A Comparison of Tests for Parameter Instability: An Examination of Asymptotic Local Power

Bruce E. Hansen

University of Rochester

April 1991

I thank Adrian Pagan for helpful discussions and the NSF for support under Grant # SES 9022176

Abstract

Many tests for parameter instability have been proposed and discussed. How should applied researchers choose between the tests? This paper attempts to construct a unified framework in which many tests can be compared. Asymptotic theory is used to develop a distributional theory for sequences of partial sample estimators and test statistics under the assumption of local parameter variation. This is a functional analog of the classic Pitman drift used to study local asymptotic power in conventional hypothesis testing. We find that the CUSUM test is essentially a test for instability in the intercept, the CUSUM of squares test is essentially a test for instability in the variance of the regression error, and the forecast Chow test used by Hendry (1989) is essentially a test for instability in the variance of the regression error as well. The post–sample prediction tests for GMM proposed by Hoffman and Pagan (1989) and Ghysels and Hall (1990) are close to the optimal test proposed here, if the tests are made robust to the choice of the sample split. The optimal tests do not require a sample split, and are best constructed using the full–sample estimates, thus being easy to apply even in non–linear contexts.

1. Introduction

A large number of tests for model stability have been proposed and analyzed by statisticians and econometricians. This analysis, however, has been fragmented, and has failed to draw direct comparisons between the various test methods. This paper attempts to meet this need by analyzing a number of stability tests within a common framework.

Section two outlines the estimators considered, the parameter process and an asymptotic distribution theory for the sequence of partial sample estimators. Explicitly considered are the class of parametric estimators which solve a system of first order conditions, which can be written as a sum across observations. This includes many common econometric estimators, such as non–linear least squares and maximum likelihood. Allowing for two–step estimators, such as generalized least squares, could be handled fairly directly (as in Hansen (1990a)), but are excluded to reduce notation and the complexity of the proofs. Extensions to handle the generalized method of moments estimator is discussed in section 5. The process allowed for the parameter vector is a generalization of the classic Pitman drift for the functional case. The parameter process is assumed to be a constant plus a process which has a weak limit in C[0,1]. This includes (standardized) polynomial trends and random walks. The null distribution (no parameter variation) obtains as a special case. To develop a distribution theory, the regularity conditions allow for heterogeneity and weak dependence, but exclude explosive processes (such as trends or unit roots). This is not simply to make the proofs easier, as the null distributions are different in the non–stationary case (see Hansen (1991)).

Section three develops the tests for parameter instability. The locally most powerful test is presented, under the assumption that the alternative is that the parameter follows a martingale. The test is an average of squared backward cumulative sums of scores (first order conditions). The locally most powerful test requires knowledge of the parameter vector; some sample estimates are required to implement the test. We consider tests based upon the full sample estimator and the sequence of partial sample estimators. The latter can be written in a recursive or a prediction format.

Section four develops a distributional theory for the test statistics. The asymptotic distribution of the test statistics under the local alternative process is derived. It is shown that all of the tests have asymptotic local power against instability in the parameters for which the tests were intended. The asymptotic power functions are displayed, and it is shown that the test based on the full–sample estimates is uniformly most powerful. Since this is also the easiest test to calculate, it emerges as the clear winner.

Section five analyzes several popular tests for parameter instability. The CUSUM test of Brown, Durbin and Evans (1975) is shown to be a test for instability in the intercept. Their CUSUM of squares test is shown to be a test for instability in the variance of the regression error. Likewise for the prediction Chow test advocated by Hendry (1989). The post–sample prediction tests for GMM proposed by Hoffman and Pagan (1989) and Ghysels and Hall (1990) are close to the optimal test proposed here, if the tests are made robust to the choice of the sample split.

Concerning notation, $\Rightarrow$ denotes weak convergence of associated probability measures with respect to the uniform metric, C[0,1] denotes the space of continuous functions on [0,1], and BM(V) denotes a Brownian motion with covariance matrix V. When applied to matrices $|A| = \max_{i,j} |A_{ij}|$, and $\|A\|_p = \left[E(\max_{i,j} |A_{ij}|)^p\right]^{1/p}$. Throughout the paper, the letters r and s denote real numbers in $[0,1]$. The notation "nr" and "ns" will be used as a shorthand for "the integer part of nr" and "the integer part of ns", respectively.

2. Local Parameter Instability and Recursive Parameter Estimation

An econometrician has some time series $\{x_1,\ldots,x_n\}$ and wishes to estimates some economic model, with unknown parameter vector $\theta \in \Theta \subset \mathbb{R}^k$. The econometrician elects to estimate the model by solving some optimization problem. In most cases, this means that the estimator solves a system of first order conditions. Assume that they can be written in the linear form

$$\text{FOC} = \sum_{i=1}^{n} m_i(\theta) \quad . \tag{1}$$

For simplicity, we only will explicitly examine one–step estimators. Allowing for two–step estimators is a straightforward generalization which provides no additional benefits while complicating the notation and proofs. The class of estimation problems covered by (1) is quite broad, including linear and non–linear regression, and maximum likelihood. To include generalized method of moments, we could generalize our results by allowing m_i to depend upon sample size as well. All of the results in this paper extend to this case, but complicate the notation while adding little to our comprehension.

The full sample estimator of the unknown parameter θ, denoted by $\hat{\theta}$, sets the FOC equal to zero. Therefore

$$\sum_{i=1}^{n} m_i(\hat{\theta}) = 0 \ .$$

Our concern is estimation in the potential presence of parameter instability. It seems natural in this context to examine the behavior of $\hat{\theta}$ as the sample size changes. It is convenient to display the dependence of $\hat{\theta}$ upon sample size by explicitly examining partial–sample estimators, which are the sequence of estimators $\{\hat{\theta}_t : t = n_1,\ldots,n\}$ based on partial–sample information. Note that there is some minimum sample size, n_1, below which the econometrician will not define a partial sample estimator. This minimum sample size needs to at least equal the number of estimated parameters; and in practice may be considerably higher, for it may not "make sense" to define a partial sample estimator with just a few degrees of freedom. For example, it may be considered reasonable that the minimum number of degrees of freedom equal no less than 10% of the total sample. Define this proportion $a = n_1/n$. The choice of a is important in practice, and this is reflected in the asymptotic theory which follows, as we assume that a remains constant as $n \to \infty$, so that n_1 is assume to grow with sample size.

These estimators will satisfy the sequence of first order conditions:

$$\sum_{i=1}^{t} m_i(\hat{\theta}_t) = 0 . \tag{2}$$

It is well understood how to develop an asymptotic distributional theory for the full–sample estimator $\hat{\theta}$ under the assumption of parameter constancy. It is also fairly well understood how to develop an asymptotic distributional theory for the partial–sample parameter sequence $\{\hat{\theta}_t\}$, using stochastic process theory, again under the assumption of parameter stability. See, for example, Andrews (1990) or Hansen (1990a).

What are the implications of parameter instability? Many forms of parameter instability are plausible. My intention is to develop a framework for parameter instability which may be thought of as a local deviation from a constant parameters. The natural way to think of local departures is to use Pitman drift. The classic Pitman drift specifies the parameter as a local approximation to the null value, such as

$$\theta = \theta_0 + \frac{1}{\sqrt{n}} c .$$

This allows construction of an asymptotic theory in which there is a smooth transition between the null and alternative. In order to think of generalizing the classic Pitman drift to incorporate parameter instability, we need to have an analogy for stochastic processes. A tractable specification is

(2) $$\theta_i = \theta_o + \frac{1}{\sqrt{n}} Z_i , \qquad Z_{nr} \Rightarrow Z(r) \in C[0,1] .$$

This specification is convenient for it allows the development of an asymptotic theory in which both constant parameter and time–varying parameters appear as special cases.

The specification of $Z(r)$ as an element of $C[0,1]$ requires the parameter process Z_i to be relatively smooth. For example, $Z(r)$ may be a linear (or polynomial) trend, a Brownian motion, or a diffusion process. Discontinuous processes, such as structural breaks, are excluded. This exclusion is probably not necessary, but facilitates the use of currently available proof techniques.

We want to consider estimation of θ using the estimator which solves (1) when the parameter follows process (2). By the latter, we will interpret this to mean that in our statements of regularity conditions, we will center functions of the data not at θ_o, as would be conventionally done in asymptotic theory, but at θ_i, the value of the parameter at time i. Define $\mathcal{N}$ to be some neighborhood of θ_o, $M_i(\theta) = \frac{\partial}{\partial \theta'} m_i(\theta)$, .

<u>Assumption A</u>.

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

(b) $\sup_{r \in [a,1]} |\hat{\theta}_{nr} - \theta_o| \rightarrow_p 0$.

(c) $V = \lim_{n \to \infty} \frac{1}{n} E \sum_{i=1}^{n} \left[m_i(\theta_i) m_i(\theta_i)' \right]$ exists and is finite.

(d) $\frac{1}{\sqrt{n}} \sum_{i=1}^{nr} m_i(\theta_i) \Rightarrow B(r) \equiv BM(V)$.

(e) $\frac{1}{n} \sum_{i=1}^{n} M_i(\theta_i) \longrightarrow M$ a.s. and $M > 0$.

(f) $\sup_{t \leq n} \frac{1}{n} \sum_{i=1}^{t} \sup_{\theta \in \mathcal{N}} \| \frac{\partial}{\partial \theta} M_i^{ab}(\theta) \| = O_p(1)$.

(g) $M_i(\theta)$ is strong mixing of size $-2p/(p-1)$, for some $p > 2$.

(h) $\sup_{i \leq n} \| \sup_{\theta \in \mathcal{N}} |M_i(\theta)| \|_{2(p+\delta)} < \infty$ for some $\delta > 0$.

(i) $\sup_{i \leq n} \|Z_i - Z_{i-1}\|_{2(p+\delta)} \leq c/n$, for some $c < \infty$.

Theorem 2.1. Under (1), (2) and assumption A, for $r \in [a,1]$,

$$\sqrt{n}\,(\hat{\theta}_{nr} - \theta_o) \Rightarrow -\frac{1}{r} M^{-1}B(r) + \frac{1}{r}\int_0^r Z$$

(Proofs are in the appendix.)

Theorem 2.1 gives a characterization of the behavior of sequential estimation in the possible presence of parameter variation. Suppose that there is no parameter instability. Then the function $Z(r)$ is identically zero, and the parameter estimates converge to a scaled version of $W(r)/r$. Note that for all r , $W(r)/r \equiv N(0,1/r)$. Thus, as expected, the variance of the partial sample estimator decreases as sample size increases. The need to bound r away from zero becomes apparent as $1/r \to \infty$, so the partial sample estimator is ill–behaved in extremely small samples.

The most interesting result of Theorem 2.1 concern the behavior of the partial sample estimator in the presence of local departures from parameter stability. A leading example is random walk parameter variation. In this case, $Z(r)$ is a Brownian motion and $\int_0^r Z$ is an *integrated* Brownian motion, a continuous time analog of an I(2) process. A straightforward calculation reveals that for all r , $\int_0^r Z \equiv N(0\ ,\ G\, r^3/3)$, for some covariance matrix G , so $r^{-1}\int_0^r Z \equiv N(0\ ,\ G\, r/3)$.

Thus this component of the asymptotic distribution has a variance which *grows* with sample size.

Consider the effect of local parameter variation in a parameter subset. That is, make the partition $\theta = (\theta^1 , \theta^2)$ and assume that Z_i^1 is a random walk, and Z_i^2 is identically zero. Then the asymptotic distribution of $\hat{\theta}_t^2$ is conventional, while the asymptotic distribution of $\hat{\theta}_t^1$ contains the divergent term. This is interesting for it suggests that local deviations from constancy in some parameters only affects the distribution of the estimators of those parameters, and not the distributions of the estimators of the other parameters. Estimation and inference can proceed on the stable parameters, as if all the other parameters were indeed stable. This fact also points out the need for extreme care in selecting tests for parameter instability. In order to detect local departures from stability, it is necessary to examine the parameter estimates of interest, for examination of a subset of the parameters will not contain information as to whether or not the other parameters are stable.

3. A General Approach to Testing for Instability

We now turn to optimal tests of parameter stability. Tests of stability may be constructed to assess the stability of the entire parameter vector, or just a subset. The former are called tests of pure parameter instability and the latter are called tests of partial parameter instability. Partition θ as (θ^1 , θ^2), and take θ^1 to be the subset of θ to test for instability. (For a test of pure parameter instability, set $\theta^1 = \theta$). Denote the dimension of θ^1 by k_1 and the dimension of θ^2 by k_2. Partition $m_i(\theta) = (m_i^1(\theta) , m_i^2(\theta))$ conformably. Denote the true (or pseudo–true) value of θ under the null hypothesis of no parameter instability as θ_0.

In order to develop a test for parameter instability, we need to be more specific about the process Z_i which describes the parameter process in (2). Write Z_i as

$$Z_i = Z_{i-1} + z_i , \qquad Z_0 = 0 ,$$

and partition Z_i and z_i in conformity with θ. It turns out to be particularly convenient to assume that Z_i^1 is a martingale with respect to some increasing sequence of sigma–fields, with covariance matrix

$$E(z_i^1 z_i^{1\prime}) = \tau^2 V_{11}^{-1}/n$$

where V_{11} is the conformable upper–left block of the covariance matrix V from Theorem 2.1.

Formulating the parameter process as a martingale allows for a wide range of behaviors under the alternative. It includes, for example, simple random walks, and parameter processes which display infrequent permanent shifts. It is in fact even possible to construct a martingale with exactly one "structural break" of unknown timing in each sample.

The null hypothesis of no parameter change is given by

$$H_0: \ \tau^2 = 0 \qquad\qquad H_1: \ \tau^2 > 0 \ .$$

In the context of maximum likelihood, Nyblom (1989) has provided the locally most powerful test statistic for H_0 against H_1. It is given by

$$L = n^{-2} \sum_{t=1}^{n} B_t^1(\theta_o)' V_{11}^{-1} B_t^1(\theta_o)$$

where

$$B_t(\theta) = \sum_{i=t}^{n} m_i(\theta) \ ,$$

and $B_t(\cdot)$ is partitioned conformably with m_i . This statistic may be derived as an approximation to the Lagrange multiplier test. In non–MLE contexts, L is not necessarily the locally most powerful test, but it can be interpreted as an "LM–like" statistic in the conventional way.

In maximum likelihood, the variables $m_i(\theta)$ are known as the scores. The variable $B_t(\theta)$ is therefore a partial sum of scores, but descending from n to 1 . We could call this a *backward cumulative score.* We will retain this terminology even in non–MLE contexts. Note that if we define a *forward cumulative score* as

$$S_t(\theta) = \sum_{i=1}^{t} m_i(\theta) \ ,$$

then we can obviously write the statistic as either a function of the forward cumulative score as well, since

$$B_t(\theta) = S_n(\theta) - S_{t-1}(\theta) \ .$$

The locally most powerful test cannot be directly implemented because the true value θ_o is unknown. Nyblom (1989) suggests using the MLE $\hat{\theta}$:

$$\hat{L} = n^{-2} \sum_{t=1}^{n} B_t^1(\hat{\theta})' \hat{V}_{11}^{-1} B_t^1(\hat{\theta}) \ .$$

Since the first order conditions guarantee that $S_n^1(\hat{\theta}) = 0$, this statistic equals

$$\hat{L} = n^{-2} \sum_{t=1}^{n} S_t^1(\hat{\theta})' \hat{V}_{11}^{-1} S_t^1(\hat{\theta}) ,$$

so when the full–sample MLE is used, the statistic may be equivalently thought of as constructed from forward or backward cumulative scores. To simplify the notation, we can write the statistic as

$$\hat{L} = n^{-2} \sum_{t=1}^{n} \hat{S}_t^{1\prime} \hat{V}_{11}^{-1} \hat{S}_t^1, \qquad \hat{S}_t^1 = \sum_{i=1}^{t} \hat{m}_i^1 , \qquad \hat{m}_i = m_i(\hat{\theta}) .$$

To render the distribution invariant to nuisance parameters and robust to heteroskedasticity, Hansen (1990a) suggests taking

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \hat{m}_i \hat{m}_i' .$$

With $\hat{V}_{11}$ the upper–left block of this matrix.

The replacement of θ_0 with the MLE $\hat{\theta}$ is a computationally simple procedure, but it is not the only option. The partial–sample estimators can also be used in some fashion. The following configurations seem promising. The *recursive* estimates of the first–order conditions are given by

$$\tilde{m}_i = m_i(\hat{\theta}_{i-1}) .$$

Denote by $n^* = n - n_1 = n - n\alpha$ the number of observations for which $\tilde{m}_i$ is defined. We have the test statistics

$$\tilde{L}_b = n^{*-2} \sum_{t=n\alpha}^{n} \tilde{B}_t^{1\prime} \tilde{V}_{11}^{-1} \tilde{B}_t^1, \qquad \tilde{B}_t = \sum_{i=t}^{n} \tilde{m}_i$$

and

$$\tilde{L}_f = n^{*-2} \sum_{t=n\alpha}^{n} \tilde{S}_t^{1\prime} \hat{V}_{11}^{-1} \tilde{S}_t^1, \qquad \tilde{S}_t = \sum_{i=n\alpha}^{t} \tilde{m}_i .$$

Note that we have specified test statistics as functions of both the forward cumulative scores as well as the backward cumulative scores. The theory suggests the use of the latter, but as we will see, some commonly used test statistics bear greater resemblance to the forward cumulative scores.

We also have the *prediction error* formulation:

$$\bar{L} = n^{*-2} \sum_{t=na}^{n} \bar{B}_t^{1\prime} \hat{V}_{11}^{-1} \bar{B}_t^1 , \qquad \bar{B}_t = \sum_{i=t}^{n} m_i(\hat{\theta}_{t-1}) .$$

Although it would be possible to define a version of $\bar{L}$ using the forward cumulative scores, there seems no particular reason to do so.

It is useful to reflect upon these four feasible formulations of the test statistic. All are based upon the average of the squared cumulative scores, where the cumulative scores are evaluated at different parameter estimates. Statistic $\hat{L}$ evaluates all the scores (first–order conditions) using the full–sample parameter estimates, which should be the most efficient estimate of the parameter value under the null hypothesis. Statistics $\tilde{L}_f$ and $\tilde{L}_b$ evaluate each score at the partial sample estimates obtained with only the previous observations. Statistic $\bar{L}$ essentially performs a sample split (for each t). Observations 1 through t–1 are used to obtain the parameter estimate $\hat{\theta}_{t-1}$, and then this estimate is used to center the cumulative scores over observations t to n .

4. Asymptotic Power Functions

None of the implementable tests discussed in section 3 is actually the locally most powerful test, because sample estimates are used. Which of the several tests proposed is the most powerful is therefore unclear *a priori*. To compare the tests, the first place to start is with an examination of local asymptotic power. This abstracts from complications due to small samples, dynamics, non–linearities and non–normal distributions. One may think of the asymptotic local power function as a first–order approximation to the actual power function.

The asymptotic power functions are simple functions of the asymptotic distributions of the test statistics under the local alternative process. Therefore, we simply provide the latter.

First, we provide the distribution of $\hat{L}$. Partition

$$M = \begin{bmatrix} M^{11} & M^{12} \\ M^{21} & M^{22} \end{bmatrix}$$

<u>Theorem 4.1</u>. Under (1), (2) and assumption A,

(a) $$\frac{1}{\sqrt{n}} \hat{S}_{nr} = \frac{1}{\sqrt{n}} \sum_{i=1}^{nr} \hat{m}_i \Rightarrow B^*(r) - MZ^*(r)$$

(b) $$\hat{L} \to_d \int_0^1 P^*(r)' P^*(r) dr ,$$

where $P^*(r) = W^*(r) + V_{11}^{-1/2}\left[M^{11}Z^{*1}(r) + M^{12}Z^{*2}(r)\right]$, $B^*(r) = B(r) - rB(1)$, a Brownian bridge, $Z^*(r) = \int_0^r Z - r\int_0^1 Z$, a tied–down integrated Brownian motion, and $W^*(r)$ is a standard Brownian bridge of dimension k_1 .

Theorem 4.1 provides the asymptotic distribution of $\hat{L}$ under the null hypothesis and local alternatives. Under the null, the distribution is $\int_0^1 W^{*\prime} W^*$, which is a multivariate generalization of the Smirnov goodness–of–fit distribution.

How can we use Theorem 4.1 to evaluate the power of the test statistic? I think that it is easiest if we simplify a moment and assume that the scores are orthogonal. This arises in linear regression, for example, if the regressors are orthogonal. In this case, M is the identity matrix, $M^{11} = I$, and $M^{12} = 0$, yielding $P^*(r) = W^*(r) + V_{11}^{-1/2} Z^{*1}(r)$. The asymptotic distribution only depends upon the non–stationary process Z^1, but not upon Z^2, so the test will have power against movements in θ^1, but not θ^2, as should be expected from the design of the test. It is important to note that when the scores are correlated (the design matrix is not orthogonal) then the test statistic will have *some* power against local departures in θ^2 as well. This power will depend upon the degree of correlation, as well as the nature of the instability in θ^2 .

We can also derive the asymptotic distributions for $\tilde{L}$ and $\bar{L}$.

<u>Theorem 4.2</u>.

(a) $$\frac{1}{\sqrt{n^*}} \tilde{S}_{nr} = \frac{1}{\sqrt{n^*}} \sum_{i=na}^{nr} \tilde{m}_i \Rightarrow \frac{1}{\sqrt{1-a}} \left[\tilde{B}(r) - \tilde{B}(a) - M[\tilde{Z}(r) - \tilde{Z}(a)]\right] ,$$

for $r \in [a,1]$;

(b) $$\tilde{L}_f \to_d \int_0^1 \tilde{P}(r)' \tilde{P}(r) dr ,$$

(c) $$\tilde{L}_b \to_d \int_0^1 \left[\tilde{P}(1) - \tilde{P}(r)\right]' \left[\tilde{P}(1) - \tilde{P}(r)\right] dr$$

where $\tilde{B}(r) = B(r) - \int_0^r \frac{1}{s} B(s) ds$, $\tilde{Z}(r) = \int_0^r Z - \int_0^r \frac{1}{s} (\int_0^s Z(\lambda) d\lambda) ds$,

$\tilde{P}(r) = W(r) + V_{11}^{-1/2}\left[M^{11}\tilde{\tilde{Z}}^1(r) + M^{12}\tilde{\tilde{Z}}^2(r)\right]$, $\tilde{\tilde{Z}}(r) = [\tilde{Z}(a + r(1-a)) - \tilde{Z}(a)]/\sqrt{1-a}$,

and $W(r) \equiv BM(I_{k_1})$.

Theorem 4.3.

(a) $$\frac{1}{\sqrt{n}}\,\bar{B}_{nr} = \frac{1}{\sqrt{n^*}}\sum_{i=nr}^{n} m_i(\hat{\theta}_{nr-1}) \implies \bar{B}(r) - M\bar{Z}(r)$$

(b) $$\bar{L} \rightarrow_d \frac{1}{(1-a)^2}\int_a^1 \bar{P}'\bar{P}$$

where $\bar{B}(r) = B(1) - r^{-1}B(r)$, $\bar{Z}(r) = \int_0^1 Z - r^{-1}\int_0^r Z$, $\bar{W}(r) = W(1) - r^{-1}W(r)$, $W(r) \equiv BM(I_{k_1})$ and $\bar{P}(r) = \bar{W}(r) + V_{11}^{-1/2}\left[M^{11}\bar{Z}^1(r) + M^{12}\bar{Z}^2(r)\right]$.

Theorems 4.2 and 4.3 show the differences in the asymptotic behavior of the various tests considered. First, the distributions under the null hypothesis are different. The null distribution of $\hat{L}$ depends upon a Brownian bridge, $\tilde{L}_f$ and $\tilde{L}_b$ depend upon a Brownian motion, and $\bar{L}$ depends upon the process $\bar{B}(r) = B(1) - \frac{1}{r}B(r)$, which I haven't seen before. I will return to a discussion of the latter process momentarily. All null distributions depend upon k_1 , the number of parameters tested for stability. The null distribution of $\hat{L}$ depends upon nothing else. The distributions of $\tilde{L}_f$, $\tilde{L}_b$ and $\tilde{L}$ depend as well upon a , the proportion of the sample excluded from testing. This slightly complicates the tabulation of critical values.

The process $\bar{B}(r)$ is fairly different from either a Brownian motion or a Brownian Bridge. Its covariance function is $E[\bar{B}(r)\bar{B}(s)] = \frac{1-s}{s}$ for $r \leq s$, which is independent of r . The fact that $Var(\bar{B}(r)) = \frac{1-r}{r}$ means that $Var(\bar{B}(r)) \to \infty$ as $r \to 0$, so the choice of the proportion parameter a appears to be quite crucial.

The power functions of the four test statistics are different in form, but similar in content. All statistics apparently have asymptotic local power against movements in the same parameters, but the expressions give little guidance as to which tests should have the greatest local power. Note as well that the power functions of $\tilde{L}_f$,

$\tilde{L}_b$ and $\tilde{L}$ depend upon a .

Since we understand what the asymptotic power functions depend upon, we can calculate the functions using simulation methods. The null distributions of $\hat{L}$, $\tilde{L}_f$, and $\tilde{L}_b$ are known; see, for example, Nyblom (1989). The distribution of $\bar{L}$ is not. I calculated the null values for $a = 0.05$, 0.1, and 0.2 (using 20,000 samples of size 1000). To calculate the power functions, I set $k_1 = 1$, and took 2,000 samples of size 500. The statistics $\tilde{L}_f$, $\tilde{L}_b$ and $\bar{L}$ were calculated using the proportions $a = 0.05$, 0.1, and 0.2 . The alternative used was a Gaussian random walk, and the power of a 5% size test was calculated for 20 values of τ , from $\tau = 1$ to $\tau = 20$. The results are displayed in Figures 1, 2 and 3. The power function of $\hat{L}$ was uniformly most powerful, so all other tests are displayed relative to it. Figure 1 displays the power functions of $\tilde{L}_f$, figure 2 displays those of $\tilde{L}_b$, and figure 3 displays those of $\bar{L}$. The figures show that $\hat{L}$ is the most powerful test and $\tilde{L}_f$ is the least powerful. The power functions of $\tilde{L}_b$ are quite close to that of $\hat{L}$, especially for small τ . Figure 2 also suggests that the choice of a is not critical for $\tilde{L}$. The power functions for $\bar{L}$ are fairly close to $\hat{L}$, but show some dependence upon the choice of a , with the power increasing with a .

Overall, the calculations show that $\hat{L}$, $\tilde{L}_b$ and $\bar{L}$ will all have similar power against the same alternatives, with $\hat{L}$ having the best performance. Since $\hat{L}$ is also the simplest statistic to calculate, this combines to yield a strong argument in its favor. The figures show that reversing the direction of the cumulative sum, to become a forward sum, will have adverse effects upon power.

For convenience, asymptotic critical values for $\hat{L}$ are provided in Table 1.

5. Popular tests for Instability

One useful feature of our present framework is that it allows us to examine popular tests for instability. We can discover their asymptotic power functions by determining on which *scores* they are based.

5.1 The CUSUM and CUSUMSQ Tests

In a seminal paper, Brown, Durbin, and Evans (or BDE) (1975) proposed the CUSUM procedure for testing for instability. In a linear regression

$$y_i = \alpha + \beta' x_i + e_i$$

BDE suggested forming the recursive residuals

$$\hat{e}_i = y_i - \hat{\alpha}_{i-1} - x_i' \hat{\beta}_{i-1}$$

where $(\hat{\alpha}_{i-1}, \hat{\beta}_{i-1})$ are the partial–sample estimates using the data up to time i. Then BDE suggested taking the forward and backward cummulative sums of scores

$$S_t = \sum_{i=n_1}^{t} \hat{e}_i \qquad B_t = \sum_{i=t}^{n} \hat{e}_i \ .$$

The authors suggested plotting S_t and/or B_t and (informally) rejecting stability if these cumulative sums were "too large".

Our analysis of this problem is now quite simple.[1] The residual is the derivative of the least squares objective function with respect to the intercept. The cumulative sums of the recursive residuals are thus the cumulative scores evaluated at the recursive estimates, and the asymptotic local power function is given in Theorem 4.2. We see that the test has good power against instability in the intercept, *but no asymptotic power against movements in any zero–mean regressor.* The test will have *some* power against movements in a non–zero–mean regressor, but the power will be a

[1]The power function of the CUSUM test has been derived independently by Kramer, Ploberger, and Alt (1987).

function of this mean. The CUSUM test was intended as a general test for instability. Its asymptotic power function tells a different story.

BDE also proposed a test involving the squared recursive residuals, often called the CUSUM of squares, or CUSUMSQ, test. The test is based upon the sequence

$$F_t = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_n^2} - 1$$

where $\hat{\sigma}_t^2 = \frac{1}{t}\Sigma_{n_1}^t \hat{e}_i^2$. Since

$$F_t = \hat{\sigma}_n^{-2}\left[\hat{\sigma}_t^2 - \hat{\sigma}_n^2\right] = \hat{\sigma}_n^{-2}\frac{1}{t}\sum_{i=n_1}^{t}(\hat{e}_i^2 - \hat{\sigma}_n^2)$$

we see that F_t is a forward cumulative sum of the centered squared cumulative residuals. But these are simply the scores with respect to the variance of the regression error in the Gaussian linear model! Thus the CUSUMSQ statistic will have asymptotic power against movements in the variance of the regression error. Due to the block–diagonality of the information matrix between the regression coefficients and the variance, the CUSUMSQ statistic will have no asymptotic power against movements in the regression coefficients. This confirms the findings of Ploberger and Kramer (1990) who used different techniques.

The bottom line is that neither test are expected to have useful power against movements in the slope parameters.

5.2 GMM Post–Sample Prediction Tests

Recently, Hoffman and Pagan (1989) and Ghysels and Hall (1990) independently proposed a test for structural stability in the context of estimation by the generalized methods of moments (GMM). To allow for GMM estimation in our framework, we have to extend the notation. The first order conditions which replace (1) are now

$$\text{FOC} = \sum_{i=1}^{n} m_{ni}(\theta)$$

so the "scores" now depend upon sample size as well. In the GMM case, θ is chosen to minimize a function of the form

$$Q_n(\theta) = g(\theta)' W \, g(\theta)$$

where

$$g(\theta) = \Sigma_1^n g_i(\theta)$$

$E\left[g_i(\theta) \mid \mathcal{F}_{i-1}\right] = 0$, and W is some weight matrix. Here, we have

$$m_{ni}(\theta) = \left[\Sigma_1^n \frac{\partial}{\partial\theta} g_i(\theta)'\right] W(\theta) \, g_i(\theta)$$

All the statistics are defined as before using m_{ni} in place of m_i . The Hoffman–Pagan and Ghysels–Hall statistic are based on quadratic forms in

$$\bar{B}_t = \Sigma_t^n m_{ni}(\hat{\theta}_{i-1})$$

where $\hat{\theta}_{i-1}$ is the partial sample estimator using information up to time t . This is the backward cumulative score using the prediction error formulation. These authors take the sample split t as known and fixed, and the authors acknowledge this as a deficiency of the test. The power of the test may still be reasonable, as we know from Theorem 4.3 that the $\bar{B}_t$ will reflect movements in the parameters vector.

These results also suggest how to improve the tests suggested by Hoffman–Pagan and Ghysels–Hall. First, the tests can be made agnostic with respect to *a priori* selection of a particular breakpoint by using the test statistic $\hat{L}$ or $\bar{L}$. Second, $\hat{L}$ should be used to reduce the computational burden. As pointed out by Ghysels–Hall, however, $\bar{L}$ would be computationally burdensome, since estimation of $\hat{\theta}_t$ for each t requires iteration. Note that $\hat{L}$ is *less* computationally demanding than even the breakpoint statistics suggested by Hoffman–Pagan and Ghysels–Hall. On *every* criterion, the statistic $\hat{L}$ dominates.

5.3 Hendry's Forecast Chow Test

Hendry (1989) has advocated the use of a particular forecast stability test, and has made it one of his standard diagnostic statistics. I will describe the test as I understand briefly. The model is

$$y_i = x_i'\beta + e_i$$

estimated by least squares. The partial sample estimators $\{\hat{\beta}_t\}$ are formed, and then for each t, the vector

$$\tilde{e}_t = \left[e_t(\hat{\beta}_{t-1}), e_{t+1}(\hat{\beta}_{t-1}), \ldots, e_n(\hat{\beta}_{t-1})\right]'$$

is formed. $\tilde{e}_t$ consists of the forecast errors from period t through n, using the partial sample estimate obtained at time $t - 1$. The idea is to see of $\tilde{e}_t$ is "close" to zero as a vector. The natural statistic for this is

$$F_t = \tilde{e}_t'\tilde{V}_t^{-1}\tilde{e}_t$$

where $\hat{V}_t$ is an estimate of the covariance matrix for $\tilde{e}_t$. This statistic is calculated for each t, scaled by the 5% (say) critical value from a chi–square (or F) distribution with $n - t$ degrees of freedom, and plotted for all $t \geq n_1$.

In large samples, the uncertainty in the partial sample estimators will diminish, so the matrices V_t will approach scaled identity matrices (of dimension $n-t$), where the scale gives the variance of the regression error. Thus in large samples, F_t is essentially a cumulative sum of squared residuals from period t to period n, where the residuals are evaluated at the parameter estimates obtained in period $t-1$. Thus F_t is essentially an element of our prediction error cumulative sum of scores, $\bar{B}_t$, the element corresponding to the variance of the regression error.

We can now see the strengths and weaknesses of this test. First, concerning size. Of course the test as currently used is hard to evaluate, since an eyeball metric is used to gauge the "significance" of the sequence of F–tests. This could be made rigorous using our theory, so is not a big concern. Less obvious, however, is the size distortion induce by the use of the chi–square/F distributional approximation. This cannot be shown to be an asymptotic approximation, but instead depends upon distributional assumptions for the errors. This could be relaxed by using a normal approximation for the centered sum of squared residuals.

Second, and more importantly, we have to examine power. The test will have good power against local movements in the variance of the regression error. But the test will have no asymptotic local power against movements in the other parameters. Although Hendry and his coauthors have been among the most vocal in their advocacy of testing the assumption of parameter stability, we can now see that the specific test advocated was not a particularly useful choice.

5. Conclusion

This paper has shown that powerful yet simple tests of the hypothesis of parameter stability are possible. The best tests are obtained by forming backward cumulative sums of scores, centered at the true parameter value. The next best test, it appears, it to center the cumulative sums at the full–sample parameter estimates. This requires no sample splitting. Only in the special case where a researcher is testing for a structural break occurring at some *known* time is a sample split necessary, or useful. Conventional wisdom based upon this special case has led most applied researchers to advocate or use tests which are based on some form of sample split. It is time to abandon this approach.

The performance of the tests was examined using the asymptotic local power function. This should be viewed as a first–order approximation to the actual power function. We found that many popular test statistics have zero asymptotic local power against alternatives of interest. This does not mean that the tests have zero power fixed alternatives of interest. They may well have power, but we would expect their power to be substantial less than the power of the tests with non–zero asymptotic local power. In particular examples this ranking may not obtain, but without further analytic study we should not expect this to be the case.

Appendix

Proof of Theorem 2.1. Take any $r \in [a,1]$. By the first order conditions

$$\text{(A1)} \qquad 0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{nr} m_i(\hat{\theta}_{nr}) .$$

Denote by m_i^a the a'th element of the vector $m_i(\cdot)$. For each a and each i, take a first–order Taylor series expansion about θ_i :

$$\text{(A2)} \quad \frac{1}{\sqrt{n}} \sum_{i=1}^{nr} m_i^a(\hat{\theta}_{nr}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{nr} m_i^a(\theta_i) + \frac{1}{n} \sum_{i=1}^{nr} M_i^a(\theta_i^*) \sqrt{n}(\hat{\theta}_{nr} - \theta_i)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{nr} m_i^a(\theta_i) + \frac{1}{n} \sum_{i=1}^{nr} M_i^a(\theta_{ir}^*) \sqrt{n}(\hat{\theta}_{nr} - \theta_0) - \frac{1}{n} \sum_{i=1}^{nr} M_i^a(\theta_{ir}^*) Z_i$$

where $M_i^a(\cdot)$ is the a'th row of $M_i(\cdot)$, and θ_{ir}^* is a random variable on a line segment joining $\hat{\theta}_{nr}$ and θ_i .

Taking another set of Taylor's series expansions,

$$\text{(A3)} \qquad \frac{1}{n} \sum_{i=1}^{nr} M_i^{ab}(\theta_{ir}^*) = \frac{1}{n} \sum_{i=1}^{nr} M_i^{ab}(\theta_i) + \frac{1}{n} \sum_{i=1}^{nr} \frac{\partial}{\partial \theta'} M_i^{ab}(\theta_{ir}^{**})(\theta_{ir}^* - \theta_i)$$

$$= \frac{1}{n} \sum_{i=1}^{nr} M_i^{ab}(\theta_i) + o_p(1) \quad \text{uniformly in } r$$

$$\to_p \quad rM^{ab} \quad \text{uniformly in } r$$

under Assumption A (b),(h). The final convergence uses Hansen (1990a), Lemma 1, with Assumption A (e). Theorem 4.1 of Hansen (1990b) shows that

$$\frac{1}{n} \sum_{i=1}^{nr} \sup_{\theta \in \mathcal{N}} \left[M_i(\theta_{ir}^*) - M \right] Z_i \to_p 0$$

uniformly in r , so

$$\text{(A4)} \qquad \frac{1}{n}\sum_{i=1}^{nr} M_i(\theta^*_{ir})Z_i = M\frac{1}{n}\sum_{i=1}^{nr} Z_i + o_p(1) \Longrightarrow M\int_0^r Z(s)ds \ .$$

(A1) – (A4) together yield

$$\sqrt{n}\left[\hat{\theta}_{nr} - \theta_o\right] = -\left[rM + o_p(1)\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i(\theta_i) - M\int_0^r Z\right]$$

$$\Longrightarrow -(rM)^{-1}\left[B(r) - M\int_0^r Z\right] = -\frac{1}{r}M^{-1}B(r) + \frac{1}{r}\int_0^r Z \ . \qquad \square$$

Proof of Theorem 4.1.

(a) By definition

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{nr} \hat{m}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i(\hat{\theta}) \ .$$

Taking element by element Taylor's series expansions

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i^a(\hat{\theta})$$

$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i^a(\theta_i) + \frac{1}{n}\sum_{i=1}^{nr} M_i^a(\theta^*_{ir})\sqrt{n}(\hat{\theta} - \theta_o) - \frac{1}{n}\sum_{i=1}^{nr} M_i^a(\theta^*_{ir})Z_i$$

Using (A3), (A4) and Theorem 2.1,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i(\hat{\theta}) \Longrightarrow B(r) + rM\left[-M^{-1}B(1) + \int_0^1 Z\right] - M\int_0^r Z \ .$$

$$= B(r) - rB(1) - M\int_0^r Z + rM\int_0^1 Z$$

$$= B^*(r) - MZ^*(r) \ .$$

(b) From part (a),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{nr} m_i^1(\hat{\theta}) \Longrightarrow B^{1*}(r) - (M^{11}\ M^{12})Z^*(r) = V_{11}^{1/2}P^*(r).$$

Therefore

$$\hat{L} = \frac{1}{n}\sum_{t=1}^{n} \left[\frac{1}{\sqrt{n}}\Sigma_1^t m_i^1(\hat{\theta})\right]'\hat{V}_{11}^{-1}\left[\frac{1}{\sqrt{n}}\Sigma_1^t m_i^1(\hat{\theta})\right]$$

$$\Rightarrow \int_0^1\left[V_{11}^{1/2}P^*(r)\right]'V_{11}^{-1}\left[V_{11}^{1/2}P^*(r)\right]dr = \int_0^1 P^*(r)'P^*(r)dr$$

by the continuous mapping theorem. □

Proof of Theorem 4.2

(a) By definition

$$\frac{1}{\sqrt{n^*}}\sum_{i=na}^{nr} \tilde{m}_i = [(1-a)n]^{-1/2}\sum_{i=na}^{nr} m_i(\hat{\theta}_{i-1}) .$$

Taking element by element Taylor's series expansions

$$[(1-a)n]^{-1/2}\sum_{i=na}^{nr} m_i^a(\hat{\theta}_{i-1})$$

$$= \frac{1}{\sqrt{1-a}}\left[\frac{1}{\sqrt{n}}\sum_{i=na}^{nr} m_i^a(\theta_i) + \frac{1}{n}\sum_{i=na}^{nr} M_i^a(\theta_{ir}^*)\sqrt{n}(\hat{\theta}_{i-1} - \theta_o) - \frac{1}{n}\sum_{i=na}^{nr} M_i^a(\theta_{ir}^*)Z_i\right]$$

where θ_{ir}^* is on a line segment joining $\hat{\theta}_{i-1}$ and θ_i . Using Theorem 2.1

$$\text{(A5)}\quad \frac{1}{n}\sum_{i=na}^{nr} M_i^a(\theta_{ir}^*)\sqrt{n}(\hat{\theta}_{i-1} - \theta_o) \Rightarrow \int_a^r M\left[-\frac{1}{s}M^{-1}B(s) + \frac{1}{s}\int_0^s Z\right]ds$$

$$= -\int_a^r s^{-1}B(s)ds + M\int_a^r(s^{-1}\int_0^s Z)ds .$$

Using (A3), (A4) and (A5),

$$\frac{1}{\sqrt{n}}\sum_{i=na}^{nr} m_i(\hat{\theta}_{i-1}) \Rightarrow B(r) - B(a) - \int_a^r s^{-1}B(s)ds + M\int_a^r(s^{-1}\int_0^s Z)ds - M\int_a^r Z$$

$$= \tilde{B}(r) - \tilde{B}(a) - M\left[\int_a^r Z - \int_a^r(s^{-1}\int_0^s Z)ds\right] .$$

(b) Set

$$S_n(r) = \frac{1}{\sqrt{n^*}} V_{11}^{-1/2} S_{na+rn^*} .$$

Then by part (a)

$$S_n(r) \Rightarrow S(r\frac{1}{\sqrt{1-a}}\left[\tilde{B}(a+r(1-a)) - \tilde{B}(a) - M\left[\int_a^{r(1-a)}Z - \int_a^{r(1-a)}(s^{-1}\int_0^s Z)ds\right]\right].$$

Set $W(r) = V_{11}^{-1/2}[\tilde{B}(a+r(1-a)) - \tilde{B}(a)]/\sqrt{1-a}$. A routine but tedious calculation can show that the process $V_{11}^{-1/2}\tilde{B}(r)$ is a mean–zero Gaussian process with the covariance function of a Brownian motion (and therefore is a standard vector Brownian motion.) One can also check that $W(r)$ is also a standard Brownian motion. Now

$$\tilde{L}_f = \sum_{t=na}^{n} \int_{\frac{t-na-1}{n^*}}^{\frac{t-na}{n^*}} dr\ \frac{1}{n^*}S_t^{1\prime}\hat{V}_{11}^{-1}S_t^1$$

$$= \sum_{t=na}^{n} \int_{\frac{t-na-1}{n^*}}^{\frac{t-na}{n^*}} S_n^1(r)'S_n^1(r)\ dr = \int_0^1 S_n^1(r)'S_n^1(r)\ dr \Rightarrow \int_0^1\tilde{P}'\tilde{P}\ .$$

(c) Follows by analysis similar to that of parts (a) and (b). □

Proof of Theorem 4.3

(a) Denote $\hat{\theta}_{nr-1}$ by $\hat{\theta}_r$. The first order conditions are that $\Sigma_1^{nr-1}m_i(\hat{\theta}_{nr-1}) = 0$, giving

$$\text{(A6)} \qquad \frac{1}{\sqrt{n}}\sum_{nr}^{n} m_i(\hat{\theta}_r) = \frac{1}{\sqrt{n}}\sum_{1}^{n} m_i(\hat{\theta}_r) \ .$$

Element by element Taylor series expansions yield

$$\text{(A7)} \quad \frac{1}{\sqrt{n}}\Sigma_1^n m_i^a(\hat{\theta}_r) = \frac{1}{\sqrt{n}}\Sigma_1^n m_i^a(\theta_i) + \frac{1}{n}\Sigma_1^n M_i^a(\theta_{ir}^*)\sqrt{n}(\hat{\theta}_r - \theta_o) - \frac{1}{n}\Sigma_1^n M_i^a(\theta_{ir}^*)Z_i$$

where θ_{ir}^* lies between θ_i and $\hat{\theta}_r$. Using (A3), (A4), and Theorem 2.1, (A7) converges weakly to

$$B(1) \;+\; M\Big[-\tfrac{1}{r}\, M^{-1}B(r) + \textstyle\int_0^r Z\Big] \;-\; M\textstyle\int_0^1 Z$$

$$= \quad B(1) - \tfrac{1}{r}B(r) - M\Big[\textstyle\int_0^1 Z - \tfrac{1}{r}\int_0^r Z\Big] \;.$$

(b) $$\bar{L} \;=\; \frac{1}{n^{*2}} \sum_{t=na}^{n} B_t^{1\prime}\hat{V}_{11}^{-1}B_t^1 \;=\; \frac{1}{(1-a)^2}\,\frac{1}{n} \sum_{t=na}^{n} \int_{\frac{t-1}{n}}^{t/n} dr\; B_t^{1\prime}\hat{V}_{11}^{-1}B_t^1$$

$$= \frac{1}{(1-a)^2} \int_a^1 \Big[\tfrac{1}{\sqrt{n}}B_{nr}^1\Big]^{\prime}\hat{V}_{11}^{-1}\Big[\tfrac{1}{\sqrt{n}}B_{nr}^1\Big] \;\Rightarrow\; \frac{1}{(1-a)^2} \textstyle\int_a^1 \bar{P}(r)^{\prime}\,\bar{P}(r)dr \;.$$

References

Andrews, D.W.K. (1990): "Tests for parameter instability and structural change with unknown change point," Cowles Discussion Paper 943.

Brown, R.L., J. Durbin, and J.M. Evans (1975): "Techniques for testing the constancy of regression relationships over time," *Journal of the Royal Statistical Society, B*, 37, 149–163.

Ghysels, E. and A. 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.

Hansen, B.E. (1990a): "Lagrange multiplier tests for parameter instability in non–linear models," University of Rochester.

Hansen, B.E. (1990b): "Convergence to stochastic integrals for dependent heterogeneous processes," University of Rochester.

Hansen, B.E. (1991): "Tests for parameter instability in regressions with I(1) processes," University of Rochester.

Hendry, D.F. (1989): *PC–Give: An Interactive Econometric Modeling System*, University of Oxford.

Hoffman, D. and A. Pagan (1989): "Post–sample prediction tests for generalized method of moments estimators," *Oxford Bulletin of Economics and Statistics*, 51, 333–343.

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

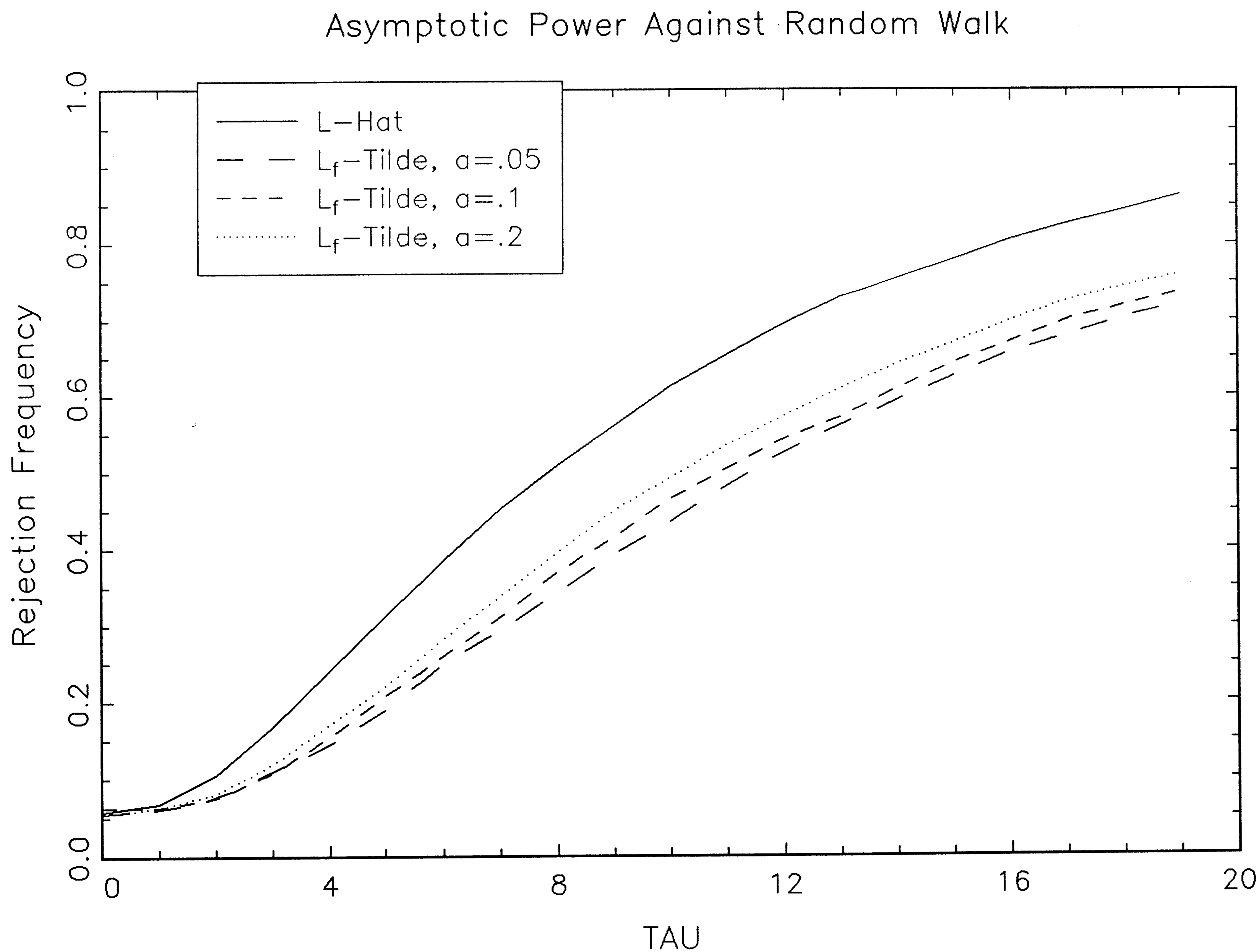
Kramer, W., W. Ploberger, and R. Alt, (1988): "Testing for structural change in dynamic models," *Econometrica*, 56, 1355–1369.

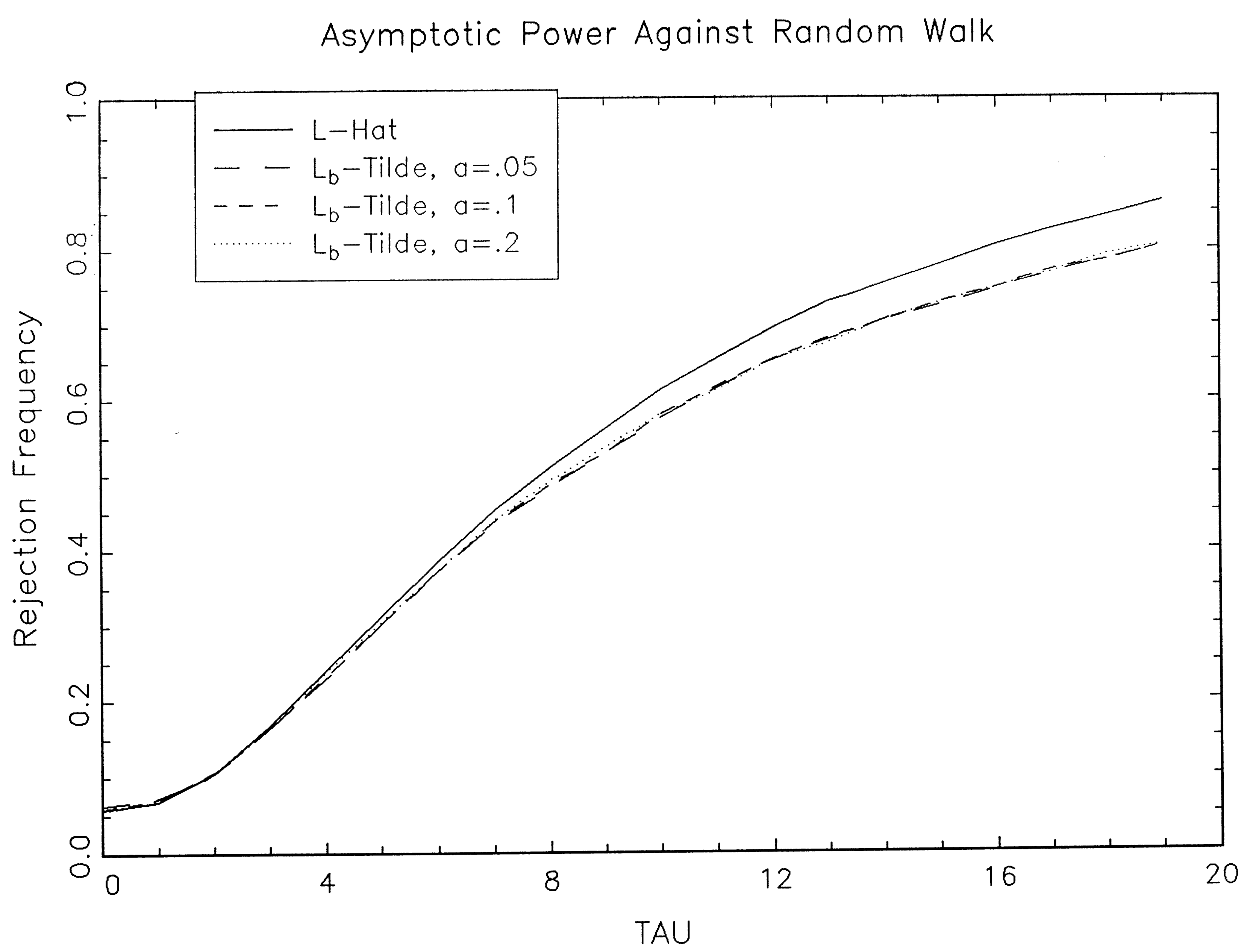
Ploberger, W. and W. Kramer (1990): "The local power of the CUSUM and CUSUM of squares tests," *Econometric Theory*, 6, 335–347.

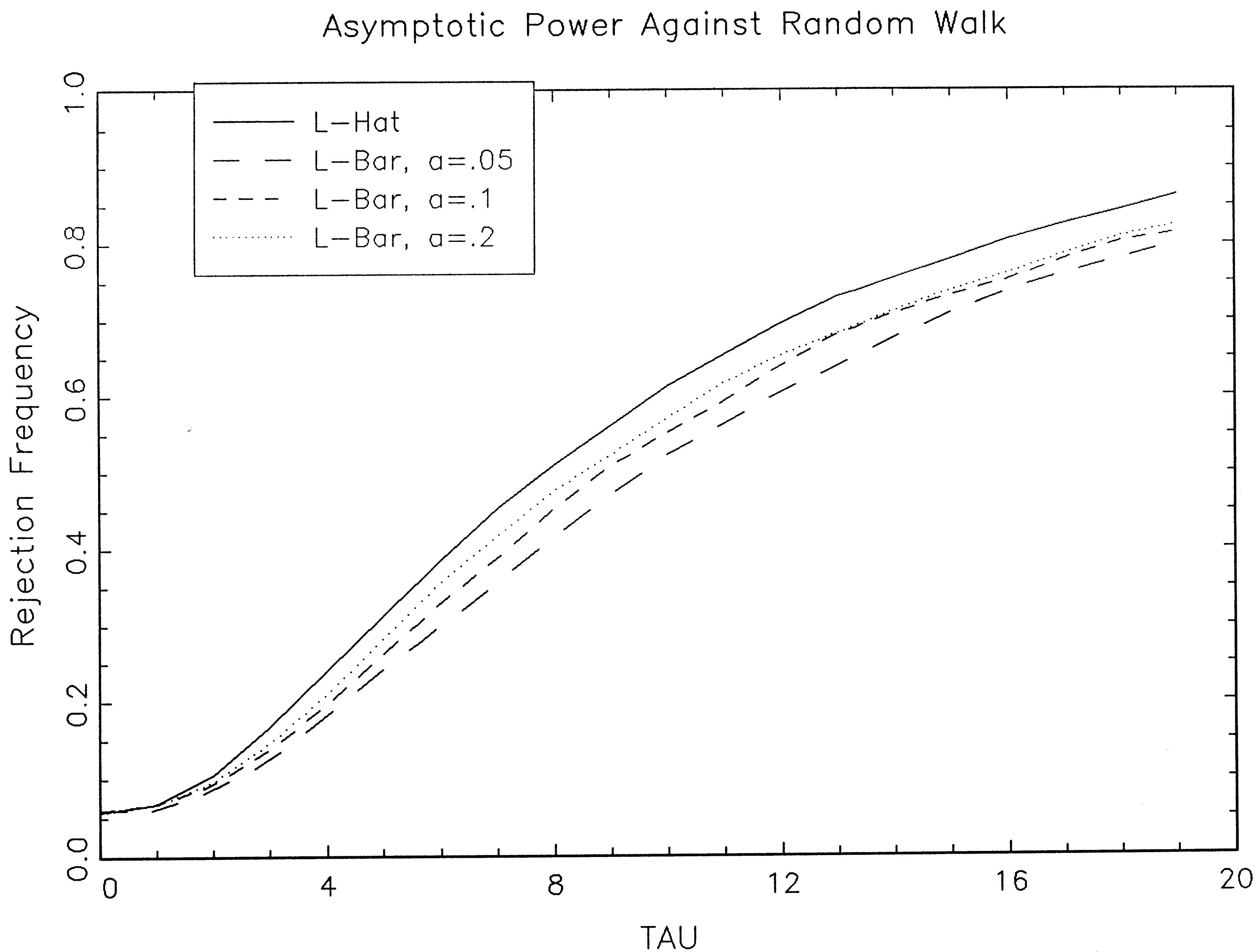
TABLE 1: ASYMPTOTIC CRITICAL VALUES FOR $\hat{L}$

	Significance Level					
Degrees of Freedom (m+1)	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

Source: Hansen (1990a), Table 1.


Asymptotic Power Against Random Walk
L-Hat
L_f-Tilde, a=.05
L_f-Tilde, a=.1
L_f-Tilde, a=.2
Rejection Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0
4
8
12
16
20
TAU


Asymptotic Power Against Random Walk
L−Hat
L_b−Tilde, a=.05
L_b−Tilde, a=.1
L_b−Tilde, a=.2
Rejection Frequency
0.0
0.2
0.4
0.6
0.8
1.0
0
4
8
12
16
20
TAU


Asymptotic Power Against Random Walk
Rejection Frequency
0.0
0.2
0.4
0.6
0.8
1.0
TAU
0
4
8
12
16
20
L−Hat
L−Bar, a=.05
L−Bar, a=.1
L−Bar, a=.2