ uniformly over h with $h'\mathcal{I}h \le r^2$ and over $\pi \in \Pi$. In consequence, there exist sequences of constants K_{3Tr} and K_{4Tr} such that $K_{3Tr} \xrightarrow{p} 1$, $K_{4Tr} \xrightarrow{p} 1$, and

$$(5.26) \qquad K_{3Tr} \le \int \exp(h'G_T - h'\mathcal{I}h/2)dQ_{r,\pi}(h)dJ(\pi)/LR_{T,r} \le K_{4Tr} .$$

For h in the support of $Q_{r,\pi}$, we have $h'\mathcal{I}h = r^2$ and $h = A_\pi\lambda$ for some $\lambda \in R^p$. For such h, $h'\mathcal{I}(I_s - A_\pi H) = 0$, since straightforward algebra shows that $A_\pi'\mathcal{I}(I_s - A_\pi H) = 0$. Thus, we have

$$\begin{aligned}(5.27)\quad &\int \exp(h'G_T - h'\mathcal{I}h/2)dQ_{r,\pi}(h)dJ(\pi) \\ &= \exp(-r^2/2)\int \exp(h'\mathcal{I}A_\pi H\bar{\theta}(\pi) + h'\mathcal{I}(I_s - A_\pi H)\bar{\theta}(\pi))dQ_{r,\pi}(h)dJ(\pi) \\ &= \exp(-r^2/2)\int \exp(h'\mathcal{I}A_\pi H\bar{\theta}(\pi))dQ_{r,\pi}(h)dJ(\pi) \\ &= \exp(-r^2/2)\int \exp(rx'(A_\pi'\mathcal{I}A_\pi)^{1/2}H\bar{\theta}(\pi))dU_p(x)dJ(\pi) \\ &= \exp(-r^2/2)\int \exp(r\bar{W}_T^{1/2}(\pi)x'\mu)dU_p(x)dJ(\pi) \\ &= \exp(-r^2/2)\int \psi_p(r\bar{W}_T^{1/2}(\pi))dJ(\pi)\ ,\end{aligned}$$

where the third equality uses Assumption 7, $\mu = (A_\pi'\mathcal{I}A_\pi)^{1/2}H\bar{\theta}(\pi)/\|(A_\pi'\mathcal{I}A_\pi)^{1/2}H\bar{\theta}(\pi)\|$, and $\bar{W}_T(\pi) = \|(A_\pi'\mathcal{I}A_\pi)^{1/2}H\bar{\theta}(\pi)\|^2$ since $A_\pi'\mathcal{I}A_\pi = \left(H\mathcal{I}^{-1}H'\right)^{-1}$ by straightforward algebra.

Equations (5.26) and (5.27) combine to give the desired result. □

PROOF OF LEMMA 2: This lemma follows straightforwardly from results in the literature. For example, it follows from equation (15.3.7), equation (15.3.9), and the last equation on p. 431 of Mardia, Kent, and Bibby (1979). Note that their equation (15.3.7) contains a typo. The expression $(p-1)/2$ should be $(p/2) - 1$ in the two places it appears. □

PROOF OF LEMMA 3: Under Assumptions 1, 2, 4, and 5,

$$(5.28)\qquad \sup|\bar{W}_T(\pi) - F_T(\pi)| \xrightarrow{p} 0 \quad \text{under} \quad P_0$$

by the proof of Theorem A-1 parts (c)-(e) of AP. By Assumptions 1, 4, and 5, the continuous mapping theorem (e.g., see Pollard (1984, p. 70)), (2.5), and (5.28), the second result of the lemma holds. Absolute continuity of $\sup F(\pi)$ follows from a result of Lifshits (1982). □

PROOF OF LEMMA 4: We make use of the following result, which follows, e.g., from Thms. 16.8 and 18.11 of Strasser (1985): If (i) $LR_{T,r} \xrightarrow{d} X_r^*$ under P_0 and (ii) $E_0X_r^* = 1$, then $\{P_r : T \geq 1\}$ are contiguous to P_0. Condition (i) holds with $X_r^* = \exp(-r^2/2)\int \psi_p(rF^{1/2}(\pi))dJ(\pi)$, where $F(\pi)$ is as in Lemma 3, by Lemma 1, Assumptions 1 and 5, and the continuous mapping theorem.

Condition (ii) is obtained as follows. Let $Z(\pi) \sim N(0, I_p)$. Then, $F(\pi)$ and $Z(\pi)'Z(\pi)$ have the same distribution by Assumption 5. We now obtain

$$E_0 X_r^* = \exp(-r^2/2) E_0 \int \left[\int \exp(r F^{1/2}(\pi) x' \mu) dU_p(x)\right] dJ(\pi)$$

(5.29)
$$= \exp(-r^2/2) \int \left[\int E_0 \exp(r x' Z(\pi)) dU_p(x)\right] dJ(\pi)$$

$$= \exp(-r^2/2) \int \left[\int \exp(r^2 x' x) dU_p(x)\right] dJ(\pi) = 1 ,$$

where the second equality holds by taking the arbitrary unit vector μ to be $Z(\pi)/|Z(\pi)|$ and applying Fubini's theorem and the third equality uses the standard formula for the moment generating function of a standard normal random vector. □

PROOF OF THEOREM 2: (Assumptions) 1(a), (b), (c), (e), and (f) follow immediately from EP1(b), (c), (d), (f), and (g) respectively. 1(d) follows with $B_T = \sqrt{T} I_s$ from EP1(a), EP1(e), and EP2(b) using the uniform WLLN given in the Theorem in Andrews (1987) adjusted to allow for non-compact parameter spaces according to footnote 1 of Andrews (1992). In particular, pointwise WLLNs hold for the inf's and sups of $g(W_t, \theta, \pi)$ over small balls in $\Theta_0 \times \Pi$ by the ergodic theorem, because such random variables are strictly stationary and absolutely regular and, hence, ergodic.

Assumptions 2 and 3 can be verified using Lemma A-1 of Andrews (1993b). We verify its conditions (a) and (b) for $Q_T(\theta, \pi) = -\frac{1}{T}\Sigma_1^T \log g(W_t, \theta, \pi)$ and $Q(\theta, \pi) = -E \log g(W_t, \theta, \pi)$. For 2, the parameter space for θ is Θ. For 3, the parameter space for θ is $\tilde{\Theta} = \Theta \cap V$, the null hypothesis parameter space. Condition (a) of Lemma A-1 requires that $Q_T(\theta, \pi)$ satisfies a uniform WLLN over $\Theta \times \Pi$. This follows by the same uniform WLLN as used above by EP1(a), EP1(e), and EP2(a). Condition (b) of Lemma A-1 holds for 2 (with parameter space Θ) and for 3 (with parameter space $\tilde{\Theta}$) by EP3.

Assumption 5 holds with $B_T = \sqrt{T}\, I_s$ by Theorem 1 of Doukhan, Massart, and Rio (1992) using EP1(a), EP1(e), and EP4. More specifically, the key condition (2.10) of Theorem 1 of Doukhan *et al.* is implied by their equations (S.1) and (2.11) by their Lemma 2. Equations (S.1) and (2.11) hold with $\phi(x) = x^{r/2}$ by EP1(a) and EP4 respectively, since $|\cdot|_{\phi,2}$ equals the L^r-norm $|\cdot|_r$ in this case.

Next we establish part (b) of the Theorem. We have: $\forall\theta \neq \theta_0$, $\forall\pi \in \Pi$,

$$(5.30) \qquad Q(\theta, \pi) - Q(\theta_0, \pi) = E \log\frac{g(W_t, \theta, \pi)}{g(W_t, \theta_0, \pi)} < \log E\frac{g(W_t, \theta, \pi)}{g(W_t, \theta_0, \pi)} = 0 ,$$

where the inequality follows from Jensen's inequality and is strict by EP3*(a). Since compactness and sequential compactness are equivalent for pseudo-metric spaces, compactness of $\Theta \times \Pi$ under d (EP3*(b)) implies compactness of Θ and Π under $d_\Theta(\theta, \bar{\theta}) = d((\theta, \bar{\pi}), (\bar{\theta}, \bar{\pi}))$ for arbitrary $\bar{\pi} \in \Pi$ and under $d_\Pi(\pi, \bar{\pi}) = d((\bar{\theta}, \pi), (\bar{\theta}, \bar{\pi}))$ for arbitrary $\bar{\theta} \in \Theta$ respectively. Thus, by continuity in π (EP2(a)) and compactness of Π, $\forall\theta \neq \theta_0 \ \exists\pi^* = \pi^*(\theta)$ such that

$$(5.31) \qquad \sup_{\pi\in\Pi}(Q(\theta, \pi) - Q(\theta_0, \pi)) = Q(\theta, \pi^*) - Q(\theta_0, \pi^*) < 0 .$$

That is, $\sup_{\pi\in\Pi}(Q(\theta, \pi) - Q(\theta_0, \pi))$ is uniquely maximized at θ_0. Since it is a continuous function of θ (by EP2(a)) and Θ is compact, EP3 holds as desired.

Part (c) of the Theorem holds by Theorem 5 of Andrews (1993a), because $\mathcal{M}$ is a type IV class of functions, as defined in Andrews (1993a), and EP4 is equivalent to Ossiander's L^r entropy condition, which holds for type IV classes of functions by Theorem 5. □

PROOF OF THEOREM 3: The proof of Theorem 1 goes through with the following changes: Throughout, B_T equals I_s, $\mathcal{I}(\theta_0, \pi)$ and $\mathcal{I}_T(\theta_0, \pi)$ equal $\frac{1}{\sigma^2}\sum_{t=1}^T \begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}\begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}'$. In consequence, $\bar{\theta}(\pi)$ equals the unrestricted LS estimator $\hat{\theta}(\pi)$ minus θ_0 and $\bar{W}_T(\pi) = W_T(\pi)$. Lemma 1 holds with "$\xrightarrow{p}$ 1" replaced by "= 1". The proof of Lemma 1 simplifies because $\mathcal{I}_T(\theta, \pi)$ does not depend on θ which yields $\omega_T(r) = 0$ and $K_{1T} = K_{2T} = K_{3Tr} = K_{4Tr} = 1$. Lemma 2 and its proof hold without change. Lemma 3 is replaced by the result that $\bar{W}_T(\pi) = W_T(\pi) \ \forall\pi \in \Pi$, as noted above, and $\sup_{\pi\in\Pi} W_T(\pi)$ has absolutely continuous distribution, which follows from a result of Lifshits (1982) given the assumption that $X_t(\pi)$ is continuous in π. Lemma 4 is not relevant and is eliminated. In the proof of Theorem 1, limits as $T \to \infty$ are deleted, (5.7) is deleted and $D_{T,r}^*$ is eliminated in (5.8) and (5.15), $G(\theta_0, \pi)$ equals $\frac{1}{\sigma^2}\sum_{t=1}^T U_t(X_t(\pi)', Z_t')'$ in (5.11), (5.12) is deleted, and (5.14) and subsequent equations hold with $K(\varepsilon)$ replaced by $K \ \forall\varepsilon > 0$. □

FOOTNOTES

[1]Andrews gratefully acknowledges research support from the National Science Foundation through grant numbers SES-9121914 and INT-9215258. Ploberger gratefully acknowledges research support from Fonds zur Foerderung der wissenschaftlichen Forschung under Project Number P2279-S0Z and Ausseninstitut der TU-Wien.

REFERENCES

Andrews, D. W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465-1471.

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241-257.

Andrews, D. W. K. (1993a): "Empirical Process Methods in Econometrics," prepared for the *Handbook of Econometrics*, Vol. IV. New York: North-Holland, forthcoming.

Andrews, D. W. K. (1993b): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821-856.

Andrews, D. W. K. and W. Ploberger (1992): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," Cowles Foundation Discussion Paper No. 1015, Yale University.

Andrews, D. W. K. and W. Ploberger (1993): "Optimal Changepoint Tests for Linear Regression," *Journal of Econometrics*, forthcoming.

Bai, J., R. Lumsdaine, and J. H. Stock (1991): "Testing for and Dating Breaks in Integrated and Cointegrated Times Series," unpublished manuscript, Kennedy School of Government, Harvard University.

Billingsley, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

Chang, J. T. and J. A. Hartigan (1993): "Admissibility of Likelihood Ratio Tests for Detecting a Change in Distribution," unpublished manuscript, Department of Statistics, Yale University.

Davies, R. B. (1977): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 64, 247-254.

Davies, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter Is Present Only under the Alternative," *Biometrika*, 74, 33-43.

Davydov, Yu A. (1973): "Mixing Conditions for Markov Chains," *Theory of Probability and Its Applications*, 28, 313-328.

Doukhan, P. (1992): "Mixing: Properties and Examples," unpublished book manuscript, Department of Mathematics, University of Paris-South, Orsay.

Doukhan, P., P. Massart, and E. Rio (1992): "Invariance Principles for Empirical Mixing Processes under Bracketing Conditions," unpublished manuscript, Department of Mathematics, University of Paris-South, Orsay, France.

Lifshits, M. A. (1982): "On the Absolute Continuity of Distributions of Functionals of Random Processes," *Theory of Probability and Its Applications*, 27, 600-607.

Mardia, K. V., J. T. Kent, and J. M. Bibby (1979): *Multivariate Analysis*. New York: Academic Press.

Mokkadem, A. (1986): "Sur le Mélange d'un processus ARMA vectoriel," *Comptes Rendus de l'Academie des Sciences Paris*, 303, Series I, no. 11, 519-521.

Mokkadem, A. (1990): "Propriétés de Mélange des Modèles Autorégressifs Polynomiaux," *Annals of the Institute of Henri Poincaré*, 26, 219-260.

Ossiander, M. (1987): "A Central Limit Theorem under Metric Entropy with L^2-Bracketing," *Annals of Probability*, 15, 897-919.

Perron, P. (1991): "A Test for Changes in a Polynomial Trend Function for Dynamic Time Series," unpublished manuscript, Department of Economics, Princeton University.

Ploberger, W., W. Krämer, and K. Kontrus (1989): "A New Test for Structural Stability in the Linear Regression Model," *Journal of Econometrics*, 40, 307-318.

Pollard, D. (1984): *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Potter, S. M. (1989): "A Nonlinear Approach to U.S. GNP," unpublished manuscript, Department of Economics, UCLA.

Strasser, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory*. New York: de Gruyter.

Tong, H. (1990): *Non-linear Time Series: A Dynamical Systems Approach*. Berlin: Springer.

Wald, A. (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426-482.