4-1-1992

Optimal Tests When a Nuisance Parameter is Present Only under the Alternative

Donald W.K. Andrews

Werner Ploberger

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale University
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1015

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

by

Donald W.K. Andrews and Werner Ploberger

April 1992

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

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

Werner Ploberger[1]
Cowles Foundation for Research in Economics
Yale University
and
Technische Universität Wien

February 1991

Revised: April 1992

ABSTRACT

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 change-point. Several other examples are discussed in the paper. The results of the paper are of interest, because the testing problem considered is non–standard and the classical asymptotic optimality results for the Wald, Lagrange multiplier (LM), and likelihood ratio (LR) tests do not apply.

A weighted average power criterion is used here to generate optimal tests. This criterion is quite 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 Wald, LM, and LR tests when standard regularity conditions hold. Nevertheless, in the non–standard cases of main interest, new optimal tests are obtained and the LR test is not found to be an optimal test.

JEL Classification No.: 211.

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 non–standard 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 common factors in autoregressive-moving average models, and of variable relevance and functional form in nonlinear 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 Wald, Lagrange multiplier (LM), and likelihood ratio (LR) tests do not hold in these non–standard problems.

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 that considered by Wald. The 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 we derive can be given a Bayesian interpretation. If one views the weight function referred to above as a prior, 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 $W_T(\pi)$ denote the standard Wald test of $\beta = 0$ 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 value after structural change, and $W_T(\pi)$ is the Wald test of $\beta = 0$ 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

$$\text{(1.1)} \qquad \text{Exp–W}_T = (1+c)^{-p/2} \int \exp\left[\frac{1}{2}\frac{c}{1+c} W_T(\pi)\right] dJ(\pi) .$$

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 and determines whether one is directing power against close or distant alternatives. The choice of c is discussed below.

Exponential LM and LR tests are defined analogously to Exp–W_T with the standard $LM_T(\pi)$ and $LR_T(\pi)$ test statistics replacing $W_T(\pi)$. 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. Simulations reported in Andrews, Lee, and Ploberger (1992) show that an optimal exponential test dominates the likelihood ratio test over a fairly wide variety of alternatives.

The general optimality properties of exponential tests are established here under a set of high–level assumptions. These assumptions are verified in the paper under primitive conditions in the leading example of one–time structural change with unknown change point. The general results are also applicable to a number of other examples as mentioned above. For brevity, primitive conditions under which the general high level assumptions hold are not given for each of these examples.

The examples covered by the general results include:

(i) Tests of One–Time Structural Change: This example is described in detail in Section 7 below.

(ii) Tests of Multiple Structural Change: This example consists of a parametric model with parameters $(\delta_1', \delta_2')'$ that may undergo m changes at unknown times $(T\pi_1, ..., T\pi_m)$ for $0 < \pi_1 < \cdots < \pi_m < 1$, where T is the sample size, $\pi_j \in (0,1)\ \forall j \leq m$, and m is some given integer. Under the alternative hypothesis, the true parameter vector is $(\delta_1', \delta_2')'$ for $t \leq T\pi_1$, $(\delta_1' + \beta_1', \delta_2')'$ for $T\pi_1 < t \leq T\pi_2, ..., (\delta_1' + \beta_{m-1}', \delta_2')'$ for $T\pi_{m-1} < t \leq T\pi_m$, and $(\delta_1' + \beta_m', \delta_2')'$ for $T\pi_m < t \leq T$. Under the null hypothesis, the parameter $\beta = (\beta_1', ..., \beta_m')' = \underset{\sim}{0}$ and the distribution of the data does not depend on $\pi = (\pi_1, ..., \pi_m)'$.

(iii) Tests of Cross–sectional Constancy: In this example, the observations are independent and the unknown parameter π partitions the sample space of some observed variable(s) into m+1 regions. In one region the model is indexed by the parameter $(\delta_1', \delta_2')'$ and in other regions it is indexed by $(\delta_1' + \beta_j', \delta_2')'$ for $j \leq m$. For example, in a linear regression model, π might be an unknown real number that partitions the regressor space into two regions according to whether a single regressor X_{1t} exceeds π or is less than π. Alternatively, π might be two unknown lines in R^2 that partition the regressor space of two regressors (X_{1t}, X_{2t}) into four regions. The extension to different parametric models is straightforward.

(iv) Tests of Threshold Effects in Threshold Autoregressive (TAR) Models: The simplest TAR model is of the form

$$(1.2) \qquad Y_t = \begin{cases} \delta_1 Y_{t-1} + U_t \text{ for } Y_{t-1} \leq \pi \\ (\delta_1 + \beta) Y_{t-1} + U_t \text{ for } Y_{t-1} > \pi \end{cases} \text{ for } t = 1, \ldots, T,$$

where $U_t \sim$ iid $N(0, \delta_2)$. This model and generalizations of it, including smooth transition AR models, have been applied in the physical and biological sciences, e.g., see Tong (1983, 1990), as well as in economics, e.g., see Potter (1989), Teräsvirta and Anderson (1991), and Hansen (1991a). A test of a threshold effect, viz., $H_0 : \beta = 0$ versus $H_1 : \beta \neq 0$, exhibits the feature that the threshold parameter π appears only under the alternative. Note that in smooth transition AR models the parameter π is not necessarily scalar. It is given by the vector of parameters that index the transition function.

(v) Tests of Common Factors: Consider a stationary ARMA (1,1) model parameterized as

$$(1.3) \qquad Y_t - \pi Y_{t-1} = (1-\pi)\delta_1 + U_t - (\pi + \beta) U_{t-1} \text{ for } t = 1, \ldots, T,$$

where $U_t \sim$ iid $N(0, \delta_2)$. Let $\theta = (\beta, \delta')'$, where $\delta = (\delta_1, \delta_2)'$. The null hypothesis for a test of a common factor in the autoregressive and moving average components of the model is given by $H_0 : \beta = 0$. Under the null, the model is $Y_t = \delta_1 + U_t$, which does not depend on π. Under the alternative, however, all four parameters β, δ_1, δ_2, and π are identified.

(vi) Tests of Variable Relevance and Functional Form in Nonlinear Models: Consider the nonlinear regression model

$$(1.4) \qquad Y_t = g(X_t, \delta_1) + \beta h(Z_t, \pi) + U_t \text{ for } t = 1, \ldots, T,$$

where $U_t \sim$ iid $N(0, \delta_2)$. For example, if Z_t is a scalar, $h(Z_t, \pi)$ might be of the Box–Cox form $(Z_t^{\pi} - 1)/\pi$. Let $\theta = (\beta, \delta')'$, where $\delta = (\delta_1', \delta_2)'$. Suppose X_t and Z_t are exogenous regressor variables. A test for the relevance of the regressors Z_t has null hypothesis $H_0 : \beta = 0$. Under the null, the parameter vector π disappears because the regressor vector Z_t does not belong in the model. The extension of this test to other parametric models is straightforward. We note that the general results below cover the case where π is finite or infinite dimensional, although the resulting optimal test statistic may be difficult to compute if π is infinite dimensional.

If one takes $Z_t = X_t$ in model (1.4), then a test of $H_0 : \beta = 0$ is a test of functional form of the nonlinear regression function and is covered by the general results of this paper. 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.

We note that several of the testing problems considered above have been analyzed recently by Hansen (1991a), though he does not address the question of choosing an optimal test. We also 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 remainder of this paper is organized as follows. Section 2 reviews related literature. Section 3 introduces the testing problem under consideration and the optimal test statistics. Section 4 provides an outline of the proof of optimality. Section 5 presents and discusses the assumptions employed, the optimality results, and the asymptotic null distribution of the optimal test statistics. Section 6 discusses the choice of a scalar constant c that indexes the class of optimal exponential test statistics. Section 7 treats the case of tests of structural change in nonlinear models with non-trending observations. Primitive conditions are given for these applications under which the assumptions of Section 5 hold.

2. RELATED LITERATURE

Here we briefly discuss results in the literature that are related to the optimality results given here. First, 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 (1989). 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.

Second, Chernoff and Zacks (1964, Sec. 8) have derived the average LM test statistic (i.e., $\int LM_T(\pi)d\pi$) via a local Bayesian approach in a very simple model of structural change with unknown changepoint. They consider iid normal variables with known variance and a mean that is subject to one–time change. They put a half–normal prior with variance v on the magnitude of change (which is appropriate for one–sided alternatives, such as positive changes in the mean), a normal prior with variance τ^2 on the pre–change value of the mean, and a uniform prior on the changepoint. They then show that the (one-sided) likelihood ratio statistic when these priors are used to specify the alternative is asymptotically equivalent to the (one–sided) average LM statistic as $v \to 0$ and $\tau^2 \to \infty$. This approach has been extended by Gardner (1969) to two–sided tests in the model above, by Sen and Srivastava (1973) to multivariate normal random variables, and by Jandhyala and MacNeill (1991) to the normal linear regression model. Farley and Hinich (1970) present closely related results for the simple linear regression model.

Although the results of Chernoff and Zacks (1964) *et al.* are useful and interesting, we believe they have several drawbacks relative to the results given in the present paper. First, the tests are designed to have power 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 non-trivial power even asymptotically. Second, the Chernoff–Zacks method requires that one puts a prior over certain nuisance parameters that appear under the null and the alternative, such as the pre–change value of the mean. The procedure used in this paper does not require a prior or weight function for these parameters. Third, with some effort the

Chernoff–Zacks method has been generalized to cover tests of one–time structural change in several models. It has not been operationalized, however, in models that are nearly as general as those covered by the results of the present paper. Nor has it been operationalized in the variety of testing problems (besides one–time change tests) to which the present paper applies.

Next, Nyblom (1989) presents optimality results for tests of parameter constancy in general parametric models. He considers martingale parameter alternatives rather than one–time change alternatives. His results are quite interesting, but can be criticized in that (1) they direct power only against very local alternatives and (2) they only apply when there are no unknown parameters under the null hypothesis, which rarely occurs in practice.

Lastly, King and Shively (1991) 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. As with the other papers discussed above, their tests direct power only against very local alternatives. In addition, the particular method of averaging power in the transformed parameter space seems subject to criticism, because it yields the fixed changepoint test with changepoint at $\pi = .5$ to be optimal and this test has relatively poor overall power properties, e.g., see Andrews, Lee, and Ploberger (1992).

3. 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 Wald*, *Lagrange multiplier* (LM), and *likelihood ratio* (LR) tests, denoted Exp–W_T, Exp–LM_T, and Exp–LR_T respectively.

We begin by introducing some notation and definitions. Let $(\Omega, \mathcal{F}, P)$ denote a probability space on which all of the random elements introduced below are defined. Let Y_T denote the data vector when the sample size is T for T = 1, 2, ... Consider a parametric family $\{f_T(y_T, \theta, \pi) : \theta \in \Theta, \pi \in \Pi\}$ of densities of 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 $f_T(\theta,\pi) = f_T(Y_T, \theta, \pi)$. 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, ..., 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. See Section 7 for a more explicit discussion of the factoring of $f_T(\theta,\pi)$ into known and unknown terms in the context of tests of structural change.

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

(3.1) $$H_0 : \theta = \theta_0, \text{ where } \theta_0 = (0', \delta_0')' \text{ for some } \delta_0 \in R^q .$$

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

(3.2) $$\begin{aligned} &H_1 : \theta = (\beta_0', \delta_0')' \text{ for some } \beta_0 \in R^p \text{ such that } \beta_0 \neq 0 \text{ and some } \delta_0 \in R^q \text{ and} \\ &\quad\text{the likelihood function depends on the parameter } \pi \text{ for some } \pi \in \Pi . \end{aligned}$$

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 non–regular and outside the domain of standard asymptotic optimality results. In particular, the standard Wald, LM, and LR statistics do not have their standard asymptotic distributions or their 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 non-random $s \times s$ diagonal matrix B_T that satisfies $[B_T^{-1}]_{jj} \to 0$ as $T \to \infty$ $\forall j \leq s$. (In models with non-trending variables, $B_T = \sqrt{T}\, I_s$, where I_s is the $s \times s$ identity matrix.) For particular weight functions (i.e., probability measures) $Q_\pi(h)$ on the values of h and a chosen weight function $J(\pi)$ on the values of π, we show that the tests Exp–K_T for K = W, LM, or LR have the greatest weighted average power asymptotically in the class of all tests of asymptotic significance level α. That is, these tests maximize

$$(3.3) \qquad \overline{\lim_{T\to\infty}} \int P(\varphi_T \text{ rejects} \mid \theta_0 + B_T^{-1}h, \pi) dQ_\pi(h) dJ(\pi)$$

over all tests φ_T of asymptotic level α (and the $\overline{\lim}_{T\to\infty}$ equals $\lim_{T\to\infty}$ for the tests Exp–K_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 $f_T(\theta_0, \pi)$, then the tests Exp–K_T for K = W, LM, or LR have greatest power asymptotically against this alternative in the class of all tests of asymptotic level α.

The asymptotically optimal test statistics Exp–W_T, Exp–LM_T, and Exp–LR_T are defined by

$$(3.4) \qquad \text{Exp–}K_T = (1+c)^{-p/2} \int \exp\left[\frac{1}{2}\frac{c}{1+c} K_T(\pi)\right] dJ(\pi) \text{ for K = W, LM, and LR,}$$

where $W_T(\pi)$, $LM_T(\pi)$, and $LR_T(\pi)$ are just the standard Wald, LM, and LR tests, respectively, of 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, $W_T(\pi)$ is just the standard Wald test of structural change occurring at the time $[T\pi]$ (see Andrews and Fair (1988) for a general treatment of such tests). One rejects H_0 if Exp–K_T exceeds a critical value $k_{T\alpha}$ that is determined using the asymptotic null distribution of Exp–K_T (or its finite sample distribution in some cases).

Note that Exp–K_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 large structural changes. In the special case where $J(\pi)$ is a pointmass at a single value π_0, Exp–K_T reduces to $(1+c)^{-p/2}\exp\left[\frac{1}{2}\frac{c}{1+c}K_T(\pi_0)\right]$, the optimal test rejects if and only if $K_T(\pi_0)$ exceeds some constant (i.e., the optimal test equals the standard Wald, LM, or LR test for fixed π_0), and the optimal test is independent of c. When $J(\pi)$ is not a pointmass distribution, however, the optimal tests Exp–K_T for K = W, LM, and LR depend on c. The larger is c, the more power is directed at alternatives for which β is large.

The limits as $c \to 0$ of the exponential Wald, LM, and LR statistics (suitably normalized) are equal to the "average Wald," "average LM," and "average LR" statistics respectively. In particular,

$$\text{(3.5)} \qquad \lim_{c\to 0} 2(\text{Exp–}K_{Tc} - 1)/c = \int K_T(\pi)dJ(\pi) \text{ for } K = W, LM, \text{ and } LR,$$

where Exp–K_{Tc} denotes the statistic Exp–K_T constructed using the constant c. Thus, the average Wald, LM, and LR statistics are the limiting cases of the exponential Wald, LM, and LR statistics, respectively, 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), Gardner (1969), Nyblom (1989), Jandhyala and MacNeill (1991), and Hansen (1990) among others.

At the other extreme, the limits as $c \to \infty$ of the exponential Wald, LM, and LR statistics (suitably normalized) are given by

(3.6) $$\lim_{c \to \infty} \log\left[(1+c)^{p/2}\text{Exp--}K_{Tc}\right] = \log\int \exp\left[\tfrac{1}{2}K_T(\pi)\right] dJ(\pi) \text{ for } K = W, \text{LM, and LR.}$$

Thus, for testing against distant alternatives, the optimal test statistics are still of an average exponential form. These statistics have not been considered previously in the literature.

We note that if the constant $c/(1+c)$, which appears in the definition of Exp–K_T, is replaced by a constant r, which can take any positive value, then the limits as $r \to \infty$ of the exponential Wald, LM, and LR statistics (suitably normalized) are the "sup Wald," "sup LM," and "sup LR" statistics respectively. More specifically, let $\Pi^* \subset \Pi$ be the support of $J(\cdot)$. Then,

(3.7) $$\lim_{r \to \infty}(\log \text{Exp--}K_T^r)/r = \sup_{\pi \in \Pi^*} K_T(\pi) \text{ for } K = W, \text{LM, and LR,}$$

where Exp–K_T^r denotes the statistic Exp–K_T with $c/(1+c)$ replaced by r and $(1+c)^{p/2}$ replaced by 1. Hence, the sup Wald, LM, and LR tests are designed for distant alternatives, but are of a more extreme form than the optimal test, which is of an exponential average form. The sup Wald, LM, and LR tests have been considered in the literature by Davies (1977, 1987), Hawkins (1987), Kim and Siegmund (1989), and Andrews (1989) among others.

4. OUTLINE OF THE PROOF OF OPTIMALITY

We now give an outline of the proof of the asymptotic optimality properties of Exp–W_T, Exp–LM_T, and Exp–LR_T. (The proof yields the asymptotic null distribution of Exp–W_T, Exp–LM_T, and Exp–LR_T as a by-product.) 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,

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

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 LR_T statistic, however, has two drawbacks. First, it depends on the unknown parameter θ_0, and second, it involves a computationally burdensome double integral. In consequence, we introduce the exponential Wald, LM, and LR statistics in place of LR_T. These statistics do not depend on θ_0 and involve only single integrals with respect to $J(\pi)$, which are usually relatively easy to compute. For example, in tests of structural change, they just equal finite sums.

The asymptotic optimality of Exp–W_T, Exp–LM_T, and Exp–LR_T (Theorem 4 below) follows from the optimality of LR_T if we can show that LR_T, Exp–W_T, Exp–LM_T, and Exp–LR_T are asymptotically equivalent (i.e., LR_T – Exp–$W_T \xrightarrow{p} 0$, Exp–W_T – Exp–$LM_T \xrightarrow{p} 0$, and Exp–LM_T – Exp–$LR_T \xrightarrow{p} 0$) under the null and under the local alternatives $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$. The proof of the latter requires several steps. First, we show that the normalized maximum likelihood (ML) estimator for each fixed π, viz., $B_T(\hat{\theta}(\pi) - \theta_0)$, is uniformly well approximated under the null, i.e., under θ_0, by its linear expansion $\overline{\theta}(\pi)$ (Lemma 1). We then show that the LR_T statistic is asymptotically equivalent under θ_0 to a double integral of an exponential function of the (unobserved) approximate ML estimator $\overline{\theta}(\pi)$, call this statistic $\overline{LR}_T$ (Lemma 2). Next, we show that if Q_π is a particular normal distribution for each π, then $\overline{LR}_T$ simplifies to a *single* integral approximate exponential Wald statistic that is based on $\overline{\theta}(\pi)$, call it Exp–$\overline{W}_T$ (Theorem 1(b)). In addition, we show that Exp–$\overline{W}_T$ is asymptotically equivalent to Exp–W_T under θ_0 (Theorem 1(c)), that Exp–W_T is asymptotically equivalent to Exp–LM_T under θ_0 (Theorem 1(d)), and that Exp–LM_T is asymptotically equivalent to

Exp–LR_T under θ_0 (Theorem 1(e)). This gives the asymptotic equivalence of LR_T, Exp–W_T, Exp–LM_T, and Exp–LR_T under the null hypothesis.

Next, we obtain straightforwardly the asymptotic distribution of the approximate exponential Wald statistic Exp–W_T (Theorem 2(a)). Given the asymptotic equivalence under θ_0 established previously, this yields the asymptotic distributions of LR_T, Exp–W_T, Exp–LM_T, and Exp–LR_T under the null (Theorem 2(b)–(e)). The convergence in distribution of LR_T under θ_0 is used to establish the contiguity of the local alternatives $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ to $\{f_T(\theta_0) : T \geq 1\}$ (Lemma 3). Contiguity plus asymptotic equivalence under θ_0 imply that LR_T, Exp–W_T, Exp–LM_T, and Exp–LR_T are asymptotically equivalent under the local alternatives $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ (Theorem 3). The latter result plus the optimality of the LR_T statistic give the asymptotic optimality of Exp–W_T, Exp–LM_T, and Exp–LR_T (Theorem 4).

5. ASSUMPTIONS AND OPTIMALITY RESULTS

First, we introduce some notation. Let $\ell_T(\theta,\pi) = \log f_T(\theta,\pi)$. Let $D\ell_T(\theta,\pi)$ denote the s–vector of partial derivatives of $\ell_T(\theta,\pi)$ with respect to θ. Let $D^2\ell_T(\theta,\pi)$ denote the $s \times s$ matrix of second partial derivatives of $\ell_T(\theta,\pi)$ with respect to θ. (Note that $D\ell_T(\theta_0, \pi)$ and $D^2\ell_T(\theta_0, \pi)$ depend on π in general even though $f_T(\theta_0, \pi)$ and $\ell_T(\theta_0, \pi)$ do not.) Let θ_0 denote the true value of θ under the null H_0.

We consider the case where the appropriate norming factors for $D\ell_T(\theta,\pi)$ and $D^2\ell_T(\theta,\pi)$ (so that each is $O_p(1)$ but not $o_p(1)$) are non–random diagonal $s \times s$ matrices B_T^{-1} and $B_T^{-1} \times B_T^{-1}$ respectively. For non–trending data, the matrix B_T is just $\sqrt{T}\, I_s$. For data with deterministic time trends, B_T is more complicated. For example, in a normal linear regression model with r non–trending regressors plus the regressors t and t^2, B_T equals $\text{diag}\{\sqrt{T}\, I_r, T, T^{3/2}\}$. Note that the matrices $\{B_T\}$ that are suitable for norming $D\ell_T(\theta,\pi)$ and $D^2\ell_T(\theta,\pi)$ 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 non–trivial 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 $f_T(\theta_0, \pi)$ for T = 1, 2,.... Let $\lambda_{min}(A)$ denote the smallest eigenvalue of a matrix A.

The likelihood function/parametric model is assumed to satisfy:

ASSUMPTION 1: (a) $f_T(\theta,\pi)$ *does not depend on* π *when* $\theta = \theta_0$.

(b) θ_0 *is an interior point of* Θ.

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

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

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

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

(g) $\sup_{\pi\in\Pi}\|B_T^{-1}D\ell_T(\theta_0, \pi)\| = O_p(1)$ *under* θ_0.

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

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 Newey (1991) or Andrews (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 is bounded away from 0 and 1. That is, one cannot consider change points that are arbitrarily close to the beginning or end of the sample. In the regime switching example with observed regimes, uniformity requires that Π is such that the probability of a regime occurring is not arbitrarily close to 0 or 1. Assumption 1(g) is implied by Assumption 4 below, so we defer discussion of it.

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

$$\ell_T(\hat{\theta}(\pi), \pi) = \max_{\theta \in \Theta} \ell_T(\theta, \pi) \quad \forall \pi \in \Pi \tag{5.1}$$

with probability that goes to one as $T \to \infty$ under θ_0. We assume that the parametric model is sufficiently regular that the ML estimator $\hat{\theta}(\pi)$ is 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.

For some applications, this assumption can be verified using results in the literature. In other cases, one can use a result given in Andrews (1989, 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.

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

$$\begin{aligned} &\tilde{\theta} \in \tilde{\Theta} = \{\theta \in \Theta : \theta = (0', \delta')' \text{ for some } \delta \in R^q\} \text{ and} \\ &\ell_T(\tilde{\theta}, \pi) = \max_{\theta \in \tilde{\Theta}} \ell_T(\theta, \pi) \end{aligned} \tag{5.2}$$

with probability that goes to one as $T \to \infty$ under θ_0. Note that since $\ell_T(\theta,\pi)$ does not depend on π when θ is in the null hypothesis, $\tilde{\theta}$ does not depend on π.

For known $\pi \in \Pi$, the standard Wald, LM, and LR test statistics for testing H_0 against H_1 (as defined in (3.1) and (3.2)) are given by

$$\begin{aligned} &W_T(\pi) = (HB_T\hat{\theta}(\pi))'\left[H\mathcal{I}_T^{-1}(\hat{\theta}(\pi),\, \pi)H'\right]^{-1}HB_T\hat{\theta}(\pi), \\ &LM_T(\pi) = \left[B_T^{-1}D\ell_T(\tilde{\theta},\pi)\right]'\mathcal{I}_T^{-1}(\tilde{\theta},\pi)B_T^{-1}D\ell_T(\tilde{\theta},\pi), \text{ and} \\ &LR_T(\pi) = -2(\ell_T(\tilde{\theta},\pi) - \ell_T(\hat{\theta}(\pi),\, \pi)), \text{ where} \\ &H = [I_p \,\vdots\, 0] \subset R^{p\times s} \text{ and } \mathcal{I}_T(\theta,\pi) = -B_T^{-1}D^2\ell_T(\theta,\pi)B_T^{-1}\,. \end{aligned} \tag{5.3}$$

Alternatively, one can define $\mathcal{I}_T(\theta,\pi)$ to be of outer product, rather than Hessian, form. Note that only the first p elements of $B_T^{-1}D\ell_T(\tilde{\theta},\pi)$ are non–zero in the definition of $LM_T(\pi)$, because $\frac{\partial}{\partial\delta}\ell_T(\tilde{\theta},\pi) = 0$ by the first order conditions for the restricted estimator $\tilde{\theta}$ (with probability that goes to one as $T \to \infty$). 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 Wald, LM, and LR statistics, Exp–W_T, Exp–LM_T, and Exp–LR_T, respectively, are defined by combining (3.4) and (5.3).

As shown in the following lemma, the (unrestricted) ML estimator $\hat{\theta}(\pi)$, suitably shifted and scaled, can be approximated under θ_0 by the score function $D\ell_T(\theta_0,\, \pi)$ suitably scaled. We refer to the latter as the *approximate* ML estimator $\overline{\theta}(\pi)$. By definition,

$$\overline{\theta}(\pi) = \mathcal{I}^{-1}(\theta_0,\, \pi)B_T^{-1}D\ell_T(\theta_0,\, \pi). \tag{5.4}$$

LEMMA 1: *Under the null hypothesis and Assumptions* 1 *and* 2,

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

This result is useful, because the large sample properties of $\overline{\theta}(\pi)$ are usually easy to determine. For each fixed π, $\overline{\theta}(\pi)$ usually satisfies the conditions for a central limit theorem. For example, in the simple case of independent identically distributed (iid) random variables (rv's), $\overline{\theta}(\pi)$ is just $1/\sqrt{T}$ times the sum of T iid rv's that typically have mean zero and finite variance.

We now show that the LR_T statistic of (4.1) is asymptotically equivalent to the following function of the approximate ML estimator $\overline{\theta}(\pi)$:

$$(5.5)\quad \overline{LR}_T = \int \exp\left[\tfrac{1}{2}\overline{\theta}(\pi)'\mathcal{I}(\theta_0,\pi)\overline{\theta}(\pi)\right]\int \exp\left[-\tfrac{1}{2}(\overline{\theta}(\pi)-h)'\mathcal{I}(\theta_0,\pi)(\overline{\theta}(\pi)-h)\right]dQ_\pi(h)dJ(\pi)\ .$$

To obtain this result, we assume Π and $\{Q_\pi : \pi \in \Pi\}$ satisfy:

ASSUMPTION 3′: $\{Q_\pi(\cdot) : \pi \in \Pi\}$ *is uniformly tight* (*i.e.*, $\forall \epsilon > 0\ \exists M < \infty$ *such that* $Q_\pi(\{h \in R^s : \|h\| \geq M\}) < \epsilon\ \ \forall \pi \in \Pi$, *where* $\|\cdot\|$ *is the Euclidean norm*).

Assumption 3′ is satisfied by the particular weight functions $\{Q_\pi(\cdot) : \pi \in \Pi\}$ that are introduced below.

LEMMA 2: *Under the null hypothesis and Assumptions* 1, 2, *and* 3′, $LR_T - \overline{LR}_T \xrightarrow{p} 0$.

Next we introduce a particular choice for the weight functions $\{Q_\pi(\cdot) : \pi \in \Pi\}$ that allows the double integral in $\overline{LR}_T$ to be reduced to a single integral. 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

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

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 $<h,\ell>_\pi = h'\mathcal{I}(\theta_0, \pi)\ell$ for $h, \ell \in R^s$; 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 $\{a_{1\pi}, ..., a_{p\pi}\}$ be some basis of $V_\pi^\perp$ and define $A_\pi = [a_{1\pi} \cdots a_{p\pi}] \in R^{s \times p}$. For example (by the proof of Lemma A–1 in the Appendix), one can take

$$(5.7) \qquad A_\pi = \begin{bmatrix} I_p \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}' \end{bmatrix}, \text{ where } \mathcal{I}(\theta_0, \pi) = \begin{bmatrix} \mathcal{I}_{1\pi} & \mathcal{I}_{2\pi} \\ \mathcal{I}_{2\pi}' & \mathcal{I}_{3\pi} \end{bmatrix}$$

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

$$(5.8) \qquad V_\pi^\perp = \left\{ h \in R^s : h = \begin{bmatrix} \lambda \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\lambda \end{bmatrix} \text{ for some } \lambda \in R^p \right\} .$$

Next, define

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

$$(5.9) \qquad = \begin{bmatrix} \left[\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\right]^{-1} & -\left[\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\right]^{-1}\mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1} \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\left[\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\right]^{-1} & \mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\left[\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'\right]^{-1}\mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1} \end{bmatrix}.$$

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

We assume:

ASSUMPTION 3: $Q_\pi = N(0, c\Sigma_\pi)$ $\forall \pi \in \Pi$ *for some positive constant* c (*that does not depend on* π).

Under Assumption 3, 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 $V_\pi^\perp$. Note that Assumption 3 implies Assumption 3′ when $\mathcal{I}(\theta_0, \pi)$ satisfies Assumption 1(f). Hence, the conclusion of Lemma 2 holds under Assumptions 1–3.

We now introduce an approximate exponential Wald statistic denoted Exp–$\overline{W}_T$. Corresponding to the Wald statistic for known π, $W_T(\pi)$, we define an (unobserved) approximate Wald statistic for known π based on the approximate ML estimator $\overline{\theta}(\pi)$ by

$$(5.10) \qquad \overline{W}_T(\pi) = (H\overline{\theta}(\pi))'\left[H\mathcal{I}^{-1}(\theta_0, \pi)H'\right]^{-1}H\overline{\theta}(\pi)\,.$$

The approximate exponential Wald test statistic is defined by combining (5.10) with

$$(5.11) \qquad \text{Exp–}\overline{W}_T = (1+c)^{-p/2}\int \exp\left[\frac{1}{2}\frac{c}{1+c}\overline{W}_T(\pi)\right]dJ(\pi)\,.$$

We require the following assumption regarding the restricted ML estimator $\tilde{\theta}$. In particular applications it can be verified using existing proofs of the consistency of ML estimators.

ASSUMPTION 4: (a) $f_T(\theta,\pi)$ *does not depend upon* π *for all* θ *in the null hypothesis, i.e.,* $\forall\theta \in \Theta \cap V$. (b) $\tilde{\theta} \xrightarrow{p} \theta_0$ *under* θ_0.

The following result shows that when $\{Q_\pi : \pi \in \Pi\}$ is taken as in Assumption 3, LR_T, $\overline{LR}_T$, Exp–$\overline{W}_T$, Exp–W_T, Exp–LM_T, and Exp–LR_T are all asymptotically equivalent under θ_0.

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

Here and in the results below that invoke Assumption 4, Assumption 4 is required only for the parts of the results that involve Exp–LM_T and Exp–LR_T.

Next, we determine the asymptotic null distribution of the approximate exponential Wald statistic Exp–W_T. In view of Theorem 1, this also yields the asymptotic null distribution of the test statistics of interest Exp–W_T, Exp–LM_T, and Exp–LR_T and of the statistic LR_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

$$\nu(\cdot) \to (1+c)^{-p/2}\int \exp\left[\frac{1}{2}\frac{c}{1+c}(H\nu_1(\pi))'(H\nu_2(\pi)H')^{-1}H\nu_1(\pi)\right]dJ(\pi) \tag{5.12}$$

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:

ASSUMPTION 5: $B_T^{-1}D\ell_T(\theta_0, \cdot) \Rightarrow G(\theta_0, \cdot)$ *under* θ_0 (*as processes indexed by* $\pi \in \Pi$) *for some mean zero* R^s*–valued Gaussian stochastic process* $\{G(\theta_0, \pi) : \pi \in \Pi\}$ *that has* $EG(\theta_0, \pi)G(\theta_0, \pi)' = \mathcal{I}(\theta_0, \pi)$ $\forall \pi \in \Pi$ *and has 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 central limit theorem for a partial sum process, as with tests of structural change, or by applying an empirical process central limit theorem, as with the other examples mentioned above.

Note that the stochastic processes $\nu_T(\pi)$ and $\nu(\pi)$ referred to in the discussion of weak convergence above correspond to $(\mathcal{I}^{-1}(\theta_0, \pi)B_T^{-1}D\ell_T(\theta_0, \pi), \mathcal{I}^{-1}(\theta_0, \pi))$ and $(\mathcal{I}^{-1}(\theta_0, \pi)G(\theta_0, \pi), \mathcal{I}^{-1}(\theta_0, \pi))$ 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 (5.12).

The asymptotic null distribution of Exp–$\overline{W}_T$, Exp–W_T, Exp–LM_T, Exp–LR_T, and LR_T is shown in the following theorem to equal that of the random variable

$$\chi(\theta_0, c) = (1+c)^{-p/2}\int \exp\Big[\frac{1}{2}\frac{c}{1+c}(H\mathcal{I}^{-1}(\theta_0, \pi)G(\theta_0, \pi))' \tag{5.13}$$
$$\times (H\mathcal{I}^{-1}(\theta_0, \pi)H')^{-1}H\mathcal{I}^{-1}(\theta_0, \pi)G(\theta_0, \pi)\Big]dJ(\pi)\ .$$

THEOREM 2: *Under the null hypothesis and Assumptions* 1–5, (a) Exp–$\overline{W}_T \xrightarrow{d} \chi(\theta_0, c)$, (b) Exp–$W_T \xrightarrow{d} \chi(\theta_0, c)$, (c) Exp–$LM_T \xrightarrow{d} \chi(\theta_0, c)$, (d) Exp–$LR_T \xrightarrow{d} \chi(\theta_0, c)$, *and* (e) $LR_T \xrightarrow{d} \chi(\theta_0, c)$.

COMMENT: 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 Wald, LM, and LR tests by simulating the distribution $\chi(\theta_0, c)$, see Section 7.2 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 (1991a, Sec. 7) for a method of simulating a single realization of $\chi(\theta^*, c)$ (which is a function of the stochastic process $G(\theta^*, \cdot)$). See Andrews, Lee, and Ploberger (1992) for a sequential method that minimizes the number of repetitions that are needed to obtain critical values via the simulation of $\chi(\theta^*, \pi)$.

Next, we establish the asymptotic equivalence of the test statistics considered above under the local alternatives. To do this, we establish the *contiguity* of the densities $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ to the densities $\{f_T(\theta_0) : T \geq 1\}$. By definition, contiguity holds if $\int_{C_T} f_T(\theta_0)d\mu_T \to 0$ implies $\int_{C_T}\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)d\mu_T \to 0$ for any sequence of (measurable) sets $\{C_T : T \geq 1\}$ such that C_T is determined by Y_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 $\int_{C_T}\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)d\mu_T$. Contiguity is used to establish the asymptotic equivalence of LR_T, $\overline{LR}_T$, Exp–$\overline{W}_T$, Exp–W_T, Exp–LM_T and Exp–LR_T under the local alternatives $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ by taking the sets C_T to equal $\{|LR_T - \overline{LR}_T| > \epsilon\}$, $\{|\overline{LR}_T - \text{Exp–}\overline{W}_T| > \epsilon\}$, etc. for arbitrary $\epsilon > 0$. Since these sets converge in probability to zero under $\{f_T(\theta_0) : T \geq 1\}$ by Theorem 1, they also do under the local alternatives if contiguity holds. We establish contiguity by using a result of LeCam that states that the convergence in distribution under $\{f_T(\theta_0) : T \geq 1\}$ of the likelihood ratio LR_T to a distribution with expectation one implies contiguity.

LEMMA 3: *Under Assumptions* 1–3 *and* 5, *the densities* $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ *are contiguous to the densities* $\{f_T(\theta_0) : T \geq 1\}$.

THEOREM 3: *Under the local alternative densities* $\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)$ *for* Y_T *for* $T \geq 1$ *and Assumptions* 1–5, (a) $LR_T - \overline{LR}_T \xrightarrow{p} 0$, (b) $\overline{LR}_T = \text{Exp–}\overline{W}_T$, (c) Exp–$\overline{W}_T$ – Exp–$W_T \xrightarrow{p} 0$, (d) Exp–W_T – Exp–$LM_T \xrightarrow{p} 0$, *and* (e) Exp–LM_T – Exp–$LR_T \xrightarrow{p} 0$.

Theorem 3 and the optimality of the LR_T test for testing the simple null $f_T(\theta_0)$ against the simple alternative $\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)$ (via the Neyman–Pearson Lemma) yield the main result of this paper – an asymptotic optimality result for the exponential Wald, LM, and LR tests.

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 $\{k_{T\alpha} : T \geq 1\}$ be a sequence of critical values (possibly random) such that the exponential Wald, LM, or LR tests, i.e.,

$$\xi_T = 1(\text{Exp–}W_T > k_{T\alpha}),\ \xi_T = 1(\text{Exp–}LM_T > k_{T\alpha}),\text{ or} \tag{5.14}$$
$$\xi_T = 1(\text{Exp–}LR_T > k_{T\alpha}),$$

respectively, has asymptotic level α. Such critical values can be determined by the method described following Theorem 2.

The following theorem establishes that the exponential Wald, LM, and LR tests have greatest asymptotic weighted average power for the weight functions $\{Q_\pi(\cdot) : \pi \in \Pi\}$ and $J(\cdot)$ against the local alternatives $\{f_T(\theta_0 + B_T^{-1}h, \pi) : T \geq 1\}$ for $h \in R^s$ and $\pi \in \Pi$ amongst all tests of asymptotic level α:

THEOREM 4: *Under Assumptions* 1–5, *for any sequence of asymptotically level* α *tests* $\{\varphi_T : T \geq 1\}$, *a sequence of asymptotically level* α *exponential Wald,* LM, *or* LR *tests* $\{\xi_T : T \geq 1\}$ *satisfies*

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

(*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 4 can be interpreted in two ways. First, it provides a greatest asymptotic weighted average power result for the exponential Wald, LM, and LR tests against the alternatives $\{f_T(\theta_0 + B_T^{-1}h, \pi) : h \in R^s, \pi \in \Pi\}$ for $T \geq 1$. Second, it shows that the exponential Wald, LM, and LR tests have the greatest asymptotic power against the single sequence of local alternatives $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ amongst all tests of asymptotic level α. This follows from Theorem 4 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. The weighted average power of a test φ_T can be re–written as an integral with respect to a normal distribution that is not singular as follows. Let $\lambda \sim N(0, c(\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}')^{-1})$ and $h = (\lambda', -\lambda'\mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1})'$. Then, λ has nonsingular variance and $h \sim Q_\pi = N(0, c\Sigma_\pi)$ as desired. In consequence, we have

$$\begin{aligned}
&\int [\int \varphi_T f_T(\theta_0 + B_T^{-1}h, \pi) d\mu_T] dQ_\pi(h) dJ(\pi) \\
(5.15)\quad &= \int \left[\int \varphi_T f_T\left(\theta_0 + B_T^{-1} \begin{bmatrix} I_p \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}' \end{bmatrix} \lambda, \pi\right) d\mu_T \right] n(\lambda; 0, c(\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}')^{-1}) d\lambda dJ(\pi),
\end{aligned}$$

where $n(\lambda; 0, c(\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}')^{-1})$ denotes a multivariate normal density with mean 0 and nonsingular covariance matrix $c(\mathcal{I}_{1\pi} - \mathcal{I}_{2\pi}\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}')^{-1}$ evaluated at the p–vector λ.

6. CHOICE OF c

In this section, we discuss the choice of the constant c that appears in the definition of the optimal exponential test statistics. Recall that c is the scale factor of the weight function Q_π that is used in constructing the exponential tests (see Assumption 3). Small values of c place most weight on alternatives for which β is small. Large values of c place most weight on alternatives for which β is large. The optimal test depends on the choice of c.

There are two ways of choosing c. One can choose some fixed value of c or one can formulate some data–dependent method of determining c. In the context of tests of structural change, see Section 7 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. In the one–time structural change case, our preferred choice of c is the limiting value $c = \infty$, see Section 7.5 below. Asymptotic critical values for this choice of c and for the other limiting case $c = 0$ are given in Section 7.2 below.

In some applications, the power properties of the optimal tests of parameter constancy may be more sensitive to the choice of c than in the structural change example. In such applications, it is useful to have available a more finely–tuned value of c. We suggest the following data–dependent method. Suppose the practitioner can supply a vector of parameters β^* against which he would like to direct power. Then, a data-dependent choice of c is given by

$$\hat{c} = \int (\beta^*)' \left[H \mathcal{I}_T^{-1}(\tilde{\theta},\pi) H' \right]^{-1} \beta^* dJ(\pi)/p. \tag{6.1}$$

(The estimator $\tilde{\theta}$ can be replaced by $\hat{\theta}(\pi)$ if desired.) This formula chooses c such that the mean of $(H\theta)' \left[H\mathcal{I}^{-1}(\theta_0, \pi)H' \right]^{-1} H\theta = \beta' \left[H\mathcal{I}^{-1}(\theta_0, \pi)H' \right]^{-1} \beta$, when (θ,π) are distributed under (Q_π, J), is approximately $(\beta^*)' \left[H\mathcal{I}^{-1}(\theta_0, \pi)H' \right]^{-1} \beta^*$. That is, c is chosen such that the mean of the "magnitudes" of the jump β under (Q_π, J) approximately equals the magnitude of the jump β^* against which one wants to direct power. The approximation arises due to the use of $\mathcal{I}_T^{-1}(\tilde{\theta},\pi)$ in place of the unobserved quantity $\mathcal{I}(\theta_0, \pi)$ in (6.1). The approximation error goes to 0 as $T \to \infty$, because $\mathcal{I}_T^{-1}(\tilde{\theta},\pi)$ consistently estimates $\mathcal{I}(\theta_0, \pi)$ uniformly over $\pi \in \Pi$. The use of $\beta' \left[H\mathcal{I}^{-1}(\theta_0, \pi)H' \right]^{-1} \beta$ to measure the "magnitude" of a jump β is natural in some sense, because for fixed π the power of a Wald, LM, or LR test depends on β only through this quantity.

The formula for $\hat{c}$ given in (6.1) is relatively easy to compute in most cases. It requires that one simulate the critical value for the test on a case by case basis, however, since the dependence of the critical value on c and the consideration of all values of c in $[0,\infty]$ precludes tabulation of critical values.

7. OPTIMAL TESTS OF STRUCTURAL CHANGE IN NONLINEAR MODELS WITH NONTRENDING OBSERVATIONS

In this section we consider tests of one–time structural change with unknown change point. The tests are designed to detect a one–time change in the value of a parameter (but they have power against more general forms of structural change, e.g., see Andrews (1989, Thm. 5 and Cor. 2) and Ploberger, Kramer and Kontrus (1989, Cor. 1)). The models we consider are dynamic nonlinear models that are suitable for ML estimation and are based on nontrending observations. The tests can be applied to both "pure" structural change problems and "partial" structural change problems. With pure structural change, the entire parameter vector is subject to change at some time point under the alternative hypothesis. With partial structural change, only a specified subvector of the parameter is subject to change under the alternative hypothesis.

A simple example where the results of this section can be applied is in a test for constancy of the intercept in an AR or ARMA model for the growth rate of a macroeconomic variable such as GNP. Tests of this sort have attracted some attention in the literature, e.g., see Perron (1991) and Bai, Lumsdaine, and Stock (1991). A more complicated example is a test of parameter constancy in a nonlinear rational expectations model estimated by GMM methods. Again, tests of this sort have attracted attention in the literature, e.g., see Nason (1991). There are endless possibilities for further applications.

7.1. The Model and the Optimal Test Statistics

We now introduce the model, the hypotheses of interest, and the optimal test statistics. The sample of observations is given by

$$(7.1) \qquad Y_T = \{(S_t, X_t) : t \leq T\},$$

where $\{S_t : t \leq T\}$ are endogenous variables and $\{X_t : t \leq T\}$ are weakly exogenous variables. (Weak exogeneity of $\{X_t : t \leq T\}$, see Engle, Hendry, and Richard (1983), means that the likelihood function for Y_T can be factored into two pieces, one of which contains conditional distributions of S_t and depends on θ and the other of which contains conditional distributions of X_t and does not depend on θ, see below.) Let

$$\{g_t(\delta_1, \delta_2) : \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\} \tag{7.2}$$
$$= \{g_t(S_t | S_1, ..., S_{t-1}; X_1, ..., X_t; \delta_1, \delta_2) : \delta_1 \in \Delta_1, \delta_2 \in \Delta_2\}$$

denote a parametric family of conditional densities (with respect to some measure) of S_t given $S_1, ..., S_{t-1}, X_1, ..., X_t$ evaluated at the rv's $S_1, ..., S_t, X_1, ..., X_t$, where $\Delta_1 \subset R^p$, $\Delta_2 \subset R^{q-p}$, and $p \leq q$. Let

$$h_t = h_t(X_t | S_1, ..., S_{t-1}; X_1, ..., X_{t-1}) \tag{7.3}$$

denote the conditional density (with respect to some measure) of X_t given $S_1, ..., S_{t-1}, X_1, ..., X_{t-1}$ evaluated at the rv's $S_1, ..., S_{t-1}, X_1, ..., X_t$. By the assumption of weak exogeneity, h_t does not depend on $\delta = (\delta_1', \delta_2')'$.

Note that $g_t(\delta_1, \delta_2)h_t$ is the conditional likelihood function for (S_t, X_t) given $(S_1, ..., S_{t-1}, X_1, ..., X_{t-1})$. Thus, when no structural change occurs, the likelihood function of the sample is

$$\prod_{t=1}^{T} g_t(\delta_1, \delta_2)h_t = \left[\prod_{t=1}^{T} g_t(\delta_1, \delta_2)\right]\left[\prod_{t=1}^{T} h_t\right]. \tag{7.4}$$

Let $\Pi \subset (0,1)$ and let $\pi \subset \Pi$. Suppose the parameter vector equals (δ_1, δ_2) for the observations $t = 1, ..., [T\pi]$ and $(\delta_1 + \beta, \delta_2)$ for the observations $t = [T\pi]+1, ..., T$, where $\beta \in B \subset R^p$ and $[\cdot]$ denotes the integer part of $\cdot$. Then, π 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. In linear models with exogenous regressors, Π does not need to be restricted in this way, see Andrews, Lee, and Ploberger (1992), since finite sample (rather than asymptotic) results can be obtained.

In the present case, the likelihood function $f_T(\theta,\pi)$ and the log likelihood function $\ell_T(\theta,\pi)$ of Sections 3 and 5 are given by

$$\begin{aligned} f_T(\theta,\pi) &= \Big[\prod_{t=1}^{[T\pi]} g_t(\delta_1, \delta_2)\Big]\Big[\prod_{t=[T\pi]+1}^{T} g_t(\delta_1+\beta, \delta_2)\Big]\Big[\prod_{t=1}^{T} h_t\Big] \text{ and} \\ \ell_T(\theta,\pi) &= \Sigma_1^{[T\pi]}\log g_t(\delta_1, \delta_2) + \Sigma_{[T\pi]+1}^{T} \log g_t(\delta_1+\beta, \delta_2) + \Sigma_1^T \log h_t\,, \end{aligned} \tag{7.5}$$

where, here and below, Σ_1^T denotes $\Sigma_{t=1}^T$. Also, the norming matrix B_T of Section 3 is taken to be $\sqrt{T}I_s$. Note that the conditional densities $\{h_t : t \leq T\}$ of $\{X_t : t \leq T\}$ drop out of the first and second partial derivatives of $\ell_T(\theta,\pi)$ with respect to θ.

The null and alternative hypotheses of interest are

$$\begin{aligned} &H_0 : \theta = \theta_0, \text{ where } \theta_0 = (0, \delta_{10}', \delta_{20}')' \text{ for some } \delta_{10} \in \Delta_1 \text{ and } \delta_{20} \in \Delta_2, \\ &H_1 : \theta = (\beta_0', \delta_{10}', \delta_{20}')', \text{ where } \beta_0 \neq 0, \beta_0 \in B, \delta_{10} \in \Delta_1, \text{ and } \delta_{20} \in \Delta_2, \\ &\qquad \text{and the point of structural change is } \pi \text{ for some } \pi \in \Pi. \end{aligned} \tag{7.6}$$

Note that the parameter δ_{20} is constant across the whole sample under H_0 and H_1. If a parameter δ_{20} appears in the conditional likelihood functions, then the problem is one of testing for partial structural change. If no parameter δ_{20} appears in the conditional likelihood functions, then the problem is one of testing for pure structural change.

Next, we define the exponential Wald, LM, and LR statistics for testing H_0 versus H_1. In addition, we define simplified asymptotically equivalent forms of the exponential Wald and LM statistics. To this end, let $\hat{\theta}(\pi) = (\hat{\beta}(\pi)', \hat{\delta}_1(\pi)', \hat{\delta}_2(\pi)')'$ be the unrestricted ML estimator of θ for fixed $\pi \in \Pi$. That is, $\hat{\theta}(\pi)$ maximizes

$$\Sigma_1^{[T\pi]}\log g_t(\delta_1, \delta_2) + \Sigma_{[T\pi]+1}^{T}\log g_t(\delta_1+\beta, \delta_2) \tag{7.7}$$

over $\theta = (\beta', \delta_1', \delta_2')' \in B \times \Delta_1 \times \Delta_2 = \Theta$. In the case of pure structural change, $\hat{\delta}_1(\pi)$ is just the standard ML estimator of δ_1 based on the observations $t = 1, \ldots, [T\pi]$ and

$\hat{\delta}_1(\pi) + \hat{\beta}(\pi)$ is the same based on the observations $t = [T\pi]+1, ..., T$. In this case, $\hat{\theta}(\pi) = (\hat{\beta}(\pi)', \hat{\delta}_1(\pi)')'$ is straightforward to compute. Returning to the general case, let $\tilde{\theta} = (\tilde{\beta}', \tilde{\delta}_1', \tilde{\delta}_2')'$ be the restricted ML estimator of θ. That is, $\tilde{\theta}$ maximizes

$$\Sigma_1^T \log g_t(\delta_1, \delta_2) \tag{7.8}$$

over $\theta = (0', \delta_1', \delta_2')' \in \tilde{\Theta}$ $(= \Theta \cap V)$. Since $\tilde{\theta}$ does not depend on π, it is generally much easier to compute than $\{\hat{\theta}(\pi) : \pi \in \Pi\}$.

By definition, we have

$$\mathcal{I}_T(\theta,\pi) = -\frac{1}{T}D^2\ell_T(\theta,\pi) = \begin{bmatrix} \mathcal{I}_T(\theta,\pi)_{11} & \mathcal{I}_T(\theta,\pi)_{12} & \mathcal{I}_T(\theta,\pi)_{13} \\ \mathcal{I}_T(\theta,\pi)_{12}' & \mathcal{I}_T(\theta,\pi)_{22} & \mathcal{I}_T(\theta,\pi)_{23} \\ \mathcal{I}_T(\theta,\pi)_{13}' & \mathcal{I}_T(\theta,\pi)_{23}' & \mathcal{I}_T(\theta,\pi)_{33} \end{bmatrix}, \text{ where}$$

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

$$\mathcal{I}_T(\theta,\pi)_{12} = -\frac{1}{T}\Sigma_{[T\pi]+1}^T \frac{\partial^2}{\partial\beta\partial\delta_1'}\log g_t(\delta_1 + \beta, \delta_2) = \mathcal{I}_T(\theta,\pi)_{11},$$

$$\mathcal{I}_T(\theta,\pi)_{13} = -\frac{1}{T}\Sigma_{[T\pi]+1}^T \frac{\partial^2}{\partial\beta\partial\delta_2'}\log g_t(\delta_1 + \beta, \delta_2),$$

$$\mathcal{I}_T(\theta,\pi)_{22} = -\frac{1}{T}\Sigma_1^{[T\pi]} \frac{\partial^2}{\partial\delta_1\partial\delta_1'}\log g_t(\delta_1, \delta_2) - \frac{1}{T}\Sigma_{[T\pi]+1}^T \frac{\partial^2}{\partial\delta_1\partial\delta_1'}\log g_t(\delta_1+\beta, \delta_2), \tag{7.9}$$

$$\mathcal{I}_T(\theta,\pi)_{23} = -\frac{1}{T}\Sigma_1^{[T\pi]} \frac{\partial^2}{\partial\delta_1\partial\delta_2'}\log g_t(\delta_1,\delta_2) - \frac{1}{T}\Sigma_{[T\pi]+1}^T \frac{\partial^2}{\partial\delta_1\partial\delta_2'}\log g_t(\delta_1+\beta,\delta_2), \text{ and}$$

$$\mathcal{I}_T(\theta,\pi)_{33} = -\frac{1}{T}\Sigma_1^{[T\pi]} \frac{\partial^2}{\partial\delta_2\partial\delta_2'}\log g_t(\delta_1,\delta_2) - \frac{1}{T}\Sigma_{[T\pi]+1}^T \frac{\partial^2}{\partial\delta_2\partial\delta_2'}\log g_t(\delta_1+\beta,\delta_2), \text{ and}$$

$$D\ell_T(\tilde{\theta},\pi) = \left[\Sigma_{[T\pi]+1}^T \frac{\partial}{\partial\delta_1'}\log g_t(\tilde{\delta}_1,\tilde{\delta}_2), \Sigma_1^T \frac{\partial}{\partial\delta_1'}\log g_t(\tilde{\delta}_1,\tilde{\delta}_2), \Sigma_1^T \frac{\partial}{\partial\delta_2'}\log g_t(\tilde{\delta}_1,\tilde{\delta}_2)\right]'.$$

We are now in a position to define $W_T(\pi), ..., LR_T(\pi)$. We also define two statistics $W_T^*(\pi)$ and $LM_T^*(\pi)$ that are simplified asymptotically equivalent versions of $W_T(\pi)$ and $LM_T(\pi)$:

$$
\begin{aligned}
W_T(\pi) &= T\hat{\beta}(\pi)'\left[H\mathcal{I}_T^{-1}(\hat{\theta}(\pi),\pi)H'\right]^{-1}\hat{\beta}(\pi),\\
W_T^*(\pi) &= T\hat{\beta}(\pi)'\left[\mathcal{I}_T^{-1}(\hat{\theta}(\pi),\pi)_{00} + \mathcal{I}_T^{-1}(\hat{\theta}(\pi),\pi)_{11}\right]^{-1}\hat{\beta}(\pi),\\
LM_T(\pi) &= \tfrac{1}{\sqrt{T}}D\ell_T(\tilde{\theta},\pi)'\mathcal{I}_T^{-1}(\tilde{\theta},\pi)\tfrac{1}{\sqrt{T}}D\ell_T(\tilde{\theta},\pi)\,,\\
LM_T^*(\pi) &= \tfrac{1}{\sqrt{T}}\Sigma_{[T\pi]+1}^{T}\frac{\partial}{\partial\delta_1'}\log g_t(\tilde{\delta}_1,\tilde{\delta}_2)\left[\mathcal{I}_T^{-1}(\tilde{\theta},\pi)_{00} + \mathcal{I}_T^{-1}(\tilde{\theta},\pi)_{11}\right]\\
&\quad\times\tfrac{1}{\sqrt{T}}\Sigma_{[T\pi]+1}^{T}\frac{\partial}{\partial\delta_1}\log g_t(\tilde{\delta}_1,\tilde{\delta}_2)\,,\ \text{and}\\
LR_T(\pi) &= -2\Big[\Sigma_1^T\log g_t(\tilde{\delta}_1,\tilde{\delta}_2) - \Sigma_1^{[T\pi]}\log g_t(\hat{\delta}_1(\pi),\hat{\delta}_2(\pi))\\
&\qquad - \Sigma_{[T\pi]+1}^{T}\log g_t(\hat{\delta}_1(\pi)+\hat{\beta}(\pi),\hat{\delta}_2(\pi))\Big],\ \text{where}\\
\mathcal{I}_T(\theta,\pi)_{00} &= -\tfrac{1}{T}\Sigma_1^{T\pi}\frac{\partial^2}{\partial\delta_1\partial\delta_1'}\log g_t(\delta_1,\delta_2).
\end{aligned}
\tag{7.10}
$$

The asymptotic equivalence of $W_T(\pi)$ and $W_T^*(\pi)$ and of $LM_T(\pi)$ and $LM_T^*(\pi)$ under the null and local alternatives is established in the Appendix following the proof of Theorem 4. As noted above, the Hessian matrices $\mathcal{I}_T(\theta,\pi)$ and $\mathcal{I}_T(\theta,\pi)_{ii}$ could be replaced by outer product matrices without affecting the asymptotic distributional or asymptotic optimality results. Note that $W_T(\pi) = W_T^*(\pi)$ and $LM_T(\pi) = LM_T^*(\pi)$ in the case of the pure structural change.

The asymptotically optimal test statistics Exp–W_T, Exp–LM_T, and Exp–LR_T are now defined by combining (3.4) and (7.10). Analogously, we define asymptotically equivalent statistics Exp–W_T^* and Exp–LM_T^* using the formula (3.4) with $K = W^*$ and $K = LM^*$, where $W_T^*(\pi)$ and $LM_T^*(\pi)$ are as in (7.10). Note that Exp–LM_T^* and Exp–LM_T are usually the easiest of the Exp–K_T statistics to compute because they only involve the computation of a single estimator $\tilde{\theta}$.

7.2. Asymptotic Critical Values

We now describe the asymptotic null distribution of Exp–W_T, Exp–W_T^*, Exp–LM_T, Exp–LM_T^*, and Exp–LR_T. Let $B_1(\cdot)$ be a p–vector of independent Brownian motions on [0,1]. Then, under H_0 and Assumption SC given below,

(7.11)
$$\text{Exp–}K_T \xrightarrow{d} \chi(\theta_0, c) \text{ for } K = W, W^*, LM, LM^*, \text{ and } LR, \text{ where}$$
$$\chi(\theta_0,c) = (1+c)^{-p/2} \int \exp\left[\frac{1}{2}\frac{c}{1+c}(B_1(\pi)-\pi B_1(1))'(B_1(\pi)-\pi B_1(1))/[\pi(1-\pi)]\right] dJ(\pi) .$$

That is, the asymptotic null distribution of Exp–K_T is an exponential average of the square of a standardized tied–down Bessel process of order p. Since $\chi(\theta_0, c)$ does not depend on θ_0 in the present case, the limit distribution of Exp–K_T is nuisance parameter free and asymptotic critical values can be tabulated.

The limit distribution of Exp–K_T under general local alternatives to H_0 (not just one–time change alternatives) can be obtained from Theorem 5 and Example 2 following Theorem 4 of Andrews (1989). It is an exponential average of the square of a noncentral standardized tied–down Bessel process of order p.

The most common case in practice is when the weight function $J(\cdot)$ is uniform on $[\pi_1, \pi_2]$ for some $0 < \pi_1 < \pi_2 < 1$. In this case we can show that the critical values based on the limit distribution $\chi(\theta_0, c)$ defined in (7.11) depend on (π_1, π_2) only through the scalar $\lambda = \pi_2(1-\pi_1)/[\pi_1(1-\pi_2)]$. This greatly simplifies the calculation and presentation of critical values for the optimal tests. In particular, we have

(7.12)
$$P\left[(1+c)^{-p/2} \int_{\pi_1}^{\pi_2} \exp\left[\frac{1}{2}\frac{c}{1+c}BB(\pi)'BB(\pi)/[\pi(1-\pi)]\right] d\pi > k_{p,\alpha}\right]$$
$$= P\left[(1+c)^{-p/2} \int_1^{\lambda} \exp\left[\frac{1}{2}\frac{c}{1+c}BM(s)'BM(s)/s\right] ds > k_{p,\alpha}\right] ,$$

where $BB(\pi) = B_1(\pi) - \pi B_1(1)$ is a p–vector of independent Brownian bridge processes on [0,1], BM(s) is a p–vector of independent Brownian motion processes on $[0,\infty)$, and $k_{p,\alpha}$

denotes the level α critical value that corresponds to $\chi(\theta_0, c)$. The proof of (7.12) follows from the proof of Corollary 1 of Andrews (1989).

Simulation results reported in Andrews, Lee, and Ploberger (1992) 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 where $c = \infty$ and $c = 0$. As noted in Section 7.5 below, we have a mild preference for the $c = \infty$ test over the $c = 0$ test.

For reporting critical values, we consider the case where $J(\cdot)$ is uniform on $[\pi_1, \pi_2]$ for some $0 < \pi_1 \leq \pi_2 < 1$. Table I reports asymptotic critical values for the $c = \infty$ statistic

$$\text{(7.13)} \qquad \log \frac{1}{1 - 2\pi_0} \int_{\pi_0}^{1-\pi_0} \exp(K_T(\pi)/2) d\pi$$

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. The Table covers values of λ between 1 and 2,401, so almost any interval of interest can be considered.

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

$$\text{(7.14)} \qquad \frac{1}{1 - 2\pi_0} \int_{\pi_0}^{1-\pi_0} K_T(\pi) d\pi .$$

Critical values for asymmetric intervals $[\pi_1, \pi_2]$ can be obtained in the same manner as above.

When the time of structural change (if it occurs) is completely unknown, we suggest taking $\pi_0 = .02$ in (7.13) or (7.14). 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., $\sup_{\pi \in [\pi_0, 1-\pi_0]} W_T(\pi)$) is defined over such a broad interval, because

$\int_0^1 \exp\left[\frac{1}{2}BB(\pi)'BB(\pi)/[\pi(1-\pi)]\right]d\pi < \infty$ 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. When the time of structural change is known to lie in some restricted interval $[\pi_1, \pi_2]$ (see the discussion in Andrews (1989, Sec. 2) regarding such cases), then the test statistic should incorporate this information to maximize power. Tables I and II allow one to obtain asymptotic critical values for a wide range of such intervals.

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 (7.13) and (7.14) 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

$$\Pi(N) = [\pi_0, 1-\pi_0] \cap \{\pi = j/N : j = 0, 1, \ldots, N\}\,. \tag{7.15}$$

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 $p \leq 4$. A single realization from the asymptotic distribution of the discretized version of (7.13) or (7.14) 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.

7.3. Assumptions and Optimality Results

Before stating assumptions, we introduce some additional notation. Let

$$\mathcal{J}(\delta_1,\delta_2) = \begin{bmatrix} \mathcal{J}_1(\delta_1,\delta_2) & \mathcal{J}_2(\delta_1,\delta_2) \\ \mathcal{J}_2(\delta_1,\delta_2)' & \mathcal{J}_3(\delta_1,\delta_2) \end{bmatrix}, \quad \mathcal{J}_1(\delta_1,\delta_2) = \lim_{T\to\infty} -\tfrac{1}{T}\Sigma_1^T E\frac{\partial^2}{\partial\delta_1\partial\delta_1'}\log g_t(\delta_1,\delta_2),$$

$$\text{(7.16)} \quad \mathcal{J}_2(\delta_1,\delta_2) = \lim_{T\to\infty} -\tfrac{1}{T}\Sigma_1^T E\frac{\partial^2}{\partial\delta_1\partial\delta_2'}\log g_t(\delta_1,\delta_2),$$

$$\mathcal{J}_3(\delta_1,\delta_2) = \lim_{T\to\infty} -\tfrac{1}{T}\Sigma_1^T E\frac{\partial^2}{\partial\delta_2\partial\delta_2'}\log g_t(\delta_1,\delta_2),$$

$$\mathcal{J} = \mathcal{J}(\delta_{10}, \delta_{20}), \text{ and } \mathcal{J}_a = \mathcal{J}_a(\delta_{10}, \delta_{20}) \text{ for } a = 1, 2, 3 .$$

Using this notation, $\mathcal{I}(\theta,\pi)$ and $\mathcal{I}(\theta_0, \pi)$ of Section 5 are given in the present case by

$$\text{(7.17)} \quad \mathcal{I}(\theta,\pi) = \begin{bmatrix} (1-\pi)\mathcal{J}_1(\delta_1+\beta,\delta_2) & (1-\pi)\mathcal{J}_1(\delta_1+\beta,\delta_2) & (1-\pi)\mathcal{J}_2(\delta_1+\beta,\delta_2) \\ (1-\pi)\mathcal{J}_1(\delta_1+\beta,\delta_2) & \pi\mathcal{J}_1(\delta_1,\delta_2)+(1-\pi)\mathcal{J}_1(\delta_1+\beta,\delta_2) & \pi\mathcal{J}_2(\delta_1,\delta_2)+(1-\pi)\mathcal{J}_2(\delta_1+\beta,\delta_2) \\ (1-\pi)\mathcal{J}_2(\delta_1+\beta,\delta_2)' & \pi\mathcal{J}_2(\delta_1,\delta_2)'+(1-\pi)\mathcal{J}_2(\delta_1+\beta,\delta_2)' & \pi\mathcal{J}_3(\delta_1,\delta_2)+(1-\pi)\mathcal{J}_3(\delta_1+\beta,\delta_2) \end{bmatrix} \text{ and}$$

$$\mathcal{I}(\theta_0, \pi) = \begin{bmatrix} (1-\pi)\mathcal{J}_1 & (1-\pi)\mathcal{J}_1 & (1-\pi)\mathcal{J}_2 \\ (1-\pi)\mathcal{J}_1 & \mathcal{J}_1 & \mathcal{J}_2 \\ (1-\pi)\mathcal{J}_2' & \mathcal{J}_2' & \mathcal{J}_3 \end{bmatrix}.$$

In turn, some algebra shows that

$$V = \{h \in R^s : h = (0', \delta_1', \delta_2')' \text{ for some } \delta_1 \in R^p \text{ and } \delta_2 \in R^{q-p}\},$$

$$V_\pi^\perp = \left\{h \in R^s : h = (\lambda', -(1-\pi)\lambda', 0')' \text{ for some } \lambda \in R^p\right\}, \text{ and}$$

(7.18)

$$Q_\pi \sim N\left[\begin{bmatrix}0\\0\\0\end{bmatrix}, c\begin{bmatrix}\frac{1}{\pi(1-\pi)}\mathcal{J}_1^{-1} & -\frac{1}{\pi}\mathcal{J}_1^{-1} & 0\\ -\frac{1}{\pi}\mathcal{J}_1^{-1} & \frac{1-\pi}{\pi}\mathcal{J}_1^{-1} & 0\\ 0 & 0 & 0\end{bmatrix}\right].$$

We use the following assumption for the problem of testing for one–time structural change (SC) in nonlinear models. It is sufficient for Assumptions 1, 2, 4, and 5 of Section 5.

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

(b) $\sup_{\delta_1\in\Delta_1,\,\delta_2\in\Delta_2}\left\|\frac{1}{T}\Sigma_1^T(\log g_t(\delta_1,\delta_2) - E\log g_t(\delta_1,\delta_2))\right\| \to 0$ a.s. *under* θ_0,

$Q(\delta_1,\delta_2) = \lim_{T\to\infty}\frac{1}{T}\Sigma_1^T E\log g_t(\delta_1,\delta_2)$ *exists uniformly over* $\delta_1 \in \Delta_1$ *and* $\delta_2 \in \Delta_2$, *and*

$\sup_{\delta_1\in\Delta_1,\,\delta_2\in\Delta_2}|Q(\delta_1,\delta_2)| < \infty$.

(c) *For all neighborhoods* Δ_{10} *of* δ_{10} *and* Δ_{20} *of* δ_{20},

$\sup_{(\delta_1,\delta_2)\subset\Delta_1\times\Delta_2/\Delta_{10}\times\Delta_{20}}(Q(\delta_1,\delta_2) - Q(\delta_{10},\delta_{20})) < 0.$

(d) θ_0 *is in the interior of* Θ.

(e) $g_t(Y_t|Y_1, ..., Y_{t-1}, X_1, ..., X_t, \delta_1, \delta_2)$ *is twice continuously partially differentiable in* (δ_1, δ_2) *for all* $\delta_1 \in \Delta_{10}$ *and* $\delta_2 \in \Delta_{20}$ *with probability one under* θ_0 *for all* $t \geq 1$, *where* Δ_{10} *and* Δ_{20} *contain neighborhoods of* δ_{10} *and* δ_{20} *respectively.*

(f) $\sup_{\delta_1\in\Delta_{10},\,\delta_2\in\Delta_{20}}\left\|\frac{1}{T}\Sigma_1^T\left[\frac{\partial^2}{\partial(\delta_1',\delta_2')'\partial(\delta_1',\delta_2')}\log g_t(\delta_1,\delta_2)\right.\right.$

$\left.\left. - E\frac{\partial^2}{\partial(\delta_1',\delta_2')'\partial(\delta_1',\delta_2')}\log g_t(\delta_1,\delta_2)\right]\right\| \to 0$ a.s. *under* θ_0 *and* $\mathcal{J}_a(\delta_1, \delta_2)$ *exists uniformly over* $(\delta_1, \delta_2) \in \Delta_{10} \times \Delta_{20}$ *and is uniformly continuous for* (δ_1, δ_2) *in* $\Delta_{10} \times \Delta_{20}$ $\forall a = 1, 2, 3$, *where* Δ_{10} *and* Δ_{20} *are as in part* (e).

(g) $\mathcal{J}$ *is positive definite.*

(h) $\left\{\frac{\partial}{\partial(\delta_1', \delta_2')'}\log g_t(\delta_{10}, \delta_{20}) : t \geq 1\right\}$ *satisfies an invariance principle with covariance matrix* $\mathcal{J}$ *under* θ_0. *That is,* $\frac{1}{\sqrt{T}}\Sigma_1^{[T\cdot]}\frac{\partial}{\partial(\delta_1', \delta_2')'}\log g_t(\delta_{10}, \delta_{20}) \Rightarrow \mathcal{K}B(\cdot)$ *as a process indexed by* $\pi \in [0,1]$, *where* $\mathcal{K}\mathcal{K}' = \mathcal{J}$ and $B(\cdot)$ *is a* q*–vector of independent Brownian motions on* [0,1].

We now comment on (Assumption) SC. Parts (a)–(c) of SC are used to verify Assumption 2 of Section 5 (i.e., uniform consistency of $\hat{\theta}(\pi)$ under θ_0). SC(a) is needed in the proofs to obtain convergence results that hold uniformly over $\pi \in \Pi$. It is not overly restrictive. For non–trending observations, SC(b) can be verified under broad assumptions regarding temporal dependence and non–identical distributions using a uniform strong law of large numbers (SLLN). A uniform SLLN can be obtained by taking any standard SLLN (e.g., see McLeish (1975) or Hansen (1991b) for dependent rv's) and strengthening it to a uniform SLLN using the results of Andrews (1987, 1992) or Pötscher and Prucha (1989, 1990). SC(c) is the standard uniqueness assumption that ensures that the ML estimator of $(\delta_{10}, \delta_{20})$ for the case of no structural change is consistent. This assumption is closely related to the identification of $(\delta_{10}, \delta_{20})$ in the model with no structural change.

Parts (a) and (d)–(g) of SC are used to verify Assumption 1 of Section 5. SC(d) and SC(e) are standard assumptions. SC(f) can be verified under broad conditions using a uniform SLLN as above. SC(g) requires the asymptotic information matrix for $(\delta_{10}, \delta_{20})$ for the model with no structural change to be positive definite. This is a standard assumption.

Parts (g) and (h) of SC are used to verify Assumption 5 of Section 5. For non-trending observations, SC(h) can be verified under broad conditions by employing a multivariate invariance principle (e.g., see Phillips and Durlauf (1986) and Eberlein (1986) for multivariate invariance principles for dependent non–identically distributed rv's). Note that univariate invariance principles can be converted into multivariate invariance principles under suitable assumptions by Lemma A–5 of Andrews (1989).

Sufficient conditions for Assumption SC in the stationary ergodic, m–th order Markov case are given in Section 7.4 below. For more general cases, Assumption SC can be verified using near epoch dependence assumptions along the lines of Assumptions A* and 1* and Theorems A–1 and 1 of Andrews (1989).

THEOREM 5: *For* $f_T(\theta,\pi)$ *as defined in* (7.5), *Assumption* SC *implies Assumptions* 1, 2, 4, *and* 5.

COMMENTS: 1. Theorem 5 implies that the asymptotic null distribution given in Theorem 2 of Section 5 and in (7.11) and the asymptotic optimality properties given in Theorem 4 of Section 5 hold for the tests based on Exp–K_T for K = W, W*, ..., LR.

2. In the present context, the weighted average power of a test φ_T can be written in terms of an integral with respect to a normal distribution that is not singular as follows:

$$\text{(7.19)} \qquad \int\left[\int \varphi_T f_T(\theta_0 + (\lambda', -(1-\pi)\lambda', 0')'/\sqrt{T}, \pi)d\mu_T\right] n\left[\lambda; 0, c\frac{1}{\pi(1-\pi)}\mathcal{J}_1^{-1}\right] d\lambda dJ(\pi) .$$

7.4. The Stationary Markov Case

In this section we provide primitive conditions for Assumption SC in the relatively simple strictly stationary ergodic Markov case. In particular, we suppose that $\{(S_t, X_t) : t \geq 1\}$ is part of a double infinite strictly stationary ergodic sequence $\{(S_t, X_t) : t = ..., 0, 1, ...\}$ and $\{S_t : t = ..., 0, 1, ...\}$ is m–th order Markov for some integer $m \geq 0$. By definition, $\{S_t : t = ..., 0, 1, ...\}$ is m–th order Markov if the conditional distribution of S_t given $\mathcal{F}_t = \sigma(..., S_{t-2}, S_{t-1}; ..., X_{t-1}, X_t)$ equals the conditional distribution of S_t given $S_{t,m} = (S_{t-m}, ..., S_{t-1})$ and $X_{t,m} = (X_{t-m}, ..., X_t)$ for all t.

The Markov assumption yields the simplification that 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 this would not be the case, because the number of relevant observed variables in the conditioning set, viz., $S_1, ..., S_{t-1}, X_1, ..., X_t$, would vary with t. Without the Markov assumption, one would need to verify Assumption SC using SLLNs and an invariance principle for nonstationary rv's, as in Andrews (1989, Assumptions A* and 1* and Theorems A–1 and 1).

The following assumption is sufficient for Assumption SC:

ASSUMPTION SC1: (a) Π *has closure contained in* (0,1).

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

(c) $\{(S_t, X_t) : t = ..., 0, 1, ...\}$ *is stationary and ergodic and* $\{S_t : t = ..., 0, 1, ...\}$ *is* m–*th order Markov.*

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

(e) $g_t(\delta_1, \delta_2)$ *is continuous in* (δ_1, δ_2) *on* $\Delta_1 \times \Delta_2$ *with probability one under* θ_0.

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

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

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

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

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

LEMMA 4: *Assumption* SC1 *implies Assumption* SC.

Note that Assumption SC1 is quite similar to standard assumptions in the literature for the consistency and asymptotic normality of ML estimators in dependent contexts. Under Assumption SC1, the test statistics Exp–K_T for K = W, W*, LM, LM*, and LR possess the null distribution and the optimality properties stated in Theorems 2 and 4.

7.5. The Choice of c

Simulation results for the exponential test statistics are reported in Andrews, Lee and Ploberger (1992) for two linear regression models each with intercept and one regressor, sample size $T = 120$, and iid $N(0, \sigma^2)$ errors. In the first "stationary" model, the regressor X_t equals $(-1)^t$. In the second "time trend" model, the regressor X_t equals $t-(T+1)/2$. The results for the stationary model are relevant to all of the nonlinear models that are covered under Assumption SC1 or SC. This is because all of these models have asymptotic local power functions that equal that for the stationary linear regression model.

Based on the results in Andrews, Lee, and Ploberger (1992), we make the recommendation that a $c = \infty$ exponential test statistic be used for testing for structural change in nonlinear models. This recommendation is based on (a) the finding that the choice of c is not crucial in these applications and (b) the choice of $c = \infty$ is somewhat preferable to smaller values of c when change occurs early or late in the sample. We also note that the only non–data–dependent values of c that yield tests that are invariant under scale transformations of the regression parameters in a linear regression model are $c = 0$ and $c = \infty$. Of these two, $c = \infty$ seems to be mildly preferable in terms of power.

APPENDIX

Let wp → 1 denote "with probability that goes to one as $T \to \infty$."

PROOF OF LEMMA 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 $\hat{\theta}(\pi)$, $D\ell_T(\hat{\theta}(\pi), \pi) = 0 \ \forall \pi \in \Pi$ wp → 1. Hence, by one-term Taylor expansions of the elements of $D\ell_T(\hat{\theta}(\pi), \pi)$ about θ_0 we get, wp → 1,

$$\text{(A.1)} \quad \begin{aligned} & 0 = B_T^{-1} D\ell_T(\hat{\theta}(\pi),\pi) = B_T^{-1} D\ell_T(\theta_0,\pi) - \mathcal{I}_{1T}(\pi) B_T(\hat{\theta}(\pi) - \theta_0) \ \ \forall \pi \in \Pi , \text{ where} \\ & \mathcal{I}_{1T}(\pi) = -\int_0^1 B_T^{-1} D^2 \ell_T(\theta_0 + \lambda(\hat{\theta}(\pi) - \theta_0), \pi) B_T^{-1} d\lambda . \end{aligned}$$

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

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

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.2) and Assumption 1(f) yield

$$\text{(A.3)} \qquad \sup_{\pi \in \Pi} \| \mathcal{I}_{1T}^{-1}(\pi) - \mathcal{I}^{-1}(\theta_0, \pi) \| = o_p(1) .$$

Equations (A.1) and (A.3) and Assumptions 1(f) and 1(g) yield

$$\text{(A.4)} \quad \begin{aligned} o_p(1) &= \sup_{\pi \in \Pi} \| B_T(\hat{\theta}(\pi) - \theta_0) - \mathcal{I}_{1T}^{-1}(\pi) B_T^{-1} D\ell_T(\theta_0, \pi) \| \\ &= \sup_{\pi \in \Pi} \| B_T(\hat{\theta}(\pi) - \theta_0) - \mathcal{I}^{-1}(\theta_0, \pi) B_T^{-1} D\ell_T(\theta_0, \pi) \| + o_p(1) . \ \square \end{aligned}$$

PROOF OF LEMMA 2: All probability calculations in this proof are made "under θ_0." For $0 < M < \infty$, define

$$\text{(A.5)} \qquad LR_T(M) = \int_\Pi \int_{\|h\| \le M} f_T(\theta_0 + B_T^{-1}h, \pi) dQ_\pi(h) dJ(\pi) / f_T(\theta_0) \quad \text{and}$$

$$\text{(A.6)} \qquad \overline{LR}_T(M) = \int_\Pi \exp\left[\tfrac{1}{2}\overline{\theta}(\pi)' \mathcal{I}(\theta_0, \pi)\overline{\theta}(\pi)\right] \int_{\|h\| \le M} \exp\left[-\tfrac{1}{2}(\overline{\theta}(\pi) - h)' \mathcal{I}(\theta_0\ \pi) \times (\overline{\theta}(\pi) - h)\right] dQ_\pi(h) dJ(\pi) .$$

Note that for any $\epsilon > 0$

$$\text{(A.7)} \qquad P(|LR_T - \overline{LR}_T| > 3\epsilon) \le P(|LR_T - LR_T(M)| > \epsilon) + P(|LR_T(M) - \overline{LR}_T(M)| > \epsilon) + P(|\overline{LR}_T - \overline{LR}_T(M)| > \epsilon) .$$

Hence, it suffices to show that (1) given any $\eta > 0$ we can choose $T^* < \infty$ and $M < \infty$ sufficiently large so that $P(|LR_T - LR_T(M)| > \epsilon) < \eta$ and $P(|\overline{LR}_T - \overline{LR}_T(M)| > \epsilon) < \eta$ for all $T \ge T^*$ and (2) $LR_T(M) - \overline{LR}_T(M) \xrightarrow{p} 0\ \forall 0 < M < \infty$.

We show (1) first. We have

$$\text{(A.8)} \qquad \begin{aligned} P(|LR_T - LR_T(M)| > \epsilon) &\le \epsilon^{-1} E|LR_T - LR_T(M)| \\ &= \epsilon^{-1} E \int_\Pi \int_{\|h\| > M} [f_T(\theta_0 + B_T^{-1}h, \pi)/f_T(\theta_0)] dQ_\pi(h) dJ(\pi) \\ &= \epsilon^{-1} \int_\Pi \int_{\|h\| > M} dQ_\pi(h) dJ(\pi) , \end{aligned}$$

where the first equality uses Assumption 1(a) and the second holds by Fubini's theorem and the fact that $E[f_T(\theta_0 + B_T^{-1}h, \pi)/f_T(\theta_0)] = 1 \quad \forall h, \forall \pi$. The right–hand side (rhs) of (A.8) can be made arbitrarily small for all T by taking M large by Assumption 3′.

Next, we have

$$\text{(A.9)} \qquad \begin{aligned} |\overline{LR}_T - \overline{LR}_T(M)| &= \int_\Pi \Big[\exp\left[\tfrac{1}{2} D\ell_T(\theta_0, \pi)' B_T^{-1} \mathcal{I}^{-1}(\theta_0, \pi) B_T^{-1} D\ell_T(\theta_0, \pi)\right] \\ &\quad \times \int_{\|h\| > M} \exp\left[-\tfrac{1}{2}(\overline{\theta}(\pi) - h)' \mathcal{I}(\theta_0, \pi)(\overline{\theta}(\pi) - h)\right] dQ_\pi(h)\Big] dJ(\pi) \\ &\le \exp\left[\tfrac{1}{2} \sup_{\pi \in \Pi} \left\| B_T^{-1} D\ell_T(\theta_0, \pi)\right\|^2 \sup_{\pi \in \Pi} \|\mathcal{I}^{-1}(\theta_0, \pi)\|\right] \cdot \int_\Pi \int_{\|h\| > M} dQ_\pi(h) dJ(\pi), \end{aligned}$$

where the inequality uses the assumption that $\mathcal{I}(\theta_0, \pi)$ is positive definite. The first term on the rhs of (A.9) is $O_p(1)$ by Assumptions 1(f) and 1(g) and the second term on the rhs can be made arbitrarily small by taking M large using Assumption 3′. In consequence, $P(|\overline{LR}_T - \overline{LR}_T(M)| > \epsilon)$ can be made arbitrarily small for all T large by taking M sufficiently large.

We now establish (2). A two term Taylor series expansion gives

$$\text{(A.10)}\quad \begin{aligned} &\ell_T(\theta_0 + B_T^{-1}h, \pi) - \ell_T(\theta_0) \\ &= h' B_T^{-1} D\ell_T(\theta_0, \pi) + \tfrac{1}{2} h' B_T^{-1} D^2\ell_T(\theta_0, \pi) B_T^{-1} h + r_{1T}(h,\pi)\ , \end{aligned}$$

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

$$\text{(A.11)}\quad \begin{aligned} &\sup_{\pi\in\Pi}\ \sup_{h:\|h\|\le M} \|r_{1T}(h,\pi)\| \\ &\le M^2 \sup_{\pi\in\Pi}\ \sup_{\theta:\|B_T(\theta-\theta_0)\|\le M} \|B_T^{-1}D^2\ell_T(\theta,\pi)B_T^{-1} - B_T^{-1}D^2\ell_T(\theta_0, \pi)B_T^{-1}\| \\ &\le M^2 \sup_{\pi\in\Pi}\ \sup_{\theta\in\Theta_0} \|B_T^{-1}D^2\ell_T(\theta,\pi)B_T^{-1} + \mathcal{I}(\theta,\pi)\| \\ &\quad + M^2 \sup_{\pi\in\Pi}\ \sup_{\theta:\|B_T(\theta-\theta_0)\|\le M} \|\mathcal{I}(\theta_0, \pi) - \mathcal{I}(\theta,\pi)\| \\ &\quad + M^2 \sup_{\pi\in\Pi} \|B_T^{-1}D^2\ell_T(\theta_0, \pi)B_T^{-1} + \mathcal{I}(\theta_0, \pi)\| \\ &= o_p(1)\ , \end{aligned}$$

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

$$\text{(A.12)}\quad \begin{aligned} &h' B_T^{-1} D^2\ell_T(\theta_0, \pi) B_T^{-1} h = -h'\mathcal{I}(\theta_0, \pi)h + r_{2T}(h,\pi), \text{ where} \\ &\sup_{\pi\in\Pi}\ \sup_{h:\|h\|\le M} |r_{2T}(h,\pi)| = o_p(1)\ , \end{aligned}$$

by Assumption 1(d). It follows from (A.11) and (A.12) that

$$\exp(r_{1T}(h,\pi) + r_{2T}(h,\pi)) = 1 + s_T(h,\pi)\ , \text{ where}$$

(A.13)

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

Combining (A.10) and (A.12) and using the definition of $\bar{\theta}(\pi)$ yields

$$\begin{aligned} \ell_T(\theta_0 + B_T^{-1}h,\ \pi) - \ell_T(\theta_0) &= h'\mathcal{I}(\theta_0,\ \pi)\bar{\theta}(\pi) - \tfrac{1}{2}h'\mathcal{I}(\theta_0,\ \pi)h + r_{1T}(h,\pi) + r_{2T}(h,\pi) \\ &= \tfrac{1}{2}\bar{\theta}(\pi)'\mathcal{I}(\theta_0,\pi)\bar{\theta}(\pi) - \tfrac{1}{2}(\bar{\theta}(\pi) - h)'\mathcal{I}(\theta_0,\pi)(\bar{\theta}(\pi) - h) + r_{1T}(h,\pi) + r_{2T}(h,\pi), \end{aligned} \tag{A.14}$$

where the second equality follows from some simple algebra.

Combining (A.5), (A.6), (A.13), (A.14), and Assumption 1(a) gives

$$\begin{aligned} LR_T(M) &= \int_\Pi\int_{\|h\|\leq M} \exp(\ell_T(\theta_0 + B_T^{-1}h,\ \pi) - \ell_T(\theta_0))dQ_\pi(h)dJ(\pi) \\ &= \int_\Pi\int_{\|h\|\leq M} \exp\Big[\tfrac{1}{2}\bar{\theta}(\pi)'\mathcal{I}(\theta_0,\ \pi)\bar{\theta}(\pi) - \tfrac{1}{2}(\bar{\theta}(\pi) - h)'\mathcal{I}(\theta_0,\ \pi)(\bar{\theta}(\pi) - h)) \\ &\qquad \times (1 + s_T(h,\pi))\Big] dQ_\pi(h)dJ(\pi) \\ &= \overline{LR}_T(M) + o_p(1)\ , \end{aligned} \tag{A.15}$$

where the third equality uses $\overline{LR}_T(M) = O_p(1)$, which follows from a close analogue to (A.9). This completes the proof. □

The following lemma is used in the proof of Theorem 1:

LEMMA A–1: *For each* $\pi \in \Pi$, *the projection matrix* $P^\perp$ *onto the orthogonal complement* $V_\pi^\perp$ *of* V *with respect to* $<\cdot,\cdot>_\pi$ *is given by*

$$P^\perp\ (= P_\pi^\perp) = A_\pi H, \text{ where } A_\pi = \begin{bmatrix} I_p \\ -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}' \end{bmatrix},\ H = [I_p \ \vdots\ 0], \text{ and } \mathcal{I}(\theta_0,\pi) = \begin{bmatrix} \mathcal{I}_{1\pi} & \mathcal{I}_{2\pi} \\ \mathcal{I}_{2\pi}' & \mathcal{I}_{3\pi} \end{bmatrix}.$$

PROOF OF LEMMA A–1: 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 $m = (m_1', m_2')' \in V_\pi^\perp$, $v'\mathcal{I}(\theta_0, \pi)m = 0 \ \forall v \in V$ iff $[0 \vdots I_q]\mathcal{I}(\theta_0, \pi)m = 0$ iff $[\mathcal{I}_{2\pi}' \vdots \mathcal{I}_{3\pi}]m = 0$ iff $m_2 = -\mathcal{I}_{3\pi}^{-1}\mathcal{I}_{2\pi}'m_1$ 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 $<\cdot,\cdot>_\pi$. □

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

$$(2\pi)^{-p/2}\det{}^{1/2}(A'\mathcal{I}A/c)\exp\left[-\frac{1}{2c}\lambda'A'\mathcal{I}A\lambda\right] \tag{A.16}$$

with respect to Lebesgue measure on R^p.

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

$$\overline{LR}_T = \int_\Pi \zeta_T(\pi)dJ(\pi)\text{ , where} \tag{A.17}$$

$$\zeta_T(\pi) = \int \exp\left[\tfrac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \tfrac{1}{2}(h-\overline{\theta})\mathcal{I}(h-\overline{\theta})\right]dQ_\pi(h)$$

$$= (2\pi)^{-p/2}\det{}^{1/2}(A'\mathcal{I}A/c) \tag{A.18}$$

$$\times \int \exp\left[\tfrac{1}{2}[\overline{\theta}'\mathcal{I}\overline{\theta} - (A\lambda-\overline{\theta})'\mathcal{I}(A\lambda-\overline{\theta}) - (A\lambda)'\mathcal{I}A\lambda/c]\right]d\lambda\ .$$

Let P and $P^\perp$ denote the projection matrices with respect to $<\cdot,\cdot>_\pi$ onto V and $V_\pi^\perp$ respectively. (Note that P and $P^\perp$ depend on π since $<\cdot,\cdot>_\pi$ and $V_\pi^\perp$ do.) The term in square brackets in the exponent on the rhs of (A.18), with $A\lambda$ replaced by h for simplicity, now simplifies as follows:

$$\overline{\theta}'\mathcal{I}\overline{\theta} - (h-\overline{\theta})'\mathcal{I}(h-\overline{\theta}) - h'\mathcal{I}h/c$$

$$= \overline{\theta}'\mathcal{I}\overline{\theta} - \left[h - \overline{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - \overline{\theta}\tfrac{c}{1+c}\right] - \tfrac{1}{1+c}\overline{\theta}'\mathcal{I}\overline{\theta}$$

$$= \tfrac{c}{1+c}(P\overline{\theta})'\mathcal{I}P\overline{\theta} + \tfrac{c}{1+c}(P^\perp\overline{\theta})'\mathcal{I}P^\perp\overline{\theta} - \left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right] \tag{A.19}$$

$$- \tfrac{c}{1+c}(P\overline{\theta})'\mathcal{I}P\overline{\theta}$$

$$= \tfrac{c}{1+c}(P^\perp\overline{\theta})'\mathcal{I}P^\perp\overline{\theta} - \left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right]\ ,$$

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

Combining (A.18) and (A.19) gives

$$\begin{aligned} \zeta_T(\pi) &= (1+c)^{-p/2}\exp\left[\frac{1}{2}\frac{c}{1+c}(P^\perp\bar{\theta})'\mathcal{I}P^\perp\bar{\theta}\right] \\ &\times \int(2\pi)^{-p/2}\det{}^{1/2}\left[A'\mathcal{I}A\frac{1+c}{c}\right]\exp\left[-\frac{1}{2}\left[A\lambda - P^\perp\bar{\theta}\frac{c}{1+c}\right]'\mathcal{I}\frac{1+c}{c}\left[A\lambda - P^\perp\bar{\theta}\frac{c}{1+c}\right]\right]d\lambda \\ &= (1+c)^{-p/2}\exp\left[\frac{1}{2}\frac{c}{1+c}(P^\perp\bar{\theta})'\mathcal{I}P^\perp\bar{\theta}\right], \end{aligned} \tag{A.20}$$

where the second equality holds because the integral of a normal density equals one.

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

$$\sup_{\pi\in\Pi}\|HB_T\hat{\theta}(\pi) - H\bar{\theta}(\pi)\| \xrightarrow{p} 0 . \tag{A.21}$$

In addition, we have

$$\begin{aligned} &\sup_{\pi\in\Pi}\|\mathcal{I}_T(\hat{\theta}(\pi), \pi) - \mathcal{I}(\theta_0, \pi)\| \\ &\le \sup_{\pi\in\Pi}\sup_{\theta\in\Theta_0}\|\mathcal{I}_T(\theta,\pi) - \mathcal{I}(\theta,\pi)\| + \sup_{\pi\in\Pi}\|\mathcal{I}(\hat{\theta}(\pi), \pi) - \mathcal{I}(\theta_0, \pi)\| \\ &= o_p(1) , \end{aligned} \tag{A.22}$$

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

For part (d), it suffices to show that

$$\sup_{\pi\in\Pi}|W_T(\pi) - LM_T(\pi)| \xrightarrow{p} 0 . \tag{A.23}$$

By (A.1) with $\hat{\theta}(\pi)$ replaced by $\tilde{\theta}$, we obtain

$$\text{(A.24)} \qquad B_T^{-1}D\ell_T(\tilde{\theta},\pi) = B_T^{-1}D\ell_T(\theta_0, \pi) - \mathcal{I}_{1T}(\pi)B_T(\tilde{\theta} - \theta_0) \;\; \forall \pi \in \Pi ,$$

where $\mathcal{I}_{1T}(\pi)$ is defined with $\tilde{\theta}$ in place of $\hat{\theta}(\pi)$. Note that (A.2) and (A.3) hold with $\mathcal{I}_{1T}(\pi)$ so defined using Assumption 4. In consequence,

$$\begin{aligned} H\mathcal{I}_T^{-1}(\tilde{\theta},\pi)B_T^{-1}D\ell_T(\tilde{\theta},\pi) &= H\mathcal{I}_{1T}^{-1}(\pi)B_T^{-1}D\ell_T(\tilde{\theta},\pi) + o_{p\pi}(1) \\ &= H\mathcal{I}_{1T}^{-1}(\pi)B_T^{-1}D\ell_T(\theta_0, \pi) + HB_T(\tilde{\theta} - \theta_0) + o_{p\pi}(1) \\ &= H\mathcal{I}^{-1}(\theta_0, \pi)B_T^{-1}D\ell_T(\theta_0,\pi) + o_{p\pi}(1) \\ &= HB_T(\hat{\theta}(\pi) - \theta_0) + o_{p\pi}(1) = HB_T\hat{\theta}(\pi) + o_{p\pi}(1) , \end{aligned} \qquad \text{(A.25)}$$

where the third equality uses $HB_T\tilde{\theta} = HB_T\theta_0 = 0$, the fourth equality holds by Lemma 1, and by definition a sequence $\{D_T(\pi) : T \geq 1\}$ is $o_{p\pi}(1)$ if $\sup_{\pi\in\Pi}\|D_T(\pi)\| = o_p(1)$. Equation (A.25) yields

$$\text{(A.26)} \qquad \begin{aligned} &\sup_{\pi\in\Pi}|W_T(\pi) - LM_T^0(\pi)| \xrightarrow{p} 0 , \text{ where} \\ &LM_T^0(\pi) = (B_T^{-1}D\ell_T(\tilde{\theta},\pi))'\mathcal{I}_T^{-1}(\tilde{\theta},\pi)H'\left[H\mathcal{I}_T^{-1}(\tilde{\theta},\pi)H'\right]^{-1}H\mathcal{I}_T^{-1}(\tilde{\theta},\pi)B_T^{-1}D\ell_T(\tilde{\theta},\pi) . \end{aligned}$$

Now, wp → 1,

$$\text{(A.27)} \qquad B_T^{-1}D\ell_T(\tilde{\theta},\pi) = -H'\lambda(\pi) \;\; \forall \pi \in \Pi$$

for some random p–vector of Lagrange multipliers $\lambda(\pi)$. In consequence, $LM_T(\pi) = LM_T^0(\pi)\ \forall \pi \in \Pi$ wp → 1 and (A.27) implies (A.23).

Next, to establish part (e), we show that

$$\text{(A.28)} \qquad \sup_{\pi\in\Pi}|LM_T(\pi) - LR_T(\pi)| \xrightarrow{p} 0 .$$

A two–term Taylor expansion of $\ell_T(\tilde{\theta},\pi)$ about $\hat{\theta}(\pi)$ gives

$$\text{(A.29)} \; \ell_T(\tilde{\theta},\pi) = \ell_T(\hat{\theta}(\pi),\pi) + (\tilde{\theta}-\hat{\theta}(\pi))'D\ell_T(\hat{\theta}(\pi),\pi) + \tfrac{1}{2}(\hat{\theta}(\pi)-\tilde{\theta})'D^2\ell_T(\theta^\dagger(\pi),\pi)(\hat{\theta}(\pi)-\tilde{\theta}),$$

where $\theta^\dagger(\pi)$ is such that $\sup_{\pi\in\Pi}\|\theta^\dagger(\pi)-\theta_0\| \xrightarrow{p} 0$. Since $D\ell_T(\hat\theta(\pi),\pi)=0\ \forall\pi\in\Pi$ wp $\to 1$, we obtain

$$\text{(A.30)}\quad \begin{aligned} LR_T(\pi) &= (B_T(\hat\theta(\pi)-\tilde\theta))'[-B_T^{-1}D^2\ell_T(\theta^\dagger(\pi),\pi)B_T^{-1}]B_T(\hat\theta(\pi)-\tilde\theta)+o_{p\pi}(1) \\ &= (B_T(\hat\theta(\pi)-\tilde\theta))'\mathcal{I}_T(\tilde\theta,\pi)B_T(\hat\theta(\pi)-\tilde\theta)+o_{p\pi}(1)\,. \end{aligned}$$

One term Taylor expansions, as is (A.1), give

$$\text{(A.31)}\qquad B_T^{-1}D\ell_T(\tilde\theta,\pi)=B_T^{-1}D\ell_T(\hat\theta(\pi),\pi)-\mathcal{I}_{1T}(\pi)B_T(\tilde\theta-\hat\theta(\pi))\,,$$

where $\mathcal{I}_{1T}(\pi)$ is as defined in (A.1) with $\hat\theta(\pi)-\theta_0$ replaced by $\tilde\theta-\hat\theta(\pi)$. In consequence,

$$\text{(A.32)}\quad B_T(\hat\theta(\pi)-\tilde\theta)=\mathcal{I}_{1T}^{-1}(\pi)B_T^{-1}D\ell_T(\tilde\theta,\pi)+o_{p\pi}(1)=\mathcal{I}_T^{-1}(\tilde\theta,\pi)B_T^{-1}D\ell_T(\tilde\theta,\pi)+o_{p\pi}(1)\,.$$

Substituting this result in (A.30) yields (A.28) as desired. □

PROOF OF THEOREM 2: First consider part (a). By Assumption 5,

$$\text{(A.33)}\quad \begin{bmatrix}\tilde\theta(\cdot)\\ \mathcal{I}^{-1}(\theta_0,\cdot)\end{bmatrix}=\begin{bmatrix}\mathcal{I}^{-1}(\theta_0,\cdot)B_T^{-1}D\ell_T(\theta_0,\cdot)\\ \mathcal{I}^{-1}(\theta_0,\cdot)\end{bmatrix}\Rightarrow\begin{bmatrix}\mathcal{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot)\\ \mathcal{I}^{-1}(\theta_0,\cdot)\end{bmatrix}\ \text{under}\ \theta_0\,.$$

By Assumptions 1(e), 1(f), and 5, $(\mathcal{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathcal{I}^{-1}(\theta_0,\cdot))$ has bounded uniformly continuous sample paths (as a function of $\pi\in\Pi$) with $G(\theta_0,\cdot)$–probability one. In consequence, using (5.12), the function $m(\cdot,\cdot)$ that maps $(\mathcal{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathcal{I}^{-1}(\theta_0,\cdot))$ into $\chi(\theta_0,c)$ is continuous with $G(\theta_0,\cdot)$–probability one and the continuous mapping theorem (e.g., see Pollard (1984, Thm. IV–12, pp. 66–70) or Billingsley (1968, Thm. 5.1, pp. 29–34)) gives the desired result:

$$\text{(A.34)}\quad \text{Exp–}W_T=m(\tilde\theta(\cdot),\mathcal{I}^{-1}(\theta_0,\cdot))\xrightarrow{d}m(\mathcal{I}^{-1}(\theta_0,\cdot)G(\theta_0,\cdot),\mathcal{I}^{-1}(\theta_0,\cdot))=\chi(\theta_0,c).$$

Parts (b)–(e) of the Theorem follow from part (a) and Theorem 1. □

PROOF OF LEMMA 3: We make use of the following assertion, which is verified below: If (i) $LR_T\xrightarrow{d}\chi(\theta_0,c)$ under θ_0 and (ii) $E\chi(\theta_0,c)=1$, then the densities

$\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi) : T \geq 1\right\}$ are contiguous to the densities $\{f_T(\theta_0) : T \geq 1\}$. Condition (i) holds by Theorem 2. Condition (ii) is obtained as follows: Let mgf(t) denote the moment generating function of a chi–square rv with p degrees of freedom. We have $\text{mgf}(t) = (1-2t)^{-p/2}$. Then,

$$E\chi(\theta_0, c) = (1+c)^{-p/2}\int_\Pi E\exp\left[\frac{1}{2}\frac{c}{1+c}(H\mathcal{I}^{-1}(\pi)G(\pi))'(H\mathcal{I}^{-1}(\pi)H')^{-1}H\mathcal{I}^{-1}(\pi)G(\pi)\right]dJ(\pi)$$

$$\text{(A.35)} \quad = (1+c)^{-p/2}\int_\Pi \text{mgf}\left[\frac{1}{2}\frac{c}{1+c}\right]dJ(\pi)$$

$$= 1,$$

where $\mathcal{I}(\pi)$ and $G(\pi)$ denote $\mathcal{I}(\theta_0, \pi)$ 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.

It remains to verify the assertion above. Let $(\Omega,\mathcal{A})$ be a measurable space. Let P_T and Q_T be a null distribution and an alternative distribution on $(\Omega,\mathcal{A})$ for $T \geq 1$. Let $E_T = (\Omega, \mathcal{A}, (P_T, Q_T))$. E_T is called a (binary) *experiment* and $\{E_T : T \geq 1\}$ is a sequence of experiments. One can define equivalent experiments and one can put a metric Δ_2 on the space $\mathcal{E}_2/\sim$ of equivalence classes of experiments, e.g., see Strasser (1985, pp. 74, 75). By Theorem 18.11 of Strasser (1985), if $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ for some experiment $E = (\Omega, \mathcal{A}, (P,Q))$, then $\{Q_T : T \geq 1\}$ is contiguous to $\{P_T : T \geq 1\}$ if and only if Q is absolutely continuous with respect to P, i.e., if and only if $\int_0^\infty x\mu_E(dx) = 1$, where μ_E is the distribution of the likelihood ratio dQ/dP under P, $\mathcal{L}(dQ/dP\,|\,P)$.

Also, by Theorem 16.8 of Strasser (1985), $(\mathcal{E}_2/\sim, \Delta_2)$ and $(\mathcal{M},\mathcal{T})$ are homeomorphic, where $\mathcal{M}$ is the set of all probability measures μ on $[0,\infty)$ with $\int_0^\infty x\mu(dx) \leq 1$ and $\mathcal{T}$ is the topology of weak convergence, with homeomorphism T defined by $T(\dot{E}) = \mathcal{L}(dQ/dP\,|\,P)$, where $\dot{E}$ is the equivalence class of experiments that contains E and $E = (\Omega, \mathcal{A}, (P,Q))$. In consequence, for any experiment $E = (\Omega, \mathcal{A}, (P,Q))$ and any sequence of experiments

$\{E_T : T \geq 1\} = \{(\Omega, \mathcal{A}, (P_T, Q_T)) : T \geq 1\}$, $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ if and only if $\mathcal{L}(dQ_T/dP_T | P_T) \Longrightarrow \mathcal{L}(dQ/dP | P)$ as $T \to \infty$, where "$\Longrightarrow$" denotes weak convergence (or equivalently, convergence in distribution). This result and the result of the previous paragraph establish the assertion above. □

PROOF OF THEOREM 3: Part (b) holds by the proof of Theorem 1(b). For the remaining parts, given any $\epsilon > 0$, consider the sets $\{|LR_T - \overline{LR}_T| > \epsilon\}$, $\{|\text{Exp–}\overline{W}_T - \text{Exp–}W_T| > \epsilon\}$, etc. for $T \geq 1$. The probabilities of these sets converge to zero as $T \to \infty$ under θ_0 by Theorem 1. Hence, by contiguity (Lemma 3), their probabilities also converge to zero under the densities $\left\{\int f_T(\theta_0 + B_T^{-1}h, \pi) dQ_\pi(h) dJ(\pi) : T \geq 1\right\}$. □

PROOF OF THEOREM 4: Let α_T be the rejection probability of φ_T under θ_0. Let $k^*_{\alpha_T} > 0$ and $\lambda \in [0,1]$ be constants such that the likelihood ratio test

$$\text{(A.36)} \qquad \gamma_T = \begin{cases} 1 & \text{if } LR_T > k^*_{\alpha_T} \\ \lambda & \text{if } LR_T = k^*_{\alpha_T} \\ 0 & \text{if } LR_T < k^*_{\alpha_T} \end{cases}$$

has rejection probability α_T under θ_0. Then, by the Neyman–Pearson Lemma (e.g., see Lehmann (1959, Thm. 3.1, p. 65)),

$$\text{(A.37)} \qquad \begin{aligned} &\int \varphi_T \left[\int f_T(\theta_0 + B_T^{-1}h, \pi) dQ_\pi(h) dJ(\pi)\right] d\mu_T \\ &\quad \leq \int \gamma_T \left[\int f_T(\theta_0 + B_T^{-1}h, \pi) dQ_\pi(h) dJ(\pi)\right] d\mu_T \end{aligned}$$

for all $T \geq 1$.

By Corollary 15.11 of Strasser (1985), if $\Delta_2(E_T, E) \to 0$ as $T \to \infty$, then the power of a sequence of likelihood ratio tests of asymptotic size α is convergent. Since $\Delta_2(E_T, E) \to 0$

by the proof of Lemma 3, the $\underline{\lim}_{T\to\infty}$ on the rhs of the inequality in the statement of Theorem 4 is actually $\lim_{T\to\infty}$.

This result, inequality (A.37), and Fubini's Theorem yield

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

$$\leq \lim_{T\to\infty} \int \gamma_T\Big[\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)\Big]d\mu_T$$

$$= \lim_{T\to\infty} \int 1(LR_T > k_{T\alpha})\Big[\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)\Big]d\mu_T \tag{A.38}$$

$$= \lim_{T\to\infty} \int 1(\text{Exp–}W_T > k_{T\alpha})\Big[\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)\Big]d\mu_T$$

$$= \lim_{T\to\infty} \int\Big[\int \xi_T f_T(\theta_0 + B_T^{-1}h, \pi)d\mu_T\Big]dQ_\pi(h)dJ(\pi)\ ,$$

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 $\text{Exp–}W_T - LR_T \xrightarrow{p} 0$ under $\int f_T(\theta_0 + B_T^{-1}h, \pi)dQ_\pi(h)dJ(\pi)$ by Theorem 3, and the third equality holds by Fubini's Theorem. The proof is analogous for $\text{Exp–}LM_T$ and $\text{Exp–}LR_T$. □

Next we show why the statistics $W_T^*(\pi)$ and $LM_T^*(\pi)$ of Section 7.1 are asymptotically equivalent to $W_T(\pi)$ and $LM_T(\pi)$ under θ_0 (and hence under local alternatives as well by Lemma 3). Let the subscript * be a deletion operator that deletes the last q rows and columns of s × s matrices, the last q rows of s–vectors and s × p matrices, and the last q columns of p × s matrices. By the formula for a partitioned inverse and some algebra, when $\mathcal{I}(\theta_0, \pi)$ is of the form in (7.17), we can show that $H\mathcal{I}^{-1}(\theta_0, \pi)H' = H_*[\mathcal{I}(\theta_0, \pi)_*]^{-1}H_*'$. Also, since $\mathcal{I}_T(\theta,\pi)_*$ is of the form $\begin{bmatrix} C & C \\ C & C+D \end{bmatrix}$, some algebra shows that $H_*[\mathcal{I}_T(\theta,\pi)_*]^{-1}H_*' = C^{-1} + D^{-1} = [\mathcal{I}_T(\theta_0, \pi)_{11}]^{-1} + [\mathcal{I}_T(\theta_0, \pi)_{00}]^{-1}$. In consequence, we obtain: Under θ_0,

$$\text{(A.39)} \qquad \sup_{\pi\in\Pi}\left\| [H\mathcal{I}_T^{-1}(\hat\theta(\pi),\pi)H']^{-1} - [\mathcal{I}_T^{-1}(\hat\theta(\pi),\pi)_{11} + \mathcal{I}_T^{-1}(\hat\theta(\pi),\pi)_{00}]^{-1}\right\| \xrightarrow{p} 0 .$$

Similarly, for $\mathcal{I}(\theta_0,\pi)$ of the form in (7.17), with some algebra, one can show that $H\mathcal{I}^{-1}(\theta_0,\pi)D\ell_T(\tilde\theta,\pi) = H_*[\mathcal{I}(\theta_0,\pi)_*]^{-1}D\ell_T(\tilde\theta,\pi)_*$. Also, wp $\to$ 1, $D\ell_T(\tilde\theta,\pi)_* = \left[\Sigma_{[T\pi]+1}^{T}\frac{\partial}{\partial\delta_1'}\log g_t(\tilde\delta_1,\tilde\delta_2),0'\right]'$ by the first order conditions for the restricted ML estimator $(\tilde\delta_1, \tilde\delta_2)$. Hence, $H\mathcal{I}^{-1}(\theta_0, \pi)D\ell_T(\tilde\theta,\pi) = H_*[\mathcal{I}(\theta_0, \pi)_*]^{-1}H_*'$ $\times \left[\Sigma_{[T\pi]+1}^{T}\frac{\partial}{\partial\delta_1}\log g_t(\tilde\delta_1, \tilde\delta_2)\right]$. This result, the results of the previous paragraph with $\hat\theta(\pi)$ replaced by $\tilde\theta$, the equivalence of $LM_T(\pi)$ and $LM_T^0(\pi)$ wp $\to$ 1 (see (A.26) and (A.27)) combine to give the desired result that $\sup_{\pi\in\Pi}\|LM_T(\pi) - LM_T^*(\pi)\| \xrightarrow{p} 0$ under θ_0.

PROOF OF THEOREM 5: Assumption 1(a) holds automatically by the definition of $f_T(\theta,\pi)$. Assumption 1(b) holds by (Assumption) SC(d). Assumption 1(c) holds by SC(e). Assumption 1(d) holds by SC(a), SC(f), Lemma A–2 of Andrews (1989) (which converts the a.s. convergence in SC(f) into uniform convergence in probability over $\pi \in \Pi$), and the choice of $B_T = \sqrt{T}I_s$. Assumption 1(e) holds by SC(a) and (f).

Assumption 1(f) is verified as follows. Let $L = [I_p \vdots 0] \subset R^{p\times q}$. Then,

$$\text{(A.40)} \qquad \mathcal{I}(\theta_0,\pi) = \begin{bmatrix}(1-\pi)L\\ I_q\end{bmatrix}\mathcal{J}[(1-\pi)L' \vdots I_q] + \pi(1-\pi)\begin{bmatrix}L\mathcal{J}L' & 0\\ 0 & 0\end{bmatrix} \text{ and}$$

$$\lambda_{\min}(\mathcal{I}(\theta_0,\pi)) = \inf_{\substack{b=(b_1', b_2')':\\ \|b\|=1}} [((1-\pi)L'b_1 + b_2)'\mathcal{J}((1-\pi)L'b_1 + b_2) + \pi(1-\pi)b_1'L\mathcal{J}L'b_1]$$

$$\text{(A.41)} \quad \geq \inf_{\substack{b=(b_1', b_2')':\\ \|b\|=1}} [\|(1-\pi)L'b_1 + b_2\|^2 + \pi(1-\pi)\|L'b_1\|^2]\lambda_{\min}(\mathcal{J})$$

$$= \inf_{\substack{b=(b_1', b_{21}', b_{22}')':\\ \|b\|=1}} \left[\left\|\begin{bmatrix}(1-\pi)b_1 + b_{21}\\ b_{22}\end{bmatrix}\right\|^2 + \pi(1-\pi)\|b_1\|^2\right]\lambda_{\min}(\mathcal{J})$$

$$= \inf_{\substack{b=(b_1', b_{21}', b_{22}')': \\ \|b\|=1}} [(1-\pi)(b_1+b_{21})'(b_1+b_{21}) + \pi b_{21}'b_{21} + b_{22}'b_{22}]\lambda_{\min}(\mathcal{J}),$$

where $b \in R^{p+q}$, $b_1, b_{21} \in R^p$, $b_{22} \in R^{q-p}$, and $b_2 = (b_{21}', b_{22}')'$. The right–hand side above is bounded away from zero for $\pi \in \Pi$ by SC(a) and (g).

Assumption 1(g) holds by SC(h) and the continuous mapping theorem using the fact that $B_T^{-1}D\ell_T(\theta_0, \pi) = \left[\frac{1}{\sqrt{T}}\Sigma_{[T\pi]+1}^T \frac{\partial}{\partial \delta_1'}\log g_t(\delta_{10}, \delta_{20}), \quad \frac{1}{\sqrt{T}}\Sigma_1^T \frac{\partial}{\partial \delta_1'}\log g_t(\delta_{10}, \delta_{20}), \frac{1}{\sqrt{T}}\Sigma_1^T \frac{\partial}{\partial \delta_2'}\log g_t(\delta_{10}, \delta_{20})\right]'$.

Assumption 2 is verified using Lemma A–1 of Andrews (1989). We need to verify its conditions (a) and (b) for $Q_T(\theta,\pi) = -\frac{1}{T}\Sigma_1^{T\pi}\log g_t(\delta_1, \delta_2) - \Sigma_{T\pi+1}^T\log g_t(\delta_1+\beta, \delta_2)$ and $Q(\theta,\pi) = -\pi Q(\delta_1, \delta_2) - (1-\pi)Q(\delta_1+\beta, \delta_2)$, where $\theta = (\beta', \delta_1', \delta_2')'$. Condition (a) requires $\sup_{\pi\in\Pi,\, \theta\in\Theta} |Q_T(\theta,\pi) - Q(\theta,\pi)| \xrightarrow{p} 0$. The latter holds by SC(a) and (b) using Lemma A–2 of Andrews (1989). Condition (b) requires that for every neighborhood Θ_0 ($\subset \Theta$) of θ_0, $\inf_{\pi\in\Pi}(\inf_{\theta\in\Theta/\Theta_0} Q(\theta,\pi) - Q(\theta_0, \pi)) > 0$. This holds by SC(a) and (c). This completes the verification of Assumption 2 of Section 5.

Assumption 4(a) holds trivially by the definition of $f_T(\theta,\pi)$ in (7.5). Assumption 4(b) is verified in exactly the same manner as Assumption 2 but with the parameter space Θ replaced by the restricted parameter space $\tilde{\Theta}$.

Next, we verify Assumption 5. Let $\mathcal{K}$, $\mathcal{J}_{21}$, and $\mathcal{J}_{22}$ be matrices such that

(A.42) $$\mathcal{K}\mathcal{K}' = \begin{bmatrix} \mathcal{J}_1^{1/2} & 0 \\ \mathcal{J}_{21} & \mathcal{J}_{22} \end{bmatrix}\begin{bmatrix} \mathcal{J}_1^{1/2} & 0 \\ \mathcal{J}_{21} & \mathcal{J}_{22} \end{bmatrix}' = \mathcal{J}.$$

That is, $\mathcal{J}_2 = \mathcal{J}_1^{1/2}\mathcal{J}_{21}'$ and $\mathcal{J}_3 = \mathcal{J}_{21}\mathcal{J}_{21}' + \mathcal{J}_{22}\mathcal{J}_{22}'$. Then, Assumption 5 holds with

(A.43) $$G(\theta_0, \pi) = ([\mathcal{J}_1^{1/2}(B_1(1) - B_1(\pi))]', [\mathcal{J}_1^{1/2}(B_1(1)]', [\mathcal{J}_{21}B_1(1) + \mathcal{J}_{22}B_2(1))]',$$

where $B(\pi) = (B_1(\pi)', B_2(\pi)')'$ is the q–vector Brownian motion of SC(h), $B_1(\pi) \in R^p$, and $B_2(\pi) \in R^{q-p}$. To see that Assumption 5 holds, note that the definition of

$B_T^{-1}D\ell_T(\theta_0, \pi)$ given above and SC(h) yield $B_T^{-1}D\ell_T(\theta_0, \cdot) \Rightarrow G(\theta_0, \cdot)$. Also,

$$\text{(A.44)} \quad EG(\theta_0,\pi)G(\theta_0,\pi)' = \begin{bmatrix} (1-\pi)\mathcal{J}_1 & (1-\pi)\mathcal{J}_1 & (1-\pi)\mathcal{J}_1^{1/2}\mathcal{J}_{21}' \\ (1-\pi)\mathcal{J}_1 & \mathcal{J}_1 & \mathcal{J}_1^{1/2}\mathcal{J}_{21}' \\ (1-\pi)\mathcal{J}_{21}\mathcal{J}_1^{1/2} & \mathcal{J}_{21}\mathcal{J}_1^{1/2} & \mathcal{J}_{21}\mathcal{J}_{21}' + \mathcal{J}_{22}\mathcal{J}_{22}' \end{bmatrix} = \mathcal{I}(\theta_0,\pi),$$

where the second equality uses (7.17) and the above expressions for $\mathcal{J}_2$ and $\mathcal{J}_3$. $G(\theta_0, \cdot)$ has bounded uniformly continuous sample paths with probability one because $B(\cdot)$ does. □

PROOF OF LEMMA 4: (Assumption) SC1(a) implies SC(a). The Markov property (SC1(c)) ensures that $\{\log g_t(\delta_1, \delta_2) : t > m\}$ is part of a double infinite stationary and ergodic sequence. Thus, using SC1(c) and (d), the ergodic theorem implies that $\frac{1}{T}\Sigma_1^T(\log g_t(\delta_1, \delta_2) - E \log g_t(\delta_1, \delta_2)) \to 0$ a.s. under θ_0 $\forall(\delta_1, \delta_2) \in \Delta_1 \times \Delta_2$. A generic uniform SLLN (e.g., Assumptions TSE–1D, BD, DM, P–SLLN, and P–SLLN2 and Theorem 6 of Andrews (1992)) strengthens this result to uniform convergence over $\Delta_1 \times \Delta_2$ a.s. using SC1(b), (c), (d), and (e). The same generic ULLN establishes the continuity of $E \log g_t(\delta_1, \delta_2)$ on $\Delta_1 \times \Delta_2$. Since $\Delta_1 \times \Delta_2$ is compact, this gives $\sup_{\pi\in\Pi}|Q(\delta_1, \delta_2)| < \infty$, where $Q(\delta_1, \delta_2) = E \log g_t(\delta_1, \delta_2)$ for any $t > m$. Thus, SC(b) holds.

Next, to establish SC(c), note that for $(\delta_1, \delta_2) \neq (\delta_{10}, \delta_{20})$,

$$\text{(A.45)} \quad \begin{aligned} Q(\delta_1, \delta_2) - Q(\delta_{10}, \delta_{20}) &= E_0 \log[g_t(\delta_1, \delta_2)/g_t(\delta_{10}, \delta_{20})] \\ &< \log E_0 g_t(\delta_1, \delta_2)/g_t(\delta_{10}, \delta_{20}) = 0 , \end{aligned}$$

where "E_0" denotes expectation under θ_0 and the inequality is an application of Jensen's inequality and is strict by SC1(f). Equation (A.45) implies that $Q(\delta_1, \delta_2)$ is uniquely minimized over $\Delta_1 \times \Delta_2$ at $(\delta_{10}, \delta_{20})$. Since $\Delta_1 \times \Delta_2$ is compact and $Q(\delta_1, \delta_2)$ is continuous on $\Delta_1 \times \Delta_2$, this gives SC(c).

Now, SC(d) holds by SC1(b), SC(e) holds by SC1(g), SC(f) holds by SC1(b), (c), (g), and (h) using a generic uniform SLLN as above, and SC(g) holds by SC1(h).

SC(h) is established as follows. Using the Markov property, $\left\{\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20}) : t > m\right\}$ is part of a doubly infinite strictly stationary ergodic sequence. Also, $\left\{\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20}), \mathcal{F}_{t-1} : t > m\right\}$ is a martingale difference sequence (MDS), because

$$E\left[\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20})\middle|\mathcal{F}_{t-1}\right] = E\left[\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20})\middle|S_{t,m}, X_{t,m}\right]$$

$$\text{(A.46)} \quad = \int \frac{\partial}{\partial(\delta_1', \delta_2')}g_t(\delta_{10}, \delta_{20})d\mu(s_t) = \frac{\partial}{\partial(\delta_1', \delta_2')}\int g_t(\delta_{10}, \delta_{20})d\mu(s_t) = 0\ ,$$

where the third inequality holds by the dominated convergence theorem using SC1(g) and the first part of SC1(h). SC(h) now follows from a multivariate invariance principle for stationary ergodic martingale difference sequences. Using Lemma A–5 of Andrews (1989), the latter holds if $\left\{\alpha'\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20}) : t > m\right\}$ satisfies a univariate invariance principle $\forall\alpha \in R^q$ with $\alpha \neq 0$ and $\eta_T(\pi) = \frac{1}{\sqrt{T}}\Sigma_1^{T\pi}\frac{\partial}{\partial(\delta_1', \delta_2')}\log g_t(\delta_{10}, \delta_{20})$ has asymptotically independent increments for $\pi \in \Pi$. Numerous univariate invariance principles exist that are applicable in the present context. For example, the univariate invariance principle of Heyde (see Hall and Heyde (1980, Thm. 5.5, pp. 141–145)), which is applied in Andrews (1989, Proof of Theorem 1), will do. To verify that $\{\eta_T(\pi) : T \geq 1\}$ has asymptotically independent increments, see Andrews (1989, Proof of Theorem 1). (Note that it is not possible to use directly a CLT for a stationarity ergodic MDS to establish this property – some complications arise.) □

FOOTNOTES

[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 and Bruce Hansen for helpful comments. Andrews gratefully acknowledges research support from the National Science Foundation through grant number SES–8821021. Ploberger gratefully acknowledges research support from Fonds zur Förderung der wissenschaftlichen Forschung under Schrödingerstipendium Project J–0469–PHY.

[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)).

REFERENCES

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

_______ (1989): "Tests for Parameter Instability and Structural Change with Unknown Change Point," Cowles Foundation Discussion Paper No. 943 (revised), Yale University.

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

Andrews, D. W. K. and R. C. Fair (1988): "Inference in Nonlinear Econometric Models with Structural Change," *Review of Economic Studies*, 55, 615–640.

Andrews, D. W. K., I. Lee, and W. Ploberger (1992): "Optimal Changepoint Tests for Linear Regression," 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.

Brown, R. L., J. Durbin, and J. M. Evans (1975): "Techniques for Testing the Constancy of Regression Relationships over Time," *Journal of the Royal Statistical Society, Series B*, 37, 149–192.

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.

Eberlein, E. (1986): "On Strong Invariance Principles under Dependence Assumptions," *Annals of Probability*, 14, 260–270.

Engle, R. F., D. F. Hendry, and J.–F. Richard (1983): "Exogeneity," *Econometrica*, 51, 277–304.

Farley, J. U. and M. J. Hinich (1970): "A Test for a Shifting Slope Coefficient in a Linear Model," *Journal of the American Statistical Association*, 65, 1320–1329.

Gardner, L. A., Jr. (1969): "On Detecting Changes in the Mean of Normal Variates," *Annals of Mathematical Statistics*, 40, 116–126.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

Hansen, B. E. (1990): "Testing for Structural Change of Unknown Form in Models with Nonstationary Regressors," unpublished manuscript, Department of Economics, University of Rochester.

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

_______ (1991b): "Strong Laws for Dependent Heterogeneous Processes," *Econometric Theory*, 7, 213–221. "Corrigenda," *Econometric Theory*, forthcoming.

Hawkins, D. L. (1987): "A Test for a Change Point in a Parametric Model Based on a Maximal Wald–type Statistic," *Sankhya*, 49, 368–376.

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.

Kim, H. J. and D. Siegmund (1989): "The Likelihood Ratio Test for a Change–point in Simple Linear Regression," *Biometrika*, 76, 409–423.

King, M. L. and T. S. Shively (1991): "Locally Optimal Testing When a Nuisance Parameter Is Present Only Under the Alternative," unpublished manuscript, Department of Economics, Monash University.

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

McLeish, D. L. (1975): "A Maximal Inequality and Dependent Strong Laws," *Annals of Probability*, 3, 829–839.

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

Newey, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equi–continuity," *Econometrica*, 59, 1161–1167.

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

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

Phillips, P. C. B. and S. N. Durlauf (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

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

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

Pötscher, B. M. and I. R. Prucha (1989): "A Uniform Law of Large Numbers for Dependent and Heterogeneous Data Processes," *Econometrica*, 57, 675–683.

_______ (1990): "Generic Uniform Convergence and Equicontinuity Concepts for Random Functions: An Exploration of the Basic Structure," Working Paper No. 90–14, Department of Economics, University of Maryland.

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

Sen, A. K. and M. S. Srivastava (1973): "On Multivariate Tests for Detecting Change in Mean," *Sankhya* A, 35, 173–185.

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

Teräsvirta, T. and H. M. Anderson (1991): "Modelling Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models," discussion paper #91–24, Department of Economics, University of California, San Diego.

Tong, H. (1983): *Threshold Models in Non–linear Time Series Analysis. Springer Lecture Notes in Statistics*, Vol. 21. Oxford: Oxford University Press.

_______ (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.

TABLE I. ASYMPTOTIC CRITICAL VALUES FOR c=∞ TEST

		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%
.500	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
.495	1.041	1.39	1.97	3.38	2.28	2.95	4.56	3.08	3.84	5.64	3.89	4.71	6.64	4.62	5.51	7.44
.490	1.083	1.40	1.98	3.40	2.29	2.96	1.381	3.09	3.84	5.62	3.90	4.72	6.67	4.62	5.50	7.44
.485	1.128	1.41	1.99	3.45	2.30	2.98	1.542	3.10	3.85	5.61	3.93	4.74	6.64	4.64	5.53	7.47
.480	1.174	1.40	1.99	3.42	2.30	2.97	1.642	3.11	3.87	5.59	3.93	4.74	6.63	4.66	5.55	7.43
.475	1.222	1.40	2.00	3.38	2.31	2.96	1.552	3.12	3.88	5.61	3.93	4.78	6.66	4.67	5.57	7.47
.470	1.272	1.41	2.00	3.37	2.31	2.96	1.562	3.12	3.88	5.64	3.94	4.79	6.67	4.70	5.58	7.44
.460	1.378	1.41	1.99	3.39	2.33	2.98	4.51	3.15	3.89	5.57	3.97	4.81	6.79	4.74	5.61	7.46
.450	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
.440	1.620	1.43	2.00	3.40	2.34	3.02	4.57	3.17	3.90	5.61	4.02	4.88	6.84	4.78	5.66	7.57
.420	1.907	1.43	1.99	3.39	2.37	3.03	4.66	3.21	3.94	5.61	4.08	4.90	6.95	4.85	5.69	7.56
.400	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
.380	2.662	1.45	2.01	3.37	2.37	3.07	4.70	3.26	4.00	5.61	4.14	4.95	6.99	4.92	5.79	7.62
.360	3.160	1.45	2.02	3.34	2.39	3.08	4.68	3.28	4.02	5.65	4.17	4.96	7.03	4.95	5.82	7.70
.350	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
.340	3.768	1.45	2.02	3.39	2.41	3.09	4.68	3.29	4.07	5.64	4.19	5.01	7.00	4.94	5.83	7.71
.320	4.516	1.46	2.02	3.40	2.43	3.10	4.68	3.31	4.11	5.64	4.21	5.05	6.97	4.99	5.85	7.88
.300	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
.280	6.612	1.47	2.03	3.43	2.46	3.11	4.76	3.38	4.11	5.68	4.26	5.07	6.96	5.04	5.89	7.86
.260	8.101	1.47	2.02	3.44	2.49	3.11	4.76	3.39	4.15	5.66	4.26	5.11	7.00	5.07	5.95	7.82
.250	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
.240	10.028	1.48	2.01	3.44	2.51	3.14	4.76	3.41	4.16	5.68	4.30	5.11	7.00	5.09	6.00	7.81
.220	12.570	1.49	2.02	3.43	2.54	3.18	4.75	3.43	4.19	5.73	4.30	5.17	7.03	5.13	6.03	7.82
.200	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
.180	20.753	1.50	2.03	3.36	2.57	3.21	4.76	3.47	4.24	5.69	4.34	5.17	7.04	5.19	6.09	7.86
.160	27.562	1.50	2.05	3.40	2.59	3.21	4.75	3.48	4.23	5.73	4.35	5.21	7.15	5.22	6.11	7.88
.150	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
.140	37.735	1.50	2.06	3.43	2.59	3.23	4.77	3.49	4.24	5.77	4.38	5.24	7.12	5.24	6.14	7.92
.120	53.778	1.51	2.08	3.39	2.59	3.23	4.74	3.51	4.26	5.75	4.40	5.24	7.13	5.26	6.14	7.95
.100	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
.080	132.250	1.54	2.09	3.39	2.62	3.27	4.76	3.55	4.30	5.85	4.45	5.25	7.08	5.34	6.19	8.02
.060	245.444	1.54	2.08	3.39	2.64	3.27	4.73	3.54	4.30	5.88	4.47	5.28	7.08	5.37	6.20	8.10
.050	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
.040	576.000	1.54	2.08	3.38	2.65	3.29	4.75	3.55	4.30	5.87	4.50	5.30	7.11	5.41	6.22	8.13
.030	1045.444	1.54	2.07	3.39	2.66	3.30	4.74	3.58	4.30	5.88	4.51	5.32	7.12	5.41	6.22	8.15
.020	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

		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%
.500	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
.495	1.041	5.29	6.19	8.34	5.93	6.95	9.06	5.95	7.09	9.45	7.35	8.51	10.71	7.91	9.06	11.48
.490	1.083	5.31	6.24	8.41	5.97	7.02	9.14	5.98	7.10	9.49	7.36	8.53	10.80	7.95	9.08	11.51
.485	1.128	5.31	6.23	8.43	6.01	7.06	9.23	6.00	7.11	9.49	7.41	8.53	10.77	7.96	9.12	11.54
.480	1.174	5.33	6.24	8.46	6.03	7.07	9.24	6.01	7.13	9.48	7.43	8.57	10.84	7.99	9.15	11.62
.475	1.222	5.36	6.27	8.46	6.06	7.10	9.20	6.04	7.16	9.54	7.46	8.60	10.88	8.03	9.18	11.63
.470	1.272	5.37	6.29	8.53	6.06	7.12	9.19	6.07	7.21	9.54	7.48	8.61	10.90	8.06	9.19	11.68
.460	1.378	5.39	6.33	8.57	6.11	7.19	9.29	6.12	7.26	9.49	7.52	8.67	10.88	8.14	9.22	11.69
.450	1.494	5.41	6.37	8.59	6.14	7.20	9.39	6.14	7.30	9.58	7.56	8.71	11.00	8.14	9.25	11.85
.440	1.620	5.45	6.38	8.62	6.16	7.23	9.44	6.17	7.34	9.58	7.59	8.76	10.99	8.19	9.30	11.87
.420	1.907	5.51	6.45	8.68	6.23	7.28	9.47	6.22	7.38	9.64	7.68	8.90	11.11	8.26	9.41	11.92
.400	2.250	5.58	6.51	8.70	6.28	7.31	9.50	6.30	7.41	9.68	7.77	8.96	11.16	8.35	9.44	12.06
.380	2.662	5.63	6.58	8.71	6.34	7.30	9.52	6.34	7.43	9.73	7.86	9.01	11.17	8.44	9.55	12.11
.360	3.160	5.65	6.62	8.74	6.38	7.33	9.47	6.41	7.49	9.81	7.89	9.00	11.25	8.50	9.56	12.20
.350	3.449	5.68	6.65	8.77	6.41	7.35	9.47	6.39	7.48	9.81	7.92	9.01	11.29	8.52	9.57	12.21
.340	3.768	5.69	6.67	8.75	6.42	7.38	9.52	6.41	7.50	9.83	7.94	9.02	11.32	8.56	9.61	12.22
.320	4.516	5.71	6.70	8.75	6.46	7.41	9.49	6.47	7.56	9.81	8.02	9.07	11.34	8.63	9.69	12.27
.300	5.444	5.73	6.74	8.78	6.48	7.46	9.55	6.53	7.62	9.87	8.04	9.14	11.38	8.67	9.75	12.42
.280	6.612	5.79	6.76	8.86	6.50	7.45	9.53	6.58	7.65	9.93	8.10	9.17	11.32	8.71	9.82	12.44
.260	8.101	5.84	6.77	8.83	6.54	7.47	9.50	6.62	7.66	10.01	8.13	9.19	11.36	8.76	9.86	12.42
.250	9.000	5.85	6.79	8.86	6.56	7.49	9.56	6.65	7.69	10.07	8.15	9.20	11.43	8.80	9.88	12.47
.240	10.028	5.87	6.81	8.85	6.59	7.50	9.54	6.66	7.69	10.06	8.16	9.23	11.44	8.84	9.91	12.54
.220	12.570	5.89	6.84	8.92	6.60	7.54	9.54	6.70	7.74	10.05	8.20	9.25	11.50	8.88	9.93	12.59
.200	16.000	5.92	6.86	8.92	6.62	7.54	9.56	6.74	7.76	10.07	8.24	9.27	11.58	8.92	9.96	12.58
.180	20.753	5.97	6.90	8.93	6.65	7.57	9.56	6.77	7.78	10.17	8.27	9.29	11.57	8.95	10.00	12.61
.160	27.562	5.98	6.90	8.98	6.69	7.63	9.54	6.82	7.82	10.17	8.30	9.33	11.60	9.00	10.04	12.58
.150	32.111	6.01	6.92	8.96	6.70	7.66	9.53	6.83	7.82	10.15	8.31	9.35	11.67	9.00	10.04	12.61
.140	37.735	6.02	6.93	8.98	6.73	7.68	9.52	6.84	7.84	10.13	8.34	9.37	11.68	9.01	10.08	12.58
.120	53.778	6.05	6.95	8.99	6.78	7.74	9.63	6.91	7.89	10.15	8.41	9.39	11.68	9.07	10.14	12.58
.100	81.000	6.09	6.98	8.99	6.82	7.77	9.61	6.95	7.93	10.25	8.44	9.42	11.66	9.13	10.19	12.61
.080	132.250	6.14	7.03	8.96	6.86	7.81	9.61	7.00	7.98	10.28	8.47	9.47	11.63	9.18	10.29	12.66
.060	245.444	6.17	7.07	8.94	6.90	7.83	9.62	7.03	8.05	10.29	8.53	9.50	11.70	9.25	10.31	12.67
.050	361.000	6.19	7.09	8.94	6.91	7.85	9.66	7.05	8.08	10.30	8.55	9.55	11.74	9.29	10.34	12.66
.040	576.000	6.20	7.09	8.94	7.00	7.94	9.82	7.82	8.81	11.10	8.59	9.57	11.73	9.33	10.38	12.68
.030	1045.444	6.21	7.10	8.96	7.03	7.95	9.80	7.84	8.83	11.10	8.60	9.61	11.74	9.36	10.41	12.67
.020	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. (CONT.)

		p=11			p=12			p=13			p=14			p=15		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.500	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
.495	1.041	8.53	9.68	12.33	9.32	10.54	12.99	9.73	10.94	13.95	10.43	11.75	14.70	11.23	12.54	15.37
.490	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
.485	1.128	8.59	9.74	12.32	9.40	10.56	13.10	9.80	11.04	13.88	10.50	11.80	14.71	11.31	12.62	15.49
.480	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	12.66	15.54
.475	1.222	8.66	9.80	12.38	9.46	10.61	13.06	9.85	11.15	13.92	10.61	11.91	14.74	11.38	12.73	15.55
.470	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
.460	1.378	8.74	9.90	12.44	9.55	10.72	13.09	9.98	11.25	13.91	10.76	11.96	14.88	11.51	12.82	15.57
.450	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
.440	1.620	8.83	9.99	12.53	9.63	10.80	13.26	10.06	11.34	14.04	10.90	12.06	14.87	11.62	12.92	15.76
.420	1.907	8.95	10.13	12.64	9.69	10.86	13.37	10.18	11.43	14.13	10.99	12.21	14.96	11.72	13.04	15.82
.400	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
.380	2.662	9.08	10.30	12.70	9.84	10.99	13.41	10.38	11.59	14.47	11.14	12.36	15.16	11.93	13.23	15.90
.360	3.160	9.18	10.38	12.78	9.92	11.11	13.47	10.45	11.67	14.46	11.21	12.40	15.28	12.00	13.26	15.92
.350	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
.340	3.768	9.23	10.42	12.77	10.00	11.18	13.47	10.53	11.76	14.46	11.25	12.47	15.29	12.07	13.33	15.93
.320	4.516	9.30	10.50	12.83	10.05	11.25	13.55	10.60	11.90	14.42	11.34	12.54	15.39	12.13	13.41	15.95
.300	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
.280	6.612	9.41	10.55	12.82	10.17	11.33	13.71	10.73	11.98	14.56	11.45	12.64	15.53	12.24	13.53	16.11
.260	8.101	9.45	10.60	12.81	10.23	11.34	13.71	10.81	12.05	14.59	11.54	12.66	15.51	12.32	13.58	16.11
.250	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
.240	10.028	9.50	10.63	12.93	10.25	11.40	13.73	10.87	12.11	14.62	11.59	12.73	15.39	12.38	13.63	16.16
.220	12.570	9.55	10.67	13.04	10.29	11.47	13.67	10.92	12.11	14.63	11.63	12.83	15.47	12.41	13.72	16.18
.200	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
.180	20.753	9.63	10.68	13.15	10.39	11.49	13.73	11.02	12.20	14.64	11.73	13.00	15.65	12.58	13.79	16.28
.160	27.562	9.68	10.69	13.22	10.43	11.55	13.84	11.07	12.25	14.63	11.79	13.06	15.70	12.66	13.82	16.38
.150	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
.140	37.735	9.72	10.77	13.19	10.47	11.57	13.80	11.12	12.30	14.65	11.86	13.11	15.82	12.70	13.88	16.44
.120	53.778	9.76	10.82	13.21	10.52	11.62	13.82	11.19	12.39	14.78	11.90	13.16	15.79	12.78	13.97	16.50
.100	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
.080	132.250	9.88	10.95	13.19	10.63	11.70	13.94	11.31	12.51	14.86	12.03	13.25	15.81	12.94	14.03	16.66
.060	245.444	9.94	10.99	13.24	10.68	11.74	14.00	11.35	12.52	15.02	12.07	13.27	15.82	12.98	14.10	16.64
.050	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
.040	576.000	10.01	11.06	13.24	10.72	11.80	13.99	11.61	12.75	15.21	12.38	13.51	15.92	13.07	14.19	16.66
.030	1045.444	10.06	11.09	13.30	10.79	11.85	14.00	11.66	12.80	15.23	12.42	13.55	15.96	13.12	14.24	16.75
.020	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

		p=16			p=17			p=18			p=19			p=20		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.500	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
.495	1.041	11.78	13.16	16.20	12.47	13.83	16.70	12.97	14.37	17.38	13.63	15.15	18.31	14.25	15.68	18.90
.490	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
.485	1.128	11.86	13.17	16.19	12.52	14.00	16.81	13.10	14.51	17.43	13.79	15.26	18.43	14.37	15.82	19.00
.480	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
.475	1.222	11.95	13.34	16.33	12.62	14.07	17.03	13.17	14.58	17.51	13.89	15.30	18.53	14.49	15.97	19.06
.470	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
.460	1.378	12.06	13.40	16.43	12.72	14.19	17.05	13.25	14.66	17.72	14.05	15.38	18.56	14.59	16.14	19.22
.450	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
.440	1.620	12.19	13.47	16.57	12.86	14.27	17.38	13.44	14.79	17.80	14.20	15.55	18.77	14.78	16.31	19.37
.420	1.907	12.32	13.64	16.67	12.96	14.38	17.33	13.56	14.93	17.83	14.30	15.66	19.05	14.88	16.42	19.39
.400	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
.380	2.662	12.48	13.85	16.80	13.18	14.52	17.33	13.79	15.17	17.99	14.50	15.89	19.09	15.09	16.60	19.47
.360	3.160	12.58	13.90	16.73	13.25	14.63	17.46	13.90	15.22	17.97	14.57	16.03	19.12	15.20	16.67	19.52
.350	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
.340	3.768	12.68	13.99	16.72	13.34	14.72	17.56	14.02	15.34	18.00	14.70	16.13	19.12	15.32	16.80	19.57
.320	4.516	12.75	14.04	16.78	13.43	14.81	17.61	14.09	15.39	18.04	14.76	16.16	19.12	15.39	16.90	19.55
.300	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
.280	6.612	12.90	14.24	16.93	13.59	14.98	17.75	14.22	15.53	18.15	14.92	16.29	19.30	15.55	17.10	19.75
.260	8.101	13.01	14.34	16.97	13.66	15.02	17.74	14.32	15.59	18.32	15.03	16.42	19.32	15.67	17.13	19.76
.250	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
.240	10.028	13.08	14.43	17.00	13.75	15.07	17.86	14.39	15.65	18.41	15.09	16.44	19.39	15.74	17.19	19.80
.220	12.570	13.13	14.47	17.00	13.84	15.11	17.84	14.47	15.70	18.45	15.15	16.51	19.41	15.84	17.23	19.95
.200	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
.180	20.753	13.27	14.54	17.13	13.94	15.21	17.82	14.62	15.85	18.52	15.32	16.63	19.62	16.02	17.43	20.15
.160	27.562	13.29	14.59	17.25	14.01	15.27	17.89	14.68	15.92	18.53	15.38	16.73	19.60	16.11	17.53	20.10
.150	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
.140	37.735	13.35	14.69	17.24	14.07	15.35	17.90	14.77	15.99	18.59	15.46	16.80	19.64	16.21	17.61	20.23
.120	53.778	13.41	14.74	17.27	14.13	15.42	17.98	14.84	16.07	18.59	15.51	16.85	19.65	16.30	17.66	20.35
.100	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
.080	132.250	13.57	14.86	17.35	14.25	15.61	18.05	14.99	16.17	18.78	15.70	16.97	19.71	16.48	17.80	20.45
.060	245.444	13.66	14.93	17.38	14.37	15.66	18.09	15.05	16.27	18.80	15.80	17.08	19.82	16.53	17.84	20.52
.050	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
.040	576.000	13.79	15.01	17.44	14.47	15.75	18.18	15.13	16.37	18.85	15.89	17.15	19.88	16.62	17.92	20.60
.030	1045.444	13.83	15.07	17.46	14.51	15.78	18.23	15.18	16.40	18.85	15.96	17.18	19.85	16.67	17.98	20.66
.020	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 TEST

		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%
.500	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
.495	1.041	2.75	3.92	6.72	4.53	5.87	9.06	6.12	7.63	11.19	7.72	9.34	13.21	9.19	10.92	14.72
.490	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
.485	1.128	2.76	3.88	6.70	4.51	5.82	8.93	6.10	7.57	11.01	7.72	9.29	13.02	9.13	10.86	14.59
.480	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
.475	1.222	2.70	3.86	6.56	4.46	5.73	8.82	6.03	7.51	10.82	7.64	9.24	12.92	9.08	10.76	14.49
.470	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
.460	1.378	2.69	3.77	6.41	4.40	5.66	8.63	5.99	7.41	10.57	7.58	9.13	12.84	9.03	10.71	14.34
.450	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
.440	1.620	2.64	3.70	6.31	4.35	5.54	8.43	5.90	7.32	10.38	7.49	9.00	12.69	8.94	10.55	13.97
.420	1.907	2.59	3.63	6.18	4.32	5.50	8.37	5.85	7.21	10.16	7.45	8.93	12.54	8.84	10.38	13.74
.400	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
.380	2.662	2.54	3.50	5.89	4.19	5.39	8.08	5.71	7.06	9.75	7.30	8.65	12.27	8.68	10.18	13.35
.360	3.160	2.48	3.43	5.82	4.14	5.31	7.90	5.66	6.96	9.66	7.24	8.47	11.98	8.59	10.01	13.21
.350	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
.340	3.768	2.45	3.40	5.65	4.08	5.22	7.77	5.59	6.83	9.46	7.18	8.39	11.72	8.52	9.91	13.05
.320	4.516	2.41	3.32	5.55	4.04	5.12	7.56	5.56	6.77	9.31	7.09	8.34	11.51	8.39	9.81	12.81
.300	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
.280	6.612	2.34	3.20	5.34	3.97	5.03	7.35	5.44	6.55	8.98	6.91	8.19	11.20	8.25	9.61	12.39
.260	8.101	2.29	3.11	5.27	3.93	5.00	7.30	5.38	6.50	8.90	6.85	8.12	11.09	8.17	9.50	12.26
.250	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
.240	10.028	2.26	3.07	5.20	3.92	4.91	7.16	5.32	6.42	8.74	6.80	8.05	10.84	8.09	9.42	12.12
.220	12.570	2.24	3.02	5.10	3.88	4.84	7.07	5.26	6.34	8.58	6.72	7.97	10.63	8.01	9.32	11.87
.200	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
.180	20.753	2.20	2.92	4.92	3.81	4.71	6.93	5.17	6.19	8.33	6.60	7.77	10.38	7.88	9.15	11.61
.160	27.562	2.17	2.90	4.78	3.78	4.65	6.80	5.13	6.11	8.25	6.53	7.68	10.24	7.80	9.04	11.41
.150	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
.140	37.735	2.15	2.85	4.67	3.73	4.59	6.67	5.07	6.05	8.14	6.47	7.63	10.15	7.74	8.95	11.20
.120	53.778	2.12	2.81	4.56	3.67	4.53	6.46	5.02	5.98	8.04	6.42	7.53	9.96	7.66	8.88	11.06
.100	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
.080	132.250	2.06	2.73	4.30	3.59	4.39	6.27	4.95	5.82	7.84	6.30	7.36	9.67	7.54	8.70	10.76
.060	245.444	2.03	2.69	4.19	3.56	4.31	6.14	4.87	5.74	7.70	6.22	7.26	9.56	7.49	8.59	10.55
.050	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
.040	576.000	2.00	2.64	4.10	3.52	4.24	5.99	4.82	5.66	7.55	6.17	7.17	9.41	7.42	8.47	10.43
.030	1045.444	1.98	2.62	4.05	3.50	4.22	5.94	4.79	5.63	7.48	6.13	7.11	9.33	7.38	8.41	10.34
.020	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

		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%
.500	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
.495	1.041	10.51	12.31	16.59	11.79	13.79	18.06	11.80	14.06	18.81	14.57	16.87	21.21	15.70	17.99	22.85
.490	1.083	10.48	12.27	16.71	11.79	13.85	18.05	11.79	14.00	18.76	14.55	16.80	21.35	15.69	17.91	22.76
.485	1.128	10.43	12.24	16.63	11.79	13.64	17.97	11.78	13.97	18.66	14.52	16.73	21.28	15.62	17.90	22.64
.480	1.174	10.40	12.18	16.57	11.76	13.82	17.88	11.74	13.94	18.54	14.50	16.75	21.26	15.59	17.85	22.63
.475	1.222	10.39	12.14	16.47	11.72	13.76	17.78	11.73	13.92	18.49	14.46	16.69	21.07	15.59	17.79	22.52
.470	1.272	10.36	12.06	16.43	11.70	13.73	17.83	11.72	13.87	18.41	14.44	16.62	21.02	15.55	17.76	22.44
.460	1.378	10.29	11.97	16.32	11.64	13.73	17.60	11.65	13.82	18.31	14.36	16.46	20.90	15.49	17.58	22.38
.450	1.494	10.24	11.98	16.22	11.58	13.64	17.48	11.58	13.76	18.14	14.31	16.41	20.81	15.43	17.48	22.36
.440	1.620	10.18	11.96	16.14	11.55	13.47	17.31	11.53	13.67	18.05	14.25	16.37	20.70	15.36	17.40	22.21
.420	1.907	10.09	11.76	15.90	11.41	13.29	17.12	11.46	13.56	17.81	14.13	16.24	20.44	15.21	17.16	21.80
.400	2.250	9.99	11.61	15.50	11.33	13.10	16.80	11.44	13.35	17.64	14.02	16.11	20.06	15.11	16.96	21.64
.380	2.662	9.92	11.45	15.19	11.23	12.90	16.56	11.31	13.24	17.41	13.91	15.89	19.82	15.00	16.87	21.49
.360	3.160	9.85	11.39	14.92	11.13	12.78	16.22	11.28	13.04	17.25	13.79	15.65	19.76	14.90	16.76	21.15
.350	3.449	9.79	11.34	14.82	11.08	12.72	16.10	11.14	12.95	17.05	13.73	15.58	19.65	14.87	16.72	21.04
.340	3.768	9.75	11.28	14.65	11.04	12.70	16.00	11.12	12.94	16.92	13.67	15.48	19.48	14.80	16.62	20.96
.320	4.516	9.65	11.15	14.37	10.96	12.56	15.79	11.07	12.82	16.64	13.51	15.34	19.37	14.65	16.39	20.72
.300	5.444	9.59	11.04	14.25	10.83	12.42	15.62	11.02	12.70	16.28	13.43	15.18	19.05	14.54	16.27	20.58
.280	6.612	9.52	10.96	14.09	10.79	12.25	15.38	10.98	12.65	16.15	13.30	14.96	18.83	14.43	16.16	20.28
.260	8.101	9.45	10.86	13.85	10.72	12.12	15.25	10.89	12.46	15.90	13.21	14.82	18.61	14.34	16.05	20.11
.250	9.000	9.41	10.81	13.68	10.66	12.08	15.18	10.85	12.39	15.79	13.16	14.77	18.50	14.28	15.97	19.98
.240	10.028	9.37	10.74	13.63	10.61	12.00	15.07	10.79	12.33	15.67	13.09	14.70	18.35	14.24	15.88	19.92
.220	12.570	9.27	10.63	13.46	10.52	11.87	14.90	10.73	12.19	15.53	13.01	14.53	18.11	14.11	15.77	19.62
.200	16.000	9.20	10.50	13.29	10.44	11.70	14.74	10.67	12.10	15.41	12.95	14.45	17.84	14.00	15.63	19.28
.180	20.753	9.12	10.38	13.15	10.38	11.62	14.61	10.61	11.99	15.21	12.87	14.29	17.62	13.93	15.48	19.05
.160	27.562	9.07	10.26	13.05	10.30	11.52	14.41	10.56	11.90	15.06	12.78	14.19	17.41	13.86	15.36	18.81
.150	32.111	9.02	10.19	12.93	10.28	11.47	14.34	10.53	11.84	15.02	12.71	14.16	17.30	13.77	15.29	18.72
.140	37.735	9.00	10.14	12.83	10.23	11.41	14.26	10.50	11.84	14.92	12.66	14.10	17.20	13.75	15.20	18.57
.120	53.778	8.92	10.03	12.64	10.15	11.29	14.04	10.47	11.77	14.74	12.61	13.98	16.95	13.64	15.05	18.33
.100	81.000	8.84	9.91	12.45	10.07	11.20	13.80	10.41	11.67	14.55	12.52	13.83	16.75	13.57	14.94	18.03
.080	132.250	8.78	9.81	12.18	9.96	11.12	13.62	10.36	11.56	14.29	12.41	13.71	16.51	13.51	14.83	17.82
.060	245.444	8.71	9.72	12.01	9.88	11.03	13.38	10.30	11.50	14.05	12.35	13.59	16.29	13.43	14.73	17.51
.050	361.000	8.67	9.66	11.92	9.84	10.97	13.23	10.25	11.46	13.99	12.30	13.53	16.20	13.40	14.65	17.41
.040	576.000	8.64	9.59	11.79	9.89	10.98	13.22	11.08	12.27	14.98	12.27	13.46	16.13	13.33	14.60	17.31
.030	1045.444	8.61	9.55	11.69	9.84	10.93	13.13	11.05	12.18	14.88	12.22	13.41	16.02	13.29	14.52	17.21
.020	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. (CONT.)

		p=11			p=12			p=13			p=14			p=15		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.600	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
.495	1.041	16.95	19.22	24.53	18.53	20.93	25.83	19.32	21.76	27.78	20.72	23.31	29.07	22.32	24.94	30.51
.490	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
.485	1.128	16.84	19.09	24.21	18.47	20.79	25.68	19.26	21.59	27.38	20.60	23.17	28.95	22.20	24.75	30.45
.480	1.174	16.85	19.07	24.19	18.43	20.72	25.58	19.22	21.56	27.30	20.65	23.10	28.84	22.15	24.67	30.27
.475	1.222	16.76	19.05	23.91	18.38	20.63	25.52	19.16	21.55	27.13	20.53	23.11	28.77	22.12	24.67	30.21
.470	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
.460	1.378	16.70	18.95	23.68	18.21	20.45	25.22	19.06	21.36	26.78	20.47	22.89	28.43	21.98	24.53	29.96
.450	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
.440	1.620	16.56	18.77	23.46	18.01	20.29	24.67	18.92	21.17	26.46	20.33	22.66	27.77	21.77	24.24	29.51
.420	1.907	16.39	18.60	23.19	17.84	20.10	24.40	18.74	20.91	26.16	20.23	22.47	27.50	21.60	24.07	29.19
.400	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
.380	2.662	16.16	18.30	22.51	17.48	19.76	23.63	18.43	20.54	25.57	19.84	22.03	26.96	21.35	23.70	28.69
.360	3.160	16.07	18.09	22.23	17.34	19.51	23.34	18.29	20.39	25.31	19.74	21.86	26.61	21.18	23.53	28.48
.350	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
.340	3.768	16.00	17.93	21.95	17.27	19.34	23.03	18.17	20.23	25.03	19.63	21.73	26.31	21.02	23.26	28.10
.320	4.516	15.88	17.78	21.70	17.15	19.21	22.74	18.09	20.09	24.77	19.47	21.53	26.07	20.89	23.06	27.61
.300	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
.280	6.612	15.68	17.42	21.05	16.94	18.78	22.26	17.87	19.78	24.07	19.21	21.14	25.44	20.57	22.71	26.97
.260	8.101	15.56	17.24	20.89	16.82	18.66	22.09	17.75	19.66	23.78	19.11	20.97	25.29	20.44	22.52	26.75
.250	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
.240	10.028	15.47	17.12	20.71	16.75	18.48	21.88	17.65	19.49	23.53	18.95	20.76	24.97	20.35	22.37	26.40
.220	12.570	15.39	17.00	20.40	16.63	18.31	21.69	17.53	19.33	23.28	18.82	20.56	24.60	20.22	22.17	26.21
.200	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
.180	20.753	15.16	16.76	19.89	16.45	18.04	21.34	17.33	19.01	22.77	18.59	20.32	24.24	19.98	21.80	25.64
.160	27.562	15.04	16.58	19.58	16.37	17.90	21.13	17.21	18.86	22.61	18.47	20.19	23.94	19.85	21.61	25.41
.150	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
.140	37.735	14.96	16.39	19.32	16.25	17.81	20.98	17.13	18.73	22.45	18.34	20.04	23.64	19.74	21.46	25.07
.120	53.778	14.84	16.23	19.06	16.15	17.67	20.77	17.05	18.62	22.19	18.21	19.90	23.28	19.61	21.26	24.83
.100	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
.080	132.250	14.65	15.96	18.62	15.90	17.38	20.33	16.84	18.34	21.65	18.00	19.56	22.81	19.33	20.94	24.32
.060	245.444	14.56	15.82	18.37	15.80	17.23	20.01	16.73	18.17	21.34	17.90	19.41	22.54	19.24	20.78	23.95
.050	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
.040	576.000	14.48	15.68	18.17	15.71	17.08	19.77	16.85	18.29	21.15	18.04	19.48	22.49	19.11	20.63	23.62
.030	1045.444	14.42	15.60	18.02	15.65	17.00	19.67	16.79	18.20	21.04	17.98	19.40	22.34	19.06	20.56	23.47
.020	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

		p=16			p=17			p=18			p=19			p=20		
		10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.500	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
.495	1.041	23.41	26.15	32.26	24.76	27.47	33.25	25.77	28.52	34.53	27.15	30.03	36.26	28.34	31.09	37.41
.490	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
.485	1.128	23.29	25.96	31.81	24.54	27.38	32.94	25.69	28.48	34.20	27.02	29.93	36.21	28.17	31.09	37.15
.480	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
.475	1.222	23.18	25.76	31.64	24.49	27.30	32.98	25.57	28.41	34.06	26.96	29.70	36.01	28.06	30.92	36.82
.470	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	36.75
.460	1.378	23.05	25.55	31.26	24.34	27.05	32.57	25.39	28.16	33.81	26.79	29.48	35.52	27.93	30.80	36.46
.450	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
.440	1.620	22.85	25.29	30.94	24.15	26.72	32.25	25.16	27.84	33.20	26.62	29.25	35.09	27.76	30.64	36.23
.420	1.907	22.70	25.01	30.65	23.90	26.51	31.86	24.96	27.53	32.96	26.48	28.94	34.76	27.59	30.32	35.73
.400	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
.380	2.662	22.33	24.69	29.88	23.54	26.04	31.17	24.70	27.00	32.32	26.01	28.50	33.89	27.17	29.81	34.95
.360	3.160	22.18	24.50	29.58	23.40	25.79	30.71	24.59	26.73	31.87	25.86	28.31	33.54	26.98	29.55	34.63
.350	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
.340	3.768	22.04	24.37	29.26	23.26	25.58	30.37	24.46	26.58	31.39	25.68	28.05	33.25	26.80	29.23	34.24
.320	4.516	21.94	24.21	28.90	23.16	25.32	30.06	24.36	26.42	30.99	25.49	27.82	32.88	26.63	29.10	33.73
.300	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
.280	6.612	21.74	23.84	28.34	22.89	24.95	29.45	24.04	26.07	30.53	25.22	27.37	31.89	26.32	28.64	33.10
.260	8.101	21.63	23.70	27.95	22.71	24.76	29.22	23.95	25.93	30.16	25.05	27.24	31.58	26.15	28.41	32.81
.250	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
.240	10.028	21.53	23.56	27.59	22.56	24.60	28.88	23.78	25.78	29.90	24.95	27.09	31.29	26.01	28.28	32.50
.220	12.570	21.40	23.33	27.27	22.42	24.43	28.60	23.65	25.60	29.63	24.84	26.86	30.97	25.92	28.09	32.15
.200	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
.180	20.753	21.18	23.01	26.90	22.17	24.06	28.09	23.41	25.29	29.11	24.54	26.53	30.64	25.68	27.66	31.64
.160	27.562	21.05	22.80	26.54	22.05	23.93	27.80	23.27	25.09	28.81	24.43	26.39	30.35	25.54	27.47	31.32
.150	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
.140	37.735	20.93	22.64	26.32	21.93	23.78	27.55	23.15	24.93	28.50	24.35	26.21	30.17	25.43	27.26	31.12
.120	53.778	20.80	22.56	26.12	21.86	23.66	27.23	23.02	24.74	28.18	24.23	26.05	29.91	25.30	27.13	30.80
.100	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
.080	132.250	20.57	22.21	25.59	21.66	23.33	26.65	22.77	24.44	27.61	23.99	25.66	29.32	25.04	26.77	30.37
.060	245.444	20.44	22.10	25.25	21.55	23.14	26.39	22.67	24.29	27.30	23.88	25.48	28.94	24.87	26.58	30.05
.050	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
.040	576.000	20.31	21.87	25.00	21.40	22.99	26.13	22.52	24.12	27.12	23.73	25.30	28.58	24.76	26.39	29.79
.030	1045.444	20.24	21.76	24.86	21.32	22.90	26.01	22.45	24.01	26.98	23.66	25.20	28.38	24.67	26.26	29.62
.020	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