OPTIMAL TESTS WHEN A NUISANCE PARAMETER IS PRESENT ONLY UNDER THE ALTERNATIVE

by

Donald W. K. Andrews and Werner Ploberger

COWLES FOUNDATION PAPER NO. 879

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

1994

Econometrica, Vol. 62, No. 6 (November, 1994), 1383–1414

OPTIMAL TESTS WHEN A NUISANCE PARAMETER IS PRESENT ONLY UNDER THE ALTERNATIVE

BY DONALD W. K. ANDREWS AND WERNER PLOBERGER[1]

This paper derives asymptotically optimal tests for testing problems in which a nuisance parameter exists under the alternative hypothesis but not under the null. For example, the results apply to tests of one-time structural change with unknown changepoint. Several other examples are discussed in the paper. The results of the paper are of interest, because the testing problem considered is nonstandard and the classical asymptotic optimality results for the Lagrange multiplier (LM), Wald, and likelihood ratio (LR) tests do not apply.

A weighted average power criterion is used here to generate optimal tests. This criterion is similar to that used by Wald (1943) to obtain the classical asymptotic optimality properties of Wald tests in "regular" testing problems. In fact, the optimal tests introduced here reduce to the standard LM, Wald, and LR tests when standard regularity conditions hold. Nevertheless, in the nonstandard cases of main interest, new optimal tests are obtained and the LR test is not found to be an optimal test.

KEYWORDS: Asymptotics, changepoint, exponential average test, multiple changepoint test, nonstandard testing problem, optimal test, structural change test, test of common factors, test of cross-sectional constancy, test of variable relevance, threshold autoregressive model.

1. INTRODUCTION

THIS PAPER CONSIDERS THE NONSTANDARD PROBLEM of testing whether a subvector 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 hypothesis. A variety of testing problems of interest in econometrics are of the above type. Examples include tests of one-time structural change, of multiple structural changes, of cross-sectional constancy of parameters, of the threshold effect in threshold autoregressive models, of variable relevance and functional form in nonlinear models, of common factors in autoregressive-moving average models, and of conditional heteroskedasticity in GARCH models. In the structural change case, for example, the parameter π that appears under the alternative, but not under the null, is the time of structural change.

The purpose of this paper is to derive asymptotically optimal tests for the testing problems described above. This is of interest because the classical asymptotic optimality properties of Lagrange multiplier (LM), Wald, and likelihood ratio (LR) tests do not hold in these nonstandard problems.

[1] This paper is an extension of an earlier working paper circulated under the title "Optimal Tests of Parameter Constancy." The authors wish to thank Inpyo Lee for computing the critical values given in Tables I and II. They also thank Christian Gourieroux, Bruce Hansen, several referees, and the co-editor for helpful comments. Andrews gratefully acknowledges research support from the National Science Foundation through Grant Numbers SES-8821021 and SES-9121914. Ploberger gratefully acknowledges research support from Fonds zur Förderung der wissenschaftlichen Forschung under Schrödinger-stipendium Project J-0469-PHY.

To derive optimal tests, we use a weighted average power criterion function similar to that used by Wald (1943). In fact, for any fixed value of π, the weight function we consider has the same contours as those considered by Wald. One difference is that we consider multiple values of π under the alternative, whereas Wald's results are applicable only for a single value. The optimal tests that are derived depend on the choice of a weight function J over the values of π. A second difference is that Wald allows for arbitrary weightings of the contours referred to above, whereas we have to specify a weight function over the contours.

The optimal tests that we derive can be given by Bayesian interpretation. If one views the weight functions referred to above as priors, then the optimal tests are of a Bayesian posterior odds ratio form or, more precisely, are asymptotically equivalent to a Bayesian posterior odds ratio. The optimal test statistics have two advantages over an actual Bayesian posterior odds ratio. First, they circumvent the need for placing a prior over those nuisance parameters that appear under both the null and the alternative. Second, they are much easier to compute, partly as a consequence of the first advantage.

The optimal tests are of an *average exponential* form. In particular, for a fixed value of π, let $LM_T(\pi)$ denote the standard LM test statistic for testing $\beta = 0$ against the alternative that $\beta \neq 0$ and that the value in Π that is true is π. For example, in the one-time structural change example, π denotes the time of structural change as a fraction of the sample size, $(\delta_1', \delta_2')'$ denotes the true parameter vector before structural change, $(\delta_1' + \beta', \delta_2')'$ denotes its values after structural change, and $LM_T(\pi)$ is the LM test statistic for testing $\beta = 0$ against the alternative that $\beta \neq 0$ and change occurs at time π. Returning to the general case, an asymptotically optimal test in terms of weighted average power in the class of all tests of asymptotic significance level α is based on the statistic

$$(1.1) \qquad \text{Exp-}LM_T = (1+c)^{-p/2} \int \exp\left(\frac{1}{2} \frac{c}{1+c} LM_T(\pi) \right) dJ(\pi).$$

Here, p is the dimension of β, $J(\cdot)$ is the weight function over values of π in Π (such as uniform on $[\pi_0, 1-\pi_0]$ for some $\pi_0 > 0$ in the one-time structural change case), and c is a scalar constant that depends on the chosen weight function over values of β and determines whether one is directing power against close or more distant alternatives. Exponential Wald and LR tests are defined analogously to $\text{Exp-}LM_T$ with the standard $W_T(\pi)$ and $LR_T(\pi)$ test statistics replacing $LM_T(\pi)$. The exponential LM and LR tests are also found to be asymptotically optimal tests.

The likelihood ratio test is of the form $\sup_{\pi \in \Pi} LR_T(\pi)$, which is not of the optimal average exponential form. It is found to be a limit of an average exponential test, but only if one considers the limit as a parameter is pushed beyond an admissible boundary. Thus, the likelihood ratio test is not found to be an optimal test using the weight functions considered here. Nevertheless, it is asymptotically admissible; see Andrews and Ploberger (1993).

The general optimality properties of exponential tests are established here under a set of high-level assumptions. Primitive sufficient conditions are given for the examples of tests of one-time structural change with unknown change point, tests of cross-sectional constancy, tests of a threshold effect in autoregressive models, tests of variable relevance, and tests of functional form. For tests of common factors in ARMA(1, 1) models, primitive sufficient conditions are given in Andrews and Ploberger (1994). We note that tests of regime switching in switching models with unobserved regimes (which includes tests of homogeneity in mixture models) are not covered by the results of this paper.[2]

The papers most closely related to the present one are Davies (1977, 1987), Hansen (1991), and King and Shively (1993). Davies (1977) has established the asymptotic optimality as the sample size T goes to infinity and the significance level α goes to zero of the likelihood ratio test (i.e., the sup LR test) in the context considered here. His results for scalar parameters are extended to vector-valued parameters in Andrews (1993). These optimality results are very weak, however, and are not indicative of finite sample performance. The reason is that the power of the likelihood ratio test with unknown π is equivalent to that with known π when $T \to \infty$ and $\alpha \to 0$. This equivalence is not found even approximately with typical sample sizes and significance levels.

Hansen (1991) does not discuss optimal choices of tests. He does, however, suggest a useful computational method for simulating critical values that can be exploited in some of the examples considered here. King and Shively (1993) consider locally mean most powerful tests for problems of the sort considered in this paper. They employ a transformation of parameters, which provides a useful alternative perspective on the testing problems under study. Their tests direct power only against very local alternatives—alternatives that are so close to the null that only trivial power is obtained asymptotically. In contrast, the alternatives considered in this paper are local, but are such that the tests have nontrivial power even asymptotically.

In the particular context of tests of structural change, there are several papers that are related to the present paper. These include Chernoff and Zacks (1964), Jandhyala and MacNeill (1991), and Nyblom (1989). For brevity, we do not discuss these papers here. (See Andrews and Ploberger (1992) for a brief discussion of them.)

The remainder of this paper is organized as follows. Section 2 introduces the testing problem under consideration and the optimal test statistics. Section 3 presents and discusses the assumptions employed, and optimality results, and the asymptotic null distribution of the optimal test statistics. Section 4 treats the example of tests of structural change in nonlinear models with stationary observations. It also treats examples that fall within the class of *empirical*

[2] Depending on the chosen parameterization of the model, the reason these models are not covered is that either the alternative hypothesis is one-sided, whereas we consider two-sided alternatives in this paper, or the information matrix for θ given $\pi \in \Pi$ is singular for θ in the null, which violates one of our regularity conditions (Assumption 1(f)).

process applications of the general results. These include tests of cross-sectional constancy, threshold effects, variable relevance, and functional form. An Appendix contains proofs of the results of the paper.

2. DEFINITION OF THE OPTIMAL TESTS

In this section we consider the general problem 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. We introduce tests that we call the (*average*) *exponential LM*, *Wald*, and *LR* tests, denoted $Exp\text{-}LM_T$, $Exp\text{-}W_T$, and $Exp\text{-}LR_T$ respectively. The asymptotic properties of these three tests are the same under the null and local alternatives. Hence, for convenience, we focus the discussion on the $Exp\text{-}LM_T$ test statistic.

We begin by introducing some notation and definitions. Let $(\Omega, \mathscr{F}, \boldsymbol{P})$ denote a probability space on which all of the random elements introduced below are defined. Let $\boldsymbol{Y}_T$ denote the data matrix when the sample size is T for $T = 1, 2, \ldots$. Consider a parametric family $\{f_T(\boldsymbol{y}_T, \theta, \pi): \theta \in \Theta, \pi \in \Pi\}$ of densities of $\boldsymbol{Y}_T$ with respect to some σ-finite measure μ_T, where $\Theta \subset R^s$ and Π is some topological space (usually a subset of Euclidean space). The likelihood function of the data is given by $f_T(\theta, \pi) = f_T(\boldsymbol{Y}_T, \theta, \pi)$. In many cases, the likelihood function $f_T(\theta, \pi)$ can be written as a product of two terms, one that depends on θ and another that does not. Often the latter term is the product over $t = 1, \ldots, T$ of the conditional distribution of some weakly exogenous variables at time t given all of the preceding variables (exogenous or not). In such cases, these conditional distributions of the weakly exogenous variables need not be known in order for one to construct the test statistics considered here. The optimality results stated below hold for any such distributions for which the assumptions on $f_T(\theta, \pi)$ hold.

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 the case of tests of one-time structural change, the parameter $\pi \in (0, 1)$ indicates the point of structural change as a fraction of the sample size, 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

$$(2.1) \qquad H_0: \beta = \mathbf{0}.$$

In the structural change case, this is the hypothesis of no structural change. The alternative hypothesis is

$$(2.2) \qquad H_1: \beta \neq \mathbf{0} \text{ and the likelihood function depends on the parameter } \pi.$$

We let θ_0 denote a parameter vector in 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 $f_T(\theta_0)$. For example, in the one-time structural change case, if no structural change occurs, the time π of structural change is redundant. It is the appearance of the parameter π under the alternative hypothesis, but not under the null, that makes the testing problem described above nonregular and outside the domain of standard asymptotic optimality results. In particular, the standard LM statistic does not have its standard asymptotic distribution or its standard asymptotic local optimality properties in the situation described above.

To derive asymptotically optimal tests of H_0, we consider local alternatives to H_0 of the form $f_T(\theta_0 + B_T^{-1}h, \pi)$ for some $\pi \in \Pi$, some $h \in R^s$, and some nonrandom $s \times s$ diagonal matrix B_T that satisfies $[B_T^{-1}]_{jj} \to 0$ as $T \to \infty$ $\forall j \leqslant s$. (In models with nontrending variables, $B_T = \sqrt{T} I_s$, where I_s is the $s \times s$ identity matrix.) For particular integrable weight functions (i.e., probability measures) $Q_\pi(h)$ on the values of h and a chosen integrable weight function $J(\pi)$ on the values of π, we show that the test *Exp-LM*$_T$ has the greatest weighted average power asymptotically in the class of all tests of asymptotic significance level α. That is, this test maximizes

$$\overline{\lim_{T\to\infty}} \int P\left(\varphi_T \text{ rejects} | \theta_0 + B_T^{-1}h, \pi\right) dQ_\pi(h)\, dJ(\pi) \tag{2.3}$$

over all tests φ_T of asymptotic level α (and the $\overline{\lim}_{T\to\infty}$ equals $\lim_{T\to\infty}$ for the test *Exp-LM*$_T$). Furthermore, if one considers the local alternative density $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h) dJ(\pi)$ to $f_T(\theta_0)$, then the test *Exp-LM*$_T$ has greatest power asymptotically against this alternative in the class of all tests of asymptotic level α.

The asymptotically optimal test statistic *Exp-LM*$_T$ is defined by

$$\textit{Exp-LM}_T = (1+c)^{-p/2} \int \exp\left(\frac{1}{2}\frac{c}{1+c} LM_T(\pi)\right) dJ(\pi), \tag{2.4}$$

where $LM_T(\pi)$ is just the standard LM test statistic for testing H_0 versus H_1 given the parameter π and $c > 0$ is a scalar constant that depends on the weight functions $Q_\pi(\cdot)$. For example, for the case of one-time structural change, $LM_T(\pi)$ is just the standard LM test of structural change occurring at the time $[T\pi]$. One rejects H_0 if *Exp-LM*$_T$ exceeds a critical value $k_{T\alpha}$ that is determined using the asymptotic null distribution of *Exp-LM*$_T$. The *Exp-W*$_T$ and *Exp-LR*$_T$ statistics are defined analogously.

Note that *Exp-LM*$_T$ depends on the weight functions $Q_\pi(\cdot)$ only through the scalar c. The larger is c, the more weight is given to alternatives for which β is large, where $\theta = (\beta', \delta_0')'$. For example, for tests of structural change, larger values of c correspond to greater weight being given to larger structural changes. In the special case where $J(\pi)$ is a pointmass at a single value π_0, *Exp-LM*$_T$ reduces to $(1+c)^{-p/2}\exp((1/2)(c/(1+c))LM_T(\pi_0))$, the optimal test rejects if and only if $LM_T(\pi_0)$ exceeds some constant (i.e., the optimal test equals the standard LM test for fixed π_0), and the optimal test is independent of c. When $J(\pi)$ is not a pointmass distribution, however, the optimal test *Exp-LM*$_T$ depends on c. The larger is c, the more power is directed at alternatives for which β is large.

The limit as $c \to 0$ of the exponential LM statistic (suitable normalized) is equal to the "average LM" statistic. In particular,

$$\lim_{c \to 0} 2(\mathit{Exp\text{-}LM}_{Tc} - 1)/c = \int LM_T(\pi)\, dJ(\pi), \tag{2.5}$$

where $\mathit{Exp\text{-}LM}_{Tc}$ denotes the statistic $\mathit{Exp\text{-}LM}_T$ constructed using the constant c. Thus, the average LM statistic is the limit of the exponential LM statistics that are designed for alternatives that are very close to the null hypothesis. For different models, the average LM statistic has been considered previously in the literature by Chernoff and Zacks (1964), Nyblom (1989), Jandhyala and MacNeill (1991), and Hansen (1992) among others.

At the other extreme, the limit as $c \to \infty$ of the exponential LM statistic (suitably normalized) is given by

$$\lim_{c \to \infty} \log\left((1+c)^{p/2}\mathit{Exp\text{-}LM}_{Tc}\right) = \log \int \exp\left(\frac{1}{2} LM_T(\pi)\right) dJ(\pi). \tag{2.6}$$

Thus, for testing against more distant alternatives the optimal test statistic is still of an average exponential form. This statistic has not been considered previously in the literature.

We note that if the constant $c/(1+c)$, which appears in the definition of $\mathit{Exp\text{-}LM}_T$, is replaced by a constant $r > 0$, then the limit as $r \to \infty$ of the exponential LM statistic (suitably normalized) is the "sup LM" statistic. More specifically, let $\Pi^* \subset \Pi$ be the support of $J(\cdot)$. Then,

$$\lim_{r \to \infty} (\log \mathit{Exp\text{-}LM}_T^r)/r = \sup_{\pi \in \Pi^*} LM_T(\pi), \tag{2.7}$$

where $\mathit{Exp\text{-}LM}_T^r$ denotes the statistic $\mathit{Exp\text{-}LM}_T$ with $c/(1+c)$ replaced by r and $(1+c)^{p/2}$ replaced by 1. Hence, the sup LM test is designed for distant alternatives, but is of a more extreme form than the optimal exponential test, since the latter requires $r < 1$. The sup LM test has been considered in the literature by Andrews (1993), among others.

3. ASSUMPTIONS AND OPTIMALITY RESULTS

This section presents the assumptions used and states the optimality results obtained in the paper. We begin by introducing some notation. 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.) Let θ_0 denote the true value of θ under the null H_0.

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. For nontrending data, the matrix B_T is just $\sqrt{T} I_s$. For data with deterministic time trends, B_T is more complicated. For example, in a normal linear regression model with r nontrending regressors plus the regressors t and t^2, B_T equals diag$\{\sqrt{T} I_r, T, T^{3/2}\}$. Note that the matrices $\{B_T\}$ that are suitable for norming $Dl_T(\theta, \pi)$ and $D^2 l_T(\theta, \pi)$ dictate the form of the local alternatives $f_T(\theta_0 + B_T^{-1} h, \pi)$ that we consider, since such alternatives are the ones for which good tests have nontrivial asymptotic power.

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 wp $\to 1$ abbreviate "with probability that goes to 1 as $T \to \infty$."

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 one under θ_0, where Θ_0 is some neighborhood of θ_0.*

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

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

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

The matrix function $\mathscr{I}(\theta, \pi)$ introduced in Assumption 1 is the asymptotic *information matrix* for θ for given π, which depends on both θ and π.

We briefly comment on Assumption 1. Assumption 1(a) specifies the crucial feature of the testing problem under consideration. Assumptions 1(b) and (c) are standard maximum likelihood (ML) assumptions (though ML regularity conditions that do not require differentiability in θ exist). For a fixed value of π, Assumption 1(d) can be verified under standard ML assumptions using a suitable weak law of large numbers (WLLN). Uniform convergence over $\pi \in \Pi$ can then be obtained, e.g., by using the generic uniform convergence results in Andrews (1987, 1992). Assumptions 1(e) and (f) also are standard ML assumptions except that they are required to hold uniformly over values of the nuisance parameter $\pi \in \Pi$. Nevertheless, Assumption 1(f) does not hold even for a fixed value of π in mixture models, or more generally, in regime switching models with unobserved regimes. The uniformity requirements in Assumptions 1(d)–(f) restrict the class Π that can be considered. For example, in the one-time structural change case, uniformity requires that the closure of Π be bounded away from 0 and 1. That is, one cannot consider changepoints that are arbitrarily close to the beginning or end of the sample. In the regime switching example with observed regimes, uniformity requires that Π be such that the probability of a regime occurring is not arbitrarily close to 0 or 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

$$(3.1) \qquad l_T(\hat{\theta}(\pi), \pi) = \max_{\theta \in \Theta} l_T(\theta, \pi) \quad \forall \pi \in \Pi \text{ wp} \to 1 \text{ under } \theta_0.$$

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

$$(3.2) \qquad \tilde{\theta} \in \tilde{\Theta} = \{\theta \in \Theta: \theta = (0', \delta')' \text{ for some } \delta \in R^q\} \quad \text{and}$$

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

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

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

For some applications, these assumptions can be verified using results in the literature. In other cases, one can use a result given in Andrews (1993, Lemma A-1), which provides sufficient conditions for uniform consistency of a family of estimators. These conditions entail uniform convergence of the criterion function to some limit function and a uniform identifiability condition on the limit function.

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

$$(3.3) \qquad LM_T(\pi) = \left(B_T^{-1} Dl_T(\tilde{\theta}, \pi)\right)' \mathscr{I}_T^{-1}(\tilde{\theta}, \pi) B_T^{-1} Dl_T(\tilde{\theta}, \pi),$$

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

$$LR_T(\pi) = -2\left(l_T(\tilde{\theta}, \pi) - l_T(\hat{\theta}(\pi), \pi)\right), \quad \text{where}$$

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

Alternatively, one can define $\mathscr{I}_T(\theta, \pi)$ to be of outer product, rather than Hessian, form. Note that only the first p elements of $B_T^{-1} Dl_T(\tilde{\theta}, \pi)$ are nonzero in the definition of $LM_T(\pi)$, because $\partial l_T(\tilde{\theta}, \pi)/\partial\delta = 0$ by the first order conditions for the restricted estimator $\tilde{\theta}$ (wp $\to$ 1). Also, note that the $LM_T(\pi)$ statistic is constructed using only the restricted ML estimator $\tilde{\theta}$ and, hence, only requires estimation of the model one time. This has considerable computational advantages, especially in nonlinear models.

The exponential statistics, *Exp-LM*$_T$, *Exp-W*$_T$, and *Exp-LR*$_T$, respectively, are defined by combining (2.4) and (3.3).

Next we introduce a particular choice for the weighted functions $\{Q_\pi(\cdot): \pi \in \Pi\}$. For each π, the chosen weight function $Q_\pi(\cdot)$ gives constant weight on the same ellipses in Θ as were considered first by Wald (1943) in his demonstration of the property of asymptotically greatest weighted average power of Wald tests for the (now standard) testing scenario where π is fixed and known. These ellipses are also the same ones over which the power of asymptotically invariant tests are required to be constant when considering locally most powerful invariant tests in the testing scenario where π is fixed and known. The chosen weight functions $Q_\pi(\cdot)$ are natural from a theoretical perspective in that they give equal weight to alternatives that are equally difficult to detect when π is known—no direction away from the null is favored over any other.

Let V denote the linear subspace of R^s defined by

$$(3.4) \qquad V = \{\theta \in R^s: \theta = (0', \delta')' \text{ for some } \delta \in R^q\}.$$

The null hypothesis can be expressed as H_0: $\theta \in \tilde{\Theta} = \Theta \cap V$. For each $\pi \in \Pi$, we consider a weight function $Q_\pi(\cdot)$ on R^s that concentrates on the orthogonal complement of V with respect to the inner product $\langle h, l\rangle_\pi = h'\mathscr{I}(\theta_0, \pi)l$ for $h, l \in R^s$; call it $V_\pi^\perp$. Since V is a q-dimensional subspace of R^s, $V_\pi^\perp$ is a p-dimensional subspace of R^s. Let $\{a_{1\pi}, \dots, a_{p\pi}\}$ be some basis of $V_\pi^\perp$ and define $A_\pi = [a_{1\pi}, \dots, a_{p\pi}] \in R^{s \times p}$. For example (by the proof of Lemma A-3 in the Appendix), one can take

$$(3.5) \qquad A_\pi = \begin{bmatrix} I_p \\ -\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}' \end{bmatrix}, \quad \text{where} \quad \mathscr{I}(\theta_0, \pi) = \begin{bmatrix} \mathscr{I}_{1\pi} & \mathscr{I}_{2\pi} \\ \mathscr{I}_{2\pi}' & \mathscr{I}_{3\pi} \end{bmatrix}$$

for $\mathscr{I}_{1\pi} \in R^{p \times p}$, $\mathscr{I}_{2\pi} \in R^{p \times q}$, and $\mathscr{I}_{3\pi} \in R^{q \times q}$. In consequence,

$$(3.6) \qquad V_\pi^\perp = \left\{ h \in R^s: h = \begin{pmatrix} \lambda \\ -\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}'\lambda \end{pmatrix} \text{ for some } \lambda \in R^p \right\}.$$

Next, define

$$(3.7) \qquad \Sigma_\pi = A_\pi \mathscr{I}(A_\pi'(\theta_0, \pi)A_\pi)^{-1}A_\pi'$$

$$= \begin{bmatrix} (\mathscr{I}_{1\pi} - \mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}')^{-1} & \Sigma_{\pi 12} \\ \Sigma_{\pi 12} & \mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}'(\mathscr{I}_{1\pi} - \mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}')^{-1}\mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1} \end{bmatrix}$$

where $\Sigma_{\pi 12} = -(\mathscr{I}_{1\pi} - \mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}')^{-1}\mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}$.

Let $N(0, \Sigma)$ denote a multivariate normal distribution with mean 0 and covariance matrix Σ (possibly singular). We make the following assumption.

ASSUMPTION 4: *$Q_\pi = N(0, c\Sigma_\pi)$ $\forall \pi \in \Pi$ for some positive constant c (that does not depend on π).*

Under Assumption 4, the weight function Q_π on R^s is a singular multivariate normal distribution with covariance matrix of rank p. (Its covariance matrix is nonsingular only in the unusual case where $p = s$. In the latter case, there are no unknown parameters under the null.) The support of Q_π is $V_\pi^\perp$.

We now determine the asymptotic null distribution of $Exp\text{-}LM_T$. Let "$\xrightarrow{d}$" denote convergence in distribution. Let "$\Rightarrow$" denote weak convergence of stochastic processes indexed by $\pi \in \Pi$. Note that the definition of weak convergence requires the specification of a metric on the appropriate space of functions on Π. Below we consider weak convergence of a process $\nu_T(\pi) = (\nu_{1T}(\pi), \nu_{2T}(\pi))$ $(\in R^s \times R^{s\times s})$ to a process $\nu(\pi) = (\nu_1(\pi), \nu_2(\pi))$. We assume that the metric on the space of functions in which $\nu_T(\cdot)$ and $\nu(\cdot)$ lie is chosen such that the function

$$(3.8) \qquad \nu(\cdot) \to (1+c)^{-p/2} \times \int \exp\left(\frac{1}{2}\frac{c}{1+c}(H\nu_1(\pi))'(H\nu_2(\pi)H')^{-1}H\nu_1(\pi)\right) dJ(\pi)$$

is continuous with $\nu(\cdot)$-probability one when $\nu(\cdot)$ has bounded uniformly continuous sample paths with probability one. This holds, for example, if the uniform metric is used, as in Pollard (1984), or if the Skorohod metric is used in the case where $\Pi \subset [0,1]$ 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 $EG(\theta_0, \pi)G(\theta_0, \pi)' = \mathscr{I}(\theta_0, \pi)$ $\forall \pi \in \Pi$ and has bounded uniformly 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 for a partial sum process, as with tests of structural change, by applying an empirical process CLT, as with the other examples of Section 4, or by applying some other functional CLT, as with the test of common factors mentioned in Section 1. In contrast to the variance function of $G(\theta_0, \cdot)$, the covariance function of $G(\theta_0, \cdot)$ is not specified in Assumption 5, because no assumptions on it are required.

Note that the stochastic processes $\nu_T(\pi)$ and $\nu(\pi)$ referred to above correspond to $(\mathscr{I}^{-1}(\theta_0, \pi)B_T^{-1}Dl_T(\theta_0, \pi), \mathscr{I}^{-1}(\theta_0, \pi))$ and $(\mathscr{I}^{-1}(\theta_0, \pi)G(\theta_0, \pi), \mathscr{I}^{-1}(\theta_0, \pi))$ respectively. Under Assumptions 1 and 5, the latter process satisfies the conditions on $\nu(\pi)$ stated above for the continuity of the function defined in (3.8).

The asymptotic null distribution of $Exp\text{-}LM_T$ is shown in the following theorem to equal that of the random variable

$$(3.9) \qquad \chi(\theta_0, c) = (1+c)^{-p/2} \times \int \exp\left(\frac{1}{2}\frac{c}{1+c}\left(H\mathscr{J}^{-1}(\theta_0,\pi)G(\theta_0,\pi)\right)' \times \left(H\mathscr{J}^{-1}(\theta_0,\pi)H'\right)^{-1}H\mathscr{J}^{-1}(\theta_0,\pi)G(\theta_0,\pi)\right) dJ(\pi).$$

THEOREM 1: *Under the null hypothesis and Assumptions* 1–5, (a) $Exp\text{-}LM_T \xrightarrow{d} \chi(\theta_0, c)$, (b) $Exp\text{-}W_T \xrightarrow{d} \chi(\theta_0, c)$, *and* (c) $Exp\text{-}LR_T \xrightarrow{d} \chi(\theta_0, c)$.

COMMENTS: 1. In many applications, e.g., structural change applications, the limit distribution $\chi(\theta_0, c)$ does not depend on θ_0. Hence, one can obtain critical values for the exponential LM test by simulating the distribution $\chi(\theta_0, c)$; see Section 4 below. In other applications, $\chi(\theta_0, c)$ does depend on θ_0. In such cases, one can obtain asymptotically valid critical values by simulating $\chi(\theta^*, c)$, where θ^* is some estimator of θ that is consistent under the null, provided $G(\theta_0, \pi)$ is continuous at θ_0 uniformly over $\pi \in \Pi$. See Hansen (1991, Sec. 7) for a method of simulating a realization of $\chi(\theta^*, c)$.

2. Here and below, Assumption 3 is not required for results that involve $Exp\text{-}W_T$.

3. The "$c = 0$" and "$c = \infty$" exponential test statistics are defined in equations (2.5) and (2.6). By the proof of Theorem 1, it is easy to see that under Assumptions 1–5 these statistics have asymptotic distributions as $T \to \infty$ given by the limits as $c \to 0$ and $c \to \infty$, respectively, of the corresponding normalized $\chi(\theta_0, c)$ rv. That is, the limits are $\int z(\pi)\, dJ(\pi)$ and $\log \int \exp(\frac{1}{2}z(\pi))\, dJ(\pi)$, respectively, where $z(\pi)$ is the quadratic form in the exponent in (3.9) (excluding $(\frac{1}{2})c/(1+c)$).

We now state the weighted average power optimality property of the exponential test. 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) and rejects H_0 with probability γ when $\varphi_T = \gamma$. The test φ_T is of asymptotic significance level α if $\int \varphi_T f_T(\theta_0)\, d\mu_T \to \alpha$ for all θ_0 that satisfy the null hypothesis H_0, where $\int \varphi_T f_T(\theta_0)\, d\mu_T$ denotes the probability of rejection of H_0 using φ_T. Similarly, the power of φ_T against the local alternative $f_T(\theta_0 + B_T^{-1}h, \pi)$ is denoted $\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi)\, d\mu_T$.

Let $\{k_{T\alpha}: T \geqslant 1\}$ be a sequence of critical values (possibly random, but with nonrandom probability limit) such that the exponential LM test, i.e., $\xi_T = 1(Exp\text{-}LM_T > k_{T\alpha})$, has asymptotic level α.

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

THEOREM 2: *Under Assumptions* 1–5, *for any sequence of asymptotically level* α *tests* $\{\varphi_T: T \geqslant 1\}$, *a sequence of asymptotically level* α *exponential LM* (*Wald or LR*) *tests* $\{\xi_T: T \geqslant 1\}$ *satisfies*

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

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

COMMENTS: 1. The asymptotic optimality result of Theorem 2 can be interpreted in two ways. First, it provides a greatest asymptotic weighted average power result for the exponential LM test against the alternatives $\{f_T(\theta_0 + B_T^{-1}h, \pi): h \in R^s,\ \pi \in \Pi\}$ for $T \geqslant 1$. Second, it shows that the exponential LM test has the greatest asymptotic power against the single sequence of local alternatives $\{\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi): T \geqslant 1\}$ amongst all tests of asymptotic level α. This follows from Theorem 2 because

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

by Fubini's Theorem.

2. Theorem 2 can be used to state some asymptotic optimality properties of the $c = 0$ and $c = \infty$ exponential test statistics defined in (2.5) and (2.6). In particular, given any $\varepsilon > 0$, the $c = 0$ test is within ε of maximizing the weighted average asymptotic (as $T \to \infty$) power for the weight function $Q_\pi (= Q_\pi^c)$ and J for all c sufficiently small (in the class of tests with the same asymptotic significance level). Thus, the $c = 0$ test is designed for testing against alternatives that are very near the null hypothesis. The $c = \infty$ test possesses analogous optimal weighted average power properties for weight functions for which c is large. Note that the weight functions give more weight to distant alternatives as c is increased and equal weight is given to all alternatives in the limit as $c \to \infty$.[3]

3. The optimality results of Theorem 2 only apply in correctly specified ML contexts. More specifically, the distribution of any weakly exogenous variables does not need to be specified, but the parametric families of conditional

[3] We note that the $c = 0$ and $c = \infty$ exponential test statistics defined in (2.5) and (2.6) differ from the statistic $Exp\text{-}LM_{Tc}$ evaluated at $c = 0$ and $c = \infty$ (which is degenerate and equals 1 and 0 respectively). When $c = 0$, the weight function $Q_\pi (= Q_\pi^c)$ puts all its mass on the null hypothesis, so it is nonsensical to try to generate an optimal test statistic for this weight function. When $c = \infty$, the weight function Q_π is flat on R^p and, hence, is not integrable. Again, it is nonsensical to try to generate an optimal test for this weight function, because the weighted average power for many tests is infinite. It is interesting, and useful, however, to find that it is possible to obtain nondegenerate limiting test statistics as $c \to 0$ and $c \to \infty$ (by normalizing the test statistics before taking the limits). The resulting tests have the optimality properties referred to above.

distributions of the remaining variables must be correctly specified. This is analogous to the case of classical tests in standard testing scenarios. On the other hand, the exponential test (or a version of it with adjusted weight matrix in $LM_T(\pi)$) may have the correct asymptotic significance level under model misspecification, even though the optimality properties of Theorem 2 do not apply.

4. The power of the optimal exponential test depends on $J(\cdot)$, usually much more so than on c. A suitable choice of $J(\cdot)$ depends on the problem at hand. For example, in structural change cases, a natural choice of $J(\cdot)$ often is a uniform weight function on some subinterval of $(0, 1)$.

5. There are two ways of choosing the constant c. One can choose some fixed value of c, such as 0 or ∞, or one can formulate some data-dependent method of determining c. See Andrew and Ploberger (1992) for one method of doing so. In the context of tests of structural change, see Section 4 below, the first method is preferable, because the power and size properties of the optimal tests are relatively insensitive to the choice of c and given a fixed value of c critical values can be tabulated.

4. EXAMPLES

4.1. *Optimal Tests of Structural Change*

In this subsection we consider tests of one-time structural change with unknown change point. Tests of this sort have attracted some attention in the empirical literature, e.g., see Perron (1991), Bai, Lumsdaine, and Stock (1991), and Nason (1991).

The sample of observations is given by $\{(Y_t, X_t): t \leqslant T\}$, where $\{Y_t: t \leqslant T\}$ are endogenous variables and $\{X_t: t \leqslant T\}$ are weakly exogenous variables.[4] We consider the relatively simple case where the data are strictly stationary, ergodic, and Markov under the null hypothesis. In particular, we suppose that $\{(Y_t, X_t): t \leqslant T\}$ is part of a doubly infinite strictly stationary ergodic sequence $\{(Y_t, X_t): t = \ldots, 0, 1, \ldots\}$ and $\{Y_t: T = \ldots, 0, 1, \ldots\}$ is mth order Markov for some integer $m \geqslant 0$. By definition, $\{Y_t: t = \ldots, 0, 1, \ldots\}$ is mth order Markov if the conditional distribution of Y_t given $\mathscr{F}_{t-1} = \sigma(\ldots, Y_{t-2}, Y_{t-1}; \ldots, X_{t-1}, X_t)$ 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. Let

$$(4.1) \quad \{g_t(\delta_1, \delta_2): \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\} = \{g_t(Y_t | Y_{t,m}; X_{t,m}; \delta_1, \delta_2): \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\}$$

denote a parametric family of conditional densities (with respect to some

[4] Weak exogeneity of $\{X_t: t \leqslant T\}$ (see Engle, Hendry, and Richard (1983)) means that the likelihood function for $\boldsymbol{Y}_T$ can be factored into two pieces, one of which contains conditional distributions of Y_t and depends on θ and the other of which contains conditional distributions of X_t and does not depend on θ; see below.

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 $p \leqslant q$. Let

$$(4.2) \qquad h_t = h_t(X_t | Y_1, \ldots, Y_{t-1}; X_1, \ldots, X_{t-1})$$

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 $Y_1, \ldots, Y_{t-1}, X_1, \ldots, X_t$. By the assumption 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 $t = 1, \ldots, [T\pi]$ and $(\delta_1 + \beta, \delta_2)$ for the observations $t = [T\pi] + 1, \ldots, T$, where $\beta \in B \subset R^p$ and $[\cdot]$ denotes the integer part of $\cdot$. Then, π indexes the point of structural change as a fraction of the sample size and $\theta = (\beta', \delta')'$ for $\delta = (\delta_1', \delta_2')'$ contains the pre- and post-change parameter values. For the nonlinear models considered here, we consider the case where Π has closure contained in $(0,1)$. That is, the point of structural change is bounded away from the beginning and end of the sample.[5] The null and alternative hypotheses of interest are H_0: $\beta = \mathbf{0}$ and H_1: $\beta \neq \mathbf{0}$ and the point of structural change is π for some $\pi \in \Pi$.

In the present case, the likelihood function is given by

$$(4.3) \qquad f_T(\theta, \pi) = \left(\prod_{t=1}^{[T\pi]} g_t(\delta_1, \delta_2) \right) \left(\prod_{t=[T\pi]+1}^{T} g_t(\delta_1 + \beta, \delta_2) \right) \left(\prod_{t=1}^{T} h_t \right).$$

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

The following assumption is sufficient for Assumptions 1–3 and 5 of Section 3. All expectations E below are taken under θ_0.

Assumption SC: (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, $\{(Y_t, X_t)$: $t = \ldots, 0, 1, \ldots\}$ is strictly stationary and ergodic, $\{Y_t$: $t = \ldots, 0, 1, \ldots\}$ is mth order Markov, and $\{X_t$: $t = \ldots, 0, 1, \ldots\}$ is weakly exogenous.*

(d) *$g_t(\delta_1, \delta_2)$ is continuous in (δ_1, δ_2) on $\Delta_1 \times \Delta_2$ 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 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.*

[5] In linear models with exogenous regressors, Π does not need to be restricted in this way; see Andrews, Lee, and Ploberger (1994), since finite sample (rather than asymptotic) results can be obtained.

(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)
$$E \sup_{\delta_1 \in \Delta_1,\, \delta_2 \in \Delta_2} \left|\log g_t(\delta_1, \delta_2)\right| < \infty,$$

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

$$E \left\| \frac{\partial}{\partial(\delta_1', \delta_2')} \log g_t(\delta_{10}, \delta_{20}) \right\|^2 < \infty, \quad \textit{and}$$

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

(g)
$$\mathscr{I} = -E \frac{\partial^2}{\partial(\delta_1', \delta_2')'\,\partial(\delta_1', \delta_2')} \log g_t(\delta_{10}, \delta_{20}) \quad \textit{is positive definite}.$$

THEOREM 3: *For* $f_T(\theta, \pi)$ *as in* (4.3), *Assumption SC implies Assumptions* 1–3 *and* 5.

COMMENTS: 1. 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 $t > m$. Without the Markov assumption, one could still verify Assumptions 1–3 and 5, but the conditions would be more complicated.

2. The proof of Theorem 3 is given in an Addendum to this paper that is available from the first author upon request.

Next, we define simplified asymptotically equivalent forms of the exponential LM and Wald statistics (for a proof of equivalency, see Andrews and Ploberger (1992)). Let

$$(4.4) \qquad LM_T^*(\pi) = \frac{1}{\sqrt{T}} \Sigma_{[T\pi]+1}^{T} \frac{\partial}{\partial \delta_1'} \log g_t\left(\tilde{\delta}_1, \tilde{\delta}_2\right)\left[\mathscr{I}_T^{-1}(\tilde{\theta}, \pi)_{00} + \mathscr{I}_T^{-1}(\tilde{\theta}, \pi)_{11}\right]$$
$$\times \frac{1}{\sqrt{T}} \Sigma_{[T\pi]+1}^{T} \frac{\partial}{\partial \delta_1} \log g_t\left(\tilde{\delta}_1, \tilde{\delta}_2\right) \quad \text{and}$$

$$W_T^*(\pi) = T\hat{\beta}(\pi)'\left[\mathscr{I}_T^{-1}\left(\hat{\theta}(\pi), \pi\right)_{00} + \mathscr{I}_T^{-1}\left(\hat{\theta}(\pi), \pi\right)_{11}\right]^{-1} \hat{\beta}(\pi),$$

where

$$\mathscr{I}_T(\theta,\pi)_{00} = -\frac{1}{T}\Sigma_1^{T\pi}\frac{\partial^2}{\partial\delta_1\,\partial\delta_1'}\log g_t(\delta_1,\delta_2) \quad \text{and}$$

$$\mathscr{I}_T(\theta,\pi)_{11} = -\frac{1}{T}\Sigma_{[T\pi]+1}^{T}\frac{\partial^2}{\partial\delta_1\,\partial\delta_1'}\log g_t(\delta_1+\beta,\delta_2).$$

The simplified exponential test statistics *Exp-*LM_T^* and *Exp-*W_T^* are defined by the formula (2.4) with $LM_T(\pi)$ replaced by $LM_T^*(\pi)$ and $W_T^*(\pi)$ respectively.

We now describe the asymptotic null distribution of *Exp-*LM_T (and of the other exponential test statistics). Let $B_1(\cdot)$ be a p-vector of independent Brownian motions on $[0,1]$. Then, under H_0 and Assumption SC, *Exp-*$LM_T \xrightarrow{d} \chi(\theta_0,c)$, where

$$\chi(\theta_0,c) = (1+c)^{-p/2} \times \int \exp\left(\frac{1}{2}\frac{c}{1+c}(B_1(\pi)-\pi B_1(1))' \times (B_1(\pi)-\pi B_1(1))/[\pi(1-\pi)]\right) dJ(\pi). \tag{4.5}$$

That is, the asymptotic null distribution of *Exp-*LM_T is an exponential average of the square of a standardized tied-down Bessel process of order p. Since $\chi(\theta_0,c)$ is nuisance parameter free, asymptotic critical values can be tabulated. The limit distribution of *Exp-*LM_T under general local alternatives to H_0 (not just one-time change alternatives) can be obtained from Theorem 4 of Andrews (1993). It is an exponential average of the square of a noncentral standardized tied-down Bessel process of order p.

Simulation results reported in Andrews, Lee, and Ploberger (1993) show that the power of the optimal tests is not sensitive to the choice of c. In consequence, it suffices to report critical values for just the two limiting cases, $c=\infty$ and $c=0$. Of these two, $c=\infty$ seems to be mildly preferable in terms of power.

For reporting critical values, we consider the case where $J(\cdot)$ is uniform on $[\pi_1,\pi_2]$ for some $0<\pi_1\leqslant\pi_2<1$. In this case we can show that the critical values based on the limit distribution $\chi(\theta_0,c)$ defined in (4.5) (or on its normalized limit as $c\to 0$ or $c\to\infty$) depend on (π_1,π_2) only through the scalar $\lambda=\pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$ (see the proof of Corollary 1 of Andrews (1993)). This greatly simplifies the presentation of critical values for the optimal tests.

Table I reports asymptotic critical values for the $c=\infty$ statistic

$$\begin{aligned}&\log\left(\frac{1}{1-2\pi_0}\int_{\pi_0}^{1-\pi_0}\exp(LM_T(\pi)/2)\,d\pi\right)\\&=\log\left(\frac{1}{T(1-2\pi_0)}\left[\sum_{t=[T\pi_0]+1}^{T-[T\pi_0]-1}\exp(LM_T(t/T)/2)\right.\right.\\&\quad+([T\pi_0]+1-T\pi_0)\{\exp(LM_T([T\pi_0]/T)/2)\\&\quad\left.\left.+\exp(LM_T((T-[T\pi_0])/T)/2)\}\right]\right)\end{aligned} \tag{4.6}$$

TABLE I

ASYMPTOTIC CRITICAL VALUES FOR $c = \infty$ STRUCTURAL CHANGE TEST

π_0	λ	$p=1$ 10%	$p=1$ 5%	$p=1$ 1%	$p=2$ 10%	$p=2$ 5%	$p=2$ 1%	$p=3$ 10%	$p=3$ 5%	$p=3$ 1%	$p=4$ 10%	$p=4$ 5%	$p=4$ 1%	$p=5$ 10%	$p=5$ 5%	$p=5$ 1%
.50	1.000	1.38	1.97	3.39	2.28	2.96	4.54	3.06	3.81	5.65	3.90	4.73	6.67	4.58	5.50	7.30
.49	1.083	1.40	1.98	3.40	2.29	2.96	4.55	3.09	3.84	5.62	3.90	4.72	6.67	4.62	5.50	7.44
.48	1.174	1.40	1.99	3.42	2.30	2.97	4.58	3.11	3.87	5.59	3.93	4.74	6.63	4.66	5.55	7.43
.47	1.272	1.41	2.00	3.37	2.31	2.96	4.54	3.12	3.88	5.64	3.94	4.79	6.67	4.70	5.58	7.44
.45	1.494	1.41	2.00	3.39	2.33	3.00	4.55	3.16	3.88	5.59	3.98	4.83	6.84	4.75	5.64	7.56
.40	2.250	1.43	2.01	3.36	2.38	3.06	4.71	3.24	3.97	5.67	4.10	4.92	6.96	4.88	5.75	7.57
.35	3.449	1.45	2.02	3.34	2.41	3.08	4.67	3.29	4.05	5.64	4.19	4.98	7.02	4.94	5.81	7.72
.30	5.444	1.47	2.02	3.41	2.45	3.12	4.76	3.34	4.09	5.63	4.24	5.06	6.95	5.01	5.87	7.87
.25	9.000	1.48	2.01	3.43	2.50	3.12	4.78	3.40	4.15	5.68	4.28	5.11	7.01	5.08	5.96	7.81
.20	16.000	1.50	2.01	3.39	2.56	3.19	4.77	3.45	4.22	5.70	4.31	5.17	7.05	5.16	6.06	7.85
.15	32.111	1.51	2.06	3.41	2.59	3.22	4.76	3.49	4.22	5.77	4.37	5.23	7.13	5.22	6.13	7.91
.10	81.000	1.52	2.08	3.41	2.59	3.25	4.76	3.53	4.28	5.74	4.43	5.24	7.12	5.30	6.17	7.90
.05	361.000	1.54	2.08	3.40	2.64	3.27	4.72	3.55	4.30	5.87	4.48	5.30	7.08	5.40	6.22	8.11
.02	2401.000	1.55	2.08	3.39	2.65	3.30	4.76	3.59	4.30	5.89	4.54	5.34	7.16	5.43	6.25	8.16

π_0	λ	$p=6$ 10%	$p=6$ 5%	$p=6$ 1%	$p=7$ 10%	$p=7$ 5%	$p=7$ 1%	$p=8$ 10%	$p=8$ 5%	$p=8$ 1%	$p=9$ 10%	$p=9$ 5%	$p=9$ 1%	$p=10$ 10%	$p=10$ 5%	$p=10$ 1%
.50	1.000	5.26	6.16	8.29	6.00	7.05	9.13	6.70	7.77	10.21	7.34	8.45	10.68	7.86	9.00	11.50
.49	1.083	5.31	6.24	8.41	5.97	7.02	9.14	6.78	7.85	10.16	7.36	8.53	10.80	7.95	9.08	11.51
.48	1.174	5.33	6.24	8.46	6.03	7.07	9.24	6.81	7.90	10.17	7.43	8.57	10.84	7.99	9.15	11.62
.47	1.272	5.37	6.29	8.53	6.06	7.12	9.19	6.82	7.95	10.25	7.48	8.61	10.90	8.06	9.19	11.68
.45	1.494	5.41	6.37	8.59	6.14	7.20	9.39	6.94	8.00	10.31	7.56	8.71	11.00	8.14	9.25	11.85
.40	2.250	5.58	6.51	8.70	6.28	7.31	9.50	7.07	8.14	10.42	7.77	8.96	11.16	8.35	9.44	12.06
.35	3.449	5.68	6.65	8.77	6.41	7.35	9.47	7.20	8.24	10.50	7.92	9.01	11.29	8.52	9.57	12.21
.30	5.444	5.73	6.74	8.78	6.48	7.46	9.55	7.32	8.36	10.55	8.04	9.14	11.38	8.67	9.75	12.42
.25	9.000	5.85	6.79	8.86	6.56	7.49	9.56	7.42	8.43	10.72	8.15	9.20	11.43	8.80	9.88	12.47
.20	16.000	5.92	6.86	8.92	6.62	7.54	9.56	7.48	8.56	10.77	8.24	9.27	11.58	8.92	9.96	12.58
.15	32.111	6.01	6.92	8.96	6.70	7.66	9.53	7.58	8.60	10.96	8.31	9.35	11.67	9.00	10.04	12.61
.10	81.000	6.09	6.98	8.99	6.82	7.77	9.61	7.68	8.68	10.97	8.44	9.42	11.66	9.13	10.19	12.61
.05	361.000	6.19	7.09	8.94	6.91	7.85	9.66	7.81	8.79	11.06	8.55	9.55	11.74	9.29	10.34	12.66
.02	2401.000	6.24	7.12	8.95	7.04	7.95	9.86	7.87	8.86	11.22	8.65	9.63	11.73	9.38	10.43	12.70

TABLE I (Continued)

π_0	λ	$p=11$ 10%	$p=11$ 5%	$p=11$ 1%	$p=12$ 10%	$p=12$ 5%	$p=12$ 1%	$p=13$ 10%	$p=13$ 5%	$p=13$ 1%	$p=14$ 10%	$p=14$ 5%	$p=14$ 1%	$p=15$ 10%	$p=15$ 5%	$p=15$ 1%
.50	1.000	8.52	9.66	12.23	9.29	10.50	12.99	9.81	11.08	14.04	10.58	11.81	14.83	11.17	12.52	15.27
.49	1.083	8.56	9.72	12.31	9.36	10.54	13.08	9.75	10.97	13.91	10.47	11.77	14.68	11.31	12.59	15.46
.48	1.174	8.62	9.75	12.39	9.43	10.59	13.12	9.83	11.09	13.96	10.56	11.83	14.73	11.36	12.65	15.54
.47	1.272	8.68	9.82	12.40	9.48	10.65	13.13	9.90	11.17	13.96	10.67	11.96	14.75	11.43	12.74	15.61
.45	1.494	8.78	9.95	12.47	9.62	10.78	13.15	10.04	11.29	13.98	10.84	11.99	14.87	11.56	12.86	15.68
.40	2.250	9.00	10.23	12.75	9.75	10.95	13.47	10.30	11.52	14.24	11.06	12.28	15.05	11.81	13.18	15.78
.35	3.449	9.20	10.40	12.82	9.97	11.15	13.48	10.49	11.71	14.45	11.24	12.45	15.28	12.02	13.28	15.91
.30	5.444	9.35	10.51	12.80	10.13	11.27	13.58	10.68	11.94	14.54	11.40	12.60	15.46	12.17	13.48	16.04
.25	9.000	9.48	10.62	12.86	10.24	11.41	13.71	10.83	12.09	14.64	11.57	12.70	15.47	12.34	13.62	16.15
.20	16.000	9.58	10.65	13.09	10.33	11.49	13.70	10.96	12.14	14.68	11.70	12.94	15.52	12.47	13.77	16.26
.15	32.111	9.69	10.75	13.21	10.45	11.55	13.83	11.10	12.28	14.64	11.84	13.09	15.76	12.69	13.84	16.37
.10	81.000	9.82	10.88	13.20	10.58	11.65	13.94	11.25	12.49	14.85	11.96	13.20	15.81	12.83	14.01	16.60
.05	361.000	9.99	11.01	13.26	10.70	11.78	13.98	11.38	12.55	15.01	12.13	13.31	15.87	13.02	14.16	16.64
.02	2401.000	10.11	11.12	13.45	10.84	11.86	14.07	11.71	12.84	15.24	12.44	13.59	15.98	13.17	14.28	16.78

π_0	λ	$p=16$ 10%	$p=16$ 5%	$p=16$ 1%	$p=17$ 10%	$p=17$ 5%	$p=17$ 1%	$p=18$ 10%	$p=18$ 5%	$p=18$ 1%	$p=19$ 10%	$p=19$ 5%	$p=19$ 1%	$p=20$ 10%	$p=20$ 5%	$p=20$ 1%
.50	1.000	11.77	13.11	16.13	12.39	13.78	16.81	12.90	14.31	17.25	13.59	15.08	18.17	14.20	15.59	18.79
.49	1.083	11.82	13.21	16.15	12.50	13.90	16.77	13.02	14.42	17.37	13.71	15.20	18.37	14.31	15.76	18.94
.48	1.174	11.90	13.21	16.25	12.57	14.02	16.95	13.14	14.56	17.47	13.85	15.29	18.54	14.42	15.90	19.02
.47	1.272	11.98	13.35	16.31	12.64	14.09	17.04	13.18	14.62	17.60	13.94	15.34	18.51	14.52	16.00	19.10
.45	1.494	12.12	13.46	16.53	12.77	14.22	17.13	13.35	14.71	17.70	14.11	15.48	18.67	14.69	16.25	19.27
.40	2.250	12.41	13.73	16.81	13.08	14.45	17.31	13.71	15.10	17.88	14.44	15.82	19.05	14.99	16.51	19.43
.35	3.449	12.64	13.93	16.70	13.32	14.66	17.55	13.96	15.29	18.00	14.66	16.08	19.14	15.25	16.76	19.61
.30	5.444	12.82	14.11	16.86	13.51	14.86	17.62	14.13	15.45	18.17	14.83	16.22	19.22	15.47	17.00	19.70
.25	9.000	13.05	14.37	16.99	13.70	15.05	17.81	14.35	15.64	18.37	15.05	16.44	19.37	15.71	17.15	19.79
.20	16.000	13.20	14.48	17.06	13.91	15.19	17.84	14.55	15.79	18.41	15.24	16.59	19.45	15.96	17.34	20.03
.15	32.111	13.33	14.63	17.25	14.05	15.30	17.93	14.74	15.93	18.55	15.43	16.73	19.58	16.16	17.57	20.18
.10	81.000	13.47	14.80	17.31	14.18	15.52	18.02	14.91	16.12	18.67	15.58	16.89	19.66	16.41	17.73	20.42
.05	361.000	13.72	14.96	17.41	14.42	15.70	18.15	15.09	16.30	18.85	15.84	17.10	19.81	16.58	17.92	20.58
.02	2401.000	13.90	15.13	17.50	14.55	15.87	18.32	15.23	16.46	18.97	16.02	17.21	19.93	16.75	18.00	20.71

TABLE II

ASYMPTOTIC CRITICAL VALUES FOR $c = 0$ STRUCTURAL CHANGE TEST

π_0	λ	$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.000	2.76	3.94	6.78	4.56	5.92	9.07	6.11	7.62	11.29	7.79	9.46	13.35	9.17	11.00	14.59
.49	1.083	2.76	3.91	6.70	4.52	5.83	9.02	6.12	7.59	11.10	7.70	9.31	13.12	9.12	10.88	14.69
.48	1.174	2.72	3.88	6.65	4.47	5.77	8.87	6.07	7.56	10.86	7.67	9.28	13.03	9.09	10.82	14.45
.47	1.272	2.71	3.81	6.50	4.44	5.71	8.72	6.01	7.49	10.72	7.61	9.22	12.92	9.03	10.74	14.42
.45	1.494	2.66	3.74	6.35	4.38	5.63	8.51	5.95	7.35	10.48	7.51	9.08	12.76	8.98	10.62	14.17
.40	2.250	2.57	3.56	6.01	4.23	5.39	8.24	5.79	7.13	9.88	7.40	8.81	12.42	8.75	10.27	13.66
.35	3.449	2.47	3.41	5.78	4.11	5.29	7.85	5.62	6.90	9.51	7.21	8.42	11.87	8.57	9.93	13.19
.30	5.444	2.37	3.24	5.42	4.02	5.06	7.42	5.50	6.66	9.18	7.02	8.27	11.35	8.30	9.69	12.63
.25	9.000	2.27	3.10	5.24	3.92	4.95	7.25	5.34	6.45	8.83	6.82	8.09	10.97	8.13	9.45	12.19
.20	16.000	2.23	2.97	5.00	3.85	4.78	7.04	5.21	6.25	8.43	6.66	7.86	10.50	7.97	9.23	11.71
.15	32.111	2.16	2.88	4.72	3.75	4.61	6.73	5.10	6.07	8.21	6.50	7.67	10.18	7.76	9.01	11.32
.10	81.000	2.09	2.77	4.43	3.63	4.46	6.31	4.97	5.91	7.95	6.36	7.44	9.79	7.60	8.78	10.90
.05	361.000	2.02	2.66	4.15	3.54	4.29	6.07	4.85	5.70	7.62	6.19	7.22	9.51	7.46	8.53	10.52
.02	2401.000	1.97	2.60	3.98	3.48	4.17	5.88	4.75	5.58	7.40	6.10	7.05	9.22	7.34	8.35	10.28
π_0	λ	$p=6$			$p=7$			$p=8$			$p=9$			$p=10$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.000	10.51	12.32	16.58	12.01	14.10	18.26	13.40	15.54	20.43	14.67	16.89	21.35	15.72	18.00	23.00
.49	1.083	10.48	12.27	16.71	11.79	13.85	18.05	13.37	15.47	20.07	14.55	16.80	21.35	15.69	17.91	22.76
.48	1.174	10.40	12.18	16.57	11.76	13.82	17.88	13.29	15.46	19.82	14.50	16.75	21.26	15.59	17.85	22.63
.47	1.272	10.36	12.06	16.43	11.70	13.73	17.83	13.13	15.32	19.58	14.44	16.62	21.02	15.55	17.76	22.44
.45	1.494	10.24	11.98	16.22	11.58	13.64	17.48	13.03	15.11	19.44	14.31	16.41	20.81	15.43	17.48	22.36
.40	2.250	9.99	11.61	15.50	11.33	13.10	16.80	12.72	14.69	18.68	14.02	16.11	20.06	15.11	16.96	21.64
.35	3.449	9.79	11.34	14.82	11.08	12.72	16.10	12.47	14.29	18.05	13.73	15.58	19.65	14.87	16.72	21.04
.30	5.444	9.59	11.04	14.25	10.83	12.42	15.62	12.20	13.95	17.43	13.43	15.18	19.05	14.54	16.27	20.58
.25	9.000	9.41	10.81	13.68	10.66	12.08	15.18	11.99	13.57	16.96	13.16	14.77	18.50	14.28	15.97	19.98
.20	16.000	9.20	10.50	13.29	10.44	11.70	14.74	11.74	13.28	16.54	12.95	14.45	17.84	14.00	15.63	19.28
.15	32.111	9.02	10.19	12.93	10.28	11.47	14.34	11.54	12.94	16.14	12.71	14.16	17.30	13.77	15.29	18.72
.10	81.000	8.84	9.91	12.45	10.07	11.20	13.80	11.32	12.63	15.55	12.52	13.83	16.75	13.57	14.94	18.03
.05	361.000	8.67	9.66	11.92	9.84	10.97	13.23	11.13	12.32	15.07	12.30	13.53	16.20	13.40	14.65	17.41
.02	2401.000	8.54	9.50	11.56	9.81	10.86	13.02	11.00	12.13	14.74	12.15	13.34	15.88	13.23	14.44	17.10

TABLE II (Continued)

π_0	λ	$p=11$ 10%	$p=11$ 5%	$p=11$ 1%	$p=12$ 10%	$p=12$ 5%	$p=12$ 1%	$p=13$ 10%	$p=13$ 5%	$p=13$ 1%	$p=14$ 10%	$p=14$ 5%	$p=14$ 1%	$p=15$ 10%	$p=15$ 5%	$p=15$ 1%
.50	1.000	17.04	19.33	24.45	18.59	20.99	25.98	19.62	22.15	28.07	21.16	23.62	29.66	22.35	25.04	30.54
.49	1.083	16.89	19.20	24.31	18.47	20.79	25.81	19.27	21.66	27.47	20.63	23.23	28.97	22.30	24.85	30.56
.48	1.174	16.85	19.07	24.19	18.43	20.72	25.58	19.22	21.56	27.30	20.55	23.10	28.84	22.15	24.67	30.27
.47	1.272	16.75	19.01	23.86	18.31	20.58	25.50	19.13	21.42	26.98	20.50	23.05	28.57	22.05	24.60	30.13
.45	1.494	16.62	18.83	23.61	18.11	20.40	24.94	19.00	21.28	26.69	20.45	22.73	28.08	21.85	24.42	29.66
.40	2.250	16.25	18.45	22.91	17.67	19.90	24.08	18.57	20.69	25.80	20.02	22.23	27.25	21.51	23.89	28.94
.35	3.449	16.01	18.02	22.23	17.29	19.45	23.18	18.23	20.32	25.17	19.69	21.78	26.49	21.12	23.41	28.31
.30	5.444	15.78	17.59	21.35	17.05	18.92	22.49	17.99	19.96	24.52	19.34	21.37	25.66	20.73	22.83	27.34
.25	9.000	15.51	17.22	20.81	16.77	18.56	21.98	17.71	19.53	23.65	19.05	20.88	25.17	20.39	22.42	26.55
.20	16.000	15.26	16.84	20.08	16.54	18.16	21.49	17.42	19.16	23.09	18.68	20.44	24.44	20.10	21.96	25.93
.15	32.111	15.00	16.46	19.44	16.31	17.85	21.03	17.18	18.80	22.55	18.41	20.12	23.77	19.79	21.56	25.26
.10	81.000	14.76	16.10	18.82	16.03	17.53	20.58	16.93	18.46	22.01	18.11	19.73	23.03	19.46	21.09	24.58
.05	361.000	14.52	15.75	18.27	15.76	17.14	19.84	16.65	18.11	21.19	17.83	19.34	22.45	19.18	20.68	23.80
.02	2401.000	14.37	15.52	17.87	15.59	16.91	19.53	16.73	18.10	20.87	17.89	19.32	22.19	19.02	20.49	23.30

π_0	λ	$p=16$ 10%	$p=16$ 5%	$p=16$ 1%	$p=17$ 10%	$p=17$ 5%	$p=17$ 1%	$p=18$ 10%	$p=18$ 5%	$p=18$ 1%	$p=19$ 10%	$p=19$ 5%	$p=19$ 1%	$p=20$ 10%	$p=20$ 5%	$p=20$ 1%
.50	1.000	23.55	26.23	32.26	24.78	27.55	33.62	25.80	28.61	34.50	27.18	30.15	36.34	28.41	31.18	37.57
.49	1.083	23.33	26.10	32.04	24.63	27.46	33.17	25.72	28.45	34.37	27.07	29.96	36.28	28.23	31.11	37.36
.48	1.174	23.24	25.82	31.66	24.55	27.33	33.02	25.63	28.45	34.08	27.00	29.80	36.11	28.09	31.08	37.04
.47	1.272	23.08	25.68	31.55	24.44	27.21	32.83	25.52	28.32	34.05	26.90	29.64	35.87	27.98	30.92	36.74
.45	1.494	22.99	25.43	31.05	24.22	26.82	32.50	25.25	28.01	33.56	26.69	29.35	35.25	27.86	30.72	36.35
.40	2.250	22.50	24.86	30.22	23.71	26.25	31.44	24.86	27.25	32.62	26.23	28.71	34.31	27.39	30.03	35.36
.35	3.449	22.11	24.44	29.41	23.32	25.65	30.60	24.53	26.69	31.61	25.79	28.20	33.42	26.87	29.37	34.47
.30	5.444	21.84	23.94	28.72	23.03	25.13	29.73	24.16	26.23	30.76	25.35	27.61	32.42	26.44	28.89	33.43
.25	9.000	21.58	23.62	27.76	22.64	24.65	29.11	23.85	25.85	30.01	24.98	27.19	31.43	26.07	28.34	32.64
.20	16.000	21.27	23.12	27.01	22.31	24.25	28.42	23.53	25.40	29.32	24.67	26.67	30.73	25.77	27.87	31.94
.15	32.111	20.98	22.73	26.43	22.00	23.87	27.69	23.22	25.03	28.63	24.40	26.30	30.23	25.50	27.38	31.19
.10	81.000	20.69	22.34	25.85	21.74	23.52	26.97	22.90	24.63	27.95	24.07	25.87	29.58	25.17	26.99	30.64
.05	361.000	20.37	21.97	25.12	21.45	23.08	26.30	22.61	24.20	27.19	23.78	25.38	28.77	24.81	26.49	29.94
.02	2401.000	20.17	21.67	24.66	21.26	22.79	25.84	22.37	23.90	26.81	23.56	25.09	28.19	24.59	26.15	29.46

for a range of values of π_0 between .02 and .5, for $p = 1, \ldots, 20$, and for significance levels $\alpha = .10, .05, .01$. Table I provides the value of λ corresponding to each value of π_0 considered (viz., $\lambda = (1-\pi_0)^2/\pi_0^2$). This allows one to obtain critical values for all intervals $\Pi = [\pi_1, \pi_2]$ whose corresponding value of $\lambda = \pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$ either is tabulated or can be interpolated from the Table.

Table II reports analogous asymptotic critical values for the $c = 0$ statistic

$$(4.7)\qquad \frac{1}{1-2\pi_0}\int_{\pi_0}^{1-\pi_0} LM_T(\pi)\,d\pi$$
$$= \frac{1}{T(1-2\pi_0)}\Bigg[\sum_{t=[T\pi_0]+1}^{T-[T\pi_0]-1} LM_T(t/T) + ([T\pi_0]+1-T\pi_0)\big\{LM_T([T\pi_0]/T) + LM_T((T-[T\pi_0])/T)\big\}\Bigg].$$

Critical values for asymmetric intervals $[\pi_1, \pi_2]$ can be obtained as above. See Section A.2 of the Appendix for a description of the computation of Tables I and II.

When the time of structural change (if it occurs) is completely unknown, we suggest taking $\pi_0 = .02$ in (4.6) or (4.7).[6] When the time of structural change is known to lie in some restricted interval $[\pi_1, \pi_2]$ (see the discussion in Andrews (1993, Sec. 2) regarding such cases), then the test statistic should incorporate this information to maximize power.

4.2. *Empirical Process Examples*

In this subsection, we describe several examples for which Assumption 5 of Section 3 can be verified using an empirical process CLT. For each example, the conditions given are sufficient for Assumptions 1–3 and 5 with Assumption 5 verified using the empirical process CLT of Ossiander (1987) or Doukhan, Massart, and Rio (1994). The proof of sufficiency is given in an Addendum to this paper that is available from the first author. The Addendum gives a general empirical process result that covers the examples here as well as a variety of others. Note that the assumption of normality of the errors that appears in the examples is used for the optimality of the test procedures, but is not needed for the tests to have correct asymptotic significance level.

[6] This choice puts little restriction on the time of change. It does not yield the same power problems as when the "sup" statistic (e.g., $\sup_{\pi \in [\pi_0, 1-\pi_0]} LM_T(\pi)$) is defined over such a broad interval, because $\int_0^1 \exp(\frac{1}{2}BB(\pi)'BB(\pi)/[\pi(1-\pi)])\,d\pi < \infty$ a.s. and $\int_0^1 BB(\pi)'BB(\pi)/[\pi(1-\pi)]\,d\pi < \infty$ a.s., whereas $\sup_{\pi \in [0,1]} BB(\pi)'BB(\pi)/[\pi(1-\pi)] = \infty$ a.s.

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 $j \leqslant m$. In this case, $\theta = (\beta', \delta')'$ for $\beta = (\beta_1', \dots, \beta_m')'$ 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: $\beta = 0$. The parameter π is present only under the alternative.

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

$$(4.8) \qquad Y_t = \begin{cases} X_t'\delta_1 + U_t & \text{for } Z_t \leqslant \pi \\ X_t'(\delta_1 + \beta) + U_t & \text{for } Z_t > \pi \end{cases} \quad (t = 1, \dots, T),$$

where $\{(Y_t, X_t, Z_t, U_t)\colon t = 1, \dots, T\}$ 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 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 $\leqslant 0$; the parameter π lies in a compact set $\Pi \subset R$; and the true parameter θ_0 lies in the interior of Θ.

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 (4.8) with $X_t = (1, Y_{t-1})'$, $Z_t = Y_{t-d}$ for some integer $d > 0$, $\{U_t\colon t = 1, \dots, T\}$ 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 with $|\delta_1| < 1$. (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 (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

$$(4.9) \qquad Y_t = g(X_t, \delta_1) + \beta h(Z_t, \pi) + U_t \quad (t = 1, \dots, T),$$

where $\{(Y_t, X_t, Z_t, U_t)\colon t = 1, \dots, T\}$ 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 $\leqslant 0$; the true parameter θ_0 lies in the interior of Θ; $g(X_t, \delta_1)$ 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 $x \geqslant 0$;

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

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

$$E \sup_{\pi \in \Pi} \left\| \frac{\partial}{\partial \pi} h(Z_t, \pi) \right\|^r < \infty$$

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

$$E(g(X_t, \delta_1) - g(X_t, \delta_{10}) + \beta h(Z_t, \pi))^2 > 0 \quad \forall \theta \in \Theta \text{ with } \theta \neq \theta_0.$$

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): 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 (4.9), Z_t is taken to be a subvector 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. The results of this paper provide optimal forms for the test statistics in these cases.

Cowles Foundation for Research in Economics, Dept. of Economics, Yale University, P.O. Box 208281, New Haven, CT 06520, U.S.A.

and

Institut für Okonometrie und Systemtheorie, Technische Universitat Wien, A-1040 Wien, Austria.

Manuscript received March, 1992; final revision received September, 1993.

APPENDIX

A.1. *Proof of Theorems 1 and 2*

We begin this appendix of proofs by outlining the sequence of results that are used below to establish Theorems 1 and 2. Consider the likelihood ratio statistic LR_T for the alternative density $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)$. By definition,

$$LR_T = \int f_T\left(\theta_0 + B_T^{-1}h, \pi\right) dQ_\pi(h)\, dJ(\pi)/f_T(\theta_0). \tag{A.1}$$

By the Neyman-Pearson Lemma, a test based on LR_T is a best test of a given significance level for testing the simple null hypothesis that $f_T(\theta_0)$ is the true density versus the simple alternative that $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)$ is true. In addition, a test based on LR_T has the best weighted average power for weight functions Q_π and J of all tests of a given significance level for testing the simple null hypothesis that $f_T(\theta_0)$ is the true density versus the alternative that $f_T(\theta_0 + B_T^{-1}h, \pi)$ is true for some $h \in R^s$ and $\pi \in \Pi$.

The asymptotic optimality of $Exp\text{-}LM_T$ (Theorem 2 above) follows from the optimality of LR_T if we can show that $LR_T - Exp\text{-}LM_T \xrightarrow{p} 0$ under the null and under the local alternatives $\{\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)\colon T \geqslant 1\}$. The proof of this result requires several steps. First we introduce two assumptions, Assumptions 4′ and 5′, which are implied by Assumptions 4 and 5 respectively.

Assumption 4′: *$\{Q_\pi(\cdot)\colon \pi \in \Pi\}$ is uniformly tight (i.e., $\forall \varepsilon > 0\ \exists M < \infty$ such that $Q_\pi(\{h \in R^s\colon \|h\| \geqslant M\}) < \varepsilon\ \forall \pi \in \Pi$, where $\|\cdot\|$ is the Euclidean norm).*

Assumption 5′: *$\sup_{\pi \in \Pi} \|B_T^{-1} Dl_T(\theta_0, \pi)\| = O_p(1)$ under θ_0.*

We define an *approximate* ML estimator $\bar{\theta}(\pi)$ and an approximate LR statistic $\overline{LR}_T$ by

$$\bar{\theta}(\pi) = \mathscr{I}^{-1}(\theta_0, \pi) B_T^{-1} Dl_T(\theta_0, \pi) \quad \text{and} \tag{A.2}$$

$$\overline{LR}_T = \int \exp\left(\frac{1}{2}\bar{\theta}(\pi)' \mathscr{I}(\theta_0, \pi)\bar{\theta}(\pi)\right) \times \int \exp\left(-\frac{1}{2}\left(\bar{\theta}(\pi) - h\right)' \mathscr{I}(\theta_0, \pi)\left(\bar{\theta}(\pi) - h\right)\right) dQ_\pi(h)\, dJ(\pi). \tag{A.3}$$

Lemma A-1: *Under the null hypothesis and Assumptions* 1, 2, *and* 5′,

$$\sup_{\pi \in \Pi} \left\| B_T\left(\hat{\theta}(\pi) - \theta_0\right) - \bar{\theta}(\pi)\right\| \xrightarrow{p} 0.$$

Lemma A-2: *Under the null hypothesis and Assumptions* 1, 2, 4′, *and* 5′, $LR_T - \overline{LR}_T \xrightarrow{p} 0$.

Lemma A-3: *For each $\pi \in \Pi$, the projection matrix $P^\perp$ onto the orthogonal complement $V_\pi^\perp$ of V with respect to $\langle \cdot, \cdot \rangle_\pi$ is given by*

$$P^\perp \left(= P_\pi^\perp\right) = A_\pi H, \quad \text{where} \quad A_\pi = \begin{bmatrix} I_p \\ -\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}' \end{bmatrix},$$

$$H = \left[I_p \,\vdots\, 0\right], \quad \text{and} \quad \mathscr{I}(\theta_0, \pi) = \begin{bmatrix} \mathscr{I}_{1\pi} & \mathscr{I}_{2\pi} \\ \mathscr{I}_{2\pi}' & \mathscr{I}_{3\pi} \end{bmatrix}.$$

We now define an approximate exponential Wald statistic $Exp\text{-}\overline{W}_T$. Let

$$\text{(A.4)}\qquad \overline{W}_T(\pi)=\left(H\bar{\theta}(\pi)\right)'\left(H\mathscr{I}^{-1}(\theta_0,\pi)H'\right)^{-1}H\bar{\theta}(\pi)\quad\text{and}$$

$$\text{(A.5)}\qquad Exp\text{-}\overline{W}_T=(1+c)^{-p/2}\int\exp\left(\frac{1}{2}\frac{c}{1+c}\overline{W}_T(\pi)\right)dJ(\pi).$$

Lemmas A-1 to A-3 and the definition of $Exp\text{-}\overline{W}_T$ are used to establish the following result.

THEOREM A-1: *Under the null hypothesis and Assumptions* 1–4 *and* 5′, (*a*) $LR_T-\overline{LR}_T\xrightarrow{p}0$, (*b*) $\overline{LR}_T=Exp\text{-}\overline{W}_T$, (*c*) $Exp\text{-}\overline{W}_T-Exp\text{-}W_T\xrightarrow{p}0$, (*d*) $Exp\text{-}W_T-Exp\text{-}LM_T\xrightarrow{p}0$, *and* (*e*) $Exp\text{-}LM_T-Exp\text{-}LR_T\xrightarrow{p}0$.

Theorem 1 follows easily from Theorem A-1 and Assumption 5.

PROOF OF THEOREM 1: By Assumption 5,

$$\text{(A.6)}\qquad \begin{pmatrix}\bar{\theta}(\cdot)\\ \mathscr{I}^{-1}(\theta_0,\cdot)\end{pmatrix}=\begin{pmatrix}\mathscr{I}^{-1}(\theta_0,\cdot)B_T^{-1}Dl_T(\theta_0,\cdot)\\ \mathscr{I}^{-1}(\theta_0,\cdot)\end{pmatrix}\Rightarrow\begin{pmatrix}\mathscr{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot)\\ \mathscr{I}^{-1}(\theta_0,\cdot)\end{pmatrix}\quad\text{under }\theta_0.$$

By Assumptions 1(e), 1(f), and 5, $(\mathscr{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathscr{I}^{-1}(\theta_0,\cdot))$ has bounded uniformly continuous sample paths (as a function of $\pi\in\Pi$) with $G(\theta_0,\cdot)$-probability one. In consequence, using (3.8), the function $m(\cdot,\cdot)$ that maps $(\mathscr{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathscr{I}^{-1}(\theta_0,\cdot))$ into $\chi(\theta_0,c)$ is continuous with $G(\theta_0,\cdot)$-probability one. The continuous mapping theorem (e.g., see Pollard (1984, Thm. IV-12, pp. 66–70) or Billingsley (1968, Thm. 5.1, pp. 29–34)) now gives:

$$\text{(A.7)}\qquad Exp\text{-}\overline{W}_T=m\left(\bar{\theta}(\cdot),\mathscr{I}^{-1}(\theta_0,\cdot)\right)\xrightarrow{d}m\left(\mathscr{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathscr{I}^{-1}(\theta_0,\cdot)\right)=\chi(\theta_0,c).$$

Theorem 1 now follows from Theorem A-1. *Q.E.D.*

PROOF OF LEMMA A-1: All probability calculations in this proof are made "under θ_0." By Assumptions 1(b), 1(c), and 2 and the definition of the ML estimator $\hat{\theta}(\pi)$, $Dl_T(\hat{\theta}(\pi),\pi)=0\ \forall\pi\in\Pi$ wp $\to$ 1. Hence, by one-term Taylor expansions of the elements of $Dl_T(\hat{\theta}(\pi),\pi)$ about θ_0 we get, wp $\to$ 1,

$$\text{(A.8)}\qquad 0=B_T^{-1}Dl_T\left(\hat{\theta}(\pi),\pi\right)=B_T^{-1}Dl_T(\theta_0,\pi)-\mathscr{I}_{1T}(\pi)B_T\left(\hat{\theta}(\pi)-\theta_0\right)\quad\forall\pi\in\Pi,$$

where

$$\mathscr{I}_{1T}(\pi)=-\int_0^1B_T^{-1}D^2l_T\left(\theta_0+\lambda\left(\hat{\theta}(\pi)-\theta_0\right),\pi\right)B_T^{-1}\,d\lambda.$$

The matrix $\mathscr{I}_{1T}(\pi)$ satisfies

$$\begin{aligned}\text{(A.9)}\qquad &\sup_{\pi\in\Pi}\left\|\mathscr{I}_{1T}(\pi)-\mathscr{I}(\theta_0,\pi)\right\|\\ &\leqslant\sup_{\pi\in\Pi}\left\|\int_0^1\left[-B_T^{-1}D^2l_T\left(\theta_0+\lambda\left(\hat{\theta}(\pi)-\theta_0\right),\pi\right)B_T^{-1}\right.\right.\\ &\qquad\left.\left.-\mathscr{I}\left(\theta_0+\lambda\left(\hat{\theta}(\pi)-\theta_0\right),\pi\right)\right]d\lambda\right\|\\ &\quad+\sup_{\pi\in\Pi}\left\|\int_0^1\left[\mathscr{I}\left(\theta_0+\lambda\left(\hat{\theta}(\pi)-\theta_0\right),\pi\right)-\mathscr{I}(\theta_0,\pi)\right]d\lambda\right\|\\ &=o_p(1),\end{aligned}$$

where the inequality follows from the triangle inequality and the equality holds using Assumptions 1(d) and 2 for the first term and Assumptions 1(e) and 2 for the second term.

Equation (A.9) and Assumption 1(f) yield

$$\text{(A.10)} \qquad \sup_{\pi \in \Pi} \left\| \mathscr{I}_{1T}^{-1}(\pi) - \mathscr{I}^{-1}(\theta_0, \pi) \right\| = o_p(1).$$

Equations (A.8) and (A.10) and Assumptions 1(f) and 5′ yield

$$\text{(A.11)} \qquad o_p(1) = \sup_{\pi \in \Pi} \left\| B_T\left(\hat{\theta}(\pi) - \theta_0\right) - \mathscr{I}_{1T}^{-1}(\pi) B_T^{-1} Dl_T(\theta_0, \pi) \right\|$$

$$= \sup_{\pi \in \Pi} \left\| B_T\left(\hat{\theta}(\pi) - \theta_0\right) - \mathscr{I}^{-1}(\theta_0, \pi) B_T^{-1} Dl_T(\theta_0, \pi) \right\| + o_p(1). \qquad Q.E.D.$$

Proof of Lemma A-2: All probability calculations in this proof are made "under θ_0." For $0 < M < \infty$, define

$$\text{(A.12)} \qquad LR_T(M) = \int_\Pi \int_{\|h\| \leqslant M} f_T\left(\theta_0 + B_T^{-1} h, \pi\right) dQ_\pi(h)\, dJ(\pi) / f_T(\theta_0) \quad \text{and}$$

$$\text{(A.13)} \qquad \overline{LR}_T(M) = \int_\Pi \exp\left(\frac{1}{2} \bar{\theta}(\pi)' \mathscr{I}(\theta_0, \pi) \bar{\theta}(\pi)\right)$$

$$\times \int_{\|h\| \leqslant M} \exp\left(-\frac{1}{2}\left(\bar{\theta}(\pi) - h\right)' \mathscr{I}(\theta_0 \pi)\left(\bar{\theta}(\pi) - h\right)\right) dQ_\pi(h)\, dJ(\pi).$$

Note that for any $\varepsilon > 0$

$$\text{(A.14)} \qquad P\left(|LR_T - \overline{LR}_T| > 3\varepsilon\right) \leqslant P\left(|LR_T - LR_T(M)| > \varepsilon\right) + P\left(|LR_T(M) - \overline{LR}_T(M)| > \varepsilon\right)$$

$$+ P\left(|\overline{LR}_T - \overline{LR}_T(M)| > \varepsilon\right).$$

Hence, it suffices to show that (i) given any $\eta > 0$ we can choose $T^* < \infty$ and $M < \infty$ sufficiently large so that $P(|LR_T - LR_T(M)| > \varepsilon) < \eta$ and $P(|\overline{LR}_T - \overline{LR}_T(M)| > \varepsilon) < \eta$ for all $T \geqslant T^*$ and (ii) $LR_T(M) - \overline{LR}_T(M) \xrightarrow{p} 0 \ \forall 0 < M < \infty$.

We show (i) first. We have

$$\text{(A.15)} \qquad P\left(|LR_T - LR_T(M)| > \varepsilon\right)$$

$$\leqslant \varepsilon^{-1} E|LR_T - LR_T(M)|$$

$$= \varepsilon^{-1} E \int_\Pi \int_{\|h\| > M} \left[f_T\left(\theta_0 + B_T^{-1} h, \pi\right) / f_T(\theta_0)\right] dQ_\pi(h)\, dJ(\pi)$$

$$= \varepsilon^{-1} \int_\Pi \int_{\|h\| > M} dQ_\pi(h)\, dJ(\pi),$$

where the first equality uses Assumption 1(a) and the second holds by Fubini's theorem and the fact that $E[f_T(\theta_0 + B_T^{-1} h, \pi)/f_T(\theta_0)] = 1 \ \forall h, \forall \pi$. The right-hand side of (A.15) can be made arbitrarily small for all T by taking M large by Assumption 4′.

Next, we have

$$\text{(A.16)}\qquad |\overline{LR}_T - \overline{LR}_T(M)|$$

$$= \int_\Pi \Bigg[\exp\left(\tfrac{1}{2} Dl_T(\theta_0,\pi)' B_T^{-1} \mathscr{I}^{-1}(\theta_0,\pi) B_T^{-1} Dl_T(\theta_0,\pi)\right)$$

$$\times \int_{\|h\|>M} \exp\left(-\tfrac{1}{2}(\bar\theta(\pi)-h)'\mathscr{I}(\theta_0,\pi)(\bar\theta(\pi)-h)\right) dQ_\pi(h)\Bigg] dJ(\pi)$$

$$\leqslant \exp\left(\frac{1}{2}\sup_{\pi\in\Pi}\left\|B_T^{-1}Dl_T(\theta_0,\pi)\right\|^2 \sup_{\pi\in\Pi}\left\|\mathscr{I}^{-1}(\theta_0,\pi)\right\|\right)\cdot\int_\Pi\int_{\|h\|>M} dQ_\pi(h)\,dJ(\pi),$$

where the inequality uses the assumption that $\mathscr{I}(\theta_0,\pi)$ is positive definite. The first term on the right-hand side of (A.16) is $O_p(1)$ by Assumptions 1(f) and 5′ and the second term on the right-hand side can be made arbitrarily small by taking M large using Assumption 4′. In consequence, $P(|\overline{LR}_T - \overline{LR}_T(M)| > \varepsilon)$ can be made arbitrarily small for all T large by taking M large.

We now establish (ii). A two term Taylor series expansion gives

$$\text{(A.17)}\qquad l_T(\theta_0 + B_T^{-1}h,\pi) - l_T(\theta_0)$$

$$= h'B_T^{-1}Dl_T(\theta_0,\pi) + \frac{1}{2}h'B_T^{-1}D^2l_T(\theta_0,\pi)B_T^{-1}h + r_{1T}(h,\pi),$$

where the remainder term $r_{1T}(h,\pi)$ satisfies

$$\text{(A.18)}\qquad \sup_{\pi\in\Pi}\ \sup_{h:\,\|h\|\leqslant M}\|r_{1T}(h,\pi)\|$$

$$\leqslant M^2 \sup_{\pi\in\Pi}\ \sup_{\theta:\,\|B_T(\theta-\theta_0)\|\leqslant M}\left\|B_T^{-1}D^2l_T(\theta,\pi)B_T^{-1} - B_T^{-1}D^2l_T(\theta_0,\pi)B_T^{-1}\right\|$$

$$\leqslant M^2 \sup_{\pi\in\Pi}\ \sup_{\theta\in\Theta_0}\left\|B_T^{-1}D^2l_T(\theta,\pi)B_T^{-1} + \mathscr{I}(\theta,\pi)\right\|$$

$$+ M^2 \sup_{\pi\in\Pi}\ \sup_{\theta:\,\|B_T(\theta-\theta_0)\|\leqslant M}\left\|\mathscr{I}(\theta_0,\pi) - \mathscr{I}(\theta,\pi)\right\|$$

$$+ M^2 \sup_{\pi\in\Pi}\left\|B_T^{-1}D^2l_T(\theta_0,\pi)B_T^{-1} + \mathscr{I}(\theta_0,\pi)\right\|$$

$$= o_p(1),$$

where the equality uses Assumptions 1(d) and 1(e). In addition,

$$\text{(A.19)}\qquad h'B_T^{-1}D^2l_T(\theta_0,\pi)B_T^{-1}h = -h'\mathscr{I}(\theta_0,\pi)h + r_{2T}(h,\pi), \quad \text{where}$$

$$\sup_{\pi\in\Pi}\ \sup_{h:\,\|h\|\leqslant M}|r_{2T}(h,\pi)| = o_p(1),$$

by Assumption 1(d). It follows from (A.18) and (A.19) that

$$\text{(A.20)}\qquad \exp\left(r_{1T}(h,\pi) + r_{2T}(h,\pi)\right) = 1 + s_T(h,\pi), \quad \text{where}$$

$$\sup_{\pi\in\Pi}\ \sup_{h:\,\|h\|\leqslant M}|s_T(h,\pi)| = o_p(1).$$

Combining (A.17) and (A.19) and using the definition of $\bar{\theta}(\pi)$ yields

$$\text{(A.21)} \quad l_T(\theta_0 + B_T^{-1}h, \pi) - l_T(\theta_0)$$
$$= h'\mathscr{I}(\theta_0, \pi)\bar{\theta}(\pi) - \tfrac{1}{2}h'\mathscr{I}(\theta_0, \pi)h + r_{1T}(h, \pi) + r_{2T}(h, \pi)$$
$$= \tfrac{1}{2}\bar{\theta}(\pi)'\mathscr{I}(\theta_0, \pi)\bar{\theta}(\pi) - \tfrac{1}{2}(\bar{\theta}(\pi) - h)'\mathscr{I}(\theta_0, \pi)(\bar{\theta}(\pi) - h)$$
$$+ r_{1T}(h, \pi) + r_{2T}(h, \pi),$$

where the second equality follows from some simple algebra.

Combining (A.12), (A.13), (A.20), (A.21), and Assumption 1(a) gives

$$\text{(A.22)} \quad LR_T(M)$$
$$= \int_\Pi \int_{\|h\| \leq M} \exp\left(l_T(\theta_0 + B_T^{-1}h, \pi) - l_T(\theta_0)\right) dQ_\pi(h)\, dJ(\pi)$$
$$= \int_\Pi \int_{\|h\| \leq M} \exp\left(\tfrac{1}{2}\bar{\theta}(\pi)'\mathscr{I}(\theta_0, \pi)\bar{\theta}(\pi) - \tfrac{1}{2}(\bar{\theta}(\pi) - h)'\mathscr{I}(\theta_0, \pi)(\bar{\theta}(\pi) - h)\right)$$
$$\times (1 + s_T(h, \pi))\, dQ_\pi(h)\, dJ(\pi)$$
$$= \overline{LR}_T(M) + o_p(1),$$

where the third equality uses $\overline{LR}_T(M) = O_p(1)$, which follows from a close analogue to (A.16). This completes the proof. *Q.E.D.*

Proof of Lemma A-3: Let A denote A_π. Since $HA = I_p$, $(AH)AH = AH$ and AH is an oblique projection matrix. For $v = (0, \delta')' \in V$, $AHv = 0$, so AH projects onto a space orthogonal to V. On the other hand, for $m = (m_1', m_2')' \in V_\pi^\perp$, $v'\mathscr{I}(\theta_0, \pi)m = 0\ \forall v \in V$ iff $[0 \vdots I_q]\mathscr{I}(\theta_0, \pi)m = 0$ iff $[\mathscr{I}_{2\pi}' \vdots \mathscr{I}_{3\pi}]m = 0$ iff $m_2 = -\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}'m_1$ iff $m = Am_1$. In consequence, $AHm = AHAm_1 = Am_1 = m\ \forall m \in V_\pi^\perp$. That is, AH projects onto the entire orthogonal complement of V with respect to $\langle \cdot, \cdot \rangle_\pi$. *Q.E.D.*

Proof of Theorem A-1: Part (a) holds by Lemma A-2. Next, consider part (b). Let $\lambda \sim N(0, c(A'\mathscr{I}A)^{-1})$ and $h = A\lambda$, where $A = A_\pi$ and $\mathscr{I} = \mathscr{I}(\theta_0, \pi)$. Then, $h \sim Q_\pi = N(0, cA(A'\mathscr{I}A)^{-1}A')$ as desired. The density of λ is

$$\text{(A.23)} \quad (2\tilde{\pi})^{-p/2}\det{}^{1/2}(A'\mathscr{I}A/c)\exp\left(-\tfrac{1}{2c}\lambda'A'\mathscr{I}A\lambda\right)$$

with respect to Lebesgue measure on R^p, where $\tilde{\pi} = \text{pi} = 3.14\ldots$.

For notational simplicity, let $\bar{\theta} = \bar{\theta}(\pi)$ and $\mathscr{I} = \mathscr{I}(\theta_0, \pi)$. Then,

$$\text{(A.24)} \quad \overline{LR}_T = \int_\Pi \zeta_T(\pi)\, dJ(\pi), \quad \text{where}$$

$$\text{(A.25)} \quad \zeta_T(\pi) = \int \exp\left(\tfrac{1}{2}\bar{\theta}'\mathscr{I}\bar{\theta} - \tfrac{1}{2}(h - \bar{\theta})'\mathscr{I}(h - \bar{\theta})\right) dQ_\pi(h)$$
$$= (2\tilde{\pi})^{-p/2}\det{}^{1/2}(A'\mathscr{I}A/c)$$
$$\times \int \exp\left(\tfrac{1}{2}\left[\bar{\theta}'\mathscr{I}\bar{\theta} - (A\lambda - \bar{\theta})'\mathscr{I}(A\lambda - \bar{\theta}) - (A\lambda)'\mathscr{I}A\lambda/c\right]\right) d\lambda.$$

Let P and $P^\perp$ denote the projection matrices with respect to $\langle \cdot, \cdot \rangle_\pi$ onto V and $V_\pi^\perp$ respectively. (Note that P and $P^\perp$ depend on π since $\langle \cdot, \cdot \rangle_\pi$ and $V_\pi^\perp$ do.) The term in square brackets in the exponent on the right-hand side of (A.25), with $A\lambda$ replaced by h for simplicity, now simplifies as follows:

$$\text{(A.26)}\quad \begin{aligned} &\bar\theta'\mathscr{I}\bar\theta - (h-\bar\theta)'\mathscr{I}(h-\bar\theta) - h'\mathscr{I}h/c \\ &= \bar\theta'\mathscr{I}\bar\theta - \left(h - \bar\theta\frac{c}{1+c}\right)'\mathscr{I}\frac{1+c}{c}\left(h-\bar\theta\frac{c}{1+c}\right) - \frac{1}{1+c}\bar\theta'\mathscr{I}\bar\theta \\ &= \frac{c}{1+c}(P\bar\theta)'\mathscr{I}P\bar\theta + \frac{c}{1+c}(P^{\perp}\bar\theta)'\mathscr{I}P^{\perp}\bar\theta \\ &\quad - \left(h - P^{\perp}\bar\theta\frac{c}{1+c}\right)'\mathscr{I}\frac{1+c}{c}\left(h - P^{\perp}\bar\theta\frac{c}{1+c}\right) \\ &\quad - \frac{c}{1+c}(P\bar\theta)'\mathscr{I}P\bar\theta \\ &= \frac{c}{1+c}(P^{\perp}\bar\theta)'\mathscr{I}P^{\perp}\bar\theta - \left(h - P^{\perp}\bar\theta\frac{c}{1+c}\right)'\mathscr{I}\frac{1+c}{c}\left(h - P^{\perp}\bar\theta\frac{c}{1+c}\right), \end{aligned}$$

where the second equality uses the fact that $(P\bar\theta)'\mathscr{I}h = 0\ \forall h \in V_\pi^{\perp}$.

Combining (A.25) and (A.26) gives

$$\text{(A.27)}\quad \begin{aligned} \zeta_T(\pi) &= (1+c)^{-p/2}\exp\left(\frac{1}{2}\frac{c}{1+c}(P^{\perp}\bar\theta)'\mathscr{I}P^{\perp}\bar\theta\right) \\ &\quad \times \int (2\bar\pi)^{-p/2}\det{}^{1/2}\left(A'\mathscr{I}A\frac{1+c}{c}\right) \\ &\quad \times \exp\left(-\frac{1}{2}\left(A\lambda - P^{\perp}\bar\theta\frac{c}{1+c}\right)'\mathscr{I}\frac{1+c}{c}\left(A\lambda - P^{\perp}\bar\theta\frac{c}{1+c}\right)\right)d\lambda \\ &= (1+c)^{-p/2}\exp\left(\frac{1}{2}\frac{c}{1+c}(P^{\perp}\bar\theta)'\mathscr{I}P^{\perp}\bar\theta\right), \end{aligned}$$

where the second equality holds because $P^{\perp}\bar\theta = AH\bar\theta$ and the integral of a normal density equals one.

Using Lemma A-3, $(P^{\perp}\bar\theta)'\mathscr{I}P^{\perp}\bar\theta = (H\bar\theta)'A'\mathscr{I}AH\bar\theta$. Hence, for part (b), it remains to show that $A'\mathscr{I}A = [H\mathscr{I}^{-1}H']^{-1}$. By simple algebra, the left-hand side equals $\mathscr{I}_{1\pi} - \mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}'$. The right-hand side equals the inverse of the upper $p \times p$ submatrix of $\mathscr{I}(\theta_0, \pi)^{-1}$, which equals $\mathscr{I}_{1\pi} - \mathscr{I}_{2\pi}\mathscr{I}_{3\pi}^{-1}\mathscr{I}_{2\pi}'$ by the formula for a partitioned inverse. The proof of part (b) is now complete.

To establish part (c) of the Theorem, note that $HB_T\theta_0 = 0$. In consequence, Lemma A-1 implies that

$$\text{(A.28)}\quad \sup_{\pi\in\Pi}\left\|HB_T\hat\theta(\pi) - H\bar\theta(\pi)\right\| \xrightarrow{p} 0.$$

In addition, we have

$$\text{(A.29)}\quad \begin{aligned} &\sup_{\pi\in\Pi}\left\|\mathscr{I}_T\big(\hat\theta(\pi),\pi\big) - \mathscr{I}(\theta_0,\pi)\right\| \\ &\leqslant \sup_{\pi\in\Pi}\sup_{\theta\in\Theta_0}\left\|\mathscr{I}_T(\theta,\pi) - \mathscr{I}(\theta,\pi)\right\| + \sup_{\pi\in\Pi}\left\|\mathscr{I}\big(\hat\theta(\pi),\pi\big) - \mathscr{I}(\theta_0,\pi)\right\| \\ &= o_p(1), \end{aligned}$$

where the inequality holds wp $\to 1$ using Assumption 2 and the equality holds by Assumptions 1(d), 1(e), and 2. Using Assumption 1(f), this establishes part (c).

For parts (d) and (e), it suffices to show that

$$\text{(A.30)}\quad \sup_{\pi\in\Pi}|W_T(\pi) - LM_T(\pi)| \xrightarrow{p} 0 \quad\text{and}\quad \sup_{\pi\in\Pi}|LM_T(\pi) - LR_T(\pi)| \xrightarrow{p} 0.$$

For brevity, the proof of (A.30) is omitted. See Andrews and Ploberger (1992) for a complete proof.

Except for the "$\sup_{\pi \in \Pi}$," the proof of (A.30) is similar to proofs in the literature, e.g., see Andrews (1993, proof of Theorem 3). *Q.E.D.*

We turn now to the proof of Theorem 2. First, we establish the *contiguity* of the densities $\{\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)\colon T \geqslant 1\}$ to the densities $\{f_T(\theta_0)\colon T \geqslant 1\}$. By definition, contiguity holds if $\int_{C_T} f_T(\theta_0)\, d\mu_T \to 0$ implies $\int_{C_T} \int f_T(\theta_0 + B_T^{-1}h, \pi) \times dQ_\pi(h)\, dJ(\pi)\, d\mu_T \to 0$ for any sequence of (measurable) sets C_T, where $\int_{C_T} f_T(\theta_0)\, d\mu_T$ denotes the probability of C_T when Y_T has density $f_T(\theta_0)$ and likewise for $\int_{C_T} \int f_T(\theta_0 + B_T^{-1}h, \pi) \times dQ_\pi(h)\, dJ(\pi)\, d\mu_T$.

LEMMA A-4: *Under Assumptions* 1, 2, 4, *and* 5, *the densities* $\{\int f_T(\theta_0 + B_T^{-1}h, \pi) \times dQ_\pi(h)\, dJ(\pi)\colon T \geqslant 1\}$ *are contiguous to the densities* $\{f_T(\theta_0)\colon T \geqslant 1\}$.

THEOREM A-2: *Under the local alternative densities* $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)$ *for* Y_T *for* $T \geqslant 1$ *and Assumptions* 1–5, (*a*) $LR_T - \overline{LR}_T \xrightarrow{p} 0$, (*b*) $\overline{LR}_T = Exp\text{-}\overline{W}_T$, (*c*) $Exp\text{-}\overline{W}_T - Exp\text{-}W_T \xrightarrow{p} 0$, (*d*) $Exp\text{-}W_T - Exp\text{-}LM_T \xrightarrow{p} 0$, *and* (*e*) $Exp\text{-}LM_T - Exp\text{-}LR_T \xrightarrow{p} 0$.

PROOF OF THEOREM 2: Let α_T be the rejection probability of φ_T under θ_0. Let $k^*_{\alpha_T} > 0$ and $\lambda_T \in [0,1]$ be constants such that the likelihood ratio test

$$\text{(A.31)} \qquad \gamma_T = \begin{cases} 1 & \text{if } LR_T > k^*_{\alpha_T}, \\ \lambda_T & \text{if } LR_T = k^*_{\alpha_T}, \\ 0 & \text{if } LR_T < k^*_{\alpha_T}, \end{cases}$$

has rejection probability α_T under θ_0. Then, by the Neyman-Pearson Lemma (e.g., see Lehmann (1959, Theorem 3.1, p. 65)), for all $T \geqslant 1$,

$$\text{(A.32)} \qquad \int \varphi_T \left[\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi) \right] d\mu_T$$
$$\leqslant \int \gamma_T \left[\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi) \right] d\mu_T.$$

By Corollary 15.11 of Strasser (1985), the $\underline{\lim}_{T\to\infty}$ on the right-hand side of the inequality in the statement of Theorem 2 is actually $\lim_{T\to\infty}$, because $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ (in his notation) by the proof of Lemma A-4 below.

This result, inequality (A.32), and Fubini's Theorem yield

$$\text{(A.33)} \qquad \overline{\lim_{T\to\infty}} \int \left[\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi)\, d\mu_T \right] dQ_\pi(h)\, dJ(\pi)$$
$$\leqslant \lim_{T\to\infty} \int \gamma_T \left[\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi) \right] d\mu_T$$
$$= \lim_{T\to\infty} \int 1(LR_T > k_{T\alpha}) \left[\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi) \right] d\mu_T$$
$$= \lim_{T\to\infty} \int 1(Exp\text{-}LM_T > k_{T\alpha}) \left[\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi) \right] d\mu_T$$
$$= \lim_{T\to\infty} \int \left[\int \xi_T f_T(\theta_0 + B_T^{-1}h, \pi)\, d\mu_T \right] dQ_\pi(h)\, dJ(\pi),$$

where the first equality holds because $k^*_{\alpha_T} - k_{T\alpha} \xrightarrow{p} 0$ and LR_T has an absolutely continuous asymptotic distribution under $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)$, the second equality holds because $Exp\text{-}LM_T - LR_T \xrightarrow{p} 0$ under $\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi)$ by Theorem A-2, and the third equality holds by Fubini's Theorem. The proof is analogous for $Exp\text{-}W_T$ and $Exp\text{-}LR_T$. *Q.E.D.*

PROOF OF LEMMA A-4: We make use of the following result, which follows, e.g., from Theorems 16.8 and 18.11 of Strasser (1985) (as described in Andrews and Ploberger (1992, proof of Lemma 3)): If (i) $LR_T \xrightarrow{d} \chi(\theta_0, c)$ under θ_0 and (ii) $E\chi(\theta_0, c) = 1$, then the densities $\{\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi): T \geqslant 1\}$ are contiguous to the densities $\{f_T(\theta_0): T \geqslant 1\}$. Condition (i) holds by Theorem 1. Condition (ii) is obtained as follows: Let $\text{mgf}(t)$ denote the moment generating function of a chi-square random variable with p degrees of freedom. We have $\text{mgf}(t) = (1-2t)^{-p/2}$. Then,

$$\text{(A.34)} \quad E\chi(\theta_0, c) = (1+c)^{-p/2} \times \int_\Pi E \exp\left(\frac{1}{2}\frac{c}{1+c}\left(H\mathscr{I}^{-1}(\pi)G(\pi)\right)'\left(H\mathscr{I}^{-1}(\pi)H'\right)^{-1}H\mathscr{I}^{-1}(\pi)G(\pi)\right) dJ(\pi)$$
$$= (1+c)^{-p/2}\int_\Pi \text{mgf}\left(\frac{1}{2}\frac{c}{1+c}\right) dJ(\pi)$$
$$= 1,$$

where $\mathscr{I}(\pi)$ and $G(\pi)$ denote $\mathscr{I}(\theta_0, \pi)$ and $G(\theta_0, \pi)$, respectively, the first equality holds by Fubini's Theorem, and the second equality uses the fact that the quadratic form in the exponent has chi-square distribution with p degrees of freedom for each fixed π. *Q.E.D.*

PROOF OF THEOREM A-2: Part (b) holds by the proof of Theorem A-1(b). For the remaining parts, given any $\varepsilon > 0$, consider the sets $\{|LR_T - \overline{LR}_T| > \varepsilon\}, \{|Exp\text{-}\overline{W}_T - Exp\text{-}W_T| > \varepsilon\}$, etc. for $T \geqslant 1$. The probabilities of these sets converge to zero as $T \to \infty$ under θ_0 by Theorem A-1. Hence, by contiguity (Lemma A-4), their probabilities also converge to zero under the densities $\{\int f_T(\theta_0 + B_T^{-1}h, \pi)\, dQ_\pi(h)\, dJ(\pi): T \geqslant 1\}$. *Q.E.D.*

A.2. Construction of Tables I and II

The values reported in Tables I and II are estimates of the desired asymptotic critical values obtained by (i) approximating the distribution of the integrals over $[\pi_0, 1-\pi_0]$ in (4.6) and (4.7) by averages over a fine grid of points $\Pi(N)$ and (ii) simulating the resultant averages by Monte Carlo. The grid $\Pi(N)$ is defined by

$$\text{(A.35)} \quad \Pi(N) = [\pi_0, 1-\pi_0] \cap \{\pi = j/N: j = 0, 1, \ldots, N\}.$$

The value of N was chosen to be 3,600 based on a comparison of the approximations generated by this method for the "sup" statistic with the numerical results for the "sup" statistic given by DeLong (1981) for $p \leqslant 4$. A single realization from the asymptotic distribution of the discretized version of (4.6) or (4.7) was obtained by simulating a p-vector $B_p(\cdot)$ of independent Brownian motions on [0, 1] at the discrete points in $\Pi(N)$ and then computing the discrete average of the appropriate function of $(B_p(\pi) - \pi B_p(1))'(B_p(\pi) - \pi B_p(1))/[\pi(1-\pi)]$. The number of repetitions R used was 10,000. The error in the rejection probabilities due to simulation has mean 0 and standard error approximately equal to $(\alpha(1-\alpha)/R)^{1/2}$. For $\alpha = .01, .05, .10$, the standard errors are .001, .002, and .003 respectively.

REFERENCES

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

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

ANDREWS, D. W. K., I. LEE, AND W. PLOBERGER (1994): "Optimal Changepoint Tests for Linear Regression," *Journal of Econometrics*, forthcoming.

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.

——— (1993): "Admissability of the Likelihood Ratio Test When a Nuisance Parameter Is Present Only under the Alternative," Cowles Foundation Discussion Paper No. 1058, Yale University.

——— (1994): "Testing for Serial Correlation Against an ARMA(1, 1) Process," unpublished manuscript, Cowles Foundation, Yale University.

BAI, J., R. LUMSDAINE, AND J. H. STOCK (1991): "Testing for and Dating Breaks in Integrated and Cointegrated Time Series," unpublished manuscript, Kennedy School of Government, Harvard University.

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

CHERNOFF, H., AND S. ZACKS (1964): "Estimating the Current Mean of a Normal Distribution Which Is Subject to Changes in Time," *Annals of Mathematical Statistics*, 35, 999–1028.

DAVIES, R. B. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

——— (1987): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

DELONG, D. M. (1981): "Crossing Probabilities for a Square Root Boundary by a Bessel Process," *Communications in Statistics—Theory and Methods*, A10(21), 2197–2213.

DOUKHAN, P., P. MASSART, AND E. RIO (1994): "Invariance Principles for Absolutely Regular Empirical Processes," *Annales de l'Institut Henri Poincaré*, 30, forthcoming.

ENGLE, R. F., D. F. HENDRY, AND J.-F. RICHARD (1983): "Exogeneity," *Econometrica*, 51, 277–304.

HANSEN, B. E. (1991): "Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis," Working Paper No. 296, Rochester Center for Economic Research, University of Rochester.

——— (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

JANDHYALA, V. K., AND I. B. MACNEILL (1991): "Tests for Parameter Changes at Unknown Times in Linear Regression Models," *Journal of Statistical Planning and Inference*, 27, 291–316.

KING, M. L., AND T. S. SHIVELY (1993): "Locally Optimal Testing When a Nuisance Parameter Is Present Only Under the Alternative," *Review of Economics and Statistics*, 75, 1–7.

LEHMANN, E. L. (1959): *Testing Statistical Hypotheses*. New York: Wiley.

NASON, J. M. (1991): "The Permanent Income Hypothesis When the Bliss Point Is Stochastic," unpublished manuscript, Department of Economics, University of British Columbia.

NYBLOM, J. (1989): "Testing for the Constancy of Parameters over Time," *Journal of the American Statistical Association*, 84, 223–230.

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.

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

POTTER, S. M. (1995): "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics*, 10, forthcoming.

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.