LAGRANGE MULTIPLIER TESTS FOR PARAMETER INSTABILITY IN NON–LINEAR MODELS

Bruce E. Hansen[1]

University of Rochester

October 1990

[1]Paper presented at the Sixth World Congress of the Econometric Society. I thank Don Andrews and Adrian Pagan for helpful discussions.

ABSTRACT

This paper examines Lagrange multiplier tests for parameter instability for general estimation problems, including maximum likelihood and generalized methods of moments. It is shown that tests constructed for various specifications of the alternative hypothesis have similar forms and sampling behavior. The two main alternatives utilized previously in the literature — based upon (i) a single shift of unknown timing ; and (ii) random walk coefficients — are studied for the first time in a unified framework. Striking similarities are found among the test statistics. All the test statistics considered in this paper are computationally simple, unlike many other tests suggested in the literature. This is because the Lagrange multiplier tests only require estimation under the null hypothesis of no parameter variation.

An asymptotic theory for the test statistics is developed under the null hypothesis. The theory allows for a wide range of estimators and data processes. Complete yet simple proofs of all results are provided. The distributions of the test statistics are non–standard, and can be represented as functions of squared Brownian bridges, or, in some cases, an infinite series of weighted chi–square random variables. Tables of critical values are provided. A Monte Carlo experiment suggests that the tests perform well in small samples.

Keywords: Brownian bridge, CUSUM tests, Lagrange multiplier test, non–linear estimator, parameter instability, structural change.

JEL Classification # 211

Address: Bruce E. Hansen
Department of Economics
University of Rochester
Rochester, NY 14627
(716) 275–7307

1. Introduction

Testing for structural change and parameter instability has a long history in econometrics and statistics, and for good reason. Parameters are frequently reduced forms of unobservables. Although it may be reasonable to believe that these unobservables have been roughly constant over the sample period, it is rarely *a priori* obvious. In many cases it is desirable to subject parameter constancy to a specification test.

The most commonly applied test for parameter change is the sample split test, often referred to as the Chow test (Chow, 1960). The procedure is to split the sample at some predetermined point, and compare the estimates obtained from the pre–split and post–split samples via an F test. We call this a pointwise F test. The simplicity of the test is its strongest attribute, and hence its wide application. On the other hand, the need for a predetermined (exogenous) split point is a severe weakness. If the split is selected conditional on the data (or real–world events), then the conventional distributional theory is invalid and may be quite misleading. If the split point is chosen arbitrarily, for example at the sample midpoint, then the test has low point against structural breaks which occur early or late in the sample period.

There are two solutions to this problem which have received considerable attention in the literature. The first treats the sample split point as an unknown parameter. The likelihood ratio (LR) test for non–constancy is found to be the maximum of all pointwise F tests. The second solution treats the parameter of interest as a random walk, and tests the hypothesis that the variance of the random walk innovations is zero. It appears from the literature and casual inspection that these approaches are distinct and produce different tests. In this paper, we show that in fact these statistics bear striking similarities in construction and performance. We also argue that Lagrange multiplier (LM) tests, rather than LR or Wald tests, are appropriate in tests for parameter instability in non–linear models. The theory is more elegant, and implementation much simpler.

The first approach mentioned above dates back to Quandt (1960). The asymptotic theory was unknown until in a recent flurry of papers simultaneously solved the problem; see Kim and Siegmund (1989), Chu (1989), Andrews (1990b), and Banerjee, Lumsdaine and Stock (1990). This test requires re–estimating the model for each sub–sample. In a linear regression model, this is not particularly costly or difficult. This does not, however, carry over to non–linear estimators. In many contexts, it may be prohibitively costly to re–estimate the model over each sub–sample. It may in fact be nearly impossible in problems which have ill–behaved likelihoods for which numerical convergence is a tricky business.

There is no need, however, to use pointwise F tests when pointwise LM tests are available. These have been studied in Andrews (1990b), but not particularly emphasized. LM test statistics only require one estimation (under the null of no coefficient variation), and therefore are cheap to compute, given that the model has been estimated anyway. The distributional theory is also easier to derive for sequences of LM tests, for there is no need to worry about uniform convergence of partial sample estimators.

The second approach to test for coefficient variation was first proposed by Gardner (1969) as a Bayes test for structural change in the mean of normal random variables. Several researchers later independently proposed the statistic as a test for the presence of random walk coefficients in the linear regression model with independent normal innovations. These were Pagan and Tanaka (1981), Nyblom and Makelainen (1983), and King (1987). The statistic is also discussed in Tanaka (1983) and Nicolls and Pagan (1985). Nyblom and Makelainen (1983) and Leybourne and McCabe (1989) found the asymptotic distribution of the test statistic for location models. Nebeya and Tanaka (1988) used fairly sophisticated mathematics to find the asymptotic distribution for linear regression with deterministically trended regressors and iid errors. A fairly general theory for maximum likelihood estimators is provided by Nyblom (1989).

We extend this literature to encompass general parametric econometric estimators, including maximum likelihood (ML), ordinary linear squares (OLS), generalized least squares (GLS), instrumental variables (IV), and generalized method of moments (GMM). Tests for pure and partial parameter instability are allowed. One interesting finding is that one version of this test is exactly the average of the pointwise LM tests for structural change (which is in linear models the pointwise F test), providing a partial unification of these two literatures.

The rest of this paper is organized as follows. Section 2 introduces the general model and the test statistics. Section 2 develops an asymptotic distribution theory. Central limit theorems are given which are not less restrictive than current results, but with somewhat simpler conditions and proofs. Explicit conditions for satisfaction of the general conditions are given. The limit theory for the stability tests are given. Section 4 reports a Monte Carlo study. The appendix contains the proofs of the theorems.

Regarding notation, we use "$[\cdot]$" to denote "integer part" and "$\Longrightarrow$" to denote weak convergence of probability measures (as defined in Billingsley, 1968).

2. General Theory

2.1 A Unified Framework for Estimation

Consider estimation of an unknown parameter $\theta \in \mathbb{R}^k$ by minimizing a criterion function $C_n(\theta, \hat{\tau})$ with respect to θ. Denote the estimate of θ by $\hat{\theta}$. The random variable $\hat{\tau}$ is some preliminary estimate used to define $\hat{\theta}$. Assume that $C_n(\cdot,\cdot)$ is differentiable with respect to θ so that $\hat{\theta}$ is equivalently described as the solution to the k first order conditions (FOC):

$$0 = \mathrm{FOC}(\hat{\theta},\hat{\tau}) .$$

Quite often[2], the FOC can be written in the form

$$\mathrm{FOC}(\theta, \tau) = Q_n(\theta,\tau)\, \Sigma_{i=1}^n m_i(\theta,\tau) \tag{1}$$

where $Q_n(\cdot,\cdot)$ is a sequence of $k \times q$ $(q \geq k)$ functions of $\{x_1, \cdots, x_n\}$. The variables $m_i(\theta,\tau)$ are functions of $\{x_1, \cdots, x_i\}$. We can write the FOC as

$$0 = \Sigma_{i=1}^n m_{ni}(\hat{\theta},\hat{\tau}) \quad , \quad m_{ni}(\theta,\tau) = Q_n(\theta,\tau)\, m_i(\theta,\tau) .$$

In MLE, $\{m_i\}$ are known as the scores, and their sum constitutes the first–order conditions for the estimator. In non–MLE problems the array $\{m_{ni}\}$ play an identical role. Denote $\hat{m}_i = m_i(\hat{\theta},\hat{\tau})$, $\hat{m}_{ni} = m_{ni}(\hat{\theta},\hat{\tau})$ and $\hat{Q} = Q_n(\hat{\theta},\hat{\tau})$.

It will be useful to define the partial sums of the FOC:

$$S(\pi) = \sum_{i=1}^{[n\pi]} m_i(\theta_o, \tau_o) \tag{2}$$

$$\hat{S}(\pi) = \sum_{i=1}^{[n\pi]} \hat{m}_i$$

[2]See Appendix A for examples of standard econometric estimators.

$$S_n(\pi) = \sum_{i=1}^{[n\pi]} m_{ni}(\theta_o, \tau_o) = Q_n(\theta_o, \tau_o)\, S(\pi)$$

$$\hat{S}_n(\pi) = \sum_{i=1}^{[n\pi]} \hat{m}_{ni} = \hat{Q}\, \hat{S}(\pi)\,,$$

and the variance of $S_n(\pi)$:

(3) $$V(\pi) = E\left[S_n(\pi)S_n(\pi)'\right] .$$

In a correctly specified MLE problem a natural estimator of $V(\pi)$ is

$$\hat{V}(\pi) = \Sigma_{i=1}^{[n\pi]} \frac{\partial^2}{\partial\theta\partial\theta'} \log f_i(\hat{\theta})$$

Similarly, in OLS regression with serially uncorrelated, homoskedastic errors, the choice is

$$\hat{V}(\pi) = \hat{\sigma}_u^2\, \Sigma_{i=1}^{[n\pi]} x_i x_i' .$$

Such choices, however, have limited applicability due to a lack of robustness. An estimate which is robust to heteroskedasticity in all models is simply

(4) $$\hat{V}(\pi) = \Sigma_{i=1}^{[n\pi]}\, \hat{m}_{ni}\hat{m}_{ni}' = \hat{Q}_n\, \Sigma_{i=1}^{[n\pi]}\, \hat{m}_i\hat{m}_i'\, \hat{Q}_n' ,$$

while in the presence of potential autocorrelation as well a good choice is

(5) $$\hat{V}(\pi) = \sum_{a=-M}^{M} w_{aM}\, \Sigma_{i\leq[n\pi]}\, \hat{m}_{ni+a}\hat{m}_{ni}'$$

$$= \hat{Q}_n \sum_{a=-M}^{M} w_{aM}\, \Sigma_{i\leq[n\pi]}\, \hat{m}_{i+a}\hat{m}_i'\, \hat{Q}_n .$$

where w_{aM} is a kernel and M a lag truncation number selected to go to infinity slower than sample size. See Newey and West (1987), and Andrews (1990a) for a detailed discussion of choice of estimators (4)–(5) and conditions under which they are consistent.

2.2 LR and LM tests for structural change of unknown timing

Partition $\theta = (\theta^{1\prime}, \theta^{2\prime})'$. Throughout the paper, superscripts are used to denote partitions conformable with θ. We are interested in testing the constancy of the parameter θ^1 over the sample period, and therefore maintain the constancy of θ^2. Andrews and Fair (1988) call this a test of partial structural change. If there is no θ^2 they termed this a test of pure structural change. Consider the sequence $\{\theta_i^1\}_{i=1}^n$. A simple structural break at time π (actually, the time is $[n\pi]$, but we will refer to the timing of the split as π for convenience) may be parameterized:

$$(6) \qquad \theta_i^1 = \begin{cases} \theta^1 + \beta, & i \leq [n\pi] \\ \theta^1, & i > [n\pi] \end{cases} .$$

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0 .$$

A sensible test of the null of constant coefficients against a structural change at time π is the F test for the restriction $\beta = 0$. Denote this statistic by $F(\pi)$. When π is unknown, a suggestion dating back to Quandt (1960) is to use the test statistic

$$(7) \qquad \text{supF} = \sup_{\pi \in \Pi} F(\pi)$$

where Π is some region in $(0,1)$. This is precisely the LR test of H_0 against H_1 when π is known to lie in Π. It of course reduces to the pointwise F test when Π contains only a single point.

In linear models, use of the supF statistic is sensible and computationally manageable. In non–linear models, however, this test may be computationally burdensome or even impossible. Each pointwise F statistic requires reestimation of the model. Obtaining reasonably accurate convergence for each $\pi \in \Pi$ may be quite challenging in many applications; almost certainly prohibitively costly for a casual specification test.

In the present context, however, it is more convenient to utilize a Lagrange Multiplier (LM) test statistic, since LM tests only require estimation under the null hypothesis, and hence do not require reestimation for each point of sample split. LM and LR test statistics have identical asymptotic distributions under the null hypothesis and sequences of local alternatives, and hence are expected to have similar sampling behavior.

Partition $\hat{m}_i$, $\hat{m}_{ni}$, $S(\pi)$, $S_n(\pi)$, $\hat{S}(\pi)$, $\hat{S}_n(\pi)$ and $\hat{V}(\pi)$ in conformity with θ. For example

$$S_n(\pi) = (S_n^1(\pi)' , S_n^2(\pi)')' .$$

(8) $$\hat{V}(\pi) = \begin{pmatrix} \hat{V}^{11}(\pi) & \hat{V}^{12}(\pi) \\ \hat{V}^{21}(\pi) & \hat{V}^{22}(\pi) \end{pmatrix} .$$

Consider the statistic

(9) $$LM(\pi) = \hat{S}_n^1(\pi)' \left[\hat{V}_*^{11}(\pi)\right]^{-1} \hat{S}_n^1(\pi)$$

where

(10) $$\hat{V}_*(\pi) = \hat{V}(\pi) - \hat{V}(\pi)\hat{V}(1)^{-1}\hat{V}(\pi)$$

is the natural estimator of $E\left[\hat{S}_n(\pi)\hat{S}_n(\pi)'\right]$, and $\hat{V}_*(\pi)$ is partitioned as $\hat{V}(\pi)$ in (8).

For MLE problems and fixed π, $LM(\pi)$ is precisely the pointwise LM statistic, and in non–MLE problems $LM(\pi)$ is an "LM–like" statistic. We will refer to $LM(\pi)$ as the "pointwise LM statistic" for simplicity.

The LM analogue to the supF statistic (7) is

(11) $$\text{supLM} = \sup_{\pi \in \Pi} LM(\pi) .$$

In section three we derive a distributional theory for the supLM statistic.

2.3 LM tests for random parameter variation — General Case

An alternative approach is to treat the parameter variation as random. Then a test for constancy reduces to a zero restriction on the variance of the innovations moving the random parameters. Specifically, consider a parameter array $\{\theta^1_{ni}\}$, and some increasing sequence of sigma–fields $\{\mathscr{G}_{ni}\}$ to which θ^1_{ni} is adapted. Set

$$\Delta_{ni} = \theta^1_{ni} - \theta^1_{ni-1}$$

and assume

$$(12) \qquad E(\Delta_{ni} \mid \mathscr{G}_{ni-1}) = 0 , \qquad E[\Delta_{ni}\Delta_{ni}'] = \delta^2 G_{ni-1} ,$$

for some known array $\{G_{ni}\}$.

This martingale formulation, introduced by Nyblom (1989), allows for substantial flexibility. It allows, for example, θ^1 to be a random walk as specified in many papers; or, alternatively, a single structural break as pointed out by Nyblom.

The parameter θ^1 is constant if the variance of the innovations Δ_{ni} is zero, i.e., if $\delta^2 = 0$. Thus the null and alternative hypotheses are

$$H_0 : \delta^2 = 0 \qquad\qquad H_1 : \delta^2 > 0$$

For maximum likelihood problems, Nyblom (1989) has shown that the following statistic is a good approximation to the LM test of H_0 against H_1:

$$(13) \qquad L = n^{-1} \sum_{i=1}^{n} \left[\sum_{j=i}^{n} \hat{m}^1_j{}'\right] G_{ni-1} \left[\sum_{j=i}^{n} \hat{m}^1_j\right] = n^{-1} \sum_{i=1}^{n-1} \left[\sum_{j=1}^{i} \hat{m}^1_j{}'\right] G_{ni} \left[\sum_{j=1}^{i} \hat{m}^1_j\right]$$

The "LM–like" analogue to this statistic in non–MLE problems is given by the same formula, replacing $\{\hat{m}^1_i\}$ with $\{\hat{m}^1_{ni}\}$. This gives

$$(14) \qquad L = n^{-1} \sum_{i=1}^{n-1} \left[\sum_{j=1}^{i} \hat{m}^1_{nj}{}'\right] G_{ni} \left[\sum_{j=1}^{i} \hat{m}^1_{nj}\right] = \int_0^1 \hat{S}^1_n(\pi)' \, G_n(\pi) \, \hat{S}^1_n(\pi) \, d\pi ,$$

where $G_n(\pi) = G_{[n\pi]}$.

2.4 LM tests for random parameter variation — Constant Hazard

The test statistic L given in (14) depends upon the covariance array $\{G_{ni}\}$. This array needs to be chosen carefully if the test is to have a convenient limiting distribution and have power against alternatives of interest. Since G_{ni} represents the uncertainty in the parameter's stability in the i'th period, we may think of G_{ni} as the hazard associated with parameter instability. The specification that $G_{ni} = G$ be constant across i — which we may call constant hazard of instability — was implicitly assumed in Pagan and Tanaka (1981), Nyblom and Makelainen (1983) and King (1987). Nyblom (1989) notes the general model but emphasizes the constant hazard case. Then the statistic simplifies to

$$L_C = \operatorname{tr}\left\{n^{-1}\left[\sum_{i=1}^{n-1}\hat{S}_n^1(\tfrac{i}{n})\hat{S}_n^1(\tfrac{i}{n})'\right] G\right\} = \operatorname{tr}\left\{\int_0^1\hat{S}_n^1(\pi)\hat{S}_n^1(\pi)'d\pi\, G\right\}$$

A particularly useful choice for G is $\left[\hat{V}^{11}(1)\right]^{-1}$. Reasons for this choice are articulated clearly in Nyblom (1989). Primarily, this renders the asymptotic distribution of L_C invariant to nuisance parameters. The statistic then becomes

$$L_C = \operatorname{tr}\left\{\int_0^1\hat{S}_n^1\hat{S}_n^{1\prime}\ \hat{V}^{11}(1)^{-1}\right\} . \tag{15}$$

2.5 LM tests for random parameter variation — Weighted Hazard

The intuitive argument for L_C is that the statistic is constructed assuming that the hazard of parameter instability is constant across observations. Examining (14), however, we see that this results in a test statistic which places unequal weights across $\hat{S}_n^1(\pi)'\ G\ \hat{S}_n^1(\pi)$. This is because the expectation of this variable varies over π. It may therefore be difficult to detect instability which arises near the beginning or end of the sample. It seems reasonable to consider tests which select $G_n(\pi) = \hat{V}_*^{11}(\pi)^{-1}$, where $\hat{V}_*(\pi)$ (given in (10)) is a natural estimator of $V_*(\pi) = E[\hat{S}_n(\pi)\hat{S}_n(\pi)]$. This specification weights the early and late observations more heavily, increasing the probability of detecting parameter variation which occurs away from the middle of the sample. This gives a new test statistic

$$(16) \qquad L_W = n^{-1}\sum_{i=1}^{n-1} \hat{S}_n^1(\tfrac{i}{n})' \left[\hat{V}_*^{11}(\tfrac{i}{n})\right]^{-1} \hat{S}_n^1(\tfrac{i}{n}) = \int_0^1 \hat{S}_n^1(\pi)' \left[\hat{V}_*^{11}(\pi)\right]^{-1} \hat{S}_n^1(\pi)\, d\pi .$$

A connection arises with our earlier analysis of tests for structural change of unknown timing. A comparison with (9) reveals that

$$L_W = \int_0^1 LM(\pi)\, d\pi \; .$$

That is, this LM statistic for random coefficient variation is the exactly the *average* pointwise LM statistic for structural change. In contrast, the LM statistic for structural change of unknown timing (10) is

$$\text{supLM} = \sup_{\pi \in \Pi} LM(\pi) \, .$$

Considering $\{LM(\pi)\}$ as an element in the function space $C[0,1]$, L_W can be viewed as the L^1–norm of $\{LM(\pi)\}$, while supLM is (approximately) the L^∞–norm of $\{LM(\pi)\}$. Hence the only difference between the statistics is the choice of norm. Although the two tests were derived via different methods, we see that they share a common format.

3. Inference

3.1 Central Limit Theory

Define $\mathcal{T}$ to be some neighborhood of τ_o , $\mathcal{N}$ to be some neighborhood of (θ_o,τ_o) , $m_{i\theta}(\theta,\tau) = \partial/\partial\theta'\ m_i(\theta,\tau)$, and $m_{i\tau}(\theta,\tau) = \partial/\partial\tau'\ m_i(\theta,\tau)$.

A1. (a) θ_0 lies in the interior of Θ , a bounded subset of $\mathbb{R}^k$;

(b) $\hat{\theta} \longrightarrow_p \theta_o$;

(c) $\sqrt{n}(\hat{\tau}-\tau) = O_p(1)$;

(d) $m_\theta(\theta,\tau) = \lim n^{-1}\Sigma_1^n E\, m_{i\theta}(\theta,\tau)$ exists uniformly over $\mathcal{N}$ and is continuous in $\mathcal{N}$;

(e) $m_\tau(\tau) = \lim n^{-1}\Sigma_1^n E\, m_{i\tau}(\theta_o,\tau)$ exists uniformly over $\mathcal{T}$ and is continuous in $\mathcal{T}$;

(f) $m_\tau(\tau_o) = 0$;

(g) $Q(\theta,\tau) = \text{plim}\ Q_n(\theta,\tau)$ exists uniformly over $\mathcal{N}$ and is continuous in $\mathcal{N}$;

(h) $\mathcal{J} = \lim_{n\to\infty} n^{-1}E\left[\Sigma_1^n m_i(\theta_o,\tau_o)\ \Sigma_1^n m_i(\theta_o,\tau_o)'\right]$ exists and is finite.

(i) $Q(\theta_o,\tau_o)m_\theta(\theta_o,\tau_o) > 0$.

Set $M = m_\theta(\theta_o,\tau_o)$, $Q = Q(\theta_o,\tau_o)$, $J = QM$, and $V = Q\mathcal{J}Q'$.

We use the following smoothness condition:

Definition. $\{A_n(\theta) : n \geq 1\}$ is *stochastically equicontinuous* on Θ if: $\forall\epsilon > 0\ \exists\delta > 0$ such that

$$\limsup_{n\uparrow\infty} P\left[\sup_{\theta\in\Theta}\ \sup_{\theta'\in B(\theta,\delta)} |A_n(\theta') - A_n(\theta)| > \epsilon\right] < \epsilon ,$$

where $B(\theta,\delta)$ denotes a ball of radius δ around θ .

<u>A2</u> (a) $n^{-1/2}S(1) \rightarrow_d N(0,\mathscr{J})$;

(b) For all $(\theta,\tau) \in \mathscr{N}$, $n^{-1}\Sigma_1^n m_{i\theta}(\theta,\tau) \rightarrow_p m_\theta(\theta,\tau)$;

(c) $\{n^{-1}\Sigma_1^n m_{i\theta}(\theta,\tau)\}$ is stochastically equicontinuous on $\mathscr{N}$;

(d) For all $\tau \in \mathscr{T}$, $n^{-1}\Sigma_1^n m_{i\tau}(\theta_o,\tau) \rightarrow_p m_\tau(\tau)$;

(e) $\{n^{-1}\Sigma_1^n m_{i\tau}(\theta_o,\tau)\}$ is stochastically equicontinuous on $\mathscr{T}$.

<u>Theorem 1</u>. A1 *and* A2 *imply that* $\sqrt{n}\,(\hat{\theta} - \theta_o) \rightarrow_d N(0\,,\,J^{-1}VJ^{-1\prime})$.

<u>Remark 1</u>. Theorem 1 covers the estimators considered in Gallant (1987), Gallant and White (1988) and Andrews and Fair (1988), but in a somewhat simpler fashion.

<u>Remark 2</u>. A1 (f) is critical. This assumption allows $\hat{\tau}$ to enter as a preliminary nuisance estimate, without affecting the limiting distribution of $\hat{\theta}$. This of course does not hold for all two–step estimators, although it holds for the examples given in Appendix A.

<u>Remark 3</u>. A2 (c)(e) impose stochastic equicontinuity, which is necessary for the application of uniform laws of large numbers. This is essentially a smoothness condition upon the functions. See Andrews (1990c) for a detailed discussion.

<u>Remark 4</u>. The CLT and WLLNs in A2 can be found from more primitive assumptions. Consider

<u>Condition SE</u>: $\{x_i\}$ is stationary and ergodic.

<u>Condition MIX</u>: $\{x_i\}$ is α–mixing (strong mixing), with mixing coefficients $\{\alpha_j\}$, and $m_i(x_1,\ldots,x_i;\theta,\tau) = m_i(x_{i-L},\ldots,x_i;\theta,\tau)$ for finite L .

Set $m_i = m_i(\theta_0, \tau_0)$. A2 (a) is implied by any of the following conditions:

(i) SE holds, and $\{m_i\}$ is a square integrable martingale difference sequence. (Billingsley, 1968, p. 206).

(ii) SE and MIX hold, and for some $\beta \geq 1$, $E|m_i' m_i|^{\beta} < \infty$ and $\Sigma_1^{\infty} \alpha_j^{1-1/2\beta} < \infty$. (Hall and Heyde, 1980, chapter 5).

(iii) MIX holds, and for some $\beta > 1$, $\sup_i E|m_i' m_i|^{\beta} < \infty$ and $\Sigma_1^{\infty} \alpha_j^{1-1/\beta} < \infty$. (Herrndorf, 1984).

Remark 5. The pointwise weak laws in A2(b) hold under any of the following conditions for all $(\theta, \tau) \in \mathcal{N}$:

(i) SE holds, and $E|m_{i\theta}(\theta,\tau)| < \infty$. (The Ergodic Theorem; see, for example, Stout, 1974, p. 181).

(ii) $m_{i\theta}(\theta,\tau)$ is uniformly integrable and an L^1–mixingale. (Andrews, 1988).

(iii) MIX holds, and $m_{i\theta}(\theta,\tau)$ is uniformly integrable. (Andrews, 1988).

Remark 6. A2(d) holds under the conditions of remark 6 for $m_{i\tau}(\theta_0, \tau)$, over $\mathcal{T}$.

3.2 Asymptotic Distribution of the Test Statistics

We need the following strengthening of stochastic equicontinuity, introduced by Andrews (1990c):

Definition. $\{A_n(\theta) : n \geq 1\}$ is *strongly stochastically equicontinuous* on Θ if $\{ \sup_{i \geq n} |A_i(\theta)| : n \geq 1 \}$ is stochastically equicontinuous on Θ .

We now strengthen assumption A2:

A3. (a) $n^{-1/2}S(\pi) \Rightarrow B(\pi) \equiv BM(J)$;

(b) For all $(\theta,\tau) \in \mathcal{N}$, $n^{-1}\Sigma_1^n m_{i\theta}(\theta,\tau) \rightarrow_{a.s} m_\theta(\theta,\tau)$;

(c) $\{n^{-1}\Sigma_1^n m_{i\theta}(\theta,\tau)\}$ is strongly stochastically equicontinuous on $\mathcal{N}$;

(d) For all $\tau \in \mathcal{T}$, $n^{-1}\Sigma_1^n m_{i\tau}(\theta_o,\tau) \rightarrow_{a.s.} m_\tau(\tau)$;

(e) $\{n^{-1}\Sigma_1^n m_{i\tau}(\theta_o,\tau)\}$ is strongly stochastically equicontinuous on $\mathcal{T}$;

The symbol BM(J) in A3(a) denotes a vector Brownian motion with covariance matrix J .

We first give a useful pair of lemmas.

Lemma 1. *If* $n^{-1}\Sigma_1^n X_i \rightarrow_{a.s.} 0$, *then* $\sup_{0\leq\pi\leq 1} |n^{-1}\Sigma_{i=1}^{[n\pi]} X_i| \rightarrow_p 0$.

Lemma 2. *If* $X_i(\varphi)$ *is indexed by* $\varphi \in \psi$, *and* $\sup_{\varphi\in\psi} |n^{-1}\Sigma_{i=1}^n X_i(\varphi)| \rightarrow_{a.s.} 0$, *then*

$$\sup_{0\leq\pi\leq 1} \sup_{\varphi\in\psi} |n^{-1}\Sigma_{i=1}^{[n\pi]} X_i(\varphi)| \rightarrow_p 0 .$$

Remark 7. Note that if $n^{-1}\Sigma_1^n X_i \rightarrow_{a.s.} \mu$, Lemma 1 implies that

$$n^{-1}\Sigma_{i=1}^{[n\pi]} X_i \rightarrow_p r\mu, \text{ uniformly in } r .$$

We now derive the distribution of the partial sum of first order conditions.

Theorem 2. A1 *and* A3 *imply that* $n^{-1/2}\hat{S}_n(\pi) \Rightarrow V^{1/2}\bar{W}(\pi)$

where $\bar{W}(\pi) = W(\pi) - \pi W(1)$, $W(\pi) \equiv BM(I_k)$;

This allows us to find the asymptotic distribution of L_C .

A4. $n^{-1}\hat{V}(1) \rightarrow_p V$.

Corollary 1. A1, A3 *and* A4 *imply that* $L_C \Rightarrow \int_0^1 \bar{W}(\pi)'\bar{W}(\pi)\,d\pi = \int_0^1 \bar{W}'\bar{W}$.

Remark 8. Corollary 1 covers most parametric econometric estimators and thus substantially generalizes the results in Nyblom (1989) who examined correctly specified MLE models. The limit distribution $\int_0^1 \bar{W}'\bar{W}$ is non–standard. When $k_1 = 1$, this is the limiting distribution of the Von Mises goodness of fit statistic, which has been examined thoroughly in Anderson and Darling (1952). One representation of this distribution is

$$\int_0^1 \bar{W}'\bar{W} = \sum_{j=1}^{\infty} (\pi j)^{-2} \chi_j^2(k_1) \tag{17}$$

where the $\chi_j^2(k_1)$ are iid chi–squares with k_1 degrees of freedom. An expansion of the distribution function is given in Nyblom (1989). Table 1 provides tabulated critical values for $k_1 = 1, \dots, 20$.

Remark 9. The strong stochastic equicontinuity conditions A3(c)(e) is necessary for application of a uniform strong law, which is necessary for the weak convergence result.

Remark 10. The invariance principle A3(a) is implied by either condition (i) or (ii) in remark 4.

Remark 11. The pointwise strong laws in A3(b) hold under any of the following conditions for all $(\theta,\tau) \in \mathcal{N}$:

(i) SE holds, and $E|m_{i\theta}(\theta,\tau)| < \infty$. (The Ergodic Theorem).

(ii) $\sup_i E|m_{i\theta}(\theta,\tau)|^q < \infty$ for some $q > 1$, and $m_{i\theta}(\theta,\tau)$ is an L^q–mixingale with summable mixingale coefficients. (Hansen, 1990, corollary 2).

(iii) MIX holds. $\sup_i E|m_{i\theta}(\theta,\tau)|^q < \infty$ for some $q > 1$, $\Sigma_1^\infty \alpha_m^{1-1/2r} < \infty$ for some $r > q$. (Hansen, 1990, corollary 3).

Remark 12. A3(d) holds under the conditions of remark 10 for $m_{i\tau}(\theta_0,\tau)$ over $\mathcal{T}$.

A5. $$n^{-1}\hat{V}(1) \rightarrow_{a.s.} V .$$

Theorem 3. *Under* A1, A3, A5, *and for* Π *a convex subset of* (0,1),

$$LM(\pi) \Rightarrow Q_{k_1}(\pi) = \bar{W}(\pi)'\bar{W}(\pi)/[\pi(1-\pi)] \quad \text{on } \pi \in \Pi .$$

Corollary 2. *Under* A1, A3, A5, *and for* Π *a convex subset of* (0,1),

$$\text{supLM} = \sup_{\pi\in\Pi} LM(\pi) \Rightarrow \sup_{\pi\in\Pi} Q_{k_1}(\pi) .$$

Corollary 3. *Under* A1, A3 *and* A5, $L_W = \int_0^1 LM(\pi)d\pi \Rightarrow \int_0^1 Q_{k_1}(\pi)\, d\pi$.

Remark 13. The process $Q_{k_1}(\pi)$ is known as a squared standardized tied–down Bessel process of order k_1. For fixed π, $Q_{k_1}(\pi) \equiv \chi^2(k_1)$. This process is completely parameterized by k_1, the dimension of θ^1. Therefore the distributions in the corollaries are completely parameterized by k_1 as well, facilitating tabulation of critical values.

Remark 14. A result similar to Corollary 2 appears in Andrews (1990b, Theorem 4(b)). The conditions given in A1, A3 and A5 are somewhat simpler (and the proof is substantially simpler) than those of Andrews. This is because we focus attention on the LM statistic (9), and thus are not required to guarantee uniform convergence of partial sample estimators. As noted in Anderson and Darling (1952) and Andrews (1990b), the restriction of Π to a convex subset of (0,1) is essential, for $Q_k(\pi)$ is ill–behaved at $\pi = 0$ and $\pi = 1$. Andrews has suggested the informal rule $\Pi = [.15 , .85]$, and has tabulated critical values of the distribution for this choice.

Remark 15. The average LM statistic L_W does not require any trimming as does supLM. This is because the aberrant behavior of $Q_k(\pi)$ is confined to the endpoints, and is smoothed out by the integral.

Remark 16. The statistic limit distribution in Corollary 3 is equivalently

$$\int_0^1 Q_{k_1} = \sum_{j=1}^{k_1} \int_0^1 Q_j ,$$

where the $\int_0^1 Q_j$ are independently distributed as $\int_0^1 Q_1(\pi)d\pi$. The latter has been examined in Anderson and Darling (1952, pp. 202–204), yielding the representation

$$(18) \qquad \int_0^1 Q_{k_1} = \sum_{j=1}^{\infty} [j(j+1)]^{-1}\chi_j^2(k_1)$$

where $\chi_j^2(k_1)$ are iid chi–squares with k_1 degrees of freedom. Table 2 provides tabulated critical values for $k_1 = 1 , \dots , 20$.

3. Monte Carlo Evidence

This section presents evidence from Monte Carlo experiments regarding size and power of the test statistics. A simple linear regression with iid normal errors is used, identical to that in Andrews (1990b). Three test statistics are compared: supLM (= supF), L_C, and L_W. The model used is

$$y_i = x_i'\beta_i + u_i , \quad x_i = (1 \ \ (-1)^i)' , \quad u_i \text{ iid } N(0,1) , \ i = 1,\dots,n .$$

Table 3 reports rejection frequencies of the three tests using the asymptotic 1%, 5%, and 10% significance levels and sample sizes of $n = 30$, 60 , and 120 . 20,000 replications are used. Regardless of sample size, L_C has virtually no size distortion. SupLM over–rejects for $n = 30$, but has good size for $n = 60$ and 120 . L_W displays strong over–rejection for small samples, diminishing with sample size.

To compare power properties, we consider both one–time shifts in β_i :

$$\beta_i = \begin{cases} (0\ ,\ 0)' & ,\ i \le [n\pi^*] \\ \gamma & \ \ i > [n\pi^*] \end{cases}$$

and β_i following a random walk :

$$\beta_i = \beta_i + \gamma u_i \quad , \qquad u_i \text{ iid } N(0, 1/n)\ .$$

The test statistics are invariant to the angle between γ and $\frac{1}{n}\Sigma_1^n x_i = (1\ 0)'$. We vary (1) the sample size $n = 30$, 60 , and 120 ; (2) the magnitude $\|\gamma\| = b/\sqrt{n}$; and (3) the time of shift $\pi^* = .1$, .3 , and .5 . For the random walk alternative (RW) we vary the sample size and magnitude as for the one–time shift alternative.

We report size–corrected power (percentage rejections in 5000 trials) in table 4. Critical values were calculated from the simulated finite–sample null distributions with 20,000 replications. Against single shifts in the tail of the sample ($\pi^* = .1$) , supLM has greater power for $n = 30$, while L_W has greater power for $n = 60$ and 120 . For single shifts at the sample mid–point ($\pi^* = .5$) the statistics are ranked: L_C , L_W , supLM. For $\pi^* = .3$ and the random walk alternative, the three statistics have quite similar power.

We can summarize our findings as follows. All three statistics are computationally feasible, with decent size and power properties. L_C is the easiest and cheapest to compute, has the least size distortion, and best power against mid–point shifts, yet has lower power against tail shifts, especially in small samples. L_W has the best power against tail shifts for larger samples, yet has the worst size distortion. SupLM has the best power against tail shifts for $n = 30$, and has moderate size distortion. It appears that this evidence is not sufficient to discriminate between these statistics.

Appendix A: Examples

OLS: θ minimizes $\Sigma_1^n(y_i - x_i'\theta)^2$;

$$m_i(\theta,\tau) = m_{ni}(\theta,\tau) = x_i(y_i - x_i'\theta)$$

$$Q_n(\theta,\tau) = I_k$$

GLS: θ minimizes $\Sigma_1^n[(y_i - x_i'\theta)/\sigma(x_i,\hat{\tau})]^2$;

$$m_i(\theta,\tau) = m_{ni}(\theta,\tau) = x_i(y_i - x_i'\theta)/\sigma^2(x_i,\tau)$$

$$Q_n(\theta,\tau) = I_k$$

MLE θ minimizes $n^{-1}\Sigma_1^n \log f_i(\theta)$;

$$m_i(\theta,\tau) = m_{ni}(\theta,\tau) = \frac{\partial}{\partial\theta}\log f_i(\theta)$$

$$Q_n(\theta,\tau) = I_k$$

IV θ minimizes $\left[\Sigma_1^n(y_i - x_i'\theta)z_i'\right]\left[\Sigma_1^n z_i z_i'\right]^{-1}\left[\Sigma_1^n z_i(y_i - x_i'\theta)\right]$;

$$m_i(\theta,\tau) = z_i(y_i - x_i'\theta)$$

$$m_{ni}(\theta,\tau) = \left[\Sigma_1^n x_i z_i'\right]\left[\Sigma_1^n z_i z_i'\right]^{-1} z_i(y_i - x_i'\theta)$$

$$Q_n(\theta,\tau) = Q_n(\tau) = \left[\Sigma_1^n x_i z_i'\right]\left[\Sigma_1^n z_i z_i'\right]^{-1} .$$

GMM θ minimizes $\Sigma_{i=1}^n g_i(\theta)'\,\Omega_n^{-1}\,\Sigma_{i=1}^n g_i(\theta)$;

$$m_i(\theta,\tau) = g_i(\theta)$$

$$m_{ni}(\theta,\tau) = \left[\Sigma_{i=1}^n \frac{\partial}{\partial\theta} g_i(\theta)\right]'\,\Omega_n^{-1} g_i(\theta)$$

$$Q_n(\theta,\tau) = \left[\Sigma_{i=1}^n \frac{\partial}{\partial\theta} g_i(\theta)\right]'\,\Omega_n^{-1} .$$

Appendix B. Proofs of the Theorems

Proof of Theorem 1. The first order conditions for $\hat\theta$ are

(B1) $$0 = Q_n(\hat\theta,\hat\tau)\hat S(1) = n^{-1/2}Q_n(\hat\theta,\hat\tau)\,\Sigma_{i=1}^n m_i(\hat\theta,\hat\tau)\,.$$

By A1(b)(c)(g),

(B2) $$\hat Q_n = Q_n(\hat\theta,\hat\tau) \rightarrow_p Q_n(\theta_o,\tau_o) = Q\,.$$

Here and in the sequel we use superscripts to denote elements of a vector or rows of a matrix. For example, m_i^a denotes the a'th element of the vector m_i, and $m_{i\theta}^a$ denotes the a'th row of $m_{i\theta}$.

Element by element mean value expansions of $\Sigma_{i=1}^n m_i(\hat\theta,\hat\tau)$ about θ_o give $\forall$a

(B3) $$n^{-1/2}\Sigma_{i=1}^n m_i^a(\hat\theta,\hat\tau) = n^{-1/2}\Sigma_{i=1}^n m_i^a(\theta_o,\hat\tau) + n^{-1}\Sigma_{i=1}^n m_{i\theta}^a(\theta^*,\hat\tau)\sqrt{n}(\hat\theta-\theta_o)$$

where θ^* is a random variable on a line segment joining $\hat\theta$ and θ_o. Below we show

(B4) $$n^{-1}\Sigma_{i=1}^n m_{i\theta}^a(\theta^*,\hat\tau) = m_\theta^a(\theta_o,\tau_o) + o_p(1)$$

and

(B5) $$n^{-1/2}\Sigma_{i=1}^n m_i(\theta_o,\hat\tau) \rightarrow_d N(0\,,\mathscr{I})\,.$$

(B1)–(B5) combine to give

$$0 = \left[Q + o_p(1)\right]\left[N(0\,,\mathscr{I}) + \left[m_\theta(\theta_o,\tau_o) + o_p(1)\right]\sqrt{n}(\hat\theta-\theta) + o_p(1)\right]$$

Since $QM > 0$ by A1(i) we have

$$\sqrt{n}(\hat\theta-\theta_o) \rightarrow_d -(QM)^{-1}Q\,N(0\,,\mathscr{I}) \equiv J^{-1}N(0\,,V)\,.$$

To show (B4), take each element for $b = 1, \ldots, k$:

$$|n^{-1}\Sigma_{i=1}^n m_{i\theta}^{ab}(\theta^*,\hat\tau) - m_\theta^{ab}(\theta_o,\tau_o)|$$

$$\le |n^{-1}\Sigma_{i=1}^n m_{i\theta}^{ab}(\theta^*,\hat\tau) - m_\theta^{ab}(\theta^*,\hat\tau)| + |m_\theta^{ab}(\theta^*,\hat\tau) - m_\theta^{ab}(\theta_o,\tau_o)|$$

$$\leq \sup_{(\theta,\tau)\in\mathcal{N}} |n^{-1}\Sigma_{i=1}^n m_{i\theta}^{ab}(\theta,\tau) - m_\theta^{ab}(\theta,\tau)| + o_p(1) \rightarrow_p 0 .$$

The second inequality uses A1(d)(b)(c). The final limit is a uniform weak law of large numbers (U–WLLN), which holds under A2(b)(c) by Theorem 1 of Andrews (1990c).

To show (B5), take the element by element expansions

$$n^{-1/2}\Sigma_{i=1}^n m_i^a(\theta_o,\hat\tau) = n^{-1/2}\Sigma_{i=1}^n m_i^a(\theta_o,\tau_o) + n^{-1}\Sigma_{i=1}^n m_{i\tau}^a(\theta_o,\tau^*)\sqrt{n}(\hat\tau-\tau_o)$$

where τ^* is a random variable on a line segment joining $\hat\tau$ and τ_o . Now for each b ,

$$|n^{-1}\Sigma_{i=1}^n m_{i\tau}^{ab}(\theta_o,\tau^*)| \leq |n^{-1}\Sigma_{i=1}^n m_{i\tau}^{ab}(\theta_o,\tau^*) - m_\tau^{ab}(\tau^*)| + |m_\tau^{ab}(\tau^*) - m_\tau^{ab}(\tau_o)|$$

$$\leq \sup_{\tau\in\mathcal{T}} |n^{-1}\Sigma_{i=1}^n m_{i\tau}^{ab}(\theta_o,\tau) - m_\tau^{ab}(\tau)| + o_p(1) \rightarrow_p 0$$

by A1(e)(c) and again applying Theorem 1 of Andrews (1990c) under A2(d)(e). Thus

$$n^{-1/2}\Sigma_{i=1}^n m_i(\theta_o,\hat\tau) = n^{-1/2}\Sigma_{i=1}^n m_i(\theta_o,\tau_o) + o_p(1) \rightarrow_d N(0, \mathcal{J})$$

by A2(a). This completes the proof. □

Proof of Lemmas 1 and 2. Lemma 1 follows from Lemma 2 so we prove the latter.

Fix $\epsilon,\eta > 0$. Set $\bar X_i = \sup_{\varphi\in\psi} |i^{-1}\Sigma_{j=1}^i X_j(\varphi)|$, and $X^* = \max_{i\geq 1} \bar X_i$. Note that $X^* < \infty$ (a.s.) by strong convergence. Choose $\delta > 0$ so that $P\{X^* > \epsilon/\delta\} \leq \eta/2$. Now

$$P\{\max_{1\leq i\leq n} \sup_{\varphi\in\psi} |n^{-1}\Sigma_1^i X_j(\varphi)| > 2\epsilon\}$$

$$\leq P\{\max_{1\leq i\leq [n\delta]} \sup_{\varphi\in\psi} |n^{-1}\Sigma_1^i X_j(\varphi)| > \epsilon\} + P\{\max_{[n\delta]\leq i\leq n} \sup_{\varphi\in\psi} |n^{-1}\Sigma_1^i X_j(\varphi)| > \epsilon\} .$$

The first term on the right equals

$$P\{\max_{1\leq i\leq [n\delta]} \tfrac{i}{n}\bar X_i > \epsilon\} = P\{\delta X^* > \epsilon\} \leq \eta/2 ,$$

and the second term is bounded by

$$P\{\max_{[n\delta]\leq i\leq n} \bar{X}_i > \epsilon\} \leq \eta/2$$

for n sufficiently large by Andrews (1990b, lemma A–1). □

Proof of Theorem 2. Element by element expansions give

(B6) $$n^{-1/2}\hat{S}^a(\pi) = n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i^a(\hat{\theta},\hat{\tau})$$

$$= n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i^a(\theta_o,\hat{\tau}) + n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\theta}^a(\theta_\pi^*,\hat{\tau})\sqrt{n}(\hat{\theta}-\theta_o) .$$

Note that the random variable θ_π^* (which lies on the line segment joining θ_o with $\hat{\theta}$) may depend upon π. Below we show

(B7) $$n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\theta}^a(\theta_\pi^*,\hat{\tau}) \rightarrow_p \pi m_\theta^a(\theta_o,\tau_o)$$, uniformly in π

and

(B8) $$n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i(\theta_o,\hat{\tau}) \Rightarrow B(\pi) \equiv BM(\mathcal{J}).$$

A calculation similar to that in the proof of Theorem 1 gives

(B9) $$\sqrt{n}(\hat{\theta}-\theta_o) \Rightarrow -J^{-1}QB(1) .$$

Combining (B6)–(B9), we find

(B10) $$n^{-1/2}\hat{S}_n(\pi) = Q(\hat{\theta},\hat{\tau})\, n^{-1/2}\hat{S}(\pi) \Rightarrow Q\left[B(\pi) - \pi MJ^{-1}QB(1)\right]$$

$$= QB(\pi) - \pi QB(1) = B_Q(\pi) - B_Q(1) = V^{1/2}\left[W(\pi) - \pi W(1)\right] ,$$

where $B_Q(\pi) = QB(\pi) \equiv BM(Q\mathcal{J}Q')= BM(V)$, and $W(\pi) = V^{-1/2}B_Q(\pi) \equiv BM(I_k)$.

To show (B7), take the b'th element,

$$\sup_{\pi\in[0,1]} |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\theta}^{ab}(\theta_\pi^*,\hat{\tau}) - \pi m_\theta^{ab}(\theta_o,\tau_o)|$$

$$\leq \sup_\pi |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\theta}^{ab}(\theta_\pi^*,\hat{\tau}) - \pi m_\theta^{ab}(\theta_\pi^*,\hat{\tau})|$$

$$+ \sup_\pi |\pi m_\theta^{ab}(\theta_\pi^*,\hat{\tau}) - \pi m_\theta^{ab}(\theta_o,\tau_o)|$$

$$\leq \sup_{\pi} \sup_{(\theta,\tau)\in\mathcal{N}} |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\theta}^{ab}(\theta,\tau) - \pi m_{\theta}^{ab}(\theta,\tau)| + \sup_{\pi} |m_{\theta}^{ab}(\theta_{\pi}^{*},\hat{\tau}) - \pi m_{\theta}^{ab}(\theta_o,\tau_o)| .$$

The first term is $o_p(1)$ by Lemma 2 and A3(b)(c). The second term is $o_p(1)$ by A1(d)(b)(c).

To show (B8), take the expansion

$$n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i^a(\theta_o,\hat{\tau}) = n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i^a(\theta_o,\tau_o) + n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\tau}^a(\theta_o,\tau_{\pi}^{*})\sqrt{n}(\hat{\tau}-\tau_o) .$$

The b'th element of $n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\tau}^a(\theta_o,\tau_{\pi}^{*})$ satisfies

$$\begin{aligned} \sup_{\pi} |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\tau}^{ab}(\theta_o,\tau_{\pi}^{*})| &\leq \sup_{\pi} |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\tau}^{ab}(\theta_o,\tau_{\pi}^{*}) - \pi m_{\tau}^{ab}(\theta_o,\tau_{\pi}^{*})| \\ &\quad + \sup_{\pi} |\pi m_{\tau}^{ab}(\theta_o,\tau_{\pi}^{*})| \\ &\leq \sup_{\pi} \sup_{\tau\in\mathcal{T}} |n^{-1}\Sigma_{i=1}^{[n\pi]} m_{i\tau}^{ab}(\theta_o,\tau) - \pi m_{\tau}^{ab}(\theta_o,\tau)| + o_p(1) = o_p(1) \end{aligned}$$

by similar arguments and A1(f). Thus

$$n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i(\theta_o,\hat{\tau}) = n^{-1/2}\Sigma_{i=1}^{[n\pi]} m_i(\theta_o,\tau_o) + o_p(1) \Rightarrow B(\pi)$$

by A3(a). This completes the proof. □

Proof of Corollary 1. Set $e = (I_{k_1}\ 0)'$. Then by Theorem 2

$$n^{-1/2}\hat{S}_{1n}(\pi) = n^{-1/2}e'\hat{S}_n(\pi) \Rightarrow (e'Ve)^{1/2}\bar{W}(\pi) = V_{11}^{1/2}\bar{W}(\pi) .$$

Combined with A4 we find

$$\begin{aligned} L_C &= \int_0^1 \hat{S}_{1n}(\pi)'\hat{V}_{11}(1)^{-1}\hat{S}_{1n}(\pi)\, d\pi \\ &\Rightarrow \int_0^1 \bar{W}(\pi)'V_{11}^{1/2}(e'Ve)^{-1}V_{11}^{1/2}\bar{W}(\pi)\, d\pi = \int_0^1 \bar{W}(\pi)'\bar{W}(\pi)\, d\pi . \end{aligned}$$

□

Proof of Theorem 3. By A5, and Lemma 1,

$$n^{-1}\hat{V}(\pi) \rightarrow_p \pi V \text{ , uniformly in } \pi \text{ ,} \quad 0 \le \pi \le 1 \text{ .}$$

Similarly,

$$n^{-1}\hat{V}^*(\pi) \rightarrow_p (\pi V) - (\pi V)V^{-1}(\pi V) = [\pi(1-\pi)]\, V \text{ , uniformly in } \pi \text{ .}$$

Thus by the continuous mapping theorem,

$$LM(\pi) = \hat{S}_{1n}(\pi)' \left[\hat{V}^*_{11}(\pi)\right]^{-1} \hat{S}_{1n}(\pi)$$

$$\Rightarrow \ ^{-}(\pi)' V_{11}^{1/2} \left[e'[\pi(1-\pi)]Ve\right]^{-1} V_{11}^{1/2} \bar{W}(\pi) = \bar{W}(\pi)'\bar{W}(\pi)/[\pi(1-\pi)]$$

in any region for which $1/[\pi(1-\pi)]$ is continuous. □

Proof of Corollary 2. Follows from Theorem 3 , the continuous mapping theorem, and the fact that the supremum is a continuous map. □

Proof of Corollary 3. From Theorem 3, for any $\epsilon > 0$,

$$\int_{\epsilon}^{1-\epsilon} LM(\pi)\, d\pi \Rightarrow \int_{\epsilon}^{1-\epsilon} Q_{k_1}(\pi)\, d\pi \text{ .}$$

The result follows if the random variable $\int_0^1 Q_{k_1}(\pi)\, d\pi$ is well defined, which was shown in Anderson and Darling (1952) (see their Theorem 4.1 and the example on pages 202–204). □

References

Anderson, T.W. and D.A. Darling (1952): "Asymptotic theory of certain 'goodness of fit' criteria based on stochastic processes," *Annals of Mathematical Statistics*, 193–212.

Andrews, D.W.K (1988): "Laws of large numbers for dependent non–identically distributed random variables," *Econometric Theory*, 4, 458–467.

______________, (1990a): "Heteroskedasticity and autocorrelation consistent covariance matrix estimation," *Econometrica*, (forthcoming).

______________, (1990b): "Tests for parameter instability and structural change with unknown change point," Cowles Discussion Paper 943.

______________, (1990c): "Generic uniform convergence," Cowles Discussion Paper 940.

______________, and R.C. Fair (1988): "Inference in nonlinear econometric models with structural change," *Review of Economic Studies*, 55, 615–639.

Banerjee, A., R.L. Lumsdaine, and J.H. Stock (1990): "Recursive and sequential tests of the unit root and trend break hypothesis", unpublished manuscript.

Billingsley, P. (1968): *Convergence of Probability Measures.* NY: John Wiley & Sons.

Chow, G.C. (1960): "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, 28, 591–605.

Chu, C–S J. (1989): "New tests for parameter constancy in stationary and nonstationary regression models," manuscript, UCSD.

Gallant, A.R. (1987). *Nonlinear statistical models.* NY: John Wiley & Sons.

Gallant, A.R. and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Basil Blackwell.

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. (1991): "Strong laws for dependent heterogeneous processes," *Econometric*

Theory, (forthcoming).

Herrndorf, N. (1984): "A functional central limit theorem for weakly dependent sequences of random variables," *The Annals of Probability*, 12, 829–839.

Kim, H–J, and D. Siegmund (1989): "The likelihood ratio test for a change–point in simple linear regression," *Biometrika*, 76, 409–23.

King, M. L. (1987): "An alternative test for regression coefficient stability," *The Review of Economics and Statistics*, 69, 379–81.

Leybourne, S.J. and B.P.M. McCabe (1989): "On the distribution of some test statistics for coefficient constancy," *Biometrika* , 76, 169–77.

Nabeya, S. and K. Tanaka (1988): "Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative," *The Annals of Statistics*, 16, 218–235.

Newey, W.K. and K.D. West (1987): "A simple, positive definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, 55, 703–708.

Nicolls, D.F. and A.R. Pagan (1985): "Varying coefficient regression," *Handbook of Statistics*, (E.J. Hannan, P.R. Krishnaiah and M.M. Rao, eds.), 5, 413–449.

Nyblom, J. (1989): "Testing for the constancy of parameters over time," *Journal of the American Statistical Association*, 84, 223–230.

Nyblom, J. and T. Makelainen (1983): "Comparisons of tests for the presence of random walk coefficients in a simple linear model," *Journal of the American Statistical Association*, 78, 856–864.

Pagan, A.R. and K. Tanaka (1981): "A further test for assessing the stability of regression coefficients," unpublished manuscript.

Quandt, R. (1960): "Tests of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association*, 55, 324–30.

Tanaka, K. (1983): "Non–normality of the Lagrange multiplier statistic for testing the constancy of regression coefficients," *Econometrica*, 51, 1577–1582.

TABLE 1: ASYMPTOTIC CRITICAL VALUES FOR L_C[3]

Degrees of Freedom (k_1)	Significance Level					
	1%	2.5%	5%	7.5%	10%	20%
1	.748	.593	.470	.398	.353	.243
2	1.07	.898	.749	.670	.610	.469
3	1.35	1.16	1.01	.913	.846	.679
4	1.60	1.39	1.24	1.14	1.07	.883
5	1.88	1.63	1.47	1.36	1.28	1.08
6	2.12	1.89	1.68	1.58	1.49	1.28
7	2.35	2.10	1.90	1.78	1.69	1.46
8	2.59	2.33	2.11	1.99	1.89	1.66
9	2.82	2.55	2.32	2.19	2.10	1.85
10	3.05	2.76	2.54	2.40	2.29	2.03
11	3.27	2.99	2.75	2.60	2.49	2.22
12	3.51	3.18	2.96	2.81	2.69	2.41
13	3.69	3.39	3.15	3.00	2.89	2.59
14	3.90	3.60	3.34	3.19	3.08	2.77
15	4.07	3.81	3.54	3.38	3.26	2.95
16	4.30	4.01	3.75	3.58	3.46	3.14
17	4.51	4.21	3.95	3.77	3.64	3.32
18	4.73	4.40	4.14	3.96	3.83	3.50
19	4.92	4.60	4.33	4.16	4.03	3.69
20	5.13	4.79	4.52	4.36	4.22	3.86

TABLE 2 : ASYMPTOTIC CRITICAL VALUES FOR L_W[4]

Degrees of Freedom (k_1)	Significance Level					
	1%	2.5%	5%	7.5%	10%	20%
1	3.83	3.05	2.48	2.14	1.93	1.40
2	5.65	4.69	4.05	3.64	3.37	2.69
3	7.33	6.35	5.50	5.01	4.69	3.89
4	8.85	7.74	6.91	6.37	5.96	5.06
5	10.1	9.08	8.23	7.67	7.25	6.23
6	11.6	10.4	9.43	8.90	8.45	7.35
7	12.7	11.6	10.7	10.1	9.65	8.50
8	14.0	12.8	11.9	11.3	10.8	9.63
9	15.4	14.1	13.0	12.4	12.0	10.7
10	16.6	15.4	14.2	13.6	13.1	11.8
11	17.9	16.5	15.4	14.7	14.2	12.9
12	19.0	17.7	16.6	15.9	15.4	14.0
13	20.3	18.9	17.8	17.0	16.5	15.1
14	21.5	20.1	18.9	18.2	17.6	16.1
15	22.9	21.3	20.1	19.3	18.8	17.2
16	24.1	22.5	21.3	20.5	19.9	18.3
17	25.5	23.7	22.4	21.6	21.0	19.3
18	26.6	25.0	23.6	22.8	22.1	20.4
19	27.7	26.1	24.7	23.8	23.2	21.5
20	28.9	27.3	25.9	25.0	24.3	22.5

[3]Critical values were calculated from 20,000 draws from distribution (19).

[4]Critical values were calculated from 20,000 draws from distribution (20).

TABLE 3[5]
SIMULATED NULL REJECTION FREQUENCY

Sample Size	Test Statistic	1%	5%	10%
n = 30	SupLM	2.5	7.4	12.0
	L_C	0.6	5.0	11.0
	L_W	2.4	8.5	14.7
n = 60	SupLM	1.7	5.6	9.5
	L_C	1.0	5.4	10.9
	L_W	1.7	7.0	12.5
n = 120	SupLM	1.2	5.1	8.9
	L_C	1.0	5.3	10.3
	L_W	1.4	6.1	11.4

TABLE 4–A
SIMULATED SIZE–ADJUSTED POWER OF 5% SIZE TEST (N = 30)

		SupLM	L_C	L_W
$\pi^* = .1$	4.8	13	10	11
	7.2	27	15	19
	9.6	51	24	33
	12.0	76	34	50
$\pi^* = .3$	4.8	29	32	32
	7.2	60	61	61
	9.6	89	87	87
	12.0	98	97	97
$\pi^* = .5$	4.8	33	40	38
	7.2	69	75	72
	9.6	93	95	93
	12.0	99	99	99
RW	4.8	19	22	22
	7.2	32	33	34
	9.6	44	44	45
	12.0	51	50	52

[5]Percentage rejections using asymptotic critical values (20,000 replications).

TABLE 4–B
SIMULATED SIZE–ADJUSTED POWER OF 5% SIZE TEST (N = 60)

		SupLM	L_C	L_W
$\pi^* = .1$	4.8	8	8	9
	7.2	13	12	15
	9.6	21	18	25
	12.0	33	26	40
$\pi^* = .3$	4.8	30	29	29
	7.2	65	63	62
	9.6	91	88	88
	12.0	99	98	97
$\pi^* = .5$	4.8	38	42	39
	7.2	75	79	77
	9.6	96	97	96
	12.0	99	99	99
RW	4.8	20	22	22
	7.2	36	36	37
	9.6	49	48	49
	12.0	59	56	57

TABLE 4–C
SIMULATED SIZE–ADJUSTED POWER OF 5% SIZE TEST (N = 120)

		SupLM	L_C	L_W
$\pi^* = .1$	4.8	9	8	9
	7.2	16	13	17
	9.6	27	21	29
	12.0	42	32	46
$\pi^* = .3$	4.8	32	33	33
	7.2	67	66	65
	9.6	93	90	90
	12.0	99	99	98
$\pi^* = .5$	4.8	39	45	42
	7.2	77	82	78
	9.6	97	97	96
	12.0	99	99	99
RW	4.8	22	23	23
	7.2	37	38	38
	9.6	52	51	51
	12.0	62	60	61