



Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework

Fallaw Sowell

Econometrica, Vol. 64, No. 5 (Sep., 1996), 1085-1107.

Stable URL:

<http://links.jstor.org/sici?sici=0012-9682%28199609%2964%3A5%3C1085%3AOTFPII%3E2.0.CO%3B2-X>

Econometrica is currently published by The Econometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/econosoc.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

OPTIMAL TESTS FOR PARAMETER INSTABILITY IN THE GENERALIZED METHOD OF MOMENTS FRAMEWORK

BY FALLAW SOWELL¹

This paper presents optimal tests for parameter instability in the GMM framework. The new tests include tests that are optimal for both one-sided and two-sided alternatives. One of the optimal tests for two-sided alternatives is the GMM generalization of the test presented in Andrews and Ploberger (1994) for the likelihood framework. The new tests include a class of optimal tests that direct the test's power to specific locations in the sample. One of these optimal tests has the attractive feature of a normal distribution under the null hypothesis.

KEYWORDS: Optimal test for parameter instability, Brownian motion, Brownian bridge, stochastic differential equation, Neyman-Pearson lemma, Radon-Nikodym derivative.

THIS PAPER PRESENTS OPTIMAL TESTS for parameter instability in the GMM framework. The locally most powerful test is presented for any parameter instability alternative that can be written as the limit of simple functions. When the magnitude and/or location of the parameter instability is unknown, the test with the greatest weighted average power is presented. Two qualifications should be noted for the optimality results. First, optimality is only an asymptotic result. GMM estimation requires general distributional assumptions. This level of generality requires the use of the asymptotic distribution for optimality results. Second, the optimality is conditional on the selected moment conditions.

The derivation of the new tests relies on the weak convergence of the stochastic processes that compose the GMM objective function. A continuous functional (mapping) from the stochastic processes that compose the GMM objective function gives a test statistic. The functional applied to the limiting processes characterizes the test statistics' distributions under the null and under local alternatives. These general results imply a continuum of optimal tests. Specific tests are presented for one-sided and two-sided alternatives that are linear in the parameters. The new tests include a class of optimal tests that direct the test's power to specific locations in the sample. These new tests present a collection of information to help judge the adequacy of fitted models.

The research closest to this paper is the path breaking work in Andrews and Ploberger (1994), hereafter denoted AP. The major differences are that AP only presented optimal tests for a special weighting distribution, AP only considered the likelihood function framework, and the proof technique in AP cannot be extended to the GMM framework. AP uses the Neyman-Pearson lemma with

¹ Helpful discussions were provided by Don Andrews, Steve Shreve, and Stan Zin. Helpful comments were provided by a co-editor, three anonymous referees, and seminar participants at the University of Virginia, the University of Washington—Seattle, Carnegie Mellon University, Yale University, the Research Triangle Econometrics group, the University of California at San Diego, and Cal Tech.

the density of the observed series. This gives the most powerful test for each finite sample for the local alternative. This method of proof cannot be generalized to the GMM framework because in a finite sample there is no density. To address the more general class of models an alternative method of proof is required. The weak convergence limits of the normalized partial sums of the sample moments under the null and local alternative hypotheses are derived. These imply two measures on the space of realizations for the normalized partial sums of the sample moments. The Neyman-Pearson lemma is then generalized to the problem of testing between these two measures. The result is a functional that implies the asymptotically most powerful test against local alternatives in the GMM framework.

Previous results on parameter instability tests in the GMM framework (e.g., Andrews and Fair (1988), Hoffman and Pagan (1989), Ghysels and Hall (1990), and Hansen (1990)) have not considered issues of optimality. Andrews (1993) does present an optimality result for the sup test. However as noted in Andrews (1993) the result is very weak, only concerns power as the size of the test goes to zero, and is not indicative of finite sample performance. Optimal tests for structural stability have been limited to the likelihood function framework. This work includes Davis (1977, 1987), Nyblom (1989), King and Shively (1993), and AP. The questions considered in Davis (1977, 1987) and in AP are special cases of the question answered in this paper. Stock (1990) used continuous mappings to unify the asymptotic theory under the null for a collection of unit root tests. Though some of the same tools are used, the current paper assumes stationarity.

The paper is organized as follows. The next section defines the class of models, presents the main asymptotic results concerning the weak convergence of stochastic processes created from sample moments, and presents a general framework to generate specification tests in the GMM framework. Section 2 uses the asymptotic results and the general framework to derive several new optimal tests for structural breaks in the GMM framework. Section 3 is a summary and presents directions for future research.

The following notation will be used in this paper: $b(s)(b^\circ(s))$ defined on $s \in [0, 1]$ denotes the univariate standard Brownian motion (bridge) process, $B_j(s)(B_j^\circ(s))$ denotes the j -dimensional vectors of independent standard Brownian motions (bridges), \Rightarrow denotes weak convergence, \rightarrow^p denotes convergence in probability, $\|\bullet\|$ denotes the Euclidian norm, and $[\bullet]$ is the greatest integer function.

1. MODEL AND ASYMPTOTIC RESULTS

The starting point for the results in this paper are the consistency of the GMM estimator and a multivariate invariance principle that applies to the sample moments. The following are not the weakest assumptions possible. Rather, these assumptions are relatively straightforward to verify and general enough to be of interest. The assumptions are strong enough to obtain the

limiting distribution under both the null hypothesis and under sequences of local alternatives.

ASSUMPTION 1: For each T , the sequence $\{x_{t,T}\}$ consists of the first T elements of an r -dimensional stationary and ergodic stochastic process $\{x_{t,T}: t = 1, 2, \dots\}$.

For notational simplicity, x_t will be used to denote $x_{t,T}$.

ASSUMPTION 2: The parameter space Θ is a compact subset of R^k .

To allow the calculation of power against local alternatives, a sequence of alternatives will be considered. This class of alternatives allows for structural changes.

ASSUMPTION 3:

$$\theta_{t,T} = \theta_0 + \frac{g\left(\eta, \pi, \frac{t}{T}\right)}{\sqrt{T}}$$

where $g(\eta, \pi, s)$, for $s \in [0, 1]$, is a k -dimensional function that can be expressed as the uniform limit of step functions, $\eta \in R^i$, $\pi \in R^j$ such that $0 < \pi_1 < \pi_2 < \dots < \pi_j < 1$, and θ_0 is in the interior of Θ .

A standard result in analysis, e.g. Kolmogorov and Fomin (1961), shows that $g(\eta, \pi, s)$ can be any measurable function; hence alternatives that imply multiple jumps or gradual shifts satisfy this assumption. If an element of $g(\eta, \pi, s)$ is zero the alternative does not involve the corresponding parameter in θ . Andrews and Fair (1988) refer to an alternative where none of the element of $g(\eta, \pi, s)$ are zero as a pure structural change and where some of the elements are zero as a partial structural change. The parameter π denotes the times of the structural changes as fractions of the sample size. The vector η parameterizes the function that defines the local alternatives. Note that the dimension of η is different from the dimension of θ , the structural parameters. For example, if the alternative of interest has n different jumps in the value of a single parameter, then $j = i = n$. There would be one parameter in η for each change in the structural parameter's value. Similarly, if the alternative of interest is a one-time jump in the values of each of the structural parameters, then $j = i = k$.

Functions of the observed data summarize the information in the model. This model may be suggested by economic theory, e.g. Euler equations, or perhaps an estimation problem, e.g. the score functions from maximum likelihood. The functions are of the form $f(x_t, \theta)$, where $f: R^r \times R^k \rightarrow R^m$. The relevant theory will imply the functions satisfy $Ef(x_t, \theta_0) = 0$. The sample estimate of $Ef(x_t, \theta)$ is the function $F_{sT}(\theta) = (1/T) \sum_{t=1}^{\lfloor sT \rfloor} f(x_t, \theta)$ evaluated at $s = 1$, where $s \in [0, 1]$. For notational simplicity, $F_T(\theta)$ will be used to denote $F_{1T}(\theta)$. The GMM

estimator of θ is selected to make an estimate of $Ef(x_t, \theta)$ close to the zero vector in some metric. A sequence of weighting matrices, $\{W_T\}$, determines this metric. The GMM estimator, $\hat{\theta}_T$, is defined as a sequence of random vectors that solves

$$\hat{\theta}_T = \underset{\theta}{\operatorname{argmin}} F_T(\theta)' W_T F_T(\theta).$$

The asymptotic variance of $F_T(\theta_0)$ is defined by $\Sigma \equiv \lim_{T \rightarrow \infty} E[TF_T(\theta_0)F_T(\theta_0)']$. The probability limit of the gradient of the sample moments will be denoted $M \equiv \operatorname{plim}_{T \rightarrow \infty} (\partial F_T(\theta_0)/\partial \theta)$. This gradient is often normalized by the symmetric matrix square root of Σ^{-1} . This matrix will be denoted $\bar{M} \equiv \Sigma^{-1/2}M$. Below, \hat{M}_T will be used to denote a consistent estimate of M .

ASSUMPTION 4: *The matrix Σ is positive definite and the matrix M has full column rank.*

An identification assumption is required to assure that the sequence of GMM estimator has a unique limit.

ASSUMPTION 5: $\lim_{T \rightarrow \infty} EF_T(\theta) = 0$, only when $\theta = \theta_0$.

The functions of the data must satisfy smoothness and boundedness regularity conditions.

ASSUMPTION 6: *$f(x, \theta)$ is continuously partially differentiable in θ in a neighborhood of θ_0 for every $\theta \in \Theta$. The functions $f(x, \theta)$ and $(\partial f(x, \theta)/\partial \theta)$ are measurable functions of x for each $\theta \in \Theta^*$, and $E \sup_{\theta \in \Theta^*} \|\partial f(x_t, \theta)/\partial \theta\| < \infty$. $Ef(x_t, \theta_0) = 0$, $Ef(x_t, \theta_0)'f(x_t, \theta_0) < \infty$, and $\sup_{\theta \in \Theta} \|f(x_t, \theta)\| < \infty$ for all $t = 1, \dots, T$ and $T = 1, 2, \dots$. Each element of $f(x_t, \theta_{t,T})$ is uniformly square integrable, for all $t = 1, \dots, T$ and $T = 1, 2, \dots$.*

Only optimal GMM is considered, i.e., attention is restricted to efficient GMM estimators. This is achieved by restricting the choice of the weighting matrix in the next assumption.

ASSUMPTION 7: *The sequence of positive definite weighting matrices $\{W_T\}_{T=k}^{\infty}$ converge in probability to Σ^{-1} .*

The symmetric matrix square root of the weighting matrix will be denoted $W_T^{1/2}$. This assumption holds for consistent estimates of the spectral density at frequency zero. Several potential estimators exist. See Hannan (1970, pp. 273–288). Satisfying Assumption 7 typically requires a two-step estimation procedure. See Hansen (1982).

The next assumption imposes restrictions on the amount of heteroskedasticity and autocorrelation allowed in the observed series. See Phillips and Durlauf

(1986) for the definitions of strong and uniform mixing for multivariate processes, which generalizes the univariate work of McLeish (1975).

ASSUMPTION 8: *Either*

1. $\{x_t\}$ is strong mixing with strong mixing coefficients $\{\alpha(n)\}$, $\sum_{n=1}^{\infty} \alpha(n)^{1-2/\beta} < \infty$ with $\beta > 2$, or

2. $\{x_t\}$ is uniform mixing with uniform mixing coefficients $\{\phi(n)\}$, $\sum_{n=1}^{\infty} \phi(n)^{1-1/\beta} < \infty$ with $\beta \geq 2$,

and the individual elements of $f(x_t, \theta_{t,T})$ have the finite absolute moment $E |f^{(i)}(x_t, \theta_{t,T})|^\beta < \infty$ for $i = 1, \dots, m$.

These assumptions define the class of models for which specification tests will be defined. The asymptotic results needed for optimal testing are the weak convergence of the partial sums of the sample moments under the null and alternative hypotheses. Theorem 1 gives the convergence under the alternative and Corollary 1 gives the convergence under the null.

THEOREM 1: *If Assumptions 1–8 are satisfied, then*

$$(1) \quad \sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) \\ \Rightarrow B_m(s) - s \bar{M}(\bar{M}' \bar{M})^{-1} \bar{M}' B_m(1) - \bar{M} \int_0^s \gamma(\eta, \pi, v) dv,$$

where $\gamma(\eta, \pi, v) = g(\eta, \pi, v) - \int_0^1 g(\eta, \pi, r) dr$.

The proofs are presented in the Appendices.

The null hypothesis for specification tests is that the model is correctly specified and the parameters of the model do not change over time. This null hypothesis is characterized in the following assumption.

ASSUMPTION 3': $\theta_{t,T} = \theta_0 \forall t, T$ and θ_0 is in the interior of Θ .

The special structure of the stochastic processes under the null hypothesis is presented in the following corollary.

COROLLARY 1: *If Assumptions 1–2, Assumption 3', and Assumptions 4–8 are satisfied, then there exists an orthonormal matrix C such that*

$$C \sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) \Rightarrow \begin{bmatrix} B_k^\circ(s) \\ B_{m-k}(s) \end{bmatrix}$$

where $B_k^\circ(s)$ and $B_{m-k}(s)$ are independent and $\bar{M}(\bar{M}' \bar{M})^{-1} \bar{M}' = C' \Lambda C$ where $CC' = I_m$ and

$$\Lambda = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

This corollary shows that under the null hypothesis the limiting continuous stochastic processes are linear combinations of k Brownian bridges, one for each parameter estimated, and $m - k$ Brownian motions, one for each over-identifying restriction.

The normalized partial sum of the sample moments converge to a stochastic process, $Z(s)$. Corollary 1 shows that if there is no structural change (i.e., under the null hypothesis), the limiting stochastic processes are linear combinations of independent standard Brownian bridges and standard Brownian motions. Theorem 1 presents the asymptotic behavior under local alternatives of parameter instability. These local alternatives only add drift to the Brownian bridges. The Brownian motions are not affected. The structure of the drift is similar to the structure of the Brownian bridges. In differential form the drift can be seen as the partial sum of deviations from the mean for the function that defined the local alternatives

$$\gamma(\eta, \pi, s)ds = g(\eta, \pi, s)ds - \left\{ \int_0^1 g(\eta, \pi, r) dr \right\} ds.$$

Theorem 1 and Corollary 1 can be used as the foundation for a general approach to asymptotic theory for specification testing in the GMM framework. To form a specification test the following steps can be followed.

1. Use the data to construct a stochastic process, e.g. $\sqrt{T} W_T^{1/2} F_{st}(\hat{\theta}_T)$, that asymptotically approaches the limiting continuous stochastic process in equation (1).

2. Choose a mapping H from \mathcal{T} the set of continuous functionals² from the space³ $D[0,1]^m$ to R such that for every $\alpha \in (0,1)$ there exists a k_α , with $|k_\alpha| < \infty$, such that

$$\Pr \left\{ H \left(C' \begin{bmatrix} B_k^\circ(s) \\ B_{m-k}(s) \end{bmatrix} \right) \leq k_\alpha \right\} = \alpha.$$

3. A test statistic is formed by applying the continuous mapping of step 2 to the stochastic process constructed in step 1, e.g. $H(\sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T))$.

4. The continuous mapping theorem and Corollary 1 can be used to show that the distribution under the null can be characterized as equivalent to the distribution of

$$H \left(C' \begin{bmatrix} B_k^\circ(s) \\ B_{m-k}(s) \end{bmatrix} \right).$$

² Continuous with respect to the uniform metric.

³ The space $D[0,1]^m$ is the m product space of $D[0,1]$, the space of functions on $[0,1]$ that are right-continuous and have left limits (Billingsley (1968, Chapter 3)).

5. The continuous mapping theorem and Theorem 1 can be used to show that the distribution of the test statistic under the local alternatives in Assumption 3 can be characterized as equivalent to the distribution of

$$H\left(C' \begin{bmatrix} B_k^\circ(s) - C_1 \bar{M} \int_0^s \gamma(\eta, \pi, \nu) d\nu \\ B_{m-k}(s) \end{bmatrix}\right)$$

where C_1 is the matrix of the first j rows of C which span the same space as the columns of \bar{M} .

This can be considered a generic approach to asymptotic theory for specification tests in the GMM framework. A functional in \mathcal{T} characterizes a test statistic. All that is needed to define a specification test is the functional. This approach is followed in the remainder of this paper. Attention is limited to the presentation of functionals.

2. OPTIMAL TESTS FOR STRUCTURAL CHANGE

Each functional in \mathcal{T} generates a statistical test. A natural question is which of these tests will have power against a particular structural break (alternative). This section uses the approach of the previous section in the derivation of optimal tests for structural breaks in the GMM framework. A functional will imply the optimal test for the structural break given by the function $g(\eta, \pi, s)$.

Under the null hypothesis of structural stability $CZ(s)$ will be a draw from $C[0, 1]^m$ using the measure defined by the stochastic differential equation

$$(2) \quad dCZ(s) = \begin{bmatrix} dB_k^\circ(s) \\ dB_{m-k}(s) \end{bmatrix}.$$

The alternative hypothesis of structural instability embodied in the local alternatives imply $CZ(s)$ will be a draw from $C[0, 1]^m$ using the measure defined by the stochastic differential equation

$$(3) \quad dCZ(s) = -C\bar{M}\gamma(\eta, \pi, s)ds + \begin{bmatrix} dB_k^\circ(s) \\ dB_{m-k}(s) \end{bmatrix}.$$

Asymptotically, testing between the null hypothesis of structural stability and the local alternative is equivalent to testing between two measures defined on the space $C[0, 1]^m$. Because the last $m - k$ rows of C are orthogonal to \bar{M} , the measures implied by the alternative and the null can only differ with respect to the Brownian bridges. In general, the alternative will place restrictions over j Brownian bridges.

To address issues of power for this asymptotic testing problem a Neyman-Pearson approach can be applied. The standard Neyman-Pearson lemma (Lehmann (1959, p. 65)) gives the form of the most powerful test when testing

between two distributions (i.e., probability measures) that possess densities. For testing problems where the competing distributions do not have densities, the next theorem uses the Radon-Nikodym derivative to generalize the ratio of the densities in the standard Neyman-Pearson lemma.

THEOREM 2 (Neyman-Pearson): *Let $X(s)$ be a draw from the sample space $C[0, 1]^m$. If the null hypothesis implies the measure μ_0 and the alternative hypothesis implies the measure μ_1 , and if μ_0 is absolutely continuous with respect to μ_1 , then the most powerful level α test will have a critical region, \mathcal{E}_* , defined by*

- (a) $\frac{d\mu_1}{d\mu_0}(X(s)) \geq k$ for each $X(s) \in \mathcal{E}_*$,
- (b) $\frac{d\mu_1}{d\mu_0}(X(s)) < k$ for each $X(s) \notin \mathcal{E}_*$, and
- (c) $\int_{\mathcal{E}_*} d\mu_0 = \alpha$,

where $(d\mu_1/d\mu_0)(X(s))$ is the Radon-Nikodym derivative of μ_1 with respect to μ_0 evaluated at $X(s)$.

Theorem 2 shows that the Radon-Nikodym derivative is the functional from $C[0, 1]^m$ that gives the most powerful test. Hence optimal testing requires the Radon-Nikodym derivative of the measure implied by the local alternative with respect to the measure implied by the null. The general form of this Radon-Nikodym derivative is given in the following theorem.

THEOREM 3: *Let μ_0 denote the probability measure over $C[0, 1]^m$ implied by the stochastic differential equation*

$$dX(s) = \begin{bmatrix} dB_k^\circ(s) \\ dB_{m-k}(s) \end{bmatrix}$$

and let μ_1 denote the probability measure over $C[0, 1]^m$ defined by the stochastic differential equation

$$dX(s) = v(s)ds + \begin{bmatrix} dB_k^\circ(s) \\ dB_{m-k}(s) \end{bmatrix}$$

where $v(s) = [v_1(s) \ v_2(s) \dots v_m(s)]'$ is a vector of functions defined on the unit interval that satisfy

1. $\int_0^1 v_i(s)^2 ds < \infty$ for $i = 1, \dots, m$, and
2. $\lim_{t \uparrow 1} (\int_0^t v_i(s) ds) / (1 - t) < \infty$ for $i = 1, \dots, k$.

The Radon-Nikodym derivative of μ_1 with respect to μ_0 is

$$\frac{d\mu_1}{d\mu_0}(X(s)) = \exp\left\{\int_0^1 \nu(s)' dX(s) - \frac{1}{2} \int_0^1 \nu(s)' \nu(s) ds\right\}$$

where $X(s)$ is a realization (or trajectory), not a random variable.

Theorems 1 to 3 immediately give the functional that implies the asymptotically most powerful test against the local alternatives given in Assumption 3.

COROLLARY 2: *The most powerful asymptotic test for the local alternative given in Assumption 3 rejects the null hypothesis of no structural break if $\zeta(\eta, \pi) \geq k_\alpha$ where*

$$\begin{aligned} \zeta(\eta, \pi) = \exp\left\{ - \int_0^1 \gamma(\eta, \pi, s)' \bar{M}' dZ(s) \right. \\ \left. - \frac{1}{2} \int_0^1 \gamma(\eta, \pi, s)' \bar{M}' \bar{M} \gamma(\eta, \pi, s) ds \right\} \end{aligned}$$

and k_α is defined so that the test has size α .

This corollary gives the functional that produces the most powerful level α test for the point alternative implied by η and π .

For composite alternatives the optimal tests are defined for weighted average power criteria functions. The optimal tests depend on the choice of the weighting densities defined over the possible alternatives. The weighting density used in AP, which is a generalization of that used in Wald (1943), is an example of this approach to deriving optimal tests. Using the Bayesian interpretation given in Andrews (1994), if the weighting densities are viewed as prior distributions over the space of local alternatives, the optimal test statistics can be interpreted as posterior odds ratios. The next corollary gives the functionals that imply the optimal test for composite alternatives.

COROLLARY 3: *Let $J(\pi)$ be a weighting distribution function on π and let $R(\eta, \pi)$ be a weighting distribution function on η for every π in the support of $J(\pi)$. The asymptotic test with the greatest weighted average power rejects the null hypothesis of no structural break if $\iint \zeta(\eta, \pi) dR(\eta, \pi) dJ(\pi) \geq k_\alpha$, where k_α is defined so the test has size α .*

This corollary gives the functional that implies the test that will asymptotically have the greatest weighted average power conditional on the weighting functions. The null and alternative distributions of the test statistics implied by these functionals can be characterized using Theorem 1 and Corollary 1. Appendix 5 gives computationally efficient forms of the terms that are needed to evaluate the test statistics and calculate critical values.

Corollary 3 implies a continuum of optimal tests for parameter instability. These tests differ in terms of either the alternatives considered or the weighting distributions used. The next two subsections derive optimal tests using Corollary 3. The examples considered are one-sided and two-sided alternatives that are linear in the parameters and weighting distributions created from normal distributions. The final subsection considers the leading case of this class of alternatives: the one-time structural break alternative.

2.1 Alternatives that are Linear in the Parameters

Consider alternatives that are linear in the parameters, i.e. $g(\eta, \pi, s) = D(\pi, s)\eta$, where $D(\pi, s)$ is a $k \times i$ matrix that does not depend on η . Define $A(\pi) = -\int_0^1 D(\pi, s)' \bar{M}' dZ(s)$ and

$$V(\pi) = \int_0^1 \left\{ \left[\left(\int_0^1 D(\pi, r) dr \right) - D(\pi, s) \right]' \times \bar{M}' \bar{M} \left[\left(\int_0^1 D(\pi, r) dr \right) - D(\pi, s) \right] \right\} ds.$$

From Corollary 3 the optimal test statistic is implied by the functional⁴

$$\int_S \exp\{\eta' A(\pi) - \frac{1}{2} \eta' V(\pi) \eta\} dR(\eta, \pi) dJ(\pi)$$

where S is the support $J(\pi)$.

The distinction between one-sided and two-sided alternatives is achieved by different weighting distributions for $R(\eta, \pi)$. In this section these distributions will be constructed from an i -dimensional multivariate normal with zero mean and variance $U(\pi)$. For two-sided alternatives the entire density is used and the optimal functional simplifies to

$$TS = \int_S \left[\frac{|U(\pi)^{-1}|^{1/2}}{|V(\pi) + U(\pi)^{-1}|^{1/2}} \times \exp\left\{\frac{1}{2} A(\pi)' (V(\pi) + U(\pi)^{-1})^{-1} A(\pi)\right\} \right] dJ(\pi).$$

For one-sided alternatives the weighting distribution will be constructed by normalizing the normal density restricted to the positive orthant of \mathcal{R}^i and the optimal functional simplifies to

$$OS = \int_S \left[\frac{2^i |U(\pi)^{-1}|^{1/2}}{|V(\pi) + U(\pi)^{-1}|^{1/2}} \exp\left\{\frac{1}{2} A(\pi)' (V(\pi) + U(\pi)^{-1})^{-1} A(\pi)\right\} \times \Phi^{(i)}\left((V(\pi) + U(\pi)^{-1})^{-1/2} A(\pi)\right) \right] dJ(\pi)$$

⁴ Note $\int_0^1 [\int_0^1 D(\pi, r) dr]' \bar{M}' dZ(s) = 0$ because the integrand does not depend on s and the GMM first order conditions.

where $\Phi^{(i)}(\bullet)$ is the i -dimensional standard normal cumulative distribution function and $(V(\pi) + U(\pi)^{-1})^{-1/2}$ is the symmetric matrix square root.

The TS and the OS functionals can both be thought of as weighted averages of the same function. The TS functional rejects the null of parameter stability when the stochastic process $(V(\pi) + U(\pi)^{-1})^{-1/2}A(\pi)$ travels too far for too long from the zero vector in \mathcal{R}^i . The TS functional assigns the same weight regardless of the direction of the deviation. The OS functional rejects the null hypothesis in favor of the alternative of an increase in a parameter value if the stochastic process travels too far for too long in the direction of the positive orthant in \mathcal{R}^i . Using the i -dimensional cumulative normal distribution function, the OS statistic assigns higher weight to deviation in the direction of the positive orthant in \mathcal{R}^i and lower weight to deviation in the direction of the negative orthant in \mathcal{R}^i .

For both functionals, TS and OS , the first term of the integrands is composed of a constant and a ratio of determinants. These terms do not depend on the realization (i.e., the data). Therefore, without loss of generality the first term can be absorbed into the weighting distribution $J(\pi)$ by simply redefining this weighting distribution. These redefinitions will be used in the remainder of this section.

To control the relative weights assigned to different alternatives, the variance of the normal weighting distribution will be restricted to the form $U(\pi) = \{(1 + \frac{1}{c})Q(\pi) - V(\pi)\}^{-1}$ where $Q(\pi)$ is an $i \times i$ symmetric positive definite matrix and c is a nonnegative scalar. The matrix $Q(\pi)$ needs to be restricted so that $Q(\pi) - V(\pi)$ is always positive semidefinite. When $Q(\pi) = V(\pi)$ this reduces to the limit of the weighting matrix used in AP for the likelihood function framework. The parameter c controls the weight assigned to different alternatives. Larger (smaller) values of c assign more weight to alternatives further (nearer) from (to) the null. The next subsection demonstrates how $Q(\pi)$ can be used to control the relative weights assigned to different locations in the sample. This variance implies the functionals

$$TS_c = \int_S \exp\left\{\frac{1}{2} \frac{c}{1+c} \tilde{A}(\pi)' \tilde{A}(\pi)\right\} dJ(\pi) \quad \text{and}$$

$$OS_c = \int_S \exp\left\{\frac{1}{2} \frac{c}{1+c} \tilde{A}(\pi)' \tilde{A}(\pi)\right\} \Phi^{(i)}\left(\sqrt{\frac{c}{1+c}} \tilde{A}(\pi)\right) dJ(\pi)$$

where $\tilde{A}(\pi) \equiv Q(\pi)^{-1/2}A(\pi)$ and $Q(\pi)^{1/2}$ is the symmetric matrix square root.

For a special case of the situation considered here, Andrews, Lee, and Ploberger (1993) reported that the power and size properties of their test statistic changed very slowly with changes in the value of c . Following their conclusions the general structure of the functional will be investigated by looking at the tests implied by the extreme values of c . As c increases more weight is assigned to alternatives further from the null hypothesis and the functionals converge to $TS_\infty \equiv \lim_{c \rightarrow \infty} TS_c = \int_S \exp\{\frac{1}{2} \tilde{A}(\pi)' \tilde{A}(\pi)\} dJ(\pi)$ and

$OS_\infty \equiv \lim_{c \rightarrow \infty} OS_c = \int_S \exp\{\frac{1}{2}\tilde{A}(\pi)' \tilde{A}(\pi)\} \Phi^{(i)}(\tilde{A}(\pi)) dJ(\pi)$. As c decreases more weight is assigned to alternatives closer to the null hypothesis and the normalized functionals converge to

$$TS_0 \equiv \lim_{c \downarrow 0} 2 \left(\frac{TS_c - 1}{c} \right) = \int_S \tilde{A}(\pi)' \tilde{A}(\pi) dJ(\pi) \quad \text{and}$$

$$OS_0 \equiv \lim_{\sqrt{c} \downarrow 0} (-2^{i-1} \sqrt{2\pi}) \frac{OS_c - 1}{\sqrt{c}} = \int_S j' \tilde{A}(\pi) dJ(\pi)$$

where j is the vector of i ones.

2.2. Optimal Tests of One-time Structural Change

A special case of the linear in the parameters alternative is the one-time structural change with the location of the break unknown. Different weighting distributions imply different optimal tests for this simple alternative. A class of weighting distributions is introduced that allows the researcher to focus the tests' power on different locations of the sample. This is achieved by using the matrix function

$$Q_a(\pi) = \bar{M}' \bar{M} (\pi(1 - \pi))^{2a}$$

in the distributions introduced in the previous subsection. The functionals for this alternative and weighting distribution are

$$TS_{\infty,a} = \int_S \exp\{\frac{1}{2}\tilde{Z}_a(\pi)' \tilde{Z}_a(\pi)\} dJ(\pi),$$

$$OS_{\infty,a} = \int_S \exp\{\frac{1}{2}\tilde{Z}_a(\pi)' \tilde{Z}_a(\pi)\} \times \Phi^{(j)}(\tilde{Z}_a(\pi)) dJ(\pi),$$

$$TS_{0,a} = \int_S \tilde{Z}_a(\pi)' \tilde{Z}_a(\pi) dJ(\pi), \quad \text{and}$$

$$OS_{0,a} = \int_S j' \tilde{Z}_a(\pi) dJ(\pi),$$

where

$$\tilde{Z}_a(\pi) = \frac{(\bar{M}' \bar{M})^{-1/2} \bar{M}' Z(\pi)}{(\pi(1 - \pi))^a}.$$

When $0 < a$ the integrands are not defined for values of π at the endpoint of the interval $[0, 1]$. This requires restricting the support of $J(\pi)$, S , to a subset of $(0, 1)$. For $a \leq 0$ the integrands are defined for all values of π and it is possible for $S = [0, 1]$.

The term $1/(\pi(1 - \pi))^a$ is a weighting function. For different values of a , more power is directed to a particular location as the function assigns relatively

more weight to that location. For positive values of a the weight function has a valley shape and assigns relatively more weight to the values near the endpoints of the sample. For negative values of a the weight function has a hill shape and assigns relatively more weight to the middle of the sample.

The asymptotic distributions of these statistics under both the null and the local alternative are easily characterized using Theorem 1 and Corollary 1. Under the null hypothesis of no structural breaks the test statistics have the following asymptotic distributions:

$$TS_{\infty,a} \sim^a \int_S \exp\left\{(1/2)B_j^\circ(\pi)'B_j^\circ(\pi)/(\pi(1-\pi))^{2a}\right\} dJ(\pi),$$

$$TS_{0,a} \sim^a \int_S B_j^\circ(\pi)'B_j^\circ(\pi)/(\pi(1-\pi))^{2a} dJ(\pi),$$

$$OS_{\infty,a} \sim^a \int_S \left[\exp\left\{(1/2)B_j^\circ(\pi)'B_j^\circ(\pi)/(\pi(1-\pi))^{2a}\right\} \times \Phi^{(j)}\left(B_j^\circ(\pi)/(\pi(1-\pi))^a\right) \right] dJ(\pi), \quad \text{and}$$

$$OS_{0,a} \sim^a \int_S j' B_j^\circ(\pi)/(\pi(1-\pi))^a dJ(\pi).$$

Critical values for these distributions can be calculated with the same simulation method used in AP. For some values of a , tables of critical values have already been calculated for $J(\pi)$, a uniform distribution. AP gives tables of critical values for the asymptotic distributions for the test statistics $TS_{0,1/2}$ and $\log(TS_{\infty,1/2})$. Nyblom (1989) gives tables of critical values for the asymptotic distributions for the statistic $TS_{0,0}$. The critical values for the asymptotic distributions for the statistics $TS_{\infty,0}$ and $\log(OS_{\infty,0})$ are reported in Table I. The fact that $\int_0^1 b^\circ(s)ds \sim N(0, 1/12)$ implies that under the null hypothesis of no structural changes the test statistic $OS_{0,0}$ converges to a draw from $N(0, j/12)$.

In applied work the value of a can be selected to focus attention on the most relevant alternatives. The tests for $a = 0$ imply the uniform weighting function for all locations and these statistics are recommended for the situation with no prior knowledge of the location of the break. Note that this is the largest value of a for which the integrand is defined over the entire unit interval. So the support of the weighting distribution $J(\pi)$ can be the entire unit interval. The $TS_{\infty,1/2}$ statistic is the GMM generalization of the LM -exp statistic introduced in AP for the likelihood function framework. This is just one of the continuum of test statistics $TS_{\infty,a}$ for $a \in \mathcal{A}$. Relative to the $a = 0$ tests, the $a = 1/2$ tests will have lower power to break near the middle of the sample but will have more power for alternatives with breaks near the endpoints. Of course if the location of the structural change can be restricted this information should be incorporated to obtain maximum power.

TABLE I
THE CRITICAL VALUES FOR DISTRIBUTIONS

p	$\int_0^1 \exp\left\{\frac{B_p^\circ(s)' B_p^\circ(s)}{2}\right\} ds$			$\log\left(\int_0^1 \exp\left\{\frac{B_p^\circ(s)' B_p^\circ(s)}{2}\right\} \Phi^{(p)}(B_p^\circ(s)) ds\right)$		
	α			α		
	0.10	0.05	0.01	0.10	0.05	0.01
1	1.21291	1.29981	1.54157	-0.3005	-0.1778	0.0978
2	1.40251	1.53357	1.90498	-0.7945	-0.6156	-0.2625
3	1.60907	1.78046	2.27492	-1.3104	-1.1110	-0.6774
4	1.83800	2.07437	2.71417	-1.8585	-1.6135	-1.1138
5	2.09766	2.38159	3.17508	-2.3935	-2.1230	-1.5666
6	2.38725	2.73231	3.74558	-2.9560	-2.6608	-2.0618
7	2.72378	3.15941	4.46667	-3.5216	-3.2019	-2.5634
8	3.09690	3.63121	5.13549	-4.0803	-3.7341	-3.0188
9	3.53434	4.20261	6.18840	-4.6379	-4.2646	-3.5295
10	4.02253	4.83677	7.09889	-5.2149	-4.8209	-4.0206

Notes: The critical values are based on 40,000 realizations. Each realization was constructed by approximating the integrals with the average over the discrete grid of 4,000 equal intervals on $[0, 1]$. For each realization the integrands were calculated by simulating the p -dimensional Brownian bridge with the partial sums of the deviations from mean for 4,000 normal random variables with variance $(4,000)^{-1}$.

The functional $OS_{0,0}$ implies an attractive univariate test for parameter stability. This test for parameter instability of a single parameter θ_i in the GMM framework is

$$St_i = \frac{\sqrt{12} e_i' \hat{M}_T' W_T \Sigma_{t=1}^T (T+1-t) f_i(\hat{\theta}_T)}{T\sqrt{T} \sqrt{e_i' \hat{M}_T' W_T \hat{M}_T e_i}}$$

where e_i is the i th column of the k dimensional identity matrix. This test can be written as a weighted average of the Sequence of LM t tests for the one-time break at all possible breaks. This test statistic has an optimality property, is easy to calculate, and has a standard normal distribution under the null hypothesis. This test has a particularly simple form for the linear regression model. For the linear regression model $y_t = \sum_{i=1}^k \beta_i x_{ti} + \epsilon_t$ with the alternative hypothesis of a one-time break in the parameter β_i , the optimal test statistic is

$$St_i = \frac{\sqrt{12}}{\hat{\sigma}_T \sqrt{\sum_{t=1}^T x_{ti}^2}} \sum_{t=1}^T \left(\frac{T+1-t}{T} \right) x_{ti} \hat{\epsilon}_t$$

where $\hat{\epsilon}_t$ are the OLS residuals and $\hat{\sigma}_T^2$ is a consistent estimate of

$$\lim_{T \rightarrow \infty} \text{var} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T \epsilon_t \right).$$

Large negative (positive) values suggest a one-time increase (decrease) in the parameter values during the sample. This test is optimal for the one-time structural break alternative; however Theorem 1 can be used to establish the larger class of alternatives for which this test has power.

3. SUMMARY AND FUTURE RESEARCH

This paper derives optimal tests for parameter instability in the GMM framework. As a demonstration of the usefulness of the theory, optimal tests were derived for one-sided and two-sided linear in the parameters alternatives. For the one-time structural change alternative these optimal tests imply a class of tests that permit a researcher to direct the power of the test to specific locations in the sample. Though these specific tests are optimal for the one-time structural break alternative, Theorem 1 can be used to demonstrate that they are consistent for a larger class of alternatives.

The tests $TS_{\infty,a}$, $TS_{0,a}$, $OS_{\infty,a}$, and $OS_{0,a}$ present a collection of information to help judge the adequacy of a fitted model. Applied researchers should select the appropriate test for their alternative: one-sided versus two-sided alternatives and alternatives that are "close to" versus "far from" the null. The univariate statistics are a collection of specification tests to judge the stability of an estimated model. These single parameter tests give the researcher an idea of which of the parameters are involved in a structural break. The multivariate statistics will have more power against alternatives which involve more than one parameter. Though these multiple parameter tests do not give the researcher an indication of which parameters are involved in the structural break. The multivariate and univariate tests should be viewed as complementary, comparable to the relationship between the F statistic and the t statistics for a linear regression model.

Several issues remain to be addressed including the tabulation of critical values for the different asymptotic distributions, the adequacy of the asymptotic approximation to the small sample distributions (see Diebold and Chen (1996)), and the implications of different weighting functions and how they affect the optimal tests.

The techniques used in this paper should generalize to additional testing problems. It should be possible to address a larger class of alternatives that includes stochastic alternatives such as those considered in Nyblom (1989) and King and Shively (1993). Also, this approach should generalize to models with nonstationary series.

Each test presented in this paper can be thought of as one element of an equivalence class of asymptotically equivalent tests. These asymptotically equivalent tests include Wald and likelihood ratio versions of the Lagrange multiplier tests presented in this paper. An open question is which of the tests in the equivalence class should be used in practice. This question can be addressed in terms of ease of calculation and the accuracy of the asymptotic approximations.

Grad. School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213-3890, U.S.A.

APPENDIX 1

PROOF⁵ OF THEOREM 1

The convergence, $\hat{\theta}_T \rightarrow^p \theta_0$, results from the identification assumption (Assumption 5) and the uniform convergence of $F_T(\theta)$ to $\lim_T EF_T(\theta)$. The uniform convergence is established by verifying Assumptions A1, B1, and A5 in Andrews (1987). A1 follows from Assumption 2. B1 follows from Assumption 8. A5 follows from Assumption 2 and Assumption 6.

Phillips and Durlauf (1986) present a multivariate generalization of the univariate results in McLeish (1975). Assumptions 1, 4, and 8 of this paper imply the assumptions of Corollary 2.2 in Phillips and Durlauf (1986); hence, the sample moments evaluated at $\theta_{i,T}$ satisfy the multivariate invariance principle.

$$\frac{1}{\sqrt{T}} W^{1/2} \sum_{t=1}^{[sT]} f(x_t, \theta_{i,T}) \Rightarrow B_m(s).$$

To reduce notation define $f_t(\theta) \equiv f(x_t, \theta)$ and $g(s) \equiv g(\eta, \pi, s)$.

Expand $f_t(\theta)$ about $\theta_{i,T}$ and evaluate the expansion at θ_0 :

$$\begin{aligned} f_t(\theta_0) &= f_t(\theta_{i,T}) + \frac{\partial f_t(\tilde{\theta}_{i,T})}{\partial \theta} (\theta_0 - \theta_{i,T}) \\ &= f_t(\theta_{i,T}) - \frac{\partial f_t(\tilde{\theta}_{i,T})}{\partial \theta} \frac{g\left(\frac{t}{T}\right)}{\sqrt{T}} \end{aligned}$$

where $\tilde{\theta}_{i,T} = [\tilde{\theta}_{i,T}^{(1)} \dots \tilde{\theta}_{i,T}^{(k)}]$ and $\tilde{\theta}_{i,T}^{(i)} = k_{i,T}^{(i)} \theta_{i,T}^{(i)} + (1 - k_{i,T}^{(i)}) \theta_0^{(i)}$ for some $k_{i,T}^{(i)} \in [0, 1]$ and each $t = 1, \dots, [sT]$ and $i = 1, \dots, k$. Because $\hat{\theta}_T$ is consistent for θ_0 , $\tilde{\theta}_{i,T} \rightarrow^p \theta_0$.

Now sum these terms from 1 to $[sT]$ and divide by T to get

$$F_{sT}(\theta_0) = \frac{1}{T} \sum_{t=1}^{[sT]} f_t(\theta_{i,T}) - \frac{1}{T} \frac{1}{\sqrt{T}} \frac{\partial f_t(\tilde{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right).$$

Multiply both sides by $\sqrt{T} W_T^{1/2}$ and let $T \rightarrow \infty$ to get

$$\sqrt{T} W_T^{1/2} \frac{1}{T} \sum_{t=1}^{[sT]} f_t(\theta_{i,T}) \Rightarrow B_m(s)$$

and

$$\sqrt{T} W_T^{1/2} \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^{[sT]} \frac{\partial f_t(\tilde{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right) \rightarrow^p \Sigma^{-1/2} M \int_0^s g(\nu) d\nu = \bar{M} \int_0^s g(\nu) d\nu$$

which gives

$$(4) \quad \sqrt{T} W_T^{1/2} F_{sT}(\theta_0) \Rightarrow B_m(s) - \bar{M} \int_0^s g(\eta, \pi, \nu) d\nu.$$

Expand $F_{sT}(\theta)$ about θ_0 and evaluate the expansion at $\hat{\theta}_T$:

$$(5) \quad F_{sT}(\hat{\theta}_T) = F_{sT}(\theta_0) + \frac{\partial F_{sT}(\bar{\theta}_{s,T})}{\partial \theta} (\hat{\theta}_T - \theta_0)$$

⁵ I thank Steve Stern and Hide Ichimura for pointing out a mistake in an earlier version of this proof.

where $\bar{\theta}_{s,T} = [\bar{\theta}_{s,T}^{(1)} \dots \bar{\theta}_{s,T}^{(k)}]$ and $\bar{\theta}_{s,T}^{(i)} = k_s^{(i)} \hat{\theta}_T^{(i)} + (1 - k_s^{(i)}) \theta_0^{(i)}$ for some $k_s^{(i)} \in [0, 1]$ and every $s \in [0, 1]$ and $i = 1, \dots, k$. Because $\hat{\theta}_T$ is consistent for θ_0 , $\bar{\theta}_{s,T} \rightarrow^p \theta_0$.

Calculate an alternative form for $(\hat{\theta}_T - \theta_0)$. Expand $f_i(\theta)$ about $\theta_{i,T}$ and evaluate the expansion at $\hat{\theta}_T$:

$$\begin{aligned} f_i(\hat{\theta}_T) &= f_i(\theta_{i,T}) + \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} (\hat{\theta}_T - \theta_{i,T}) \\ &= f_i(\theta_{i,T}) + \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} (\hat{\theta}_T - \theta_0) - \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right) \end{aligned}$$

where $\bar{\theta}_{i,T} = [\bar{\theta}_{i,T}^{(1)} \dots \bar{\theta}_{i,T}^{(k)}]$ and $\bar{\theta}_{i,T}^{(i)} = \bar{k}_{i,T}^{(i)} \theta_{i,T}^{(i)} + (1 - \bar{k}_{i,T}^{(i)}) \hat{\theta}_T^{(i)}$ for some $\bar{k}_{i,T}^{(i)} \in [0, 1]$ and each $t = 1, \dots, T$ and $i = 1, \dots, k$. Because $\hat{\theta}_T$ is consistent for θ_0 , $\bar{\theta}_{i,T} \rightarrow^p \theta_0$.

Sum these terms from 1 to T and divide by T to get

$$F_T(\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T f_i(\theta_{i,T}) + \frac{1}{T} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} (\hat{\theta}_T - \theta_0) - \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right).$$

Multiply both sides by

$$\frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T.$$

Assumptions 4 and 7 imply that there exists a T^* such that for all $T^* < T$

$$\begin{aligned} (6) \quad (\hat{\theta}_T - \theta_0) &= \left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \frac{1}{T} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} \right]^{-1} \frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \\ &\quad \times \left[-\frac{1}{T} \sum_{t=1}^T f_i(\theta_{i,T}) + \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right) \right]. \end{aligned}$$

This equality holds because with probability one the first order condition

$$F_T(\hat{\theta}_T)' W_T \frac{\partial F_T(\hat{\theta}_T)}{\partial \theta} = 0$$

will be satisfied and

$$\frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \frac{1}{T} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} \rightarrow^p M' \Sigma^{-1} M$$

where Σ is nonsingular and M is of full rank.

Substitute (6) into (5) to get

$$\begin{aligned} F_{sT}(\hat{\theta}_T) &= F_{sT}(\theta_0) + \frac{\partial F_{sT}(\bar{\theta}_{s,T})}{\partial \theta} \left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \frac{1}{T} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} \right]^{-1} \\ &\quad \times \frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \left[-\frac{1}{T} \sum_{t=1}^T f_i(\theta_{i,T}) + \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial f_i(\bar{\theta}_{i,T})}{\partial \theta} g\left(\frac{t}{T}\right) \right]. \end{aligned}$$

Multiply both sides by $\sqrt{T} W_T^{1/2}$. The convergence of the first term on the right-hand side is given by (4). The remaining terms converge as:

$$\begin{aligned} & W_T^{1/2} \frac{\partial F_{sT}(\bar{\theta}_{s,T})}{\partial \theta} \left[\frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T \frac{1}{T} \sum_{t=1}^T \frac{\partial f_t(\bar{\theta}_{t,T})}{\partial \theta} \right]^{-1} \frac{\partial F_T(\hat{\theta}_T)'}{\partial \theta} W_T^{1/2} \\ & \rightarrow {}^p s \bar{M} (\bar{M}' \bar{M})^{-1} \bar{M}', \\ & \sqrt{T} W_T^{1/2} \frac{1}{T} \sum_{t=1}^T f_t(\theta_{t,T}) \Rightarrow B_m(1), \\ & \sqrt{T} W_T^{1/2} \frac{1}{T} \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial f_t(\bar{\theta}_{t,T})}{\partial \theta} g\left(\frac{t}{T}\right) \rightarrow {}^p \Sigma^{-1/2} M \int_0^1 g(\nu) d\nu \\ & = \bar{M} \int_0^1 g(\nu) d\nu. \end{aligned}$$

These results immediately imply (1).

APPENDIX 2

PROOF OF COROLLARY 1

Under the null hypothesis of no structural breaks, i.e., $g(\eta, \pi, t/T) \equiv 0$, Theorem 1 implies

$$(7) \quad \sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) \Rightarrow B_m(s) - s \bar{M} (\bar{M}' \bar{M})^{-1} \bar{M}' B_m(1).$$

The $m \times m$ matrix $\bar{M} (\bar{M}' \bar{M})^{-1} \bar{M}'$ has rank k and is idempotent; hence $m - k$ of its eigenvalues are 0 and the remaining k eigenvalues are each 1. This matrix is symmetric with real elements so it has a spectral decomposition $\bar{M} (\bar{M}' \bar{M})^{-1} \bar{M}' = C' \Lambda C$ where $CC' = I_m$ and Λ is a diagonal matrix of the eigenvalues of $\bar{M} (\bar{M}' \bar{M})^{-1} \bar{M}'$. Without loss of generality, select a spectral decomposition where the first k -eigenvalues are 1 and the last $m - k$ are 0,

$$\Lambda = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix}.$$

Substitute the spectral decomposition into equation (7) and multiply both sides by C :

$$C \sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) \Rightarrow C B_m(s) - s \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} C B_m(1).$$

Because C is orthonormal $C B_m(s)$ is also a multivariate standard Brownian motion process. Hence

$$\begin{aligned} C \sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T) & \Rightarrow B_m(s) - s \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} B_m(1) \\ & = \begin{bmatrix} B_k^\circ(s) \\ B_{m-k}(s) \end{bmatrix} \end{aligned}$$

where $B_k^\circ(s)$ and $B_{m-k}(s)$ are independent.

APPENDIX 3

PROOF OF THEOREM 2

The proof requires showing that the power for the critical region \mathcal{E}_* is greater than or equal to the power for any other critical region. Let \mathcal{E} be any other critical region of size α for testing $H_0: \mu_0$ versus $H_1: \mu_1$.

$$\begin{aligned}
 \int_{\mathcal{E}_*} d\mu_1 - \int_{\mathcal{E}} d\mu_1 &= \int_{\mathcal{E}_* \cap \mathcal{E}} d\mu_1 + \int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_1 - \int_{\mathcal{E} \cap \mathcal{E}_*} d\mu_1 - \int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_1 \\
 &= \int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_1 - \int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_1 \\
 &\geq k \left[\int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_0 - \int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_0 \right] \\
 &= k \left[\int_{\mathcal{E}_* \cap \mathcal{E}} d\mu_0 + \int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_0 - \int_{\mathcal{E} \cap \mathcal{E}_*} d\mu_0 - \int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_0 \right] \\
 &= k \left[\int_{\mathcal{E}_*} d\mu_0 - \int_{\mathcal{E}} d\mu_0 \right] \\
 &= k[\alpha - \alpha].
 \end{aligned}$$

The first and fourth equalities occur because $\mathcal{E}_* = (\mathcal{E}_* \cap \mathcal{E}) \cup (\mathcal{E}_* \cap \mathcal{E}^c)$. The inequality holds because for the optimal test

$$\frac{d\mu_1}{d\mu_0}(X(s)) \geq k$$

for all $X(s) \in \mathcal{E}_*$ and hence

$$\int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_1 \geq k \int_{\mathcal{E}_* \cap \mathcal{E}^c} d\mu_0.$$

Similarly reasoning for $X(s) \notin \mathcal{E}_*$ shows that

$$\int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_1 \leq k \int_{\mathcal{E} \cap \mathcal{E}_*^c} d\mu_0.$$

The final equality occurs because the critical regions were determined to obtain tests with size α .

APPENDIX 4

PROOF⁶ OF THEOREM 3

Because the Brownian motions and Brownian bridges are independent, it is sufficient to show the result of the two cases of a single Brownian motion and a single Brownian bridge.

The Radon-Nikodym derivatives for these two cases have similar structures but are fundamentally different. For the Brownian motion the Radon-Nikodym derivative is between the measure induced by a standard Brownian motion with drift and the measure induced by a standard Brownian motion without drift. For the Brownian bridge the Radon-Nikodym derivative is between the measure induced by a standard Brownian bridge with drift and the measure induced by a standard Brownian bridge without drift. An important difference between these two situations is that the absolute

⁶ I thank an anonymous referee for greatly simplifying this proof.

continuity of the Brownian bridge measures requires the additional assumption that the drift for the Brownian bridge takes value zero at time $s = 1$ (in the statement of the Theorem this is restriction 2 for the functions ν_i).

The univariate Brownian motion result is presented in Theorem 3.1 of Basawa and Rao (1980, p. 206).

It remains to show the Radon-Nikodym derivative between the measure induced by a standard Brownian bridge with drift with respect to the measure induced by a standard Brownian bridge without drift.

Let $b^\circ(s)$ be the solution to the stochastic differential equation

$$(8) \quad \begin{aligned} b^\circ(0) &= 0, \\ db^\circ(s) &= -\frac{b^\circ(s)}{1-s} ds + db(s), \quad 0 \leq s \leq 1. \end{aligned}$$

In Karatzas and Shreve (1988, Section 5.6.B), it is shown that $b^\circ(s)$ is a Brownian bridge on $[0, 1]$. In particular,

$$b^\circ(1) = \lim_{s \uparrow 1} b^\circ(s) = 0 \quad \text{a.s.}$$

Now introduce a deterministic function $\nu: [0, 1] \rightarrow R$ and its indefinite integral

$$\mathcal{V}(t) = \int_0^t \nu(s) ds.$$

Define

$$\tilde{b}^\circ(t) \equiv b^\circ(t) + \mathcal{V}(t).$$

Which is the solution to the stochastic differential equation

$$(9) \quad d\tilde{b}^\circ(s) = \nu(s) ds + \frac{\mathcal{V}(s)}{(1-s)} - \frac{\tilde{b}^\circ(s)}{(1-s)} + db(s), \quad 0 \leq s \leq 1.$$

Theorem 7.19 of Liptser and Shirayev (1978) applied to the two diffusions (8) and (9) gives the Radon-Nikodym derivative:

$$(10) \quad \begin{aligned} \frac{d\mu_1}{d\mu_0} &= \exp \left\{ \int_0^1 \left(\frac{\mathcal{V}(s)}{1-s} + \nu(s) \right) db^\circ(s) \right. \\ &\quad \left. - \frac{1}{2} \int_0^1 \left(\frac{\mathcal{V}(s)}{1-s} + \nu(s) \right)^2 ds + \int_0^1 \frac{b^\circ(s)}{1-s} \left(\frac{\mathcal{V}(s)}{1-s} + \nu(s) \right) ds \right\}. \end{aligned}$$

Use Ito's formula to write

$$d \left(\frac{b^\circ(s) \mathcal{V}(s)}{1-s} \right) = \frac{\mathcal{V}(s)}{1-s} db^\circ(s) + \frac{b^\circ(s)}{1-s} \nu(s) ds + \frac{b^\circ(s) \mathcal{V}(s)}{(1-s)^2} ds,$$

so

$$\frac{b^\circ(t) \mathcal{V}(t)}{1-t} = \int_0^t \frac{\mathcal{V}(s)}{1-s} db^\circ(s) + \int_0^t \frac{b^\circ(s)}{1-s} \nu(s) ds + \int_0^t \frac{b^\circ(s) \mathcal{V}(s)}{(1-s)^2} ds$$

for all $t \in [0, 1]$. But,

$$\lim_{t \uparrow 1} \frac{b^\circ(t) \mathcal{V}(t)}{1-t} = 0$$

so

$$\int_0^1 \frac{\mathcal{V}(s)}{1-s} db^\circ(s) + \int_0^1 \frac{b^\circ(s)}{1-s} \nu(s) ds + \int_0^1 \frac{b^\circ(s) \mathcal{V}(s)}{(1-s)^2} ds = 0.$$

Substituting into (10) gives

$$\begin{aligned}\frac{d\mu_1}{d\mu_0} &= \exp \left\{ \int_0^1 \nu(s) db^\circ(s) - \frac{1}{2} \int_0^1 \left(\frac{\mathcal{V}(s)}{1-s} + \nu(s) \right)^2 ds \right\} \\ &= \exp \left\{ \int_0^1 \nu(s) db^\circ(s) - \frac{1}{2} \int_0^1 (\nu(s))^2 ds \right\}.\end{aligned}$$

The final step follows by an integration by parts.

APPENDIX 5

EVALUATION OF THE TEST STATISTICS AND ASYMPTOTIC DISTRIBUTIONS

The calculation of a test statistic requires evaluating functionals of the form $\bar{A}(\pi) = -\int_0^1 g(\eta, \pi, s)' \bar{M}' dZ(s)$ applied to the stochastic processes generated by the moments calculated from the data, i.e. where $Z(s)$ is replaced by $\sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T)$. Let $\bar{A}_T(\pi)$ denote the required functional where \bar{M} is replaced with a consistent estimate

$$\begin{aligned}\bar{A}_T(\pi) &= -\int_0^1 g(\eta, \pi, s)' \hat{M}_T' W_T^{1/2} d(\sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T)) \\ &= -\frac{1}{T\sqrt{T}} \sum_{t=1}^T g\left(\eta, \pi, \frac{t}{T}\right)' \hat{M}_T' W_T f_t(\hat{\theta}_T) \\ &= -\frac{1}{T\sqrt{T}} \text{tr} \left\{ \left(\sum_{t=1}^T f_t(\hat{\theta}_T) g\left(\eta, \pi, \frac{t}{T}\right)' \right) \hat{M}_T' W_T \right\}.\end{aligned}$$

This last form is computationally efficient to evaluate.

The other term needed to evaluate the test statistics is of the form

$$\bar{V}(\pi) = \int_0^1 \left[\left(\int_0^1 g(\eta, \pi, r) dr \right) - g(\eta, \pi, s) \right]' \bar{M}' \bar{M} \left[\left(\int_0^1 g(\eta, \pi, r) dr \right) - g(\eta, \pi, s) \right] ds.$$

To make this operational \bar{M} needs to be replaced with a consistent estimate. A computationally attractive form of this statistic is

$$\bar{V}_T(\pi) = \text{tr} \left\{ \text{diag} \left[\overline{g_i(\eta, \pi, s)^2} - \overline{(g_i(\eta, \pi, s))^2} \right] \hat{M}_T' W_T \hat{M}_T \right\}.$$

where $\text{diag}[\overline{g_i(\eta, \pi, s)^2} - \overline{(g_i(\eta, \pi, s))^2}]$ is a diagonal matrix with element

$$\int_0^1 g_i(\eta, \pi, s)^2 ds - \left(\int_0^1 g_i(\eta, \pi, s) ds \right)^2$$

on its i th position on the major diagonal where $g_i(\eta, \pi, s)$ is the i th element of $g(\eta, \pi, s)$.

These formulas facilitate the calculation of approximations to the asymptotic distributions under the null and alternative hypotheses. For distributions under the null, set T large⁷ and replace $\sqrt{T} W_T^{1/2} F_{sT}(\hat{\theta}_T)$ with a discrete approximation of

$$C' \begin{bmatrix} B_k(s) \\ 0_{m-k} \end{bmatrix}.$$

⁷ For Table I in this paper $T = 4000$ was used. In AP $T = 3600$ was used.

Because the Brownian motions do not contribute to the asymptotic distribution, they should not be simulated. For the distribution under the alternative, again set T large and replace $\sqrt{T}W_T^{1/2}F_{sT}(\hat{\theta}_T)$ with a discrete approximation of

$$C' \begin{bmatrix} B_k(s) - C_1 \bar{M} \int_0^s \gamma(\eta, \pi, \nu) d\nu \\ 0_{m-k} \end{bmatrix}.$$

REFERENCES

- ANDREWS, DONALD W. K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.
- (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.
- (1994): "The Large Sample Correspondence between Classical Hypothesis Tests and Bayesian Posterior Odds Tests," *Econometrica*, 62, 1207–1232.
- ANDREWS, DONALD W. K., AND RAY C. FAIR (1988): "Inference in Nonlinear Econometric Models With Structural Change," *Review of Economic Studies*, 55, 615–640.
- ANDREWS, DONALD W. K., L. LEE, AND WERNER PLOBERGER (1993): "Optimal Changepoint Test for Linear Regression," Cowles Foundation Discussion Paper, Yale University, forthcoming *Journal of Econometrics*.
- ANDREWS, DONALD W. K., AND WERNER PLOBERGER (1994): "Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative," *Econometrica*, 62, 1383–1414.
- BASAWA, ISHWAR V., AND B. L. S. PRAKASA RAO (1980): *Statistical Inference for Stochastic Processes*. New York: Academic Press Inc.
- BILLINGSLEY, PATRICK (1968): *Convergence of Probability Measures*. New York: John Wiley and Sons.
- DAVIS, 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.
- DIEBOLD, FRANCIS X., AND CELIA CHEN (1996): "Testing Structural Stability with Endogenous Break Points: A Size Comparison of Analytic and Bootstrap Procedures," *Journal of Econometrics*, 70, 221–241.
- GHYSELS, ERIC, AND ALASTAIR HALL (1990): "A Test for Structural Stability of Euler Conditions Parameters Estimated via the Generalized Method of Moments Estimator," *International Economic Review*, 31, 355–364.
- HANNAN, E. J. (1970): *Multiple Time Series*. New York: John Wiley and Sons, Inc.
- HANSEN, BRUCE E. (1990): "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," University of Rochester, Department of Economics.
- HANSEN, LARS PETER (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.
- HOFFMAN, DENNIS, AND ADRIAN PAGAN (1989): "Post-sample Prediction Tests for Generalized Method of Moments Estimators," *Oxford Bulletin of Economics and Statistics*, 51, 333–343.
- KARATZAS, IOANNIS, AND STEVEN E. SHREVE (1988): *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- KING, MAXWELL L., AND THOMAS S. SHIVELY (1993): "Locally Optimal Testing when a Nuisance Parameter is Present Only Under the Alternative," *The Review of Economics and Statistics*, 75, 1–7.
- KOLMOGOROV, A. N., AND S. V. FOMIN (1961): *Elements of the Theory of Functions and Functional Analysis: Volume 2 Measure. The Lebesgue Integral. Hilbert Space*. Albany, New York: Graylock Press.

- LEHMANN, E. L. (1959): *Testing Statistical Hypotheses*. New York: John Wiley and Sons.
- LIPTSER, R. S., AND A. N. SHIRYAYEV (1978): *Statistics of Random Processes I*. New York: Springer-Verlag.
- MCLEISH, D. L. (1975): "Invariance Principles for Dependent Variables," *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 32, 165–178.
- NYBLOM, JUKKA (1989): "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association*, 84, 223–230.
- PHILLIPS, P. C. B., AND S. N. DURLAUF, (1986): "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.
- STOCK, JAMES H. (1990): "A Class of Tests for Integration and Cointegration," Harvard University, Kennedy School of Government.
- 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.