

Inference When a Nuisance Parameter is Not Identified Under the Null Hypothesis

Hansen, Bruce E.

Working Paper No. 296
September 1991

University of
Rochester

**Inference when a Nuisance Parameter
is not Identified Under the Null Hypothesis**

Bruce E. Hansen

Rochester Center for Economic Research
Working Paper No. 296

**Inference When a Nuisance Parameter
Is Not Identified Under the Null Hypothesis**

Bruce E. Hansen
University of Rochester

and

The Rochester Center for Economic Research
Working Paper No. 296

January 1991

Revised: September 1991

I would like to thank the NSF for financial support.

ABSTRACT

It is not uncommon to find economists testing hypotheses in models where a nuisance parameter is not identified under the null hypotheses. This paper studies the asymptotic distribution theory for such problems. The asymptotic distributions of test statistics are found to be functionals of *chi-square processes*. In general, the distributions depend upon a large number of unknown parameters. A simulation method is proposed which can calculate the asymptotic distribution. The testing method is applied to the threshold autoregressive model for GNP growth rates proposed by Potter (1991). We present formal statistical tests which (marginally) support Potter's claim that there is a statistically significant threshold effect in a univariate autoregression for U.S. GNP growth rates.

1. INTRODUCTION

This paper studies the problem of inference in the presence of nuisance parameters which are not identified under the null hypothesis. The asymptotic distributions of Wald, likelihood ratio (LR) and Lagrange multiplier-like (LM-like) statistics are obtained for parametric econometric estimators under quite general assumptions, allowing for simultaneous equations, stochastic regressors, heterogeneity, and weak dependence. The asymptotic distributions are shown to be represented by the supremum of a *chi-square process*, a stochastic process which is a quadratic form in a vector Gaussian process indexed by the nuisance parameter. This generalizes the results of Davies (1977, 1987). Unfortunately, these distributions appear to depend, in general, upon the covariance function of the chi square process, which may depend in complicated ways upon the model and data, precluding tabulation. As a proposed remedy, we develop a simulation method which approximates the asymptotic null distribution. This approximation is an improvement over the bounds of Davies (1977, 1987), whose approximation error increases with sample size in many cases of interest.

This paper is organized as follows. Section 2 gives several examples of non-identified nuisance parameters. Section 3 introduces a distinction between global estimates (where the structural and nuisance parameters are estimated jointly) and pointwise estimates (where the structural parameters are estimated for fixed nuisance parameters). Conditions for consistent pointwise estimation of the structural parameters, uniformly in the nuisance parameter, are given. Section 4 develops a theory for testing structural hypotheses when the nuisance parameter is not identified under the null hypothesis. Likelihood ratio, Wald, Lagrange multiplier (LM), and maximal pointwise Wald and LM tests statistics are examined. Section 5 develops an asymptotic distribution theory for the test statistics. This distributions are represented as functionals of *chi-square processes*, which are quadratic forms in mean-zero

Gaussian processes. In the absence of heteroskedasticity and serial correlation, these test statistics have the same asymptotic distribution. A new finding is that only the maximal pointwise Wald and LM statistics have asymptotic distributions which are robust to the presence of heteroskedasticity and serial correlation. The standard LR and Wald statistics, for example, are not robust. Section 6 develops a simulation method which can approximate the null asymptotic distribution. Section 7 extends the results to t -statistics. Section 8 shows how to apply the theory and techniques to threshold models, and reports an application to a threshold autoregressive model of GNP. All proofs are left to the appendix.

Throughout the paper " \Rightarrow " is used to denote weak convergence of probability measures with respect to the uniform metric, and $\|\cdot\|$ denotes the Euclidean metric. Sample size is n .

2. EXAMPLES

It may not be commonly understood how pervasive is the problem of unidentified nuisance parameters. I list below a few examples taken from the recent literature. In most of the following examples, the model has been parameterized so that the null and alternative hypotheses are

$$H_0 : \theta = 0 \qquad H_1 : \theta \neq 0$$

and the nuisance parameter γ is not identified under H_0 .

In this situation, an error commonly made in applied research is the unqualified reporting of t-statistics to measure the "significance" of the parameter estimate of θ . Since the t-statistic is testing the hypothesis that $\theta = 0$, under which γ is not identified, the normal approximation is not valid and inferences made from a conventional interpretation of the t-statistic may be misguided.

In the following examples, y_t , x_t , and e_t are real-valued.

1. Additive non-linearity. Gallant, (1987) p. 139.

$$y_t = g(x_t, \alpha) + \theta h(x_t, \gamma) + e_t.$$

A simple example of this is

$$y_t = \alpha_0 + \alpha_1 x_t + \theta \exp(\gamma x_t) + e_t.$$

2. Box-Cox Transformation. Box and Cox (1964).

$$\frac{y_t^{\gamma_1} - 1}{\gamma_1} = \alpha + \theta \frac{x_t^{\gamma_2} - 1}{\gamma_2} + e_t.$$

Originally introduced as a transformation of the dependent variable, the Box-Cox transformation has been used by some authors, such as Heckman and Polachek (1974), separately for each independent variable as well. In the above specification, neither γ_1 nor γ_2 is identified when $\theta = 0$.

3. Structural Change of Unknown Timing. Quandt (1960).

$$y_t = \alpha x_t + 1(t/n > \gamma)\theta x_t + e_t .$$

Here and elsewhere, $1(\cdot)$ is the indicator function. Under the hypothesis of no structural change ($\theta = 0$), the time of structural change (γ) is undefined. A distributional theory for this test has been developed recently by Andrews (1990b), Chu (1989) and Hansen (1990, 1991a).

4. Threshold models in Cross-Section regression.

$$y_t = \alpha x_t + 1(x_t > \gamma)\theta x_t + e_t .$$

This model is useful as a simple model of non-linear relationships. Under the null hypothesis of a linear relationship ($\theta = 0$), the threshold (γ) is undefined. Kim and Siegmund (1989) present a partial distributional theory for a one-regressor model.

It is also possible to test for multiple thresholds, in which case there would be several nuisance parameters undefined under the null hypothesis.

5. Threshold models in Time Series Regression. Tong (1983).

$$y_t = \alpha(L)y_t + 1(y_{t-d} \leq \gamma)\theta(L)y_t + e_t ,$$

where

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p$$

$$\theta(L) = \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$$

and (d,p,q) are known positive integers. This model is known as the self-exciting threshold autoregressive model, and is a simple way to capture non-linearities in a stationary process. The null hypothesis of linearity implies $\theta_1 = \dots = \theta_q = 0$, in which case the threshold γ is undefined. The distribution theory of the LR statistic is studied in Chan (1990) and Chan and Tong (1991).

6. Two-State Markov Trend Model. Hamilton (1989).

$$\begin{aligned}\Delta y_t &= \alpha + \theta s_t + e_t \\ s_t &= \{0, 1\} \\ P\{s_t = 1 | s_{t-1} = 1\} &= \gamma_1 \\ P\{s_t = 0 | s_{t-1} = 0\} &= \gamma_2 .\end{aligned}$$

The test of the two-state model against the standard one-state model takes the null hypothesis $\theta = 0$. Under this model, the transition probabilities $\gamma = (\gamma_1, \gamma_2)$ are undefined. Note that the nuisance parameter γ is two-dimensional. Hansen (1991b) develops a method to test the null hypothesis.

7. Common ARMA Roots.

$$y_t - \gamma y_{t-1} = e_t - (\gamma + \theta)e_{t-1} .$$

A frequent test of interest in ARMA models is whether there are canceling AR and MA roots. Under the above parameterization, this hypothesis is $\theta = 0$. Note that the common root, γ , is not identified under this hypothesis.

8. Non-Expected Utility. Epstein and Zin (1989), Giovannini and Jorion (1989).

The representative agent has the utility function

$$U_t = \left[c_t^\theta + \beta (E_t U_{t+1}^{1-\gamma})^{\theta/(1-\gamma)} \right]^{1/\theta} .$$

Here, $\gamma > 0$ is the measure of relative risk aversion, and $\rho = 1 - \theta > 0$ is the inverse of the elasticity of intertemporal substitution. When $\theta = 0$, intertemporal substitution is unit elastic and γ is undefined. Thus conventional hypothesis testing methods cannot test the unit elastic restriction.

9. Representative Agent Models.

This example is taken from Eichenbaum, Hansen, and Singleton (1988). The representative agent's utility function (I use their notation) is

$$E \sum_{t=0}^{\infty} \theta^{-1} \beta^t \left[(c_t^* \gamma l_t^{*1-\gamma})^\theta - 1 \right]$$

$$c_t^* = A(L)c_t$$

$$l_t^* = B(L)l_t$$

where c_t is consumption and l_t is leisure, and $A(L)$ and $B(L)$ are polynomials in the lag operator. Further, $B(L)$ is specified as

$$B(L) = 1 - \delta(1-\eta L)^{-1}L .$$

This model has two potential problems. First, to test if leisure enters the utility function, the relevant null is $\gamma = 1$, in which case the parameters (δ, η) are not identified. (Eichenbaum, et. al., do not ask this question, however.) Second, to test if lagged leisure is significant, the relevant hypothesis is $\delta = 0$. As the authors point out, in this case η is not identified. In fact, we can rewrite the equation as

$$(1 - \eta L)l_t^* = (1 - (\eta + \delta)L)l_t$$

so we see that the problem is exactly that of canceling ARMA roots. Since there is an unidentified nuisance parameter under the null hypothesis, the authors err in using conventional asymptotic theory to assess whether or not $\delta = 0$.

10. Nonlinear ARCH. Higgins and Bera (1989).

$$e_t | \mathcal{F}_{t-1} \equiv N(0, h_t^2) , \quad \frac{h_t^{2\gamma} - 1}{\gamma} = \alpha + \theta \frac{e_t^{2\gamma} - 1}{\gamma} .$$

The hypothesis of no ARCH effect ($\theta = 0$) renders the parameter γ unidentified. Bera and Higgins (1990) develop an appropriate test using Davies (1987).

11. GARCH and GARCH-M. Bollerslev (1986), Engle, Lilien, and Robins (1987).

$$y_t | \mathcal{F}_{t-1} \equiv N(\mu + \gamma_1 h_t, h_t^2),$$

$$h_t^2 = \alpha + \gamma_2 h_t^2 + \theta (y_{t-1} - \mu - \gamma_1 h_{t-1})^2.$$

Under the null hypothesis of no ARCH effect ($\theta = 0$), the risk premium parameter γ_1 and the GARCH parameter γ_2 are not identified. Thus conventional asymptotic distribution theory cannot be invoked to test the hypothesis of no ARCH.

12. Testing Stability Against AR(1) Alternative. Watson and Engle (1985).

$$y_t = x_t' \alpha + z_t' \beta_t + e_t$$

$$\beta_t = \mu + \gamma \beta_{t-1} + \theta u_t$$

Under the null hypothesis of no random parameter variation ($\theta = 0$) the AR parameter γ is unidentified.

13. Consistent Tests of Functional Form. Bierens (1990).

$$y_t = f(x_t, \beta) + \theta \exp(\gamma' x_t) + e_t$$

To test if $f(x_t, \beta)$ is the correct conditional mean, then one can test the hypothesis $\theta = 0$, under which γ is not identified.

14. Neural Network Tests. White (1989), Stinchcombe and White (1991), and Lee, White, and Granger (1989).

$$y_t = f(x_t, \beta) + \theta \sum_{i=1}^m \psi(\gamma_i x_t) + e_t$$

where $\psi(\cdot)$ is the logistic function. To test if $f(x_t, \beta)$ is the correct conditional mean, then one can test the hypothesis $\theta = 0$, under which γ is not identified.

3. CONSISTENCY

The econometric model is assumed to be described by the parameters (θ, γ) . We will call $\theta \in \Theta \subset \mathbb{R}^k$ the structural parameter vector. We will call $\gamma \in \Gamma$ the nuisance parameter vector, where Γ is some metric space with metric $\rho(\cdot)$. There is some sequence of random criterion functions $Q_n(\theta, \gamma) : \Theta \times \Gamma \rightarrow \mathbb{R}$. One estimation strategy is to find the global maximum of $Q_n(\theta, \gamma)$ over $\Theta \times \Gamma$.

Def. The *Global Estimates* are $(\hat{\theta}, \hat{\gamma}) = \underset{\theta \in \Theta \quad \gamma \in \Gamma}{\text{Argmax}} Q_n(\theta, \gamma)$.

It will be useful in the sequel to define estimates of θ obtained from maximization of $Q_n(\theta, \gamma)$ over $\theta \in \Theta$, while holding γ fixed.

Def. The *Pointwise Estimate* for given $\gamma \in \Gamma$ is $\hat{\theta}(\gamma) = \underset{\theta \in \Theta}{\text{Argmax}} Q_n(\theta, \gamma)$.

One useful fact relating these estimates is that $\hat{\theta} = \hat{\theta}(\hat{\gamma})$.

Assumption 1.

- (i) Θ and Γ are compact ;
- (ii) $Q(\theta, \gamma) = \lim_{n \rightarrow \infty} EQ_n(\theta, \gamma)$ is continuous in (θ, γ) uniformly over $\Theta \times \Gamma$;
- (iii) $Q_n(\theta, \gamma) \rightarrow_p Q(\theta, \gamma)$ for all $(\theta, \gamma) \in \Theta \times \Gamma$;
- (iv) $Q_n(\theta, \gamma) - Q(\theta, \gamma)$ is stochastically equicontinuous in (θ, γ) over $\Theta \times \Gamma$;
- (v) For all $\gamma \in \Gamma$, $Q(\theta, \gamma)$ is uniquely maximized over $\theta \in \Theta$ at θ_0 .

Theorem 1. Under assumption 1,

- (i) $\hat{\theta}(\gamma) \rightarrow_p \theta_0$ uniformly in $\gamma \in \Gamma$;
- (ii) $\hat{\theta} \rightarrow_p \theta_0$.

The assumption that Γ is compact is critical, while the assumption that Θ is compact is made for convenience. It could be relaxed as in, say, Richardson and Bhattacharyya (1990). Assumption 1 (iii) is a statement of pointwise weak convergence. Suppose Q_n takes the form

$$(1) \quad Q_n(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^n q(\theta, \gamma, x_i) .$$

If $\{x_i\}$ is α -mixing and $q(\theta, \gamma, x_i)$ is uniformly integrable for all (θ, γ) , then the weak large of large numbers due to Andrews (1988) gives assumption 1 (iii).

The concept of stochastic equicontinuity is used here and in the sequel to obtain uniform convergence, and therefore warrants some discussion. Stochastic equicontinuity¹ is essentially a smoothness condition. If Q_n takes form (1), then Andrews (1990a, Lemma 2) has shown that a sufficient condition for assumptions 1 (ii) and (iv) is the Lipschitz condition

$$\text{For all } \delta > 0, \quad |q(\theta', \gamma', x_i) - q(\theta, \gamma, x_i)| \leq B(x_i)h(\delta) \text{ for all } \|\theta - \theta'\| < \delta \\ \text{and } \rho(\gamma, \gamma') < \delta, \text{ where } \lim_{\delta \downarrow 0} h(\delta) = 0 \text{ and } \sup_{n \geq 1} \frac{1}{n} \sum_{i=1}^n EB(x_i) < \infty .$$

The most unusual assumption is 1 (v). It states that at the global maximum, the limiting criterion function does not depend upon γ . This is equivalent to the statement that the nuisance parameter is not asymptotically identified.

¹ $\{G_n(\lambda)\}$ is *stochastically equicontinuous* on Λ if for all $\epsilon > 0$ and $\eta > 0$ there exists some $\delta > 0$ such that $\lim_n P \left[\sup_{\lambda \in \Lambda} \sup_{\rho(\lambda, \lambda') < \delta} |G_n(\lambda') - G_n(\lambda)| > \epsilon \right] < \eta$, where $\rho(\cdot, \cdot)$ denotes the distance metric defined on Λ .

4. Hypothesis Testing

4.1 Null Hypothesis

The econometrician is interested in testing the following hypothesis concerning the structural parameters:

$$H_0 : h(\theta) = 0$$

where $h: \Theta \rightarrow \mathbb{R}^q$ is continuously differentiable. Set $h_\theta(\theta) = \partial h(\theta)/\partial \theta'$, and $h_\theta = h_\theta(\theta_0)$. Assume that $\text{rank}(h_\theta) = q$.

We assumed in section 3 that the nuisance parameters are not identified asymptotically. We now assume that γ is not identified even in finite samples for any θ which satisfies the null hypothesis. This is true of all the examples listed in section 2.

Assumption 2. For $\theta \in \Theta_0 = \{\theta \in \Theta : h(\theta) = 0\}$, $Q_n(\theta, \gamma)$ does not depend upon γ .

We can define a criterion function restricted to satisfy H_0 , and an estimate obtained by maximizing this function.

Def The *Restricted Estimate* of θ is $\tilde{\theta} = \underset{\theta \in \Theta_0}{\text{Argmax}} Q_n(\theta)$, where $Q_n(\theta) = Q_n(\theta, \gamma)$ for $\theta \in \Theta_0$.

A standard argument gives the consistency of $\tilde{\theta}$.

Theorem 2. Under assumption 1 and H_0 , $\tilde{\theta} \xrightarrow{p} \theta_0$.

4.2 Alternative Hypotheses

The alternative hypothesis of interest is

$$H_1 : h(\theta) \neq 0 \quad , \quad \gamma \in \Gamma .$$

Since γ is not identified under the null hypothesis in the sequel it will be convenient to specify as well the set of pointwise alternative hypotheses:

$$H_1(\gamma) : h(\theta) \neq 0 \quad , \quad \gamma \text{ given .}$$

Testing H_0 against $H_1(\gamma)$ for any particular γ does not lead to any particular difficulties, for the nuisance parameter is effectively eliminated by fixing it at a predetermined value. In the structural change application, for example, $H_1(\gamma)$ would be the hypothesis of a single structural change of known timing, while H_1 specifies the timing as unknown.

4.3 Likelihood Ratio Tests

If $Q_n(\theta, \gamma)$ is the log-likelihood function, then appropriate statistics for the tests of H_0 against H_1 and H_0 against $H_1(\gamma)$ are given by

$$LR_n = 2n \left[Q_n(\hat{\theta}, \hat{\gamma}) - Q_n(\tilde{\theta}) \right]$$

and

$$LR_n(\gamma) = 2n \left[Q_n(\hat{\theta}(\gamma), \gamma) - Q_n(\tilde{\theta}) \right] ,$$

respectively. The connection between the statistics may be seen in the following result.

Theorem 3. $LR_n = \sup_{\gamma \in \Gamma} LR_n(\gamma) .$

4.4 Wald Tests

Define the random functions

$$S_n(\theta, \gamma) = \frac{\partial}{\partial \theta} Q_n(\theta, \gamma)$$

$$M_n(\theta, \gamma) = -\frac{\partial^2}{\partial \theta \partial \theta'} Q_n(\theta, \gamma)$$

$$\Omega_n(\theta, \gamma) = M_n(\theta, \gamma)^{-1} V_n(\theta, \gamma) M_n(\theta, \gamma)^{-1},$$

where $V_n(\theta, \gamma)$ is some estimate of $\text{var}(\sqrt{n} S_n(\theta, \gamma))$.

The pointwise Wald statistics for testing H_0 against $H_1(\gamma)$ are

$$W_n(\gamma) = n h[\hat{\theta}(\gamma)]' \left[h_{\theta}(\hat{\theta}(\gamma)) \Omega_n(\hat{\theta}(\gamma), \gamma) h_{\theta}(\hat{\theta}(\gamma))' \right]^{-1} h[\hat{\theta}(\gamma)].$$

The standard Wald statistic for the test of H_0 against H_1 is given by

$$W_n = W_n(\hat{\gamma}) = n h(\hat{\theta})' \left[h_{\theta}(\hat{\theta}) \Omega_n(\hat{\theta}, \hat{\gamma}) h_{\theta}(\hat{\theta})' \right]^{-1} h(\hat{\theta}).$$

A reasonable alternative statistic is the largest pointwise Wald statistic:

$$\text{Sup}W_n = \sup_{\gamma \in \Gamma} W_n(\gamma).$$

4.5 Lagrange Multiplier Tests

The Lagrange multiplier (LM) statistic for the test of H_0 against H_1 is not defined, for γ is not identified under the null. The sequence of pointwise LM statistics for the test of H_0 against $H_1(\gamma)$, however, are well defined and given by

$$\text{LM}_n(\gamma) = n S_n(\tilde{\theta}, \gamma)' M_n(\tilde{\theta}, \gamma)^{-1} h_{\theta}(\tilde{\theta})' \left[h_{\theta}(\tilde{\theta}) \Omega_n(\tilde{\theta}, \gamma) h_{\theta}(\tilde{\theta})' \right]^{-1} h_{\theta}(\tilde{\theta}) M_n(\tilde{\theta}, \gamma)^{-1} S_n(\tilde{\theta}, \gamma).$$

We can consider two LM-like statistics which generate feasible tests of H_0 against H_1 . The first uses the estimate of γ obtained under global maximization:

$$\text{LM}_n = \text{LM}(\hat{\gamma}),$$

while the second takes the largest pointwise LM statistic:

$$\text{SupLM}_n = \sup_{\gamma \in \Gamma} \text{LM}_n(\gamma).$$

5. DISTRIBUTIONAL THEORY

The key to unlocking a valid distributional theory for the test statistics of H_0 against H_1 is the fact that the test statistics can be written as supremum over the stochastic processes $LR_n(\gamma)$, $W_n(\gamma)$, and $LM_n(\gamma)$, for which we can find functional distributions via empirical process theory. We will be using the following concepts.

Def. $G(\lambda)$ is a *mean zero vector Gaussian process* in $\lambda \in \Lambda$, if for all $\lambda \in \Lambda$, $E[G(\lambda)] = 0$, and all the finite dimensional distributions of $G(\cdot)$ are multivariate normal. The *covariance function* of $G(\lambda)$ is given by $K(\lambda_1, \lambda_2) = E[G(\lambda_1)G(\lambda_2)']$.

Def. $Z(\lambda)$ is a *chi-square process* in $\lambda \in \Lambda$ of order q , if $Z(\cdot)$ can be represented as $Z(\lambda) = G(\lambda)'K(\lambda, \lambda)^{-1}G(\lambda)$, where $G(\lambda)$ is a mean zero q -vector Gaussian process with covariance function $K(\cdot, \cdot)$.

Mean-zero Gaussian processes and chi-square processes are completely described by their covariance functions.

Let θ_c be some neighborhood of θ_0 , and $\mathcal{N} = \theta_c \times \Gamma$.

Assumption 3.

- (i) $M(\theta, \gamma) = \lim_{n \rightarrow \infty} EM_n(\theta, \gamma)$ and $V(\theta, \gamma) = \lim_{n \rightarrow \infty} n E S_n(\theta, \gamma)S_n(\theta, \gamma)'$ are continuous in (θ, γ) uniformly over \mathcal{N} ;
- (ii) $[M_n(\theta, \gamma), V_n(\theta, \gamma)] \rightarrow_p [M(\theta, \gamma), V(\theta, \gamma)]$ for all $(\theta, \gamma) \in \mathcal{N}$;
- (iii) $M_n(\theta, \gamma) - M(\theta, \gamma)$ and $V_n(\theta, \gamma) - V(\theta, \gamma)$ are stochastically equicontinuous in (θ, γ) over \mathcal{N} ;

(iv) $M(\gamma) = M(\theta_0, \gamma)$ and $V(\gamma) = V(\theta_0, \gamma)$ are positive definite uniformly over $\gamma \in \Gamma$;

(v) $\sqrt{n} S_n(\theta_0, \gamma) \Rightarrow S(\gamma)$ on $\gamma \in \Gamma$, where $S(\cdot)$ is a mean zero Gaussian process with the covariance function

$$K(\gamma_1, \gamma_2) = \lim_{n \rightarrow \infty} n E \left[S_n(\theta_0, \gamma_1) S_n(\theta_0, \gamma_2)' \right] .$$

Let $\bar{S}(\gamma)$ be a mean zero Gaussian process with the covariance function

$$\bar{K}(\gamma_1, \gamma_2) = h_\theta M(\gamma_1)^{-1} K(\gamma_1, \gamma_2) M(\gamma_2)^{-1} h_\theta' ,$$

and let $C(\gamma)$ be a chi-square process with covariance function $\bar{K}(\cdot, \cdot)$.

Theorem 4. Under assumptions 1, 2, and 3,

(a) $\sqrt{n}(\hat{\theta}(\gamma) - \theta_0) \Rightarrow M(\gamma)^{-1} S(\gamma)$ on $\gamma \in \Gamma$;

(b) $W_n(\gamma) \Rightarrow C(\gamma)$ on $\gamma \in \Gamma$;

(c) $LM_n(\gamma) \Rightarrow C(\gamma)$ on $\gamma \in \Gamma$;

(d) $LR_n(\gamma) \Rightarrow C^*(\gamma) = \bar{S}(\gamma)' \left[h_\theta M(\gamma)^{-1} h_\theta' \right]^{-1} \bar{S}(\gamma)$ on $\gamma \in \Gamma$.

Assumption 3 is standard for central limit theory, except that all the results are assumed uniformly over $\gamma \in \Gamma$. The pointwise laws of large numbers and stochastic equicontinuity assumptions may be verified as discussed after assumption 2.

One important condition is 3 (iv). If $M(\gamma)$ or $V(\gamma)$ is singular for some values of γ , then the theory developed here will break down. This possibility will depend upon the particular application. In this event, we will need to redefine Γ to exclude singular values from consideration. One example where this arises is in structural change problems. If the timing of structural change is considered to be

some fraction of the sample size, then this fraction has to be bounded away from zero and one. See Andrews (1990b) on this point.

Assumption 3 (v) may appear unconventional. It is simply the function space generalization of the statement that for each $\gamma \in \Gamma$, $\sqrt{n} S_n(\theta_0, \gamma)$ converges in distribution to a normal random vector. Sufficient conditions are given, for example, in Andrews (1990c). If Q_n takes form (1), $\{\partial/\partial\theta q(\theta_0, \gamma, x_1)\}$ needs to be uniformly L^r -integrable for some $r > 2$, be near epoch dependent of size -1 on some series which is strong mixing of size $-2r/(r-2)$, and satisfy some form of smoothness condition with respect to γ .

The absence of serial correlation and heteroskedasticity frequently implies $V(\gamma) = M(\gamma)^{-1}$, in which case the processes $C^*(\gamma)$ and $C(\gamma)$ are identical, and the LR, Wald, and LM processes have the same asymptotic probability measures. This is analogous to conventional theory. In this case all the test statistics are asymptotically similar, as shown in the following theorem.

Theorem 5. *Under assumptions 1, 2, and 3, and if $V(\gamma) = M(\gamma)^{-1}$, then*

- (a) $\hat{\gamma} \rightarrow_d \underset{\gamma \in \Gamma}{\text{Argmax}} C(\gamma)$;
- (b) $LR_n, W_n, \text{Sup}W_n, LM_n, \text{Sup}LM_n \rightarrow_d \text{Sup}C = \underset{\gamma \in \Gamma}{\text{Sup}} C(\gamma)$.

Note that the parameter estimate $\hat{\gamma}$ fails to converge in probability. Instead, it converges in distribution to a random variable, as is common among unidentified parameter estimates.

In general, however, the equivalence between the LR, Wald and LM tests does not hold. The following theorem gives the asymptotic distribution theory allowing for heteroskedasticity and serial correlation.

Theorem 6. *Under assumptions 1, 2, and 3,*

- (a) $\hat{\gamma} \rightarrow_d \underset{\gamma \in \Gamma}{\text{Argmax}} C^*(\gamma) ;$
- (b) $\text{LR}_n \rightarrow_d \underset{\gamma \in \Gamma}{\text{Sup}} C^*(\gamma) ;$
- (c) $W_n, \text{LM}_n \rightarrow_d C(\underset{\gamma \in \Gamma}{\text{Argmax}} C^*(\gamma)) ;$
- (d) $\text{Sup}W_n, \text{SupLM}_n \rightarrow_d \text{Sup}C .$

It is not surprising to find that the likelihood ratio statistic is not robust to heteroskedasticity or serial correlation, since this occurs in standard models. What is surprising, however, is that the Wald statistic is not robust as well, even though a robust covariance matrix estimate is used. The problem is due to the unidentified nuisance parameter, which is not consistent under the null hypothesis. The only test statistics with distributions robust to heteroskedasticity and serial correlation are the maximal Wald and Lagrange multiplier statistics. It is interesting to note that these are the statistics studied in most of the earlier theoretical literature, such as Davies (1977, 1987), Chan (1990), Andrews (1990b), and Hansen (1990, 1991a). In contrast, the most commonly reported test statistics in applications are LR statistics and t-statistics (which are signed square roots of Wald statistics), which do not share this robustness property, as shown in the next section.

6. T-TESTS

The theory developed in sections 4 and 5 apply to two-sided hypothesis tests. Often, the hypothesis test only concerns one element of θ and the alternative hypothesis is one-sided, i.e.

$$H_0 : e' \theta = 0$$

$$H_1 : e' \theta > 0$$

where $e = (1 \ 0 \ 0 \ \dots \ 0)'$. In this case, it is desirable to develop one-sided versions of the test statistics and asymptotic distributions.

We can define the sequence of pointwise t-statistics

$$t_n(\gamma) = \frac{e' \hat{\theta}(\gamma)}{e' \Omega_n(\hat{\theta}(\gamma), \gamma) e}.$$

The standard t-statistic is the pointwise t-statistic evaluated at the global estimate $\hat{\gamma}$:

$$t_n = t_n(\hat{\gamma}).$$

We can also define the maximal pointwise t-statistic:

$$\text{Sup}T_n = \sup_{\gamma \in \Gamma} t_n(\gamma).$$

We can also define a one-sided version of the LM test. The sequence of pointwise one-sided LM statistics are

$$t_n^*(\gamma) = \frac{e' M_n(\tilde{\theta}, \gamma)^{-1} S_n(\tilde{\theta}, \gamma)}{e' \Omega_n(\tilde{\theta}, \gamma) e},$$

giving the test statistics

$$t_n^* = t_n^*(\hat{\gamma})$$

and

$$\text{SupT}_n^* = \sup_{\gamma \in \Gamma} t_n^*(\gamma).$$

Similar arguments as those of the previous section enable us to obtain the following result. The proofs are quite similar and omitted.

Theorem 7. *Under assumptions 1, 2, and 3,*

- (a) $t_n(\gamma), t_n^*(\gamma) \Rightarrow t(\gamma)$
- (b) $t_n, t_n^* \rightarrow_d t(\text{Argmax}_{\gamma \in \Gamma} C^*(\gamma))$
- (c) $\text{SupT}_n, \text{SupT}_n^* \rightarrow_d \sup_{\gamma \in \Gamma} t(\gamma)$

where $t(\gamma)$ is a Gaussian process with covariance function

$$E[t(\gamma_1)t(\gamma_2)] = \frac{e'M(\gamma_1)^{-1}K(\gamma_1, \gamma_2)M(\gamma_2)^{-1}e}{\left[e'M(\gamma_1)^{-1}V(\gamma_1)M(\gamma_1)^{-1}e\right]^{1/2} \left[e'M(\gamma_2)^{-1}V(\gamma_2)M(\gamma_2)^{-1}e\right]^{1/2}}.$$

7. OBTAINING THE DISTRIBUTION SupC

7.1 Previous Literature

The process $C(\gamma)$ is completely described by its covariance function $\bar{K}(\cdot, \cdot)$, so the asymptotic distribution of the test statistics, SupC, is fully described by the pair (\bar{K}, Γ) . Unfortunately, the function \bar{K} is context-dependent, precluding tabulation.

In some cases, the distribution simplifies. In the structural change applications with weakly dependent data, Andrews (1990b), Chu (1989) and Hansen (1990) found that the covariance function is that of a vector Brownian bridge. If the regressors are trended, Hansen (1991a) found that the covariance function is different, but dependent upon only a small number of parameters. In the one-dimensional threshold model, a Brownian bridge result was obtained by Chan (1990) and Kim and Siegmund (1989) under different assumptions. If there is more than one regressor, however, Chan and Tong (1991) find that the covariance function is more complicated.

It is possible to construct simple examples, however, which show that the covariance function need not be particularly simple. Take, for example, the model

$$y_t = \theta \exp(\gamma x_t) + e_t, \quad e_t \text{ iid } (0, \sigma^2),$$

where x_t is stationary, independent of $\{e_t\}$. The test statistics have the asymptotic distribution $\sup_{\gamma \in \Gamma} C(\gamma)$, where $C(\gamma)$ is a chi-square process with covariance function

$$K(\gamma_1, \gamma_2) = E\left[\exp[(\gamma_1 + \gamma_2)x_t]\right] = \psi(\gamma_1 + \gamma_2).$$

The function $\psi(\cdot)$ is the moment generating function of x_t . In this simple example, the covariance function depends upon the entire distribution of the regressor!

Davies (1977, 1987) attempts to circumvent this problem by finding a bound for

the asymptotic distribution of the test statistic. His bound, however, depends upon the assumption that $C(\gamma)$ has a continuous derivative except possibly for a finite number of jumps. This is critical to his approach since his bound utilizes the total variation of the process $C(\gamma)$. Unfortunately, in some of the examples in section 2, the chi-square process $C(\gamma)$ may be *nowhere differentiable*, and thus have infinite total variation. Although the number of jumps may be finite for any given sample size, this number will tend to infinity as the sample size increases, so this approximation may become arbitrarily poor in large samples.

7.2 Distribution Theory Under Uncorrelated Errors

We now place more structure on the problem. Assume that Q_n takes the form

$$(2) \quad Q_n(\theta, \gamma) = \frac{1}{n} \sum_{i=1}^n q_i(\theta, \gamma; x_i) .$$

Set $q_i(\theta, \gamma) = q_i(\theta, \gamma; x_i)$, and $s_i(\theta, \gamma) = \frac{\partial}{\partial \theta} q_i(\theta, \gamma)$. $s_i(\gamma) = s_i(\hat{\theta}(\gamma), \gamma)$.

Assumption 5. $E[s_i(\theta_o, \gamma)s_j(\theta_o, \gamma)'] = 0$, for all $i \neq j$, $\gamma \in \Gamma$.

This rules out serial correlation, but not heteroskedasticity. In this context we can use the following estimator of $V(\gamma)$:

$$V_n(\hat{\theta}(\gamma), \gamma) = \frac{1}{n} \sum_{i=1}^n s_i(\hat{\theta}(\gamma), \gamma)s_i(\hat{\theta}(\gamma), \gamma)' .$$

To simplify the notation, set

$$s_i(\gamma) = s_i(\hat{\theta}(\gamma), \gamma) ,$$

$$M_n(\gamma) = M_n(\hat{\theta}(\gamma), \gamma),$$

$$\Omega_n(\gamma) = \Omega_n(\hat{\theta}(\gamma), \gamma) = M_n(\gamma)^{-1}V_n(\hat{\theta}(\gamma), \gamma)M_n(\gamma)^{-1}$$

and

$$\hat{h}_\theta(\gamma) = h_\theta(\hat{\theta}(\gamma)).$$

Define $\mathcal{F}_n = \sigma(x_1, \dots, x_n)$. Now imagine the following experiment. Draw n iid standard normal random variables $\{u_i\}$ and construct the $q \times 1$ process

$$\hat{S}_n(\gamma) = h_\theta(\gamma) M_n(\gamma)^{-1} \left[\frac{1}{n} \sum_{i=1}^n s_i(\gamma) u_i \right].$$

Conditional on \mathcal{F} , $\sqrt{n} \hat{S}_n(\gamma)$ is a mean-zero Gaussian process with covariance function

$$\bar{K}_n(\gamma_1, \gamma_2) = h_\theta(\gamma_1) M_n(\gamma_1)^{-1} \left[\frac{1}{n} \sum_{i=1}^n s_i(\gamma_1) s_i(\gamma_2)' \right] M_n(\gamma_2)^{-1} h_\theta(\gamma_2).$$

Now construct the process

$$\hat{C}_n(\gamma) = n \hat{S}_n(\gamma)' \left[h_\theta(\gamma) \Omega_n(\gamma) h_\theta(\gamma)' \right]^{-1} \hat{S}_n(\gamma).$$

Conditional on \mathcal{F} , $\hat{C}_n(\gamma)$ is a chi-square process with covariance function $\bar{K}(\cdot, \cdot)$.

Thus as $n \rightarrow \infty$,

$$\hat{C}_n(\gamma) \Rightarrow C(\gamma),$$

$$\text{Sup} \hat{C}_n = \text{Sup}_{\gamma \in \Gamma} \hat{C}_n(\gamma) \Rightarrow \text{Sup} C.$$

Now repeat the experiment with another n independent draws $\{u_i\}$.

Conditional on \mathcal{F} , the $\text{Sup} \hat{C}_n$ are mutually independent, and as $n \rightarrow \infty$, their empirical distribution approaches $\text{Sup} C$. Since this distribution is independent of \mathcal{F} , the dependence upon the data is eliminated in large samples. Thus the upper tail of the empirical distribution of the $\text{Sup} \hat{C}_n$ will provide an asymptotically valid method to determine critical values.

This method easily generates p -values. Suppose the test statistic calculated from the data is T_n , which has the asymptotic null distribution $\text{Sup} C$. Generate $\{u_i\}$ and construct the statistics $\text{Sup} \hat{C}_n$ and

$$\hat{p} = 1(\text{Sup}\hat{C}_n > T_n) .$$

If n is large, the variables \hat{p} will be approximately iid Bernouli draws whose expectation is the p-value of the test statistic T_n . With a relatively small number of replications, p-value estimation is quite precise. For example, using 500 replications and the null of a p-value of 5%, the standard error is approximately 1%.

This simulation method does not require numerical optimization of the non-linear model. The parameter values and first order conditions are held fixed at the global estimates. It is not a trivial calculation, however, since the maximal value of the function $C_n(\gamma)$ may have to be found by a grid search.

7.3 Distributional Theory Under Homoskedasticity

In some cases it is possible to simplify a few of the calculations and improve the approximation of the simulation method in small samples. Consider the example of additive non-linearity discussed in section 2, additionally assuming homoskedastic errors:

$$y_i = \theta_1 h(x_i, \gamma) + g(x_i, \theta_2) + \epsilon_i$$

$$E(\epsilon_i | x_i) = 0$$

$$E(\epsilon_i^2 | x_i) = \sigma^2$$

$$H_0 : e' \theta = 0 \quad , \quad e = (1 \ 0 \ \dots \ 0)' .$$

If the model is estimated by non-linear least squares we have

$$S_n(\theta, \gamma) = -\frac{2}{n} \sum_{i=1}^n h(x_i, \gamma) \hat{\epsilon}_i \quad , \quad \hat{\epsilon}_i = y_i - g(x_i, \alpha) - \theta h(x_i, \gamma)$$

The process $\sqrt{n}S_n(\hat{\theta}, \gamma)$ converges weakly to a process $S(\gamma)$ with covariance function

$$K(\gamma_1, \gamma_2) = 4\sigma^2 \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E[h(x_i, \gamma_1)h(x_i, \gamma_2)'] .$$

Therefore, a good approximation to the limit process $C(\gamma)$ can be found in this context by the following simplified version of the simulation method outlined in section 6.2. Draw iid standard normal random variables $\{u_i\}$ and construct

$$\hat{S}_n^+(\gamma) = e' M_n(\gamma)^{-1} \frac{1}{n} \sum_{i=1}^n h(x_i, \gamma) u_i ,$$

which, conditional upon \mathcal{F} , is a Gaussian process with covariance function

$$\bar{K}_n(\gamma_1, \gamma_2) = e' M_n(\gamma)^{-1} \left[\frac{1}{n} \sum_{i=1}^n h(x_i, \gamma_1) h(x_i, \gamma_2)' \right] M_n(\gamma)^{-1} e .$$

Therefore the process

$$C_n^+(\gamma) = n \hat{S}_n^+(\gamma)' \left[e' M_n(\gamma) e \right]^{-1} \hat{S}_n^+(\gamma) / \hat{\sigma}^2$$

where $\hat{\sigma}^2 = n^{-1} \sum_{i=1}^n \hat{\epsilon}_i^2$, has the asymptotic distribution of $C(\gamma)$. Replication of $\text{Sup}_\gamma C_n^+(\gamma)$ should provide asymptotically valid draws from the null distribution.

The construction of section 7.2 is similar in many respects to the heteroskedasticity-consistent covariance matrix estimate of White (1980). It is frequently found in simulations that this estimator requires large samples for the asymptotic theory to provide a good approximation to the finite sample distributions. One would expect similar behavior for the statistics reported here. It seems reasonable, therefore, to use methods such as that outlined in this section, when homoskedasticity is not too wild an assumption.

7.4 Distributional Theory Under Autocorrelation

If $Q_n(\theta, \gamma)$ is a correctly specified likelihood, the scores $\{s_i(\theta, \gamma)\}$ should be a martingale difference sequence and hence uncorrelated. In some applications (such as GMM), this may not be part of the maintained hypothesis. It is known in this event that conventional standard error estimates are invalid and some correction needs to be made. A similar problem arises in the current context.

We again assume that the criterion function takes the form (2), but allow for serial correlation in the scores s_i . In this context we have

$$K(\gamma_1, \gamma_2) = \sum_{k=-n}^n \frac{1}{n} \sum_{i=1}^n E \left[s_i(\theta_0, \gamma_1) s_{i+k}(\theta_0, \gamma_2)' \right].$$

A natural sample estimate of this function is

$$K_n(\gamma_1, \gamma_2) = \sum_{k=-m}^m w(k/m) \frac{1}{n} \sum_{i=1}^n E \left[s_i(\theta_0, \gamma_1) s_{i+k}(\theta_0, \gamma_2)' \right].$$

where $w(k/m)$ is a weight function, such as the Bartlett kernel $w(x) = 1 - |x|$, and m is a bandwidth parameter.

We perform an experiment similar to that of section 7.2. Draw a sample of iid standard normal random variables $\{u_i\}$ and construct the process

$$\hat{S}_n(\gamma) = h_\theta(\gamma) M_n(\gamma)^{-1} \frac{1}{n} \sum_{i=1}^n s_i(\gamma) \left[u_i + u_{i-1} + \cdots + u_{i-m} \right].$$

It is a straightforward calculation to show that, conditional upon the data, $\hat{S}_n(\gamma)$ is a Gaussian process with covariance function

$$\bar{K}_n(\gamma_1, \gamma_2) = h_\theta(\gamma_1) M_n(\gamma_1)^{-1} K_n(\gamma_1, \gamma_2) M_n(\gamma_1)^{-1} h_\theta(\gamma_1)'$$

where the sample function $K_n(\cdot, \cdot)$ is computed using the Bartlett kernel. So long as $\bar{K}_n(\cdot, \cdot)$ is consistent for $K(\cdot, \cdot)$, $\hat{S}_n(\gamma)$ is asymptotically distributed as $\bar{S}(\gamma)$.

It similarly follows that conditional on \mathcal{F} , the process

$$\hat{C}_n(\gamma) = n \hat{S}_n(\gamma)' \left[h_{\theta}(\gamma) \Omega_n(\gamma) h_{\theta}(\gamma)' \right]^{-1} \hat{S}_n(\gamma) .$$

is a chi-square process with covariance function $\bar{K}(\cdot, \cdot)$. Under these assumptions we find

$$\hat{C}_n(\gamma) \Rightarrow C(\gamma) ,$$

$$\text{Sup} \hat{C}_n = \text{Sup}_{\gamma \in \Gamma} \hat{C}_n(\gamma) \Rightarrow \text{Sup} C .$$

As before, repeated draws from $\text{Sup} \hat{C}_n$ allow for the calculation of critical values and p-values.

8. TESTING FOR THRESHOLDS

8.1 Threshold Models

A fairly general formulation for linear threshold models may be given as

$$y_t = \theta_1' x_{1t} + \theta_2' x_{2t} 1(x_{3t} > \gamma) + e_t,$$

$$E(e_t | x_{1t}, x_{2t}, x_{3t}) = 0$$

where $1(\cdot)$ is the indicator function. In most models, the x_{2t} is a sub-vector of x_{1t} , and x_{3t} is a scalar element of x_{1t} . This is a simple way to capture non-linear regression effects.

The null hypothesis of frequent interest is linearity:

$$H_0 : \theta_2 = 0$$

under which the threshold parameter γ is not identified. This model was a primary motivation for the work of Davies (1977, 1987) and special cases have been studied by Kim and Seigmund (1989), Chan (1990) and Chan and Tong (1991).

It is fairly straightforward to apply the tests developed in this paper to the threshold model. For any given γ , the model is linear and can be estimated by ordinary least squares (OLS). A practical issue is the selection of Γ . Since the function $\hat{\theta}(\gamma)$ and the associated test statistics will be discontinuous functions, with jumps at the values $\gamma = x_{3t}$, it makes sense to select Γ to consist of (at most) the sample values of $\{x_{3t}\}$. Further, to exclude the possibility of near-singularities, it is necessary to select Γ to exclude values of $\{x_{3t}\}$ too far in the tails of its empirical distribution. Following the advice of Andrews (1990b) in the context of testing for structural change, I suggest the informal rule of using the values of $\{x_{3t}\}$ between the 15th and 85th percentile of its empirical distribution.

8.2 Monte Carlo Experiment

Take the simple threshold model

$$y_t = \theta_1 x_t + e_t \quad \text{if } x_t \leq \gamma$$

$$y_t = \theta_2 x_t + e_t \quad \text{if } x_t > \gamma$$

$$x_t \equiv N(0,1) \quad , \quad e_t \equiv N(0,1) \quad .$$

I used Monte Carlo methods to evaluate the small sample distribution of the test statistics for the hypothesis $H_0: \theta_1 = \theta_2$. Sample size was set to 100 and 1000 replications were made.

For each replication, a new sample $\{y_t, x_t\}$ was drawn. The region Γ was chosen as suggested in the previous subsection, taking the 15th to 85th percentile of the empirical distribution of x_t . The $\text{Sup}W_n$ and $\text{Sup}LM_n$ statistics were calculated, both under the assumption of homoskedasticity and allowing for heteroskedasticity as in White (1980) (thus generating four statistics). The p-values were calculated as discussed in section 7. The statistics calculated using the standard covariance matrices used the method of section 7.3 which assumes homoskedastic errors, while the statistics calculated with the White heteroskedastic-consistent covariance matrices used the method of section 7.2. A statistic was considered "significant" at the 10% (5%) level if the calculated p-value were less than .10 (.05).

First, the model was evaluated under the null hypothesis, setting $\theta_1 = \theta_2 = 1$. Table 1 reports the estimates of the upper tails (10% and 5%) of the distributions of the four test statistics. For comparison, the tail values from the chi-square distribution with one degree of freedom is also given. The chi-square values (which we could call the naive critical values) are noticeably lower than the Monte Carlo values. This is not surprising, given our theory, but reinforces the argument that unidentified parameters may have important effects upon the correct sampling theory.

Table 1
Upper Tails of Test Statistics Under Null

	<u>10%</u>	<u>5%</u>
<u>Standard</u>		
SupW	5.3	6.8
SupLM	5.0	6.3
<u>Hetero Consistent</u>		
SupW	5.7	7.5
SupLM	5.0	6.3
χ_1^2	2.7	3.8

Table 2
Rejection Frequencies Under Null

	<u>10%</u>	<u>5%</u>
<u>Standard</u>		
SupW	.122	.058
SupLM	.108	.054
<u>Hetero Consistent</u>		
SupW	.149	.075
SupLM	.105	.051

Table 3
Rejection Frequencies Under Alternative

	<u>10%</u>	<u>5%</u>
<u>Standard</u>		
SupW	.774	.673
SupLM	.757	.638
<u>Hetero Consistent</u>		
SupW	.792	.689
SupLM	.750	.602

Table 2 presents the frequency at which the p-values rejected the null hypothesis. This is often called the "nominal size" of the test, since it is the true size of the test when asymptotic critical values are used. The results are quite favorable. Both LM tests have nominal sizes very close to their asymptotic value. The Wald tests are somewhat liberal (reject too often), especially the heteroskedasticity-corrected Wald test.

Next, the power of the test was examined. For these calculations, I set $\theta_1 = 1$, $\theta_2 = 1.3$, and $\gamma = 0$. The same testing methods were employed as under the null model. Rejection frequencies are presented in table 3. All tests are easily able to reject the null in favor of the alternative. No size adjustment was performed, but casual inspection of the rejection frequencies suggests that the tests have very similar power in this example.

8.3 Self-Exciting Threshold Autoregressive Models

The self-exciting threshold autoregressive model (or SETAR) is a special case of the general threshold model which has received considerable attention recently in the non-linear time-series literature. The model may be written as

$$(3) \quad \begin{aligned} y_t &= \mu_1 + \alpha_1(L)y_{t-1} + e_t && \text{if } y_{t-d} \leq \gamma \\ y_t &= \mu_2 + \alpha_2(L)y_{t-1} + e_t && \text{if } y_{t-d} > \gamma \end{aligned}$$

where the error e_t is assumed to be a martingale difference with respect to the past history of the scalar process $\{y_t\}$. The lag polynomials $\alpha_1(L)$ and $\alpha_2(L)$ are of order p . The delay parameter d is an integer satisfying $1 \leq d \leq p$. The relevant null is linearity:

$$H_0 : \mu_1 = \mu_2, \quad \alpha_1(L) = \alpha_2(L)$$

under which the threshold γ is not identified. If we consider the delay parameter

d also as a parameter to be estimated, then it is also not identified under the null hypothesis.

The likelihood ratio test for this model (under the assumption of Gaussian errors) has been studied by Chan (1990) and Chan and Tong (1991). This test is equivalent in this context to the $\text{Sup}W_n$ test, when calculated using the standard covariance matrix. Chan (1990) obtains the asymptotic distribution in a form similar to Theorem 5 above, and Chan and Tong (1991) present some special cases. They show that when there is only one regressor (i.e., there is no intercept, and $p = q = 1$), then the relevant Gaussian process is a Brownian bridge. When $p > 1$, however, they find that the covariance function is more complicated, precluding tabulation. Instead, they give some informal rules in a couple special cases which they obtained from simulation evidence. Our testing method, on the other hand, requires no *ad hoc* rule, and is easy to apply in this context.

In addition, our method allows the delay parameter, d , to enter the specification in a consistent and rigorous manner. We can estimate d in the same way we estimate γ , by choosing the model with the lowest sum of squared errors. Being explicit about the way we choose d means that we have to treat it as an unidentified parameter under the null hypothesis. The fact that d takes only a finite (and small) number of possible values does not invalidate the asymptotic theory in this context.

Another variant of the SETAR model is the smooth transition threshold autoregressive model (STAR) of Chan and Tong (1985). This model generalizes the SETAR model by replacing the indicator function by a smooth transition function. This introduces a smoothing parameter which is also not identified under the null hypothesis of linearity.

8.4 GNP Growth Rates

We now apply this testing methodology to a real-world problem. The goal is to obtain a useful characterization of U.S. GNP growth rates, considered as a univariate process. Many macroeconomists have been satisfied with a low-order autoregressive model, but other possibilities have been suggested. Neftci (1984) found evidence for asymmetries in the business cycle. Stock (1987) found evidence for time-deformation nonlinearities. Hamilton (1989) suggested a Markov switching model. Although using distinct models, each of these researchers presented evidence that GNP growth rates are more than just a simple autoregressive process.

Potter (1991) fit a SETAR model to postwar quarterly GNP growth rates. To select the threshold and delay parameters, Potter did not directly minimize the sum of squared errors, but instead used informal graphical methods. Still, these (unidentified) parameters are selected conditional upon the data (rather than from *a priori* theory), and therefore conventional asymptotic theory cannot properly assess the significance of the nonlinear specification. Our methods, however, allow us to directly assess the statistical significance of his SETAR vis-a-vis a linear model.

I used the real GNP series (seasonally adjusted) from Citibase for the period 1947–1990. The data were transformed into annualized quarterly growth rates. (That is, $\Delta y_t = 400(\ln Y_t - \ln Y_{t-1})$, where Y_t is real GNP in period t). Potter (1991) suggested that this series is fit well by a SETAR with lagged first, second and fifth differences. My estimates for the associated autoregressive model with no threshold is

$$\Delta y_t = 1.99 + 0.32 \Delta y_{t-1} + 0.13 \Delta y_{t-2} - 0.09 \Delta y_{t-5} + \hat{e}_t$$

(0.57) (0.09) (0.08) (0.06)

$$R^2 = 0.16 .$$

(Heteroskedastic-consistent standard errors in parenthesis.)

To fit a TAR model, it is necessary to choose the region Γ over which to let the unidentified parameters vary (in this case, γ and d). As before, I let γ vary from the 15th to the 85th percentile of the empirical distribution of Δy_t . For the delay parameter, d , I let it take on the values 1, 2 and 5.

Maximizing the sum of squared errors, the estimates for the threshold parameters are $\hat{d} = 2$ and $\hat{\gamma} = 0.27$. This is surprisingly close to the values $d = 2$ and $\gamma = 0$ chosen by Potter's informal identification methods. With these parameters, we find the following estimates for the TAR model

Regime 1 ($\Delta y_{t-2} \leq 0.266$)

$$\Delta y_t = -3.21 + 0.51 \Delta y_{t-1} - 0.93 \Delta y_{t-2} + 0.38 \Delta y_{t-5} + \hat{e}_t .$$

(1.78) (0.19) (0.26) (0.20)

Regime 2 ($\Delta y_{t-2} > 0.266$)

$$\Delta y_t = 2.14 + 0.30 \Delta y_{t-1} + 0.18 \Delta y_{t-2} - 0.16 \Delta y_{t-5} + \hat{e}_t .$$

(0.73) (0.10) (0.09) (0.07)

$$R^2 = 0.26 .$$

The SupW and SupLM tests for the null hypothesis of linearity, with and without heteroskedasticity corrections, are reported in Table 4. If no correction is made for heteroskedasticity, both the SupW and SupLM tests reject linearity at the 5% level. When corrected for heteroskedasticity, however, the SupLM test ceases to be statistically significant. This is difficult to explain since the innovations do not appear to be heteroskedastic.

Table 4.
Tests of a SETAR vs a TAR

	<u>Statistic</u>	<u>P-value</u>
<u>Standard</u>		
SupW	21.0	0.02
SupLM	18.7	0.04
<u>Hetero Consistent</u>		
SupW	18.0	0.06
SupLM	14.1	0.21

Figure 1 displays a plot of the Wald test statistics for $d = 2$ and the sequence of possible values of γ . The plot shows that the value $\hat{\gamma} \approx 0$ is clearly chosen by the data. The figure also displays the 10%, 5% and 1% critical values obtained from the simulated distribution for the test statistic. Note that these critical values are substantially higher than those from a conventional chi-square table. To make this point clear, figure 2 displays plots of three probability densities: the $\chi^2(4)$, the $\chi^2(8)$ and the density estimated for SupW for this data set. The latter was estimated from the simulated empirical distribution using a normal kernel. The $\chi^2(4)$ would be the appropriate asymptotic distribution if the parameters d and γ were known *a priori*, and are implicitly those used in common practice. The $\chi^2(8)$ is also displayed to counter any illusion that an appropriate rule-of thumb might be to double the degrees of freedom. The estimated density function is substantially to the right of the $\chi^2(8)$. This figure makes plain the fact that unidentified nuisance parameters should not be ignored when making inferences. The distributions are non-standard, and may be dramatically so.

This empirical exercise has uncovered the following surprising results. First, the threshold parameters chosen by Potter (1991) for this series using informal methods are essential the same as those chosen by a classic least squares criterion. Second, even after the selection of this potentially unidentified parameters is taken into account, the

linear model can be (marginally) rejected in favor of a SETAR alternative. This lends considerable support to the view that non-linearities are important in properly specified conditional expectations for macroeconomic time series.

APPENDIX

We will frequently use the following result due to Andrews (1990a).

Lemma A. If Λ is compact, $G_n(\lambda) - G(\lambda)$ is stochastically equicontinuous and for all

$$\lambda \in \Lambda, \quad G_n(\lambda) \xrightarrow{p} G(\lambda), \quad \text{then} \quad \sup_{\lambda \in \Lambda} \|G_n(\lambda) - G(\lambda)\| \xrightarrow{p} 0.$$

Proof of Theorem 1.

(i) Fix $\epsilon > 0$. By assumption 1(v), for all $\gamma \in \Gamma$, there exists some $\delta(\gamma) > 0$ such that

$$\inf_{\|\theta - \theta_0\| \geq \epsilon} [Q(\theta_0, \gamma) - Q(\theta, \gamma)] = \delta(\gamma).$$

Note that the region $\{\theta: \|\theta - \theta_0\| \geq \epsilon\}$ is compact under assumption 1(i). By the maximum theorem, $\delta(\gamma)$ is continuous on Γ , so $\{\delta(\gamma) : \gamma \in \Gamma\}$ is compact, thus

$$\delta = \min_{\gamma \in \Gamma} \delta(\gamma) > 0,$$

and

$$\min_{\gamma \in \Gamma} \inf_{\|\theta - \theta_0\| \geq \epsilon} [Q(\theta_0, \gamma) - Q(\theta, \gamma)] = \delta.$$

It therefore follows that $\{[Q(\theta_0, \gamma) - Q(\hat{\theta}(\gamma), \gamma)] < \delta\}$ implies $\{\|\hat{\theta}(\gamma) - \theta_0\| < \epsilon\}$, and thus

$$P\{\sup_{\gamma \in \Gamma} \|\hat{\theta}(\gamma) - \theta_0\| \leq \epsilon\} \geq P\{\sup_{\gamma \in \Gamma} [Q(\theta_0, \gamma) - Q(\hat{\theta}(\gamma), \gamma)] < \delta\}.$$

The result follows if

$$\sup_{\gamma \in \Gamma} [Q(\theta_0, \gamma) - Q(\hat{\theta}(\gamma), \gamma)] \xrightarrow{p} 0.$$

Indeed,

$$\begin{aligned}
0 &\leq \sup_{\gamma \in \Gamma} \left[Q(\theta_0, \gamma) - Q(\hat{\theta}(\gamma), \gamma) \right] \\
&= \sup_{\gamma \in \Gamma} \left[Q(\theta_0, \gamma) - Q_n(\hat{\theta}(\gamma), \gamma) + Q_n(\hat{\theta}(\gamma), \gamma) - Q(\hat{\theta}(\gamma), \gamma) \right] \\
&\leq \sup_{\gamma \in \Gamma} \left[Q(\theta_0, \gamma) - Q_n(\theta_0, \gamma) + Q_n(\hat{\theta}(\gamma), \gamma) - Q(\hat{\theta}(\gamma), \gamma) \right] \\
&\leq \sup_{\gamma \in \Gamma} |Q(\theta_0, \gamma) - Q_n(\theta_0, \gamma)| + \sup_{\gamma \in \Gamma} |Q(\hat{\theta}(\gamma), \gamma) - Q_n(\hat{\theta}(\gamma), \gamma)| \\
&\leq 2 \sup_{\gamma \in \Gamma} \sup_{\theta \in \Theta} |Q(\theta, \gamma) - Q_n(\theta, \gamma)| \xrightarrow{p} 0,
\end{aligned}$$

by Lemma A.

$$(ii) \quad |\hat{\theta} - \theta_0| = |\hat{\theta}(\hat{\gamma}) - \theta_0| \leq \sup_{\theta \in \Gamma} |\hat{\theta}(\gamma) - \theta_0| \xrightarrow{p} 0$$

by part (i). \square

Proof of Theorem 3.

$$\begin{aligned}
LR_n &= 2n \left[Q_n(\hat{\theta}, \hat{\gamma}) - Q_n(\check{\theta}) \right] \\
&= 2n \left[\sup_{\gamma \in \Gamma} Q_n(\hat{\theta}(\gamma), \gamma) - Q_n(\check{\theta}) \right] \\
&= \sup_{\gamma \in \Gamma} 2n \left[Q_n(\hat{\theta}(\gamma), \gamma) - Q_n(\check{\theta}) \right] = \sup_{\gamma \in \Gamma} LR_n(\gamma). \quad \square
\end{aligned}$$

Proof of Theorem 4 (a): Here and elsewhere superscripts will denote elements of vectors and matrices. For example, S_n^a will denote the a 'th element of the vector S_n and M_n^{ab} will denote the a - b 'th element of the matrix M_n .

For each $\gamma \in \Gamma$, the first order conditions for $\hat{\theta}(\gamma)$ are

$$0 = S_n(\hat{\theta}(\gamma), \gamma)$$

Expand each first order condition (for each value of γ) about θ_0

$$(A1) \quad 0 = S_n^a(\hat{\theta}(\gamma), \gamma) = S_n^a(\theta_0, \gamma) - M_n^a(\theta^*(\gamma), \gamma) (\hat{\theta}(\gamma) - \theta_0)$$

where $\theta^*(\gamma)$ is on a line segment joining θ_0 and $\hat{\theta}(\gamma)$. Now

$$(A2) \quad \sup_{\gamma \in \Gamma} |M_n^{ab}(\theta^*(\gamma), \gamma) - M^{ab}(\theta_0, \gamma)| \\ \leq \sup_{\gamma \in \Gamma} |M_n^{ab}(\theta^*(\gamma), \gamma) - M^{ab}(\theta^*(\gamma), \gamma)| + \sup_{\gamma \in \Gamma} |M^{ab}(\theta^*(\gamma), \gamma) - M^{ab}(\theta_0, \gamma)| \\ \leq \sup_{(\theta, \gamma) \in \mathcal{N}} |M_n^{ab}(\theta, \gamma) - M^{ab}(\theta, \gamma)| + o_p(1) \leq o_p(1).$$

The second inequality exploits the assumed continuity of $M(\cdot, \cdot)$ (assumption 3 (i)) and the fact that $\theta^*(\gamma) \rightarrow_p \theta_0$ uniformly in γ . The final inequality follows from Lemma A. Stacking the row vectors $M_n^a(\theta^*(\gamma), \gamma)$ into a matrix $M_n^*(\gamma)$, (A1), (A2), assumption 3 (iv)(v) and the continuous mapping theorem (CMT) give

$$\sqrt{n}(\hat{\theta}(\gamma) - \theta_0) = M_n^*(\gamma)^{-1} \sqrt{n} S_n(\theta_0, \gamma) \Rightarrow M(\gamma)^{-1} S(\gamma) \quad \square$$

Proof of Theorem 4(b): Since $\hat{\theta}(\gamma) \rightarrow_p \theta_0$ uniformly in γ , and $h_{\theta}(\cdot)$ is continuous in θ_c ,

$$(A3) \quad h_{\theta}(\hat{\theta}(\gamma)) \rightarrow_p h_{\theta}(\theta_0) = h_{\theta},$$

uniformly in γ . Similarly,

$$(A4) \quad \Omega_n(\hat{\theta}(\gamma), \gamma) = M_n(\hat{\theta}(\gamma), \gamma)^{-1} V(\hat{\theta}(\gamma), \gamma) M_n(\hat{\theta}(\gamma), \gamma)^{-1} \\ \rightarrow_p M(\gamma)^{-1} V(\gamma) M(\gamma)^{-1} = \Omega(\gamma),$$

uniformly in $\gamma \in \Gamma$.

Expand each element of the vector $h(\hat{\theta}(\gamma))$ about θ_0 :

$$\sqrt{nh}^a(\hat{\theta}(\gamma)) = \sqrt{nh}^a(\theta_0) + h_{\theta}^a(\theta^*(\gamma)) \sqrt{n}(\hat{\theta}(\gamma) - \theta_0)$$

$$= h_{\theta}^a(\theta^*(\gamma))\sqrt{n}(\hat{\theta}(\gamma) - \theta_0) ,$$

Thus by (A3), part (a), and the CMT,

$$(A5) \quad \sqrt{n}h(\hat{\theta}(\gamma)) \Rightarrow h_{\theta}(\theta_0) M(\gamma)^{-1}S(\gamma) .$$

which when combined with (A4), yields the result. \square

Proof of Theorem 4 (c): Expand each element of $S_n(\tilde{\theta}, \gamma)$ about θ_0 :

$$S_n^a(\tilde{\theta}, \gamma) = S_n(\theta_0, \gamma) - M_n^a(\theta^*(\gamma), \gamma)(\tilde{\theta} - \theta_0)$$

where $\theta^*(\gamma)$ lies on a line segment between θ_0 and $\tilde{\theta}$. Since $\tilde{\theta} \rightarrow_p \theta_0$, the argument of (A2) allows us to rewrite this as

$$(A6) \quad S_n(\tilde{\theta}, \gamma) = S_n(\theta_0, \gamma) - M_n^*(\gamma)(\tilde{\theta} - \theta_0)$$

where $M_n^*(\gamma) \rightarrow_p M(\gamma)$, uniformly in $\gamma \in \Gamma$.

Expand each element of $h(\tilde{\theta})$ about θ_0 :

$$h^a(\tilde{\theta}) = h^a(\theta_0) + h_{\theta}^a(\theta^*)(\tilde{\theta} - \theta_0) ,$$

or, stacking equations, and using $h(\tilde{\theta}) = h(\theta_0) = 0$,

$$(A7) \quad 0 = h_{\theta}^*(\tilde{\theta} - \theta_0) ,$$

where $h_{\theta}^* \rightarrow_p h_{\theta}$.

For n sufficiently large, there exists a Lagrange multiplier vector $\tilde{\lambda}$ which satisfies

$$(A8) \quad S_n(\tilde{\theta}, \gamma) = \tilde{h}_{\theta}' \tilde{\lambda} ,$$

where $\tilde{h}_{\theta} = h_{\theta}(\tilde{\theta})$. (A8), (A6), (A7), and assumption 3 (v) combine to yield,

$$(A9) \quad \begin{aligned} \sqrt{n} h_{\theta}^* M_n^*(\gamma)^{-1} \tilde{h}_{\theta}' \tilde{\lambda} &= \sqrt{n} h_{\theta}^* M_n^*(\gamma)^{-1} S_n(\tilde{\theta}, \gamma) \\ &= \sqrt{n} h_{\theta}^* M_n^*(\gamma)^{-1} S_n(\theta_0, \gamma) - \sqrt{n} h_{\theta}^* (\tilde{\theta} - \theta_0) \end{aligned}$$

$$= \mathbf{h}_\theta^* \mathbf{M}_n^*(\gamma)^{-1} \sqrt{n} \mathbf{S}_n(\theta_o, \gamma) \Rightarrow \mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{S}(\gamma) .$$

Therefore

$$(A10) \quad \sqrt{n} \mathbf{S}_n(\tilde{\theta}, \gamma) = \sqrt{n} \tilde{\mathbf{h}}'_\theta \tilde{\lambda} = \tilde{\mathbf{h}}'_\theta \left[\mathbf{h}_\theta^* \mathbf{M}_n^*(\gamma)^{-1} \tilde{\mathbf{h}}'_\theta \right]^{-1} \mathbf{h}_\theta^* \mathbf{M}_n^*(\gamma)^{-1} \sqrt{n} \mathbf{S}_n(\theta_o, \gamma) \\ \Rightarrow \mathbf{h}'_\theta \left[\mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \right]^{-1} \mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{S}(\gamma)$$

and

$$(A11) \quad \sqrt{n} \tilde{\mathbf{h}}'_\theta \mathbf{M}_n(\tilde{\theta}, \gamma)^{-1} \mathbf{S}_n(\tilde{\theta}, \gamma) \Rightarrow \mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{S}(\gamma) \equiv \bar{\mathbf{S}}(\gamma) .$$

(A4) and (A11) combined with the CMT complete the proof.

Proof of Theorem 4 (d). First note that by a second-order Taylor's expansion of $Q_n(\tilde{\theta})$ about $\hat{\theta}(\gamma)$, and the first order conditions for estimation of $\hat{\theta}(\gamma)$,

$$(A12) \quad \text{LR}_n(\gamma) = 2n \left[Q_n(\hat{\theta}(\gamma), \gamma) - Q_n(\tilde{\theta}) \right] \\ = -2n \mathbf{S}_n(\hat{\theta}(\gamma), \gamma) (\tilde{\theta} - \hat{\theta}(\gamma)) + (\tilde{\theta} - \hat{\theta}(\gamma))' \mathbf{M}_n(\theta^*(\gamma), \gamma) (\tilde{\theta} - \hat{\theta}(\gamma)) \\ = \sqrt{n} (\tilde{\theta} - \hat{\theta}(\gamma))' \mathbf{M}_n^*(\gamma) \sqrt{n} (\tilde{\theta} - \hat{\theta}(\gamma)),$$

where $\mathbf{M}_n^*(\gamma) = \mathbf{M}_n(\theta^*(\gamma), \gamma) \xrightarrow{p} \mathbf{M}(\gamma)$, uniformly in $\gamma \in \Gamma$.

Expanding each element of $\mathbf{S}_n(\tilde{\theta}, \gamma)$ about $\hat{\theta}(\gamma)$ and stacking we find

$$(A13) \quad \mathbf{S}_n(\tilde{\theta}, \gamma) = \mathbf{S}_n(\hat{\theta}(\gamma), \gamma) - \mathbf{M}^{**}(\gamma) \sqrt{n} (\tilde{\theta} - \hat{\theta}(\gamma)) = -\mathbf{M}^{**}(\gamma) \sqrt{n} (\tilde{\theta} - \hat{\theta}(\gamma)) ,$$

where $\mathbf{M}^*(\gamma) \xrightarrow{p} \mathbf{M}(\gamma)$, uniformly in $\gamma \in \Gamma$. (A13) and (A10) give

$$(A14) \quad \sqrt{n} (\tilde{\theta} - \hat{\theta}(\gamma)) = -\mathbf{M}^{**}(\gamma)^{-1} \mathbf{S}_n(\tilde{\theta}, \gamma) \\ \Rightarrow -\mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \left[\mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \right]^{-1} \mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{S}(\gamma) .$$

(A12) and (A14) combine to yield

$$\text{LR}_n(\gamma) \Rightarrow \mathbf{S}(\gamma)' \mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \left[\mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \right]^{-1} \mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{S}(\gamma) \\ = \bar{\mathbf{S}}(\gamma)' \left[\mathbf{h}_\theta \mathbf{M}(\gamma)^{-1} \mathbf{h}'_\theta \right]^{-1} \bar{\mathbf{S}}(\gamma) = \mathbf{C}^*(\gamma) . \quad \square$$

Proof of Theorem 4. A corollary to Theorem 3 is that $\hat{\gamma} = \underset{\gamma \in \Gamma}{\text{Argmax}} \text{LR}_n(\gamma)$.

Part (a) follows from $\text{LR}_n(\gamma) \Rightarrow C(\gamma)$ and the CMT. Now

$$W_n = W_n(\hat{\gamma}) \Rightarrow C[\underset{\gamma \in \Gamma}{\text{Argmax}} C(\gamma)] = \sup_{\gamma \in \Gamma} C(\gamma),$$

and similarly for LM_n . The results for $\text{Sup}W_n$ and LM_n follow directly from Theorem 4 (b) (c) and the CMT. \square

Proof of Theorem 6.

$$(a) \quad \hat{\gamma} = \underset{\gamma \in \Gamma}{\text{Argmax}} \text{LR}_n(\gamma) \Rightarrow \underset{\gamma \in \Gamma}{\text{Argmax}} C^*(\gamma);$$

$$(b) \quad \text{LR}_n = \sup_{\gamma \in \Gamma} \text{LR}_n(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} C^*(\gamma);$$

$$(c) \quad W_n = W_n(\hat{\gamma}) \Rightarrow C[\underset{\gamma \in \Gamma}{\text{Argmax}} C^*(\gamma)];$$

$$\text{LM}_n = \text{LM}_n(\hat{\gamma}) \Rightarrow C[\underset{\gamma \in \Gamma}{\text{Argmax}} C^*(\gamma)];$$

$$(d) \quad \text{Sup}W_n = \sup_{\gamma \in \Gamma} W_n(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} C(\gamma);$$

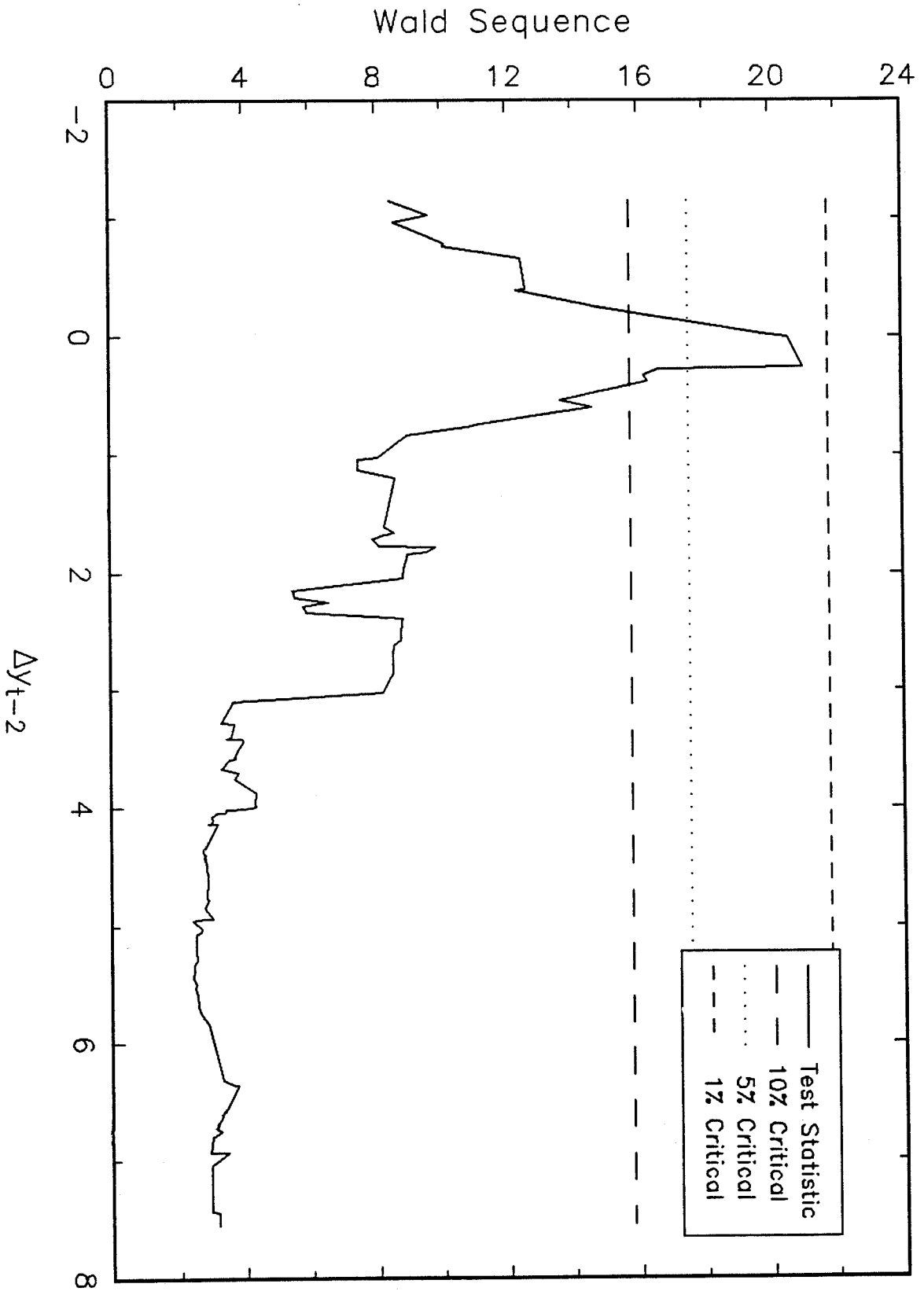
$$(e) \quad \text{SupLM}_n = \sup_{\gamma \in \Gamma} \text{LM}_n(\gamma) \Rightarrow \sup_{\gamma \in \Gamma} C(\gamma). \quad \square$$

REFERENCES

- Andrews, D.W.K. (1988): "Laws of large numbers for dependent non-identically distributed random variables," *Econometric Theory*, 4, 458–467.
- Andrews, D.W.K. (1990a): "Generic uniform convergence," Cowles Foundation Discussion Paper # 940.
- Andrews, D.W.K. (1990b): "Tests for parameter instability and structural change with unknown change point," Cowles Foundation Discussion Paper # 943.
- Andrews, D.W.K. (1990c): "An empirical process central limit theorem for dependent non-identically distributed random variables," Cowles Foundation Discussion Paper # 907–R.
- Bera, A.K. and M.L. Higgins (1990): "A test for conditional heteroskedasticity in time series models," University of Western Ontario.
- Box, G.E.P. and D.R. Cox (1964): "An analysis of transformations," *Journal of the Royal Statistical Society, Series B*, 26, 211–252.
- Chan, K.S. (1990): "Testing for threshold autoregression," *The Annals of Statistics*, 18, 1886–1894.
- Chan, K.S. and H. Tong (1985): "On estimating thresholds in autoregressive models," *Journal of Time Series Analysis*, 7, 179–190.
- Chan, K.S. and H. Tong (1990): "On likelihood ratio tests for autoregression," *Journal of the Royal Statistical Society, Series B*, 52, 469–476.
- Chu, C–S J. (1989): "New tests for parameter constancy in stationary and nonstationary regression models," unpublished manuscript, UCSD.
- Davies, R.B. (1977): "Hypothesis testing when a nuisance parameter in present only under the alternative," *Biometrika*, 64, 247–254.
- Davies, R.B. (1987): "Hypothesis testing when a nuisance parameter in present only under the alternative," *Biometrika*, 74, 33–43.
- Eichenbaum, M.S., L.P. Hansen, and K.J. Singleton (1988): "A time series analysis of representative agent models of consumption and leisure choice under uncertainty," *Quarterly Journal of Economics*, 53, 51–78.
- Engle, R.F., D.M. Lilien, and R.P. Robins (1987): "Estimating time varying risk premia in the term structure: The Arch–M model," *Econometrica*, 55, 391–407.
- Epstein, L. and S. Zin (1989): "Substitution, risk aversion and the temporal behavior of consumption and asset returns I: A theoretical framework," *Econometrica*, 57, 937–969.
- Giovannini, A. and P. Jorion (1989): "The time-variation of risk and return in the foreign exchange and stock markets," *Journal of Finance*, 44, 307–325.

- Gallant, A.R. (1987): *Nonlinear Statistical Models*, NY: John Wiley and Sons.
- J.D. Hamilton (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.
- Hansen, B.E. (1990): "Lagrange multiplier tests for parameter instability in non-linear models," unpublished manuscript, University of Rochester.
- Hansen, B.E. (1991a): "Tests for parameter instability in regressions with I(1) processes" unpublished manuscript, University of Rochester.
- Hansen, B.E. (1991b): "The likelihood ratio test under non-standard conditions: testing the Markov trend model of GNP," University of Rochester Discussion Paper 279.
- Heckman, J. and S. Polachek (1974): "Empirical evidence on the functional form of the earnings–schooling relationship," *Journal of the American Statistical Association*, 69, 350–354.
- Higgins, M.L. and A.K. Bera (1989): "A class of nonlinear ARCH models," unpublished manuscript, University of Wisconsin–Milwaukee.
- Kim, H–J, and D. Siegmund (1989): "The likelihood ratio test for a change–point in simple linear regression," *Biometrika*, 76, 409–423.
- Neftci, S.N. (1984): "Are economic time series asymmetric over the business cycle?" *Journal of Political Economy*, 92, 307–328.
- Potter, S.M. (1991): "A nonlinear approach to U.S. GNP," University of Wisconsin, Madison.
- Quandt, R. (1960): "Tests of the hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association*, 55, 324–330.
- Richardson, G.D. and B.B. Bhattacharyya (1990): "Strong consistency in nonlinear regressions," *Econometric Theory*, 6, 459–465.
- Stock, J.H. (1987): "Measuring business cycle time," *Journal of Political Economy*, 95, 1240–1261.
- Tong, H. (1983): *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics*, 21, Berlin: Springer.
- White, H. (1980): "A heteroskedastic–consistent covariance matrix estimator and a direct test for heteroskedasticity", *Econometrica*, 48, 817–838.

U.S. GNP: 1947-1990



Asymptotic Distribution

