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

By Donald W. K. Andrews[1] and Werner Ploberger[2]

Yale University and Technische Universität Wien

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 nonstandard 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 (1994) (denoted AP) developed a class of tests, called average 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, namely, 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) cross-sectional constancy in nonlinear models, (ii) threshold effects in autoregressive models and (iii) variable relevance in nonlinear models, such as Box–Cox transformed regressor models. Primitive sufficient conditions for tests of changepoints are given in AP. 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 and (ii) white noise versus first-order generalized autoregressive conditional heteroskedasticity.

Next, we consider finite sample admissibility of the LR test for the Gaussian linear regression model with known variance. 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).

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.

2.1. *Notation and definitions.* Let $\underset{\sim}{Y}_T$ denote the data matrix when the sample size is T for $T = 1, 2, \ldots$. Consider a parametric family $\{f_T(y_T, \theta, \pi): \theta \in \Theta, \pi \in \Pi\}$ 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 $f_T(\theta, \pi) = f_T(\underset{\sim}{Y}_T, \theta, \pi)$.

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 prechange parameter vector, $\delta_1 + \beta$ is a postchange parameter vector and δ_2 is a parameter vector that is constant across regimes.

The null hypothesis of interest is

(2.1) $$H_0: \beta = \underset{\sim}{0}.$$

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

(2.2) $$H_1: \beta \neq \underset{\sim}{0}$$

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 $l_T(\theta, \pi) = \log f_T(\theta, \pi)$. Let $Dl_T(\theta, \pi)$ denote the s-vector of partial derivatives of $l_T(\theta, \pi)$ with respect to θ. Let $D^2 l_T(\theta, \pi)$ denote the $s \times s$ matrix of second partial derivatives of $l_T(\theta, \pi)$ with respect to θ. [Note that $Dl_T(\theta_0, \pi)$ and $D^2 l_T(\theta_0, \pi)$ depend on π in general even though $f_T(\theta_0, \pi)$ and $l_T(\theta_0, \pi)$ do not.]

We consider the case where the appropriate norming factors for $Dl_T(\theta, \pi)$ and $D^2 l_T(\theta, \pi)$ [so that each is $O_p(1)$, but not $o_p(1)$] are nonrandom 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 nontrending 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 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 the following assumption.

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 1 under θ_0, where Θ_0 is some neighborhood of θ_0.

(d) $\sup_{\pi \in \Pi, \ \theta \in \Theta_0} | - B_T^{-1} D^2 l_T(\theta, \pi) B_T^{-1} - \mathbf{I}(\theta, \pi)| \to_p 0$ under θ_0 for some nonrandom $s \times s$ matrix function $\mathbf{I}(\theta, \pi)$ and some sequence of nonrandom diagonal $s \times s$ matrices $\{B_T \colon T \geq 1\}$ that satisfies $[B_T]_{jj} \to \infty$ as $T \to \infty$ $\forall \ j \leq s$.

(e) $\mathbf{I}(\theta, \pi)$ is continuous in (θ, π) on $\Theta_0 \times \Pi$.

(f) $\mathbf{I}(\theta_0, \pi)$ is positive definite for all $\pi \in \Pi$.

The matrix function $\mathbf{I}(\theta, \pi)$ 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.

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

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

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

$$\tilde{\theta} \in \tilde{\Theta} = \{ \theta \in \Theta : \theta = (0', \delta')' \text{ for some } \delta \in R^q \},$$

(2.4)
$$l_T(\tilde{\theta}, \pi) = \sup_{\theta \in \tilde{\Theta}} l_T(\theta, \pi) \quad \text{wp} \to 1 \text{ under } \theta_0.$$

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. $\sup_{\pi \in \Pi} \|\hat{\theta}(\pi) - \theta_0\| \to_p 0$ under θ_0.

ASSUMPTION 3. $\tilde{\theta} - \theta_0 \to_p 0$ under θ_0.

The parameter space Π is assumed to satisfy the next assumption.

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 $\to_d$ 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 $B_T^{-1} Dl_T(\theta_0, \pi)$ $(\in R^s)$ 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 $\mathbf{E}$ of R^s-valued functions on Π. We assume d is chosen such that (i) the function

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

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

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

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 the following assumption.

ASSUMPTION 5. $B_T^{-1} Dl_T(\theta_0, \cdot) \Rightarrow G(\theta_0, \cdot)$ 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 $EG(\theta_0, \pi)G(\theta_0, \pi)' = \mathbf{I}(\theta_0, \pi) \, \forall \, \pi \in \Pi$ and has continuous sample paths (as functions of π for fixed θ_0) with probability 1.

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 the following statement.

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

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

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

$$\mathrm{LR}_T(\pi) = -2\left(l_T(\tilde{\theta}, \pi) - l_T(\hat{\theta}(\pi), \pi)\right),$$

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

$$\mathrm{LM}_T(\pi) = \left[B_T^{-1}Dl_T(\tilde{\theta}, \pi)\right]'\mathbf{I}_T^{-1}(\tilde{\theta}, \pi)B_T^{-1}Dl_T(\tilde{\theta}, \pi),$$

where

$$H = \left[I_p : 0\right] \subset R^{p \times s} \quad \text{and} \quad \mathbf{I}_T(\theta, \pi) = -B_T^{-1}D^2l_T(\theta, \pi)B_T^{-1}.$$

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

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

(2.8) $$\sup_{\pi \in \Pi} \mathrm{LR}_T(\pi), \quad \sup_{\pi \in \Pi} W_T(\pi) \quad \text{and} \quad \sup_{\pi \in \Pi} \mathrm{LM}_T(\pi).$$

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, but with nonrandom probability limit in this case) 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

$$(2.9) \qquad \xi_T = 1\left(\sup_{\pi \in \Pi} \mathrm{LR}_T(\pi) > k_{T_\alpha} \right)$$

or where ξ_T is defined analogously with $\mathrm{LR}_T(\pi)$ replaced by $W_T(\pi)$ or $\mathrm{LM}_T(\pi)$.

Under Assumptions 1–5, the asymptotic null distribution of

$$\sup_{\pi \in \Pi} \mathrm{LR}_T(\pi), \quad \sup_{\pi \in \Pi} W_T(\pi) \quad \text{and} \quad \sup_{\pi \in \Pi} \mathrm{LM}_T(\pi)$$

is that of

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

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 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

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

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. Inequality (2.11) requires that this distinction between φ_T and ξ_T does not disappear as $T \to \infty$.

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 $\{\xi_T \colon T \geq 1\}$. Then there exists an $r_0 < \infty$ such that, for all $r \geq r_0$,

$$\limsup_{T \to \infty} \int \left[\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi)\, d\mu_T \right] dQ_{r,\pi}(h)\, dJ(\pi)$$

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

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

REMARKS.

1. The proof of Theorem 1 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 diverges to infinity as $r \to \infty$.
2. Theorem 1 holds for any weight functions J that satisfy Assumption 6.
3. Theorem 1 holds for *any* sequence of asymptotically distinct tests $\{\varphi_T \colon T \geq 1\}$—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.
4. Assumption 3 is not required in Theorem 1 for the case of the sup Wald test.

3. Examples.

3.1. *Empirical process examples.* This section provides primitive sufficient conditions for Assumptions 1–5 for empirical process 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 $\beta = (\beta_1', \ldots, \beta_m')'$ 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 $H_0 \colon \beta = 0$.

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

$$(3.1) \qquad Y_t = \begin{cases} X_t'\delta_1 + U_t, & \text{for } Z_t \leq \pi, \\ X_t'(\delta_1 + \beta) + U_t, & \text{for } Z_t > \pi, \end{cases} \quad \text{for } t = 1, \ldots, T,$$

where $\{(Y_t, X_t, Z_t, U_t) \colon t = 1, \ldots, T\}$ 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;

$$\inf_{\pi \in \Pi} \lambda_{\min}\left(E\left(\begin{matrix} X_t 1(Z_t > \pi) \\ X_t \end{matrix} \right) \left(\begin{matrix} X_t 1(Z_t > \pi) \\ X_t \end{matrix} \right)' \right) > 0;$$

the parameter $\theta = (\beta', \delta_1', \delta_2)'$ lies in a compact set $\Theta \subset R^{2p+1}$ that excludes δ_2 values less than or equal to 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 (3.1) with $X_t = (1, Y_{t-1})'$, $Z_t = Y_{t-d}$ for some integer $d > 0$, $\{U_t: t = 1, \dots, T\}$ are iid, (Y_0, Y_{1-d}) have distributions that correspond to a stationary startup of the AR model when $\beta = 0$ and Θ and Π are as defined above with $p = 2$ and $|\delta_1| < 1$. 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 (1995)]. 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

$$(3.2) \qquad Y_t = g(X_t, \delta_1) + \beta h(Z_t, \pi) + U_t \quad \text{for } t = 1, \dots, T,$$

where $\{(Y_t, X_t, Z_t, U_t): t = 1, \dots, T\}$ 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 less than or equal to 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 $\delta_1 \; \forall \; \theta \in \Theta_0$ with probability 1 under θ_0, where Θ_0 is some neighborhood of θ_0; $h(Z_t, \pi)$ is differentiable in π with probability 1 under $\theta_0 \; \forall \; \pi \in \Pi$; $E \sup_{\theta \in \Theta} g^2(X_t, \delta_1) < \infty$; $E \sup_{\pi \in \Pi} h^2(Z_t, \pi) \log^+(|h(Z_t, \pi)|) < \infty$, where $\log^+(x) = \max\{\log(x), 0\}$ for $x \geq 0$;

$$E \sup_{\theta \in \Theta_0} \left\| \frac{\partial}{\partial \delta_1} g(X_t, \delta_1) \right\|^2 < \infty;$$

$$E \sup_{\theta \in \Theta_0} \left\| \frac{\partial^2}{\partial \delta_1 \, \partial \delta_1'} g(X_t, \delta_1) \right\|^2 < \infty;$$

$$E \sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \pi} h(Z_t, \pi) \right\|^r \leq \infty \quad \text{for some } r > 2;$$

$$\inf_{\pi \in \Pi} \lambda_{\min} \left(E \left(\begin{array}{c} h(Z_t, \pi) \\ (\partial/\partial \delta_1) g(X_t, \delta_{10}) \end{array} \right) \left(\begin{array}{c} h(Z_t, \pi) \\ (\partial/\partial \delta_1) g(X_t, \delta_{10}) \end{array} \right)' \right) > 0$$

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: $\beta = 0$.

EXAMPLE 4 (Functional form). Example 3 covers tests of functional form. For example, in (3.2), if Z_t is taken to be a subvector of X_t, a test of H_0: $\beta = 0$ is a test of functional form of the regression function. Neural network tests of functional form and some consistent tests of model specification are of this type.

We now introduce the requisite definitions and assumptions used for the empirical process examples. The data are given by $\{(Y_t, X_t): t = 1, \ldots, T\}$ which are part of a strictly stationary, absolutely regular process $\{(Y_t, X_t): t = \ldots, 0, 1, \ldots\}$, where $\{Y_t\}$ is an mth order Markov sequence of random variables and $\{X_t\}$ is a sequence of weakly exogeneous variables. By definition, $\{Y_t: t = \ldots, 0, 1, \ldots\}$ is *mth order Markov* if the conditional distribution of Y_t given $\mathbf{F}_{t-1} = \sigma(\ldots, Y_{t-2}, Y_{t-1}; \ldots, X_{t-1}, X_t)$ equals the conditional distribution of Y_t given $Y_{t,m} = (Y_{t-m}, \ldots, Y_{t-1})$ and $X_{t,m} = (X_{t-m}, \ldots, X_t)$ for all t. By definition, $\{X_t\}$ is *weakly exogeneous* if the conditional distribution of X_t given $Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_{t-1}$ does not depend on the unknown parameters θ and π $\forall\, t \geq 1$.

By definition, a sequence $\{W_t: t = \ldots, 0, 1, \ldots\}$ 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 $\mathbf{A}$ and $\mathbf{B}$, define

$$(3.3) \qquad \beta(\mathbf{A}, \mathbf{B}) = \tfrac{1}{2}\sup \sum_{(i,j)\in I\times J} |P(A_i \cap B_j) - P(A_i)P(B_j)|,$$

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 $\mathbf{A}$ and $\mathbf{B}$ measurable, respectively. Let $\mathbf{F}_t = \sigma(\ldots, W_{t-1}, W_t)$ and $\mathbf{F}^t = \sigma(W_t, W_{t+1}, \ldots)$, where $\sigma(\cdot)$ denotes a σ-field. Then

$$(3.4) \qquad \beta(s) = \sup_t \beta(\mathbf{F}_t, \mathbf{F}^{t+s}).$$

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 (1994). 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 $0 < \rho < 1$ by Mokkadem (1986).

Let $W_t = (Y_{t-m}', \ldots, Y_t', X_{t-m}', \ldots, X_t')'$. Let

$$(3.5) \qquad g(W_t, \theta, \pi) = g(Y_t | Y_{t-m}, \ldots, Y_{t-1}, X_{t-m}, \ldots, X_t; \theta, \pi),$$

for $\theta \in \Theta$ and $\pi \in \Pi$ denote a parametric family of conditional densities (with respect to some measure) of Y_t given $Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_t$ evaluated at the random variables $Y_1, \ldots, Y_t, X_1, \ldots, X_t$. The parameter space Θ is a subset of

R^s and Π also is a subset of Euclidean space. Let

$$(3.6) \qquad h_t = h(X_t | Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_{t-1})$$

denote the conditional density (with respect to some measure) of X_t given $Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_{t-1}$ evaluated at the random variables $Y_1, \ldots, Y_{t-1}$, $X_1, \ldots, X_t$.

The likelihood and log-likelihood functions of the sample are

$$(3.7) \qquad \begin{aligned} f_T(\theta, \pi) &= \prod_{t=1}^{T} g(W_t, \theta, \pi) \prod_{t=1}^{T} h_t \quad \text{and} \\ l_T(\theta, \pi) &= \sum_{t=1}^{T} \log g(W_t, \theta, \pi) + \sum_{t=1}^{T} h_t. \end{aligned}$$

The information matrix for θ given π is defined to be

$$(3.8) \qquad \mathbf{I}(\theta, \pi) = -E \frac{\partial^2}{\partial \theta \, \partial \theta'} \log g(W_t, \theta, \pi).$$

Below, we use the concept of L^r-continuity. Let $f(W_t, \tau)$ be a vector-valued function of a vector $\tau \in \mathbf{T}$. We say that f is L^r-*continuous at* τ_0 if

$$(3.9) \qquad E \sup_{\tau \in \mathbf{T}: \, \|\tau - \tau_0\| < \delta} \|f(W_t, \tau) - f(W_t, \tau_0)\|^r \to 0 \quad \text{as } \delta \to 0,$$

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 (3.9) is less than or equal to $c(\delta)$ $\forall \, \delta$ 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., $E \sup_{\tau \in \mathbf{T}: \, \|\tau - \tau_0\| < \delta} \|f(W_t, \tau) - f(W_t, \tau_0)\|^r < \infty$ for some $\delta > 0$] by the dominated convergence theorem.

To obtain the weak convergence property of Assumption 5, we use a bracketing empirical process central limit theorem (CLT) of Doukhan, Massart and Rio (1995). The latter is a generalization to strictly stationary absolutely regular processes of an empirical process CLT of Ossiander (1987) for iid processes.

Throughout this section, we assume $g(\cdot, \theta, \pi), (\partial/\partial \theta) \log g(\cdot, \theta, \pi)$ and $(\partial^2/\partial \theta \, \partial \theta') \log g(\cdot, \theta, \pi)$ 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. All expectations E below are taken under θ_0.

The following Assumptions EP1–EP4' are sufficient for Assumptions 1–5 of Section 2.

ASSUMPTION EP1. (a) Under θ_0, $\{(Y_t, X_t): t = \ldots, 0, 1, \ldots\}$ 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 $r > 2$, $\{Y_t: t = \ldots, 0, 1, \ldots\}$ is mth order Markov and $\{X_t: t = \ldots, 0, 1, \ldots\}$ 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 1 under θ_0, where Θ_0 is some neighborhood of θ_0.

(e) $E \sup_{\theta \in \Theta, \pi \in \Pi} |\log g(W_t, \theta, \pi)| < \infty$, $E \sup_{\pi \in \Pi} \|(\partial/\partial\theta)\log g(W_t, \theta_0, \pi)\|^r < \infty$, for r as in part (a), and $E \sup_{\theta \in \Theta_0, \pi \in \Pi} \|(\partial^2/\partial\theta\,\partial\theta')\log g(W_t, \theta, \pi)\| < \infty$. (If $\{(Y_t, X_t): t = \ldots, 0, 1, \ldots\}$ 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 Assumption EP4 below. If $\{(Y_t, X_t): t = \ldots, 0, 1, \ldots\}$ has geometrically declining β-mixing numbers [i.e., $\beta(s) = O(\rho^s)$ for some $0 < \rho < 1$], then $E \sup_{\pi \in \Pi} \|(\partial/\partial\theta)\log g(W_t, \theta_0, \pi)\|^r < \infty$ 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 greater than 2 in Assumption EP4 below.)

(f) $\mathbf{I}(\theta, \pi)$ is continuous in (θ, π) on $\Theta_0 \times \Pi$.

(g) $\mathbf{I}(\theta_0, \pi)$ is positive definite for all $\pi \in \Pi$.

ASSUMPTION EP2. (a) Θ is compact.

(b) $\log g(W_t, \theta, \pi)$ is L^1-continuous in (θ, π) on $\Theta \times \Pi$ under θ_0.

(c) $(\partial^2/\partial\theta\,\partial\theta')\log g(W_t, \theta, \pi)$ is L^1-continuous in (θ, π) on $\Theta_0 \times \Pi$ under θ_0.

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

ASSUMPTION EP4.   (a) Π is compact.

(b) $(\partial/\partial\theta)\log g(W_t, \theta_0, \pi)$ 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).

We note that Assumptions EP1–EP4 are satisfied in Examples 1–4 above. Assumption EP3 can be verified in the examples by showing that $E(\log g(W_t, \theta_0, \pi) - \log g(W_t, \theta, \pi)) > 0$ $\forall \theta \in \Theta$ with $\theta \neq \theta_0$ $\forall \pi \in \Pi$. In Example 1, the assumption that Z_t has a bounded Lebesgue density is used to verify Assumption EP4 with $\psi = 1$. 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 result referred to above is summarized as follows:

THEOREM 2.   *Assumptions* EP1–EP4 *imply Assumptions* 1–5.

Note that under assumptions EP1–EP4, Assumption 5 holds with the Gaussian process $G(\theta_0, \cdot)$ having covariance function given by $EG(\theta_0, \pi_1)G(\theta_0, \pi_2)' = E(\partial/\partial\theta)\log g(W_t, \theta_0, \pi_1)(\partial/\partial\theta')\log g(W_t, \theta_0, \pi_2)$ 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-pseudometric, where r is as in Assumptions EP1(a), EP1(e) and EP4.

3.2. *Changepoint example.* Sufficient conditions for Assumptions 1–3 and 5 for changepoint tests in stationary dynamic nonlinear models are given in AP (Assumption SC). These assumptions are quite similar to standard assumptions in the literature for the consistency and asymptotic normality of ML estimators in stationary contexts. For brevity, these conditions are not restated here.

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

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

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

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.

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

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

where X_{1t} is an element of the regressor X_t. In this example, one is testing

for constancy of the parameters across two (unknown) regions. Multiple regions could be considered.

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 square (LS) estimator of $\theta = (\beta', \delta')'$,

$$\mathbf{I}_T(\theta, \pi) = \frac{1}{\sigma^2} \sum_{t=1}^{T} \begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix} \begin{pmatrix} X_t(\pi) \\ Z_t \end{pmatrix}' \quad \forall \theta \in \Theta = R^{p+q},$$

$\tilde{\theta}$ equal to the restricted LS estimator of θ and $l_T(\theta, \pi)$ and $Dl_T(\theta, \pi)$ 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. In consequence, the test statistics $\sup_{\pi \in \Pi} W_T(\pi)$, $\sup_{\pi \in \Pi} LM_T(\pi)$ and $\sup_{\pi \in \Pi} LR_T(\pi)$ yield equivalent tests.

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

$$(4.4) \qquad \delta = \int (1 - \varphi_T) \xi_T f_T(\theta_0) \, d\mu_T > 0,$$

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.

THEOREM 3. *Suppose Assumptions 4 and 6–8 hold with $\mathbf{I}_\pi = (1/\sigma^2)\sum_{t=1}^{T}(X_t(\pi)', Z_t')'(X_t(\pi)', Z_t')$ 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 $r \geq r_0$,*

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

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

REMARKS.

1. Remarks 1–3 following Theorem 1 all apply to Theorem 3 (with the references to asymptotics deleted).
2. See Andrews, Lee and Ploberger (1996) for an alternative set of finite sample optimality properties for some changepoint tests.

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 $\int f_T(\theta_0 + B_T^{-1}h, \pi) \, dQ_{r,\pi}(h) \, dJ(\pi)$. The case $r = 0$ corresponds to the null density $f_T(\theta_0)$. The likelihood ratio of P_r to P_0 is denoted

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

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

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

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

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

$$(5.3) \qquad \begin{aligned} \bar{\theta}(\pi) &= \mathbf{I}^{-1}(\theta_0, \pi) B_T^{-1} Dl_T(\theta_0, \pi), \\ \overline{W}_T(\pi) &= (H\bar{\theta}(\pi))'[H\mathbf{I}^{-1}(\theta_0, \pi)H']^{-1} H\bar{\theta}(\pi). \end{aligned}$$

The proof of Theorem 1 uses the following lemmas.

LEMMA 1.   *Under Assumptions 1, 2, 5 and 7,*

$$\exp(-r^2/2) \int \psi_p\big(r\overline{W}_T^{1/2}(\pi)\big)\, dJ(\pi)/\mathrm{LR}_{T,r} \to_p 1$$

under P_0.

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

(a) $\psi_p(\lambda) \leq C_1 + C_2 \exp(\lambda)\ \forall\, \lambda \geq 0.$
(b) $\psi_p(\lambda) \geq C_3 \exp(\lambda)\lambda^{-(p-1)/2}\ \forall\, \lambda \geq 1.$

LEMMA 3.   *Under Assumptions 1–5, $\sup|\overline{W}_T(\pi) - F_T(\pi)| \to_p 0$ and $\sup F_T(\pi) \to_d \sup F(\pi) = \sup(HG(\theta_0, \pi))'[H\mathbf{I}^{-1}(\theta_0, \pi)H']^{-1} HG(\theta_0, \pi)$ 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 $k_\alpha\ \forall\, T \geq 1$. For the case of random $k_{T\alpha}$, we must have $k_{T\alpha} \to_p k_\alpha$ 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

$$(5.4) \quad \liminf_{T\to\infty} E_r(1 - \varphi_T)\bigg/\limsup_{T\to\infty} P_r\big(\sup F_T(\pi) \leq k_\alpha\big) \to \infty \quad \text{as } r \to \infty.$$

Below we show that

$$(5.5) \qquad \begin{aligned} &\limsup_{T\to\infty} P_r\big(\sup F_T(\pi) \leq k_\alpha\big) \\ &\qquad \leq 2\exp(-r^2/2)\big[C_1 + C_2 \exp(rk_\alpha^{1/2})\big] \qquad \forall\, r > 0. \end{aligned}$$

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

$$(5.6) \qquad \begin{aligned} &\liminf_{T\to\infty} E_r(1 - \varphi_T) \\ &\qquad \geq C_4 \exp(-r^2/2)\exp\big(r(k_\alpha + \gamma)^{1/2}\big)\big[r(k_\alpha + \gamma)^{1/2}\big]^{-(p-1)/2}. \end{aligned}$$

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

$$(5.7) \quad D^*_{T,r} = \left\{ \exp(-r^2/2) \int \psi_p\left(r\overline{W}^{1/2}_T(\pi)\right) dJ(\pi)/\mathrm{LR}_{T,r} \in [1/2, 2] \right\}.$$

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

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

$$\limsup_{T \to \infty} P_r\left(\sup F_T(\pi) \leq k_\alpha\right)$$

$$= \limsup_{T \to \infty} P_r\left(\sup \overline{W}_T(\pi) \leq k_\alpha, D^*_{T,r}\right)$$

$$= \limsup_{T \to \infty} E_0 \mathrm{LR}_{T,r} \cdot 1\left(\sup \overline{W}_T(\pi) \leq k_\alpha, D^*_{T,r}\right)$$

$$\leq 2\exp(-r^2/2)$$

$$(5.8) \qquad \times \limsup_{T \to \infty} E_0 \int \psi_p\left(r\overline{W}^{1/2}_T(\pi)\right) dJ(\pi) \cdot 1\left(\sup \overline{W}_T(\pi) \leq k_\alpha\right)$$

$$\leq 2\exp(-r^2/2) \limsup_{T \to \infty} E_0 \int \left[C_1 + C_2 \exp\left(r\overline{W}^{1/2}_T(\pi)\right)\right] dJ(\pi)$$

$$\times 1\left(\sup \overline{W}_T(\pi) \leq k_\alpha\right)$$

$$\leq 2\exp(-r^2/2)\left[C_1 + C_2 \exp\left(rk^{1/2}_\alpha\right)\right],$$

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

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

where δ is as in the definition of asymptotically distinct. This follows because

$$(5.10) \qquad \begin{aligned} &\limsup_{T \to \infty} P_0\left(\sup F_T(\pi) \in (k_\alpha, k_\alpha + 2\gamma]\right) \\ &\qquad = P_0\left(\sup F(\pi) \in (k_\alpha, k_\alpha + 2\gamma]\right) \leq \delta/2, \end{aligned}$$

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

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

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

In consequence,

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

Under the assumptions, given $\gamma > 0$, $\exists \, \varepsilon_1 > 0$ and $\xi > 0$ such that

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

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

$$(5.14) \quad D_{T, r} = \Big\{ 1 - \varphi_T > \delta/8, \sup \overline{W}_T(\pi) > k_\alpha + 2\gamma, \overline{\theta} \in K(\varepsilon_1), D^*_{T, r} \Big\}.$$

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

Using Lemma 1, we now have

$$
\begin{aligned}
E_r(1 - \varphi_T) &\geq \frac{\delta}{8} P_r\left(1 - \varphi_T > \frac{\delta}{8}\right) \\[2mm]
&\geq \frac{\delta}{8} P_r(D_{T, r}) = \frac{\delta}{8} E_0 \mathrm{LR}_{T, r} \cdot 1(D_{T, r}) \\[2mm]
(5.15) \qquad &\geq \frac{\delta}{16} \exp\left(\frac{-r^2}{2}\right) E_0 1(D_{T, r}) \int \psi_p\big(r\overline{W}_T^{1/2}(\pi)\big) \, dJ(\pi) \\[2mm]
&\geq \frac{\delta}{16} \exp\left(\frac{-r^2}{2}\right) E_0 1(D_{T, r}) \int 1\big(\pi \in S(\hat{\pi}, \xi)\big) \psi_p\big(r\overline{W}_T^{1/2}(\pi)\big) \, dJ(\pi) \\[2mm]
&\geq \frac{\delta}{16} \exp\left(\frac{-r^2}{2}\right) E_0 1(D_{T, r}) \inf_{\pi \in S(\hat{\pi}, \xi)} \psi_p\big(r\overline{W}_T^{1/2}(\pi)\big) \inf_{\pi \in \Pi} J(S(\pi, \xi)).
\end{aligned}
$$

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

$$(5.16) \qquad r\overline{W}_T^{1/2}(\pi)_\omega \geq r\left(\sup_{\pi \in \Pi} \overline{W}_T(\pi)_\omega - \gamma\right)^{1/2} \geq r(k_\alpha + \gamma)^{1/2}.$$

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, (5.15) and (5.16) combine to yield

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

The desired result follows if $\liminf_{T\to\infty} P_0(D_{T,r}) \geq \delta/8 \ \forall \ r \geq 0$. Inequalities (5.9) and (5.12) yield

$$\liminf_{T\to\infty} E_0(1 - \varphi_T) 1\big(\sup F_T(\pi) > k_\alpha + 2\gamma, \bar{\theta} \in K(\varepsilon_1)\big)$$

(5.18)

$$\geq \frac{\delta}{2} - \limsup_{T\to\infty} P_0\big(\bar{\theta} \notin K(\varepsilon_1)\big) \geq \frac{\delta}{4}.$$

This result, Lemma 1, and Lemma 3 give

$$\delta/4 \leq \liminf_{T\to\infty} E_0(1 - \varphi_T) 1\big(\sup \overline{W}_T(\pi) > k_\alpha + 2\gamma, \bar{\theta} \in K(\varepsilon_1)\big)$$

$$\leq \liminf_{T\to\infty} P_0\big(1 - \varphi_T > \delta/8, \sup \overline{W}_T(\pi) > k_\alpha + 2\gamma, \bar{\theta} \in K(\varepsilon_1)\big)$$

(5.19)

$$+ \limsup_{T\to\infty} E_0(1 - \varphi_T) 1(1 - \varphi_T \leq \delta/8)$$

$$\leq \liminf_{T\to\infty} P_0(D_{T,r}) + \delta/8. \qquad \square$$

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

(5.20) $\quad \omega_T(r) = \sup_{\substack{\pi \in \Pi, h: h'\mathbf{I}h \leq r^2, \\ 0 < \lambda < 1}} \|\mathbf{I}_T\big(\theta_0 + \lambda B_T^{-1}h, \pi\big) - \mathbf{I}_T(\theta_0, \pi)\|.$

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

$$l_T\big(\theta_0 + B_T^{-1}h, \pi\big) - l_T(\theta_0) = h'G_T - h'\mathbf{I}_T h/2 + R_T,$$

(5.21) $\quad$ where $|R_T| \leq \omega_T(r)\|h\|^2$

$$\leq \omega_T(r) r^2 / \inf_{\pi \in \Pi} \lambda_{\min}(\mathbf{I}(\theta_0, \pi)) = C_r \omega_T(r)$$

for some constant $C_r < \infty$.

By Assumption 1(d) and (e), $\omega_T(r) \to_p 0 \ \forall \ r > 0$ [by adding in and subtracting out both $\mathbf{I}(\theta_0 + \lambda B_T^{-1}h, \pi)$ and $\mathbf{I}(\theta_0, \pi)$ and then applying the triangle inequality]. Hence, $\forall \ \pi \in \Pi$ and $\forall \ h$ with $h'\mathbf{I}h \leq r^2$, we have

$$K_{1T} = \exp\big(-C_r \omega_T(r)\big)$$

(5.22)

$$\leq \exp\big(h'G_T - h'\mathbf{I}_T h/2\big) / \big[f_T\big(\theta_0 + B_T^{-1}h, \pi\big) / f_T(\theta_0)\big]$$

$$\leq \exp\big(C_r \omega_T(r)\big) = K_{2T},$$

where $K_{1T} \to_p 1$ and $K_{2T} \to_p 1$. In turn, this yields

(5.23) $\quad K_{1T} \leq \int \exp\big(h'G_T - h'\mathbf{I}_T h/2\big) \, dQ_{r,\pi}(h) \, dJ(\pi) / \text{LR}_{T,r} \leq K_{2T}.$

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

$$(5.24)\quad K_{3Tr} \le \int \exp(h'G_T - h'\mathbf{I}h/2)\, dQ_{r,\pi}(h)\, dJ(\pi)/\mathrm{LR}_{T,r} \le K_{4Tr}.$$

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

$$\int \exp(h'G_T - h'\mathbf{I}h/2)\, dQ_{r,\pi}(h)\, dJ(\pi)$$

$$= \exp(-r^2/2)\int \exp\big(h'\mathbf{I}A_\pi H\bar\theta(\pi)$$

$$\qquad\qquad + h'\mathbf{I}(I_s - A_\pi H)\bar\theta(\pi)\big)\, dQ_{r,\pi}(h)\, dJ(\pi)$$

$$(5.25)\qquad = \exp(-r^2/2)\int \exp\big(h'\mathbf{I}A_\pi H\bar\theta(\pi)\big)\, dQ_{r,\pi}(h)\, dJ(\pi)$$

$$= \exp(-r^2/2)\int \exp\big(rx'(A'_\pi \mathbf{I}A_\pi)^{1/2}H\bar\theta(\pi)\big)\, dU_p(x)\, dJ(\pi)$$

$$= \exp(-r^2/2)\int \exp\big(r\overline{W}_T^{1/2}(\pi)x'\mu\big)\, dU_p(x)\, dJ(\pi)$$

$$= \exp(-r^2/2)\int \psi_p\big(r\overline{W}_T^{1/2}(\pi)\big)\, dJ(\pi),$$

where the third equality uses Assumption 7,

$$\mu = (A'_\pi \mathbf{I}A_\pi)^{1/2}H\bar\theta(\pi)/\|(A'_\pi \mathbf{I}A_\pi)^{1/2}H\bar\theta(\pi)\|$$

and

$$\overline{W}_T(\pi) = \|(A'_\pi \mathbf{I}A_\pi)^{1/2}H\bar\theta(\pi)\|^2$$

since $A'_\pi \mathbf{I}A_\pi = (H\mathbf{I}^{-1}H')^{-1}$ by straightforward algebra.

Equations (5.24) and (5.25) combine to give the desired result. □

PROOF OF LEMMA 2.   This lemma follows straightforwardly from results in the literature. For example, it follows from (15.3.7), (15.3.9), and the last equation on page 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,

$$\text{(5.26)} \qquad \sup \|\overline{W}_T(\pi) - F_T(\pi)\| \to_p 0 \quad \text{under } P_0$$

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), page 70], (2.5) and (5.26), 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, for example, from Theorems 16.8 and 18.11 of Strasser (1985): If (i) $\text{LR}_{T,r} \to_d X_r^*$ under P_0 and (ii) $E_0 X_r^* = 1$, 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

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

$$\text{(5.27)} \qquad = \exp(-r^2/2) \int \left[\int E_0 \exp(rx'Z(\pi)) \, dU_p(x) \right] dJ(\pi)$$

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

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), EP2(a), EP2(c) and EP4(a) using the uniform weak law of large numbers (WLLN) given in the theorem in Andrews (1987). In particular, pointwise WLLN's hold for the infs 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 (1993). We verify its conditions (a) and (b) for $Q_T(\theta, \pi) = -(1/T) \sum_1^T \log g(W_t, \theta, \pi)$ and $Q(\theta, \pi) = -E \log g(W_t, \theta, \pi)$. For Assumption 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), EP2(a), EP2(b) and EP4(a). Condition (b) of Lemma A-1 holds for Assumptions 2 (with parameter space Θ) and 3 (with parameter space $\tilde{\Theta}$) by EP2(a), EP3 and EP4(a).

Assumption 5 holds with $B_T = \sqrt{T} I_s$ by Theorem 1 of Doukhan, Massart and Rio (1995) using EP1(a), EP1(e) and EP4. More specifically, the key condition (2.10) of Theorem 1 of Doukhan, Massart and Rio is implied by their equation (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, using Theorem 5 of Andrews (1994). The latter follows because **M** is a type IV class of functions, as defined in Andrews (1994), and by Theorem 5 satisfies Ossiander's L^r entropy condition, which is equivalent to equation (2.11) of Doukhan, Massart and Rio. □

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. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856.

ANDREWS, D. W. K. (1994). Empirical process methods in econometrics. In *Handbook of Econometrics* **4**, Chapter 2. North-Holland, New York.

ANDREWS, D. W. K., LEE, I. and PLOBERGER, W. (1996). Optimal changepoint tests for normal linear regression. *J. Econometrics* **70**. To appear.

ANDREWS, D. W. K. and PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383–1414.

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

CHANG, J. T. and HARTIGAN, J. A. (1993). Admissibility of likelihood ratio tests for detecting a change in distribution. Unpublished manuscript, Dept. Statistics, Yale Univ.

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 Probab. Appl.* **28** 313–328.

DOUKHAN, P. (1994). *Mixing: Properties and Examples. Lecture Notes in Statist.* **85**. Springer, New York.

DOUKHAN, P., MASSART, P. and RIO, E. (1995). Principes d'invariance pour la mesure empirique d'un processus absolument régulier. *Ann. Inst. H. Poincaré Probab. Statist.* **31** 393–427.

LIFSHITS, M. A. (1982). On the absolute continuity of distributions of functionals of random processes. *Theory Probab. Appl.* **27** 600–607.

MARDIA, K. V., KENT, J. T. and BIBBY, J. M. (1979). *Multivariate Analysis*. Academic Press, New York.

MOKKADEM, A. (1986). Sur le mélange d'un processus ARMA vectoriel. *C. R. Acad. Sci. Paris Ser. I Math.* **303** 519–521.

MOKKADEM, A. (1990). Propriétés de mélange des modèles autorégressifs polynomiaux. *Ann. Inst. H. Poincaré* **26** 219–260.

OSSIANDER, M. (1987). A central limit theorem under metric entropy with L^2-bracketing. *Ann. Probab.* **15** 897–919.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

POTTER, S. M. (1995). A nonlinear approach to U.S. GNP. *Journal of Applied Econometrics* **10** 109–125.

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

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

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

COWLES FOUNDATION FOR
   RESEARCH IN ECONOMICS
YALE UNIVERSITY
NEW HAVEN, CONNECTICUT 06520-8281

INSTITUT FÜR ÖKONOMETRIE,
   OPERATIONS RESEARCH UND
   SYSTEMTHEORIE
TECHNISCHE UNIVERSITÄT WIEN
ARGENTINTERSTRASSE 8 / 119
A-1040 WIEN
AUSTRIA