Econometrica, Vol. 68, No. 3 (May, 2000), 575–603

SAMPLE SPLITTING AND THRESHOLD ESTIMATION

BY BRUCE E. HANSEN[1]

Threshold models have a wide variety of applications in economics. Direct applications include models of separating and multiple equilibria. Other applications include empirical sample splitting when the sample split is based on a continuously-distributed variable such as firm size. In addition, threshold models may be used as a parsimonious strategy for nonparametric function estimation. For example, the threshold autoregressive model (TAR) is popular in the nonlinear time series literature.

Threshold models also emerge as special cases of more complex statistical frameworks, such as mixture models, switching models, Markov switching models, and smooth transition threshold models. It may be important to understand the statistical properties of threshold models as a preliminary step in the development of statistical tools to handle these more complicated structures.

Despite the large number of potential applications, the statistical theory of threshold estimation is undeveloped. It is known that threshold estimates are super-consistent, but a distribution theory useful for testing and inference has yet to be provided.

This paper develops a statistical theory for threshold estimation in the regression context. We allow for either cross-section or time series observations. Least squares estimation of the regression parameters is considered. An asymptotic distribution theory for the regression estimates (the threshold and the regression slopes) is developed. It is found that the distribution of the threshold estimate is nonstandard. A method to construct asymptotic confidence intervals is developed by inverting the likelihood ratio statistic. It is shown that this yields asymptotically conservative confidence regions. Monte Carlo simulations are presented to assess the accuracy of the asymptotic approximations. The empirical relevance of the theory is illustrated through an application to the multiple equilibria growth model of Durlauf and Johnson (1995).

KEYWORDS: Confidence intervals, nonlinear regression, growth regressions, regime shifts.

1. INTRODUCTION

A ROUTINE PART OF AN EMPIRICAL ANALYSIS of a regression model such as $y_i = \beta' x_i + e_i$ is to see if the regression coefficients are stable when the model is estimated on appropriately selected subsamples. Sometimes the subsamples are selected on categorical variables, such as gender, but in other cases the subsamples are selected based on continuous variables, such as firm size. In the latter case, some decision must be made concerning what is the appropriate threshold (i.e., how big must a firm be to be categorized as "large") at which to split the sample. When this value is unknown, some method must be employed in its selection.

[1] This research was supported by a grant from the National Science Foundation and an Alfred P. Sloan Foundation Research Fellowship. Thanks go to Robert de Jong and James MacKinnon for insightful comments and Mehmet Caner and Metin Celebi for excellent research assistance. Special thanks go to two diligent referees whose careful readings of multiple drafts of this paper have eliminated several serious errors.

Such practices can be formally treated as a special case of the threshold regression model. These take the form

$$y_i = \theta_1' x_i + e_i, \qquad q_i \le \gamma, \tag{1}$$

$$y_i = \theta_2' x_i + e_i, \qquad q_i > \gamma, \tag{2}$$

where q_i may be called the threshold variable, and is used to split the sample into two groups, which we may call "classes," or "regimes," depending on the context. The random variable e_i is a regression error.

Formal threshold models arise in the econometrics literature. One example is the Threshold Autoregressive (TAR) model of Tong (1983, 1990), recently explored for U.S. GNP by Potter (1995). In Potter's model, y_i is GNP growth and x_i and q_i are lagged GNP growth rates. The idea is to allow important nonlinearities in the conditional expectation function without over-parameterization. From a different perspective, Durlauf and Johnson (1995) argue that models with multiple equilibria can give rise to threshold effects of the form given in model (1)–(2). In their formulation, the regression is a standard Barro-styled cross-country growth equation, but the sample is split into two groups, depending on whether the initial endowment is above a specific threshold.

The primary purpose of this paper is to derive a useful asymptotic approximation to the distribution of the least-squares estimate $\hat{\gamma}$ of the threshold parameter γ. The only previous attempt (of which I am aware) is Chan (1993) who derives the asymptotic distribution of $\hat{\gamma}$ for the TAR model. Chan finds that $n(\hat{\gamma} - \gamma_0)$ converges in distribution to a functional of a compound Poisson process. Unfortunately, his representation depends upon a host of nuisance parameters, including the marginal distribution of x_i and all the regression coefficients. Hence, this theory does not yield a practical method to construct confidence intervals.

We take a different approach, taking a suggestion from the change-point literature, which considers an analog of model (1)–(2) with $q_i = i$. Let $\delta_n = \theta_2 - \theta_1$ denote the "threshold effect." The proposed solution is to let $\delta_n \to 0$ as $n \to \infty$. (We will hold θ_2 fixed and thereby make θ_1 approach θ_2 as $n \to \infty$.) Under this assumption, it has been found (see Picard (1985) and Bai (1997)) that the asymptotic distribution of the changepoint estimate is nonstandard yet free of nuisance parameters (other than a scale effect). Interestingly, we find in the threshold model that the asymptotic distribution of the threshold estimate $\hat{\gamma}$ is of the same form as that found for the change-point model, although the scale factor is different.

The changepoint literature has confined attention to the sampling distribution of the threshold estimate. We refocus attention on test statistics, and are the first to study likelihood ratio tests for the threshold parameter. We find that the likelihood ratio test is asymptotically pivotal when δ_n decreases with sample size, and that this asymptotic distribution is an upper bound on the asymptotic distribution for the case that δ_n does not decrease with sample size. This allows

us to construct asymptotically valid confidence intervals for the threshold based on inverting the likelihood ratio statistic. This method is easy to apply in empirical work. A GAUSS program that computes the estimators and test statistics is available on request from the author or from his Web homepage.

The paper is organized as follows. Section 2 outlines the method of least squares estimation of threshold regression models. Section 3 presents the asymptotic distribution theory for the threshold estimate and the likelihood ratio statistic for tests on the threshold parameter. Section 4 outlines methods to construct asymptotically valid confidence intervals. Methods are presented for the threshold and for the slope coefficients. Simulation evidence is provided to assess the adequacy of the asymptotic approximations. Section 5 reports an application to the multiple equilibria growth model of Durlauf and Johnson (1995). The mathematical proofs are left to an Appendix.

2. ESTIMATION

The observed sample is $\{y_i, x_i, q_i\}_{i=1}^n$, where y_i and q_i are real-valued and x_i is an m-vector. The *threshold variable* q_i may be an element of x_i, and is assumed to have a continuous distribution. A sample-split or *threshold regression* model takes the form (1)–(2). This model allows the regression parameters to differ depending on the value of q_i. To write the model in a single equation, define the dummy variable $d_i(\gamma) = \{q_i \le \gamma\}$ where $\{\cdot\}$ is the indicator function and set $x_i(\gamma) = x_i d_i(\gamma)$, so that (1)–(2) equal

$$y_i = \theta' x_i + \delta_n' x_i(\gamma) + e_i \tag{3}$$

where $\theta = \theta_2$. Equation (3) allows all of the regression parameters to switch between the regimes, but this is not essential to the analysis. The results generalize to the case where only a subset of parameters switch between regimes and to the case where some regressors only enter in one of the two regimes.

To express the model in matrix notation, define the $n \times 1$ vectors Y and e by stacking the variables y_i and e_i, and the $n \times m$ matrices X and X_γ by stacking the vectors x_i' and $x_i(\gamma)'$. Then (3) can be written as

$$Y = X\theta + X_\gamma \delta_n + e. \tag{4}$$

The regression parameters are $(\theta, \delta_n, \gamma)$, and the natural estimator is least squares (LS). Let

$$S_n(\theta, \delta, \gamma) = (Y - X\theta - X_\gamma \delta)'(Y - X\theta - X_\gamma \delta) \tag{5}$$

be the sum of squared errors function. Then by definition the LS estimators $\hat{\theta}, \hat{\delta}, \hat{\gamma}$ jointly minimize (5). For this minimization, γ is assumed to be restricted to a bounded set $[\underline{\gamma}, \overline{\gamma}] = \Gamma$. Note that the LS estimator is also the MLE when e_i is iid $N(0, \sigma^2)$.

The computationally easiest method to obtain the LS estimates is through concentration. Conditional on γ, (4) is linear in θ and δ_n, yielding the condi-

tional OLS estimators $\hat{\theta}(\gamma)$ and $\hat{\delta}(\gamma)$ by regression of Y on $X_\gamma^* = [X\ X_\gamma]$. The concentrated sum of squared errors function is

$$S_n(\gamma) = S_n\big(\hat{\theta}(\gamma), \hat{\delta}(\gamma), \gamma\big) = Y'Y - Y'X_\gamma^*(X_\gamma^{*\prime}X_\gamma^*)^{-1}X_\gamma^{*\prime}Y,$$

and $\hat{\gamma}$ is the value that minimizes $S_n(\gamma)$. Since $S_n(\gamma)$ takes on less than n distinct values, $\hat{\gamma}$ can be defined uniquely as

$$\hat{\gamma} = \underset{\gamma \in \Gamma_n}{\operatorname{argmin}} S_n(\gamma)$$

where $\Gamma_n = \Gamma \cap \{q_1, \ldots, q_n\}$, which requires less than n function evaluations. The slope estimates can be computed via $\hat{\theta} = \hat{\theta}(\hat{\gamma})$ and $\hat{\delta} = \hat{\delta}(\hat{\gamma})$.

If n is very large, Γ can be approximated by a grid. For some $N < n$, let $q_{(j)}$ denote the (j/N)th quantile of the sample $\{q_1, \ldots, q_n\}$, and let $\Gamma_N = \Gamma \cap \{q_{(1)}, \ldots, q_{(n)}\}$. Then $\hat{\gamma}_N = \operatorname{argmin}_{\gamma \in \Gamma_N} S_n(\gamma)$ is a good approximation to $\hat{\gamma}$ which only requires N function evaluations.

From a computational standpoint, the threshold model (1)–(2) is quite similar to the changepoint model (where the threshold variable equals time, $q_i = i$). Indeed, if the observed values of q_i are distinct, the parameters can be estimated by sorting the data based on q_i, and then applying known methods for changepoint models. When there are tied values of q_i, estimation is more delicate, as the sorting operation is no longer well defined nor appropriate. From a distributional standpoint, however, the threshold model differs considerably from the changepoint model. One way to see this is to note that if the regressors x_i contain q_i, as is typical in applications, then sorting the data by q_i induces a trend into the regressors x_i, so the threshold model is somewhat similar to a changepoint model with trended data. The presence of trends is known to alter the asymptotic distributions of changepoint tests (see Hansen (1992, 2000) and Chu and White (1992)). More importantly, the distribution of changepoint estimates (or the construction of confidence intervals) has not been studied for this case. Another difference is that the stochastic process $R_n(\gamma) = \sum_{i=1}^n x_i e_i \{q_i \le \gamma\}$ is a martingale (in γ) when $q_i = i$ (the changepoint model), but it is not necessarily so in the threshold model (unless the data are independent across i). This difference may appear minor, but it requires the use of a different set of asymptotic tools.

3. DISTRIBUTION THEORY

3.1. *Assumptions*

Define the moment functionals

$$M(\gamma) = E(x_i x_i' \{q_i \le \gamma\}), \tag{6}$$

$$D(\gamma) = E(x_i x_i' | q_i = \gamma), \tag{7}$$

and

$$V(\gamma) = E(x_i x_i' e_i^2 | q_i = \gamma).$$

Let $f(q)$ denote the density function of q_i, γ_0 denote the true value of γ, $D = D(\gamma_0)$, $V = V(\gamma_0)$, $f = f(\gamma_0)$, and $M = E(x_i x_i')$.

ASSUMPTION 1:

1. *(x_i, q_i, e_i) is strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$.*
2. *$E(e_i|F_{i-1}) = 0$.*
3. *$E|x_i|^4 < \infty$ and $E|x_i e_i|^4 < \infty$.*
4. *For all $\gamma \in \Gamma$, $E(|x_i|^4 e_i^4 | q_i = \gamma) \le C$ and $E(|x_i|^4 | q_i = \gamma) \le C$ for some $C < \infty$, and $f(\gamma) \le \bar{f} < \infty$.*
5. *$f(\gamma)$, $D(\gamma)$, and $V(\gamma)$ are continuous at $\gamma = \gamma_0$.*
6. *$\delta_n = cn^{-\alpha}$ with $c \neq 0$ and $0 < \alpha < \frac{1}{2}$.*
7. *$c'Dc > 0$, $c'Vc > 0$, and $f > 0$.*
8. *$M > M(\gamma) > 0$ for all $\gamma \in \Gamma$.*

Assumption 1.1 is relevant for time series applications, and is trivially satisfied for independent observations. The assumption of stationarity excludes time trends and integrated processes. The ρ-mixing assumption[2] controls the degree of time series dependence, and is weaker than uniform mixing, yet stronger than strong mixing. It is sufficiently flexible to embrace many nonlinear time series processes including threshold autoregressions. Indeed, Chan (1989) has demonstrated that a strong form of geometric ergodicity holds for TAR processes and, as discussed in the proof of Proposition 1 of Chan (1993), this implies $\rho_m = O(\rho^m)$ with $|\rho| < 1$, which implies Assumption 1.1.

Assumption 1.2 imposes that (1)–(2) is a correct specification of the conditional mean. Assumptions 1.3 and 1.4 are unconditional and conditional moment bounds. Assumption 1.5 requires the threshold variable to have a continuous distribution, and essentially requires the conditional variance $E(e_i^2|q_i = \gamma)$ to be continuous at γ_0, which excludes regime-dependent heteroskedasticity.

Assumption 1.6 may appear unusual. It specifies that the difference in regression slopes gets small as the sample size increases. Conceptually, this implies that we are taking an asymptotic approximation valid for small values of δ_n. The parameter α controls the rate at which δ_n decreases to zero, i.e., how small we are forcing δ_n to be. Smaller values of α are thus less restrictive. The reason for Assumption 1.6 is that Chan (1993) found that with δ_n fixed, $n(\hat{\gamma} - \gamma_0)$ converged to an asymptotic distribution that was dependent upon nuisance parameters and thus not particularly useful for calculation of confidence sets. The difficulty is due to the $O_p(n^{-1})$ rate of convergence. By letting δ_n tend towards zero, we reduce the rate of convergence and find a simpler asymptotic distribution.

Assumption 1.7 is a full-rank condition needed to have nondegenerate asymptotic distributions. While the restriction $c'Dc > 0$ might appear innocuous, it

[2] For a definition, see Ibragimov (1975) and Peligrad (1982).

excludes the interesting special case of a "continuous threshold" model, which is (1)–(2) with $x_i = (1\ q_i)'$ and $\delta_n'\gamma_0^* = 0$ where $\gamma_0^* = (1\ \gamma_0)'$. In this case the conditional mean takes the form of a continuous linear spline. From definition (7) we can calculate for this model that $c'Dc = c'E(x_i x_i'|q_i = \gamma_0)c = c'\gamma_0^*\gamma_0^{*\prime}c = 0$. A recent paper that explores the asymptotic distribution of the least squares estimates in this model is Chan and Tsay (1998).

Assumption 1.8 is a conventional full-rank condition which excludes multicollinearity. Note that this assumption restricts Γ to a proper subset of the support of q_i. This is a technical condition which simplifies our consistency proof.

3.2. *Asymptotic Distribution*

A two-sided Brownian motion $W(r)$ on the real line is defined as

$$W(r) = \begin{cases} W_1(-r), & r < 0, \\ 0, & r = 0, \\ W_2(r), & r > 0, \end{cases}$$

where $W_1(r)$ and $W_2(r)$ are independent standard Brownian motions on $[0, \infty)$.

THEOREM 1: *Under Assumption* 1, $n^{1-2\alpha}(\hat{\gamma} - \gamma_0) \to_d \omega T$, *where*

$$\omega = \frac{c'Vc}{(c'Dc)^2 f}$$

and

$$T = \underset{-\infty < r < \infty}{\operatorname{argmax}} \left[-\frac{1}{2}|r| + W(r) \right].$$

Theorem 1 gives the rate of convergence and asymptotic distribution of the threshold estimate $\hat{\gamma}$. The rate of convergence is $n^{1-2\alpha}$, which is decreasing in α. Intuitively, a larger α decreases the threshold effect δ_n, which decreases the sample information concerning the threshold γ, reducing the precision of any estimator of γ.

Theorem 1 shows that the distribution of the threshold estimate under our "small effect" asymptotics takes a similar form to that found for changepoint estimates. For the latter theory, see Picard (1985), Yao (1987), Dümbgen (1991), and Bai (1997). The difference is that the asymptotic precision of $\hat{\gamma}$ is proportional to the matrix $E(x, x_i'|q_i = \gamma_0)$ while in the changepoint case the asymptotic precision is proportional to the unconditional moment matrix $E(x_i x_i')$. It is interesting to note that these moments are equal when x_i and q_i are independent, which would not be typical in applications.

The asymptotic distribution in Theorem 1 is scaled by the ratio ω. In the leading case of conditional homoskedasticity

$$E(e_i^2|q_i) = \sigma^2, \tag{9}$$

then $V = \sigma^2 D$ and ω simplifies to

$$\omega = \frac{\sigma^2}{(c'Dc)f}.$$

The asymptotic distribution of $\hat{\gamma}$ is less dispersed when ω is small, which occurs when σ^2 is small, $f(\gamma_0)$ is large (so that many observations are near the threshold), and/or $|c|$ is large (a large threshold effect).

The distribution function for T is known. (See Bhattacharya and Brockwell (1976).) Let $\Phi(x)$ denote the cumulative standard normal distribution function. Then for $x \geq 0$,

$$P(T \leq x) = 1 + \sqrt{\frac{x}{2\pi}} \exp\left(-\frac{x}{8}\right) + \frac{3}{2}\exp(x)\Phi\left(-\frac{3\sqrt{x}}{2}\right) - \left(\frac{x+5}{2}\right)\Phi\left(-\frac{\sqrt{x}}{2}\right).$$

and for $x < 0$, $P(T \leq x) = 1 - P(T \leq -x)$. A plot of the density function of T is given in Figure 1.

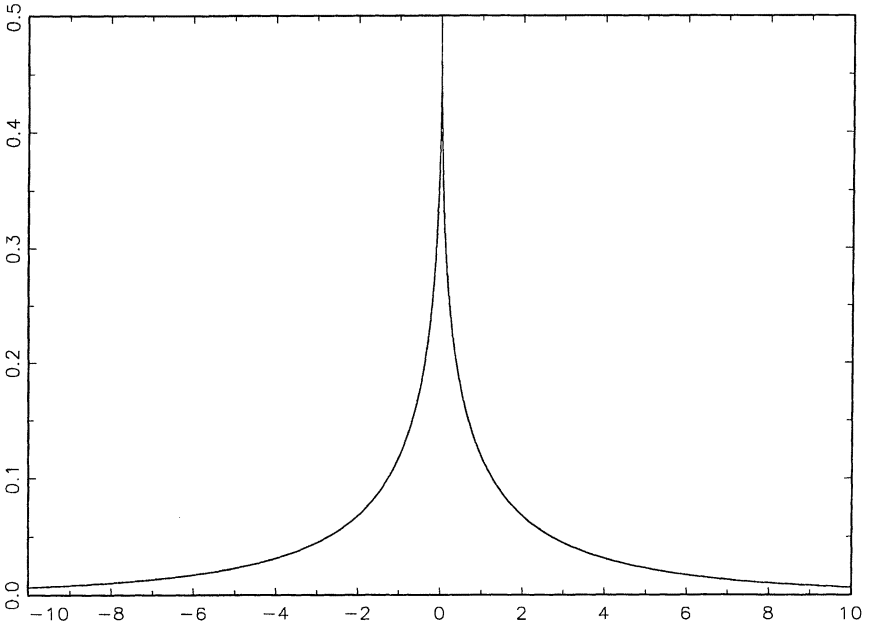

FIGURE 1.—Asymptotic density of threshold estimator.

3.3. *Likelihood Ratio Test*

To test the hypothesis H_0: $\gamma = \gamma_0$, a standard approach is to use the likelihood ratio statistic under the auxiliary assumption that e_i is iid $N(0, \sigma^2)$. Let

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})}.$$

The likelihood ratio test of H_0 is to reject for large values of $LR_n(\gamma_0)$.

THEOREM 2: *Under Assumption* 1,

$$LR_n(\gamma_0) \to_d \eta^2 \xi,$$

where

$$\xi = \max_{s \in R}[2W(s) - |s|]$$

and

$$\eta^2 = \frac{c'Vc}{\sigma^2 c'Dc}.$$

The distribution function of ξ *is* $P(\xi \le x) = (1 - e^{-x/2})^2$.

If homoskedasticity (9) holds, then $\eta^2 = 1$ and the asymptotic distribution of $LR_n(\gamma_0)$ is free of nuisance parameters. If heteroskedasticity is suspected, η^2 must be estimated. We discuss this in the next section.

Theorem 2 gives the large sample distribution of the likelihood ratio test for hypotheses on γ. The asymptotic distribution is nonstandard, but free of nuisance parameters under (9). Since the distribution function is available in a simple closed form, it is easy to generate p-values for observed test statistics. Namely,

$$p_n = 1 - \left(1 - \exp\left(-\frac{1}{2}LR_n(\gamma_0)^2\right)\right)^2$$

is the asymptotic p-value for the likelihood ratio test. Critical values can be calculated by direct inversion of the distribution function. Thus a test of H_0: $\gamma = \gamma_0$ rejects at the asymptotic level of a if $LR_n(\gamma_0)$ exceeds $c_\xi(1-a)$, where $c_\xi(z) = -2\ln(1 - \sqrt{z})$. Selected critical values are reported in Table I.

TABLE I

ASYMPTOTIC CRITICAL VALUES

	.80	.85	.90	.925	.95	.975	.99
$P(\xi \le x)$	4.50	5.10	5.94	6.53	7.35	8.75	10.59

3.4. *Estimation of* η^2

The asymptotic distribution of Theorem 2 depends on the nuisance parameter η^2. It is therefore necessary to consistently estimate this parameter. Let $r_{1i} = (\delta_n' x_i)^2(e_i^2/\sigma^2)$ and $r_{2i} = (\delta_n' x_i)^2$. Then

$$\eta^2 = \frac{E(r_{1i}|q_i = \gamma_0)}{E(r_{2i}|q_i = \gamma_0)} \tag{10}$$

is the ratio of two conditional expectations. Since r_{1i} and r_{2i} are unobserved, let $\hat{r}_{1i} = (\hat{\delta}' x_i)^2(\hat{e}_i^2/\hat{\sigma}^2)$ and $\hat{r}_{2i} = (\hat{\delta}' x_i)^2$ denote their sample counterparts.

A simple estimator of the ratio (10) uses a polynomial regression, such as a quadratic. For $j = 1$ and 2, fit the OLS regressions

$$\hat{r}_{ji} = \hat{\mu}_{j0} + \hat{\mu}_{j1} q_i + \hat{\mu}_{j2} q_i^2 + \hat{\varepsilon}_{ji},$$

and then set

$$\hat{\eta}^2 = \frac{\hat{\mu}_{10} + \hat{\mu}_{11}\hat{\gamma} + \hat{\mu}_{12}\hat{\gamma}^2}{\hat{\mu}_{20} + \hat{\mu}_{21}\hat{\gamma} + \hat{\mu}_{22}\hat{\gamma}^2}.$$

An alternative is to use kernel regression. The Nadaraya-Watson kernel estimator is

$$\hat{\eta}^2 = \frac{\sum_{i=1}^n K_h(\hat{\gamma} - q_i)\hat{\tau}_{1i}}{\sum_{i=1}^n K_h(\hat{\gamma} - q_i)\hat{\tau}_{2i}}$$

where $K_h(u) = h^{-1}K(u/h)$ for some bandwidth h and kernel $K(u)$, such as the Epanechnikov $K(u) = \frac{3}{4}(1 - u^2)\{|u| \le 1\}$. The bandwidth h may be selected according to a minimum mean square error criterion (see Hardle and Linton (1994)).

4. CONFIDENCE INTERVALS

4.1. *Threshold Estimate*

A common method to form confidence intervals for parameters is through the inversion of Wald or t statistics. To obtain a confidence interval for γ, this would involve the distribution T from Theorem 1 and an estimate of the scale parameter ω. While T is parameter-independent, ω is directly a function of δ_n and indirectly a function of γ_0 (through $D(\gamma_0)$). When asymptotic sampling distributions depend on unknown parameters, the Wald statistic can have very poor finite sample behavior. In particular, Dufour (1997) argues that Wald statistics have particularly poorly-behaved sampling distributions when the parameter has a region where identification fails. The threshold regression model is an example where this occurs, as the threshold γ is not identified when

$\delta_n = 0$. These concerns have encouraged us to explore the construction of confidence regions based on the likelihood ratio statistic $LR_n(\gamma)$.

Let C denote the desired asymptotic confidence level (e.g. $C = .95$), and let $c = c_\xi(C)$ be the C-level critical value for ξ (from Table I). Set

$$\hat{\Gamma} = \{\gamma : LR_n(\gamma) \le c\}.$$

Theorem 2 shows that $P(\gamma_0 \in \hat{\Gamma}) \to C$ as $n \to \infty$ under the homoskedasticity assumption (9). Thus $\hat{\Gamma}$ is an asymptotic C-level confidence region for γ. A graphical method to find the region $\hat{\Gamma}$ is to plot the likelihood ratio $LR_n(\gamma)$ against γ and draw a flat line at c. (Note that the likelihood ratio is identically zero at $\gamma = \hat{\gamma}$.) Equivalently, one may plot the residual sum of squared errors $S_n(\gamma)$ against γ, and draw a flat line at $S_n(\hat{\gamma}) + \hat{\sigma}^2 c$.

If the homoskedasticity condition (9) does not hold, we can define a scaled likelihood ratio statistic:

$$LR_n^*(\gamma) = \frac{LR_n(\gamma)}{\hat{\eta}^2} = \frac{S_n(\gamma) - S_n(\hat{\gamma})}{\hat{\sigma}^2 \hat{\eta}^2}$$

and an amended confidence region

$$\hat{\Gamma}^* = \{\gamma : LR_n^*(\gamma) \le c\}.$$

Since $\hat{\eta}^2$ is consistent for η^2, $P(\gamma_0 \in \hat{\Gamma}^*) \to C$ as $n \to \infty$ whether or not (9) holds, so $\hat{\Gamma}^*$ is a heteroskedasticity-robust asymptotic C-level confidence region for γ.

These confidence intervals are asymptotically correct under the assumption that $\delta_n \to 0$ as $n \to \infty$, which suggests that the actual coverage may differ from the desired level for large values of δ_n. We now consider the case of $\alpha = 0$, which implies that δ_n is fixed as n increases. We impose the stronger condition that the errors e_i are iid $N(0, \sigma^2)$, strictly independent of the regressors x_i and threshold variable q_i.

THEOREM 3: *Under Assumption* 1, *modifying part* 6 *so that* $\alpha = 0$, *and the errors* e_i *are iid* $N(0, \sigma^2)$ *strictly independent of the regressors* x_i *and threshold variable* q_i, *then*

$$P(LR_n(\gamma_0) \ge x) \le P(\xi \ge x) + o(1).$$

Theorem 3 shows that at least in the case of iid Gaussian errors, the likelihood ratio test is asymptotically conservative. Thus inferences based on the confidence region $\hat{\Gamma}$ are asymptotically valid, even if δ_n is relatively large. Unfortunately, we do not know if Theorem 3 generalizes to the case of non-normal errors or regressors that are not strictly exogenous. The proof of Theorem 3 relies on the Gaussian error structure and it is not clear how the theorem would generalize.

4.2. *Simulation Evidence*

We use simulation techniques to compare the coverage probabilities of the confidence intervals for γ. We use the simple regression model (3) with iid data and $x_i = (1 \;\; z_i)'$, $e_i \sim N(0,1)$, $q_i \sim N(2,1)$, and $\gamma = 2$. The regressor z_i was either iid $N(0,1)$ or $z_i = q_i$. The likelihood ratio statistic is invariant to θ. Partitioning $\delta_n = (\delta_1 \;\; \delta_2)'$ we set $\delta_1 = 0$ and assessed the coverage probability of the confidence interval $\hat{I}$ as we varied δ_2 and n. We set $\delta_2 = .25$, .5, 1.0, 1.5, and 2.0 and $n = 50$, 100, 250, 500, and 1000. Using 1000 replications Table II reports the coverage probabilities for nominal 90% confidence intervals.

The results are quite informative. For all cases, the actual coverage rates increase as n increases or δ_2 increases, which is consistent with the prediction of Theorem 3. For small sample sizes and small threshold effects, the coverage rates are lower than the nominal 90%. As expected, however, as the threshold effect δ_2 increases, the rejection rates rise and become quite conservative.

4.3. *Slope Parameters*

Letting $\underline{\theta} = (\theta, \delta_n)$ and $\hat{\underline{\theta}} = (\hat{\theta}, \hat{\delta})$. Lemma A.12 in the Appendix shows that

$$\sqrt{n}\left(\hat{\underline{\theta}} - \underline{\theta}\right) \to_d Z \sim N(0, V_\theta) \tag{11}$$

where V_θ is the standard asymptotic covariance matrix if $\gamma = \gamma_0$ were fixed. This means that we can approximate the distribution of $\hat{\underline{\theta}}$ by the conventional normal approximation as if γ were known with certainty. Let $\hat{\Theta}(\gamma)$ denote the conventional asymptotic C-level confidence region for $\underline{\theta}$ constructed under the assumption that γ is known. (11) shows that $P(\underline{\theta} \in \hat{\Theta}(\hat{\gamma})) \to C$ as $n \to \infty$.

In finite samples, this procedure seems likely to under-represent the true sampling uncertainty, since it is not the case that $\hat{\gamma} = \gamma_0$ in any given sample. It may be desirable to incorporate this uncertainty into our confidence intervals for $\underline{\theta}$. This appears difficult to do using conventional techniques, as $\hat{\underline{\theta}}(\gamma)$ is not differentiable with respect to γ, and $\hat{\gamma}$ is non-normally distributed. A simple yet constructive technique is to use a Bonferroni-type bound. For any $\rho < 1$, let

TABLE II

CONFIDENCE INTERVAL CONVERGENCE FOR γ AT 10% LEVEL

	$x_i = q_i$					$x_i \sim N(0,1)$				
$\delta_2 =$	.25	.5	1.0	1.5	2.0	.25	.5	1.0	1.5	2.0
$n = 50$	.86	.87	.93	.97	.99	.90	.87	.93	.93	.97
$n = 100$	.82	.90	.96	.98	.99	.84	.86	.92	.96	.95
$n = 250$	.83	.93	.97	.98	.99	.80	.92	.94	.96	.98
$n = 500$	.90	.93	.97	.98	.99	.81	.93	.95	.96	.98
$n = 1000$	.90	.93	.98	.99	.99	.86	.93	.94	.96	.97

$\hat{\Gamma}(\rho)$ denote the confidence interval for γ with asymptotic coverage ρ. For each $\gamma \in \hat{\Gamma}(\rho)$ construct the pointwise confidence region $\hat{\Theta}(\gamma)$ and then set

$$\hat{\Theta}_\rho = \bigcup_{\gamma \in \hat{\Gamma}(\rho)} \hat{\Theta}(\gamma).$$

Since $\hat{\Theta}_\rho \supset \hat{\Theta}(\hat{\gamma})$, it follows that $P(\theta \in \hat{\Theta}_\rho) \geq P(\theta \in \hat{\Theta}(\hat{\gamma})) \to C$ as $n \to \infty$.

This procedure is assessed using a simple Monte Carlo simulation. In Table III we report coverage rates of a nominal 95% confidence interval ($C = .95$) on δ_2. The same design is used as in the previous section, although the results are reported only for the case x_i independent of q_i and a more narrow set of n and δ_2 to save space. We tried $\rho = 0, .5, .8$, and .95. As expected, the simple rule $\hat{\Theta}_0$ is somewhat liberal for small θ_2 and n, but is quite satisfactory for large n or δ_2. In fact, all choices for ρ lead to similar results for large δ_2. For small δ_2 and n, the best choice may be $\rho = .8$, although this may produce somewhat conservative confidence intervals for small δ_2.

In summary, while the naive choice $\hat{\Theta}_0 = \hat{\Theta}(\hat{\gamma})$ works fairly well for large n and/or large threshold effects, it has insufficient coverage probability for small n or threshold effect. This problem can be solved through the conservative procedure $\hat{\Theta}_\rho$ with $\rho > 0$, and the choice $\rho = .8$ appears to work reasonably well in a simulation.

5. APPLICATION: GROWTH AND MULTIPLE EQUILIBRIA

Durlauf and Johnson (1995) suggest that cross-section growth behavior may be determined by initial conditions. They explore this hypothesis using the Summers-Heston data set, reporting results obtained from a *regression tree* methodology. A regression tree is a special case of a multiple threshold regression. The estimation method for regression trees due to Breiman et al. (1984) is somewhat ad hoc, with no known distributional theory. To illustrate the usefulness of our estimation theory, we apply our techniques to regressions similar to those reported by Durlauf-Johnson.

TABLE III

CONFIDENCE INTERVAL CONVERGENCE FOR δ_2 AT 5% LEVEL

	$n = 100$				$n = 250$				$n = 500$			
$\delta_2 =$	.25	.5	1.0	2.0	.25	.5	1.0	2.0	.25	.5	1.0	2.0
$\hat{\Theta}_0$	.90	.93	.96	.95	.90	.95	.95	.94	.91	.97	.94	.94
$\hat{\Theta}_{.5}$	.90	.95	.96	.95	.94	.96	.96	.94	.94	.98	.94	.95
$\hat{\Theta}_{.8}$	.95	.97	.97	.96	.97	.98	.96	.94	.97	.98	.95	.95
$\hat{\Theta}_{.95}$	.99	.99	.95	.94	.99	.99	.97	.94	.99	.99	.95	.95

The model seeks to explain real GDP growth. The specification is

$$\begin{aligned}\ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960} &= \zeta + \beta \ln(Y/L)_{i,1960} + \pi_1 \ln(I/Y)_i \\ &\quad + \pi_2 \ln(n_i + g + \delta) + \pi_3 \ln(SCHOOL)_i + e_i,\end{aligned}$$

where for each country i:

- $(Y/L)_{i,t}$ = real GDP per member of the population aged 15–64 in year t;
- $(I/Y)_i$ = investment to GDP ratio;
- n_i = growth rate of the working-age population;
- $(SCHOOL)_i$ = fraction of working-age population enrolled in secondary school.

The variables not indexed by t are annual averages over the period 1960–1985. Following Durlauf-Johnson, we set $g + \delta = 0.05$.

Durlauf-Johnson estimate (12) for four regimes selected via a regression tree using two possible threshold variables that measure initial endowment: per capita output Y/L and the adult literacy rate LR, both measured in 1960. The authors argue that the error e_i is heteroskedastic so present their results with heteroskedasticity-corrected standard errors. We follow their lead and use heteroskedasticity-consistent procedures, estimating the nuisance parameter η^2 using an Epanechnikov kernel with a plug-in bandwidth.

Since the theory outlined in this paper only allows one threshold and one threshold variable, we first need to select among the two threshold variables, and verify that there is indeed evidence for a threshold effect. We do so by employing the heteroskedasticity-consistent Lagrange multiplier (LM) test for a threshold of Hansen (1996). Since the threshold γ is not identified under the null hypothesis of no threshold effect, the p-values are computed by a bootstrap analog, fixing the regressors from the right-hand side of (12) and generating the bootstrap dependent variable from the distribution $N(0, \hat{e}_i^2)$, where $\hat{e}_i$ is the OLS residual from the estimated threshold model. Hansen (1996) shows that this bootstrap analog produces asymptotically correct p-values. Using 1000 bootstrap replications, the p-value for the threshold model using initial per capital output was marginally significant at 0.088 and that for the threshold model using initial literacy rate was insignificant at 0.214. This suggests that there might be a sample split based on output.

Figure 2 displays a graph of the normalized likelihood ratio sequence $LR_n^*(\gamma)$ as a function of the threshold in output. The LS estimate of γ is the value that minimizes this graph, which occurs at $\hat{\gamma} = \$863$. The 95% critical value of 7.35 is also plotted (the dotted line), so we can read off the asymptotic 95% confidence set $\hat{\Gamma}^* = [\$594, \$1794]$ from the graph from where $LR_n^*(\gamma)$ crosses the dotted line. These results show that there is reasonable evidence for a two-regime

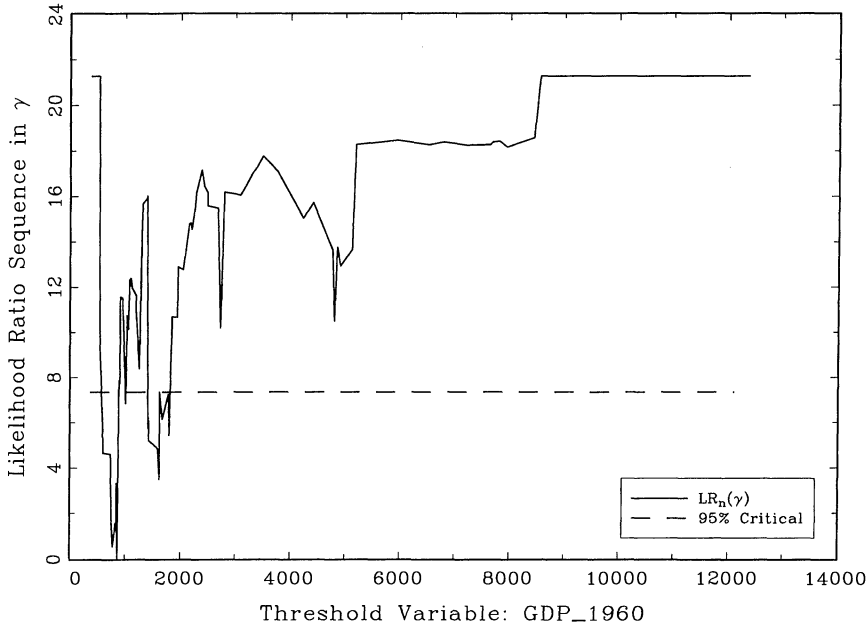

FIGURE 2.—First sample split: Confidence interval construction for threshold.

specification, but there is considerable uncertainty about the value of the threshold. While the confidence interval for γ might seem rather tight by viewing Figure 2, it is perhaps more informative to note that 40 of the 96 countries in the sample fall in the 95% confidence interval, so cannot be decisively classified into the first or second regime.

If we fix γ at the LS estimate \$863 and split the sample in two based on initial GDP, we can (mechanically) perform the same analysis on each subsample. It is not clear how our theoretical results extend to such procedures, but this will enable more informative comparisons with the Durlauf-Johnson results. Only 18 countries have initial output at or below \$863, so no further sample split is possible among this subsample. Among the 78 countries with initial output above \$863, a sample split based on initial output produces an insignificant p-value of 0.152, while a sample split based on the initial literacy rate produces a p-value of 0.078, suggesting a possible threshold effect in the literacy rate. The point estimate of the threshold in the literacy rate is 45%, with a 95% asymptotic confidence interval [19%, 57%]. The graph of the normalized likelihood ratio statistic as a function of the threshold in the literacy rate is displayed in Figure 3. This confidence interval contains 19 of the 78 countries in the subsample. We could try to further split these two subsamples, but none of the bootstrap test statistics were significant at the 10% level.

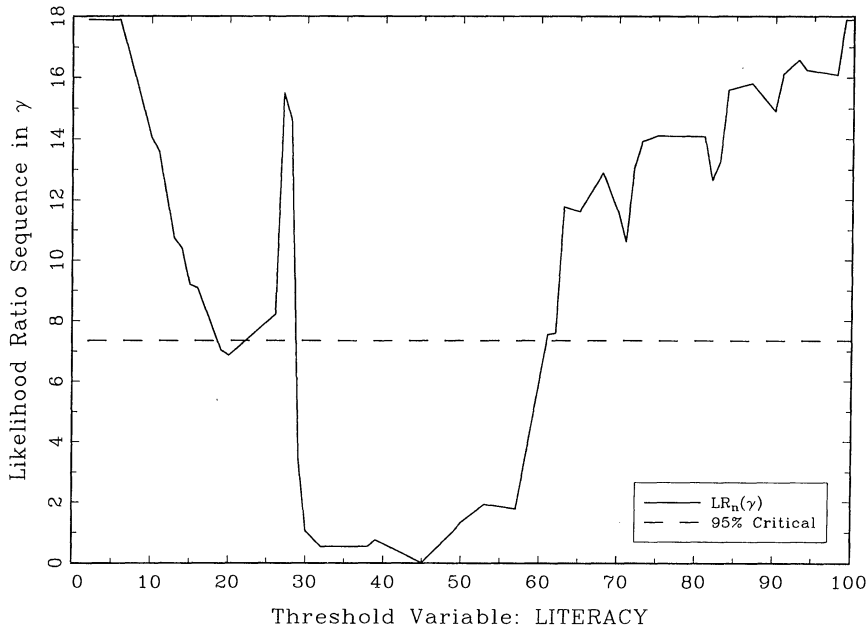

FIGURE 3.—Second sample split: Confidence interval construction for threshold.

Our point estimates are quite similar to those of Durlauf and Johnson (1995). What is different are our confidence intervals. The confidence intervals for the threshold parameters are sufficiently large that there is considerable uncertainty regarding their values, hence concerning the proper division of countries into convergence classes as well.

6. CONCLUSION

This paper develops asymptotic methods to construct confidence intervals for least-squares estimates of threshold parameters. The confidence intervals are asymptotically conservative. It is possible that more accurate confidence intervals may be constructed using bootstrap techniques. This may be quite delicate, however, since the sampling distribution of the likelihood ratio statistic appears to be nonstandard and nonpivotal. This would be an interesting subject for future research.

Dept. of Economics, Social Science Bldg., University of Wisconsin, Madison, WI 53706-1393, U.S.A.; bhansen@ssc.wisc.edu;www.ssc.wisc.edu/ ~ bhansen.

Manuscript received July, 1996; final revision received April, 1999.

APPENDIX: MATHEMATICAL PROOFS

Define $h_i(\gamma_1,\gamma_2)=|x_ie_i||d_i(\gamma_2)-d_i(\gamma_1)|$ and $k_i(\gamma_1,\gamma_2)=|x_i||d_i(\gamma_2)-d_i(\gamma_1)|$.

LEMMA A.1: *There is a* $C_1<\infty$ *such that for* $\underline{\gamma}\le\gamma_1\le\gamma_2\le\bar{\gamma}$, *and* $r\le 4$,

$$Eh_i^r(\gamma_1,\gamma_2)\le C_1|\gamma_2-\gamma_1|, \tag{12}$$

$$Ek_i^r(\gamma_1,\gamma_2)\le C_1|\gamma_2-\gamma_1|. \tag{13}$$

PROOF: For any random variable Z,

$$\frac{d}{d\gamma}E(Zd_i(\gamma))=E(Z|q_i=\gamma)f(\gamma). \tag{14}$$

Thus under Assumption 1.4,

$$\begin{aligned}\frac{d}{d\gamma}(|x_ie_i|^r d_i(\gamma))&=E(|x_ie_i|^r|q_i=\gamma)f(\gamma)\\&\le[E(|x_ie_i|^4|q_i=\gamma)]^{r/4}f(\gamma)\\&\le C^{r/4}\bar{f}\le C_1,\end{aligned}$$

setting $C_1=\max[1,C]\bar{f}$. Since $d_i(\gamma_2)-d_i(\gamma_1)$ equals either zero or one,

$$Eh_i^r(\gamma_1,\gamma_2)=E(|x_ie_i|^r d_i(\gamma_2))-E(|x_ie_i|^r d_i(\gamma_1))\le C_1|\gamma_2-\gamma_1|,$$

by a first-order Taylor series expansion, establishing (12). The proof of (13) is identical. *Q.E.D.*

LEMMA A.2: *There is a* $K<\infty$ *such that for all* $\underline{\gamma}\le\gamma_1\le\gamma_2\le\bar{\gamma}$,

$$E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(h_i^2(\gamma_1,\gamma_2)-Eh_i^2(\gamma_1,\gamma_2))\right|^2\le K|\gamma_2-\gamma_1|, \tag{15}$$

$$E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(k_i^2(\gamma_1,\gamma_2)-Ek_i^2(\gamma_1,\gamma_2))\right|^2\le K|\gamma_2-\gamma_1|. \tag{16}$$

PROOF: Lemma 3.4 of Peligrad (1982) shows that for ρ-mixing sequences satisfying Assumption 1.1, there is a $K'<\infty$ such that

$$\begin{aligned}&E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(h_i^2(\gamma_1,\gamma_2)-Eh_i^2(\gamma_1,\gamma_2))\right|^2\\&\le K'E(h_i^2(\gamma_1,\gamma_2)-Eh_i^2(\gamma_1,\gamma_2))^2\\&\le 2K'Eh_i^4(\gamma_1,\gamma_2)\\&\le 2K'C_1|\gamma_2-\gamma_1|\end{aligned}$$

where the final inequality is (12). This is (15) with $K=2K'C_1$. The proof of (16) is identical. *Q.E.D.*

Let

$$J_n(\gamma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i e_i d_i(\gamma).$$

LEMMA A.3: *There are finite constants K_1 and K_2 such that for all γ_1, $\varepsilon > 0$, $\eta > 0$, and $\delta \geq n^{-1}$, if $\sqrt{n} \geq K_2/\eta$, then*

$$P\left(\sup_{\gamma_1 \leq \gamma \leq \gamma_1 + \delta} |J_n(\gamma) - J_n(\gamma_1)| > \eta \right) \leq \frac{K_1 \delta^2}{\eta^4}.$$

PROOF: Let m be an integer satisfying $n\delta/2 \leq m \leq n\delta$, which is possible since $n\delta \geq 1$. Set $\delta_m = \delta/m$. For $k = 1, \ldots, m+1$, set $\gamma_k = \gamma_1 + \delta_m(k-1)$, $h_{ik} = h_i(\gamma_k, \gamma_{k+1})$, and $h_{ijk} = h_i(\gamma_j, \gamma_k)$. Note that (12) implies $Eh_{ik}^r \leq C_1 \delta_m$ and $Eh_{ijk}^r \leq C_1 |k-j| \delta_m$ for $r \leq 4$. Letting $H_{nk} = n^{-1} \sum_{i=1}^{n} h_{ik}$, observe that for $\gamma_k \leq \gamma \leq \gamma_{k+1}$,

$$|J_n(\gamma) - J_n(\gamma_k)| \leq \sqrt{n}\, H_{nk} \leq \sqrt{n} |H_{nk} - EH_{nk}| + \sqrt{n}\, EH_{nk}.$$

Thus

$$\sup_{\gamma_1 \leq \gamma \leq \gamma_1 + \delta} |J_n(\gamma) - J_n(\gamma_1)| \leq \max_{2 \leq k \leq m+1} |J_n(\gamma_k) - J_n(\gamma_1)| + \max_{1 \leq k \leq m} \sqrt{n} |H_{nk} - EH_{nk}| + \max_{1 \leq k \leq m} \sqrt{n}\, EH_{nk}. \tag{17}$$

For any $1 \leq j < k \leq m+1$, by Burkholder's inequality (see, e.g., Hall and Heyde (1980, p. 23)) for some $C' < \infty$, Minkowski's inequality, $Eh_{ik}^2 \leq C_1 \delta_m$, (15), $n^{-1} \leq \delta_m$ and $(k-j)^{1/2} \leq (k-j)$,

$$\begin{aligned} E|J_n(\gamma_k) - J_n(\gamma_j)|^4 &= E\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_i e_i (d_i(\gamma_k) - d_i(\gamma_j)) \right|^4 \\ &\leq C' E \left| \frac{1}{n} \sum_{i=1}^{n} h_{ijk}^2 \right|^2 \\ &\leq C' \left[EH_{ijk}^2 + \left(E \left| \frac{1}{n} \sum_{i=1}^{n} (h_{ijk}^2 - Eh_{ijk}^2) \right|^2 \right)^{1/2} \right]^2 \\ &\leq C' \left[C_1 (k-j) \delta_m + \left(\frac{K(k-j)\delta_m}{n} \right)^{1/2} \right]^2 \\ &\leq C' [C_1 + \sqrt{K}]^2 ((k-j)\delta_m)^2. \end{aligned} \tag{18}$$

The bound (18) and Theorem 12.2 of Billingsley (1968, p. 94) imply that there is a finite K'' such that

$$P\left(\max_{2 \leq k \leq m+1} |J_n(\gamma_k) - J_n(\gamma_1)| > \eta \right) \leq K'' \frac{(m\delta_m)^2}{\eta^4} = K'' \frac{\delta^2}{\eta^4}, \tag{19}$$

which bounds the first term on the right-hand side of (17).

We next consider the second term. Lemma 3.6 of Peligrad (1982) shows there is a $K''' < \infty$ such that

$$E|\sqrt{n}(H_{nk} - EHG_{nk})|^4 \le K'''\left(n^{-1}Eh_{ik}^4 + (Eh_{ik}^2)^2\right) \tag{20}$$
$$\le K'''(n^{-1}C_1\delta_m + (C_1\delta_m)^2)$$
$$\le K'''(C_1 + C_1^2)\delta_m^2,$$

where the third inequality is $n^{-1} \le \delta_m$. Markov's inequality and (20) yield

$$P\left(\max_{1\le k\le m}|\sqrt{n}(H_{nk} - EH_{nk})| > \eta\right) \le m\frac{K'''(C_1 + C_1^2)\delta_m^2}{\eta^4} \tag{21}$$
$$\le \frac{K'''(C_1 + C_1^2)\delta^2}{\eta^4},$$

where the final inequality uses $m\delta_m = \delta$ and $\delta_m \le \delta$.

Finally, note that

$$\sqrt{n}EH_{nk} = \sqrt{n}Eh_{ik} \le \sqrt{n}C_1\delta_m \le \frac{2C_1}{\sqrt{n}}, \tag{22}$$

since $\delta_m \le 2/n$. Together, (17), (19), (21), and (22) imply that when $2C_1/\sqrt{n} \le \eta$,

$$P\left(\sup_{\gamma_1\le\gamma\le\gamma_1+\delta}|J_n(\gamma) - J_n(\gamma_1| > 3\eta\right) \le \frac{[K'' + K'''(C_1 + C_1^2)]\delta^2}{\eta^4},$$

which implies the result setting $K_1 = 3^4[K'' + K'''(C_1 + C_1^2)]$ and $K_2 = 6C_1$. *Q.E.D.*

Let " $\Rightarrow$ " denote weak convergence with respect to the uniform metric.

LEMMA A.4: $J_n(\gamma) \Rightarrow J(\gamma)$, *a mean-zero Gaussian process with almost surely continuous sample paths.*

PROOF: For each γ, $x_i e_i d_i(\gamma)$ is a square integrable stationary martingale difference, so $J_n(\gamma)$ converges pointwise to a Gaussian distribution by the CLT. This can be extended to any finite collection of γ to yield the convergence of the finite dimensional distributions.

Fix $\varepsilon > 0$ and $\eta > 0$. Set $\delta = \varepsilon\eta^4/K_1$ and $\bar{n} = \max[\delta^{-1}, K_2^2/\eta^2]$, where K_1 and K_2 are defined in Lemma A.3. Then by Lemma A.3, for any γ_1, if $n \ge \bar{n}$,

$$P\left(\sup_{\gamma_1\le\gamma\le\gamma_1+\delta}|J_n(\gamma) - J_n(\gamma)| > \eta\right) \le \frac{K_1\delta^2}{\eta^4} = \delta\varepsilon.$$

This establishes the conditions for Theorem 15.5 of Billingsley (1968). (See also the proof of Theorem 16.1 of Billingsley (1968).) It follows that $J_n(\gamma)$ is tight, so $J_n(\gamma) \Rightarrow J(\gamma)$. *Q.E.D.*

LEMMA A.5: $\hat{\gamma} \to_p \gamma_0$.

PROOF: Lemma 1 of Hansen (1996) shows that Assumption 1 is sufficient for

$$M_n(\gamma) = \frac{1}{n}X_\gamma'X_\gamma = \frac{1}{n}\sum_{i=1}^n x_i x_i' d_i(\gamma) \to_p M(\gamma), \tag{23}$$

uniformly over $\gamma \in R$. Observe that (23) implies $M_n = (1/n)X'X \to_p M$.

Let $X_0 = X_{\gamma_0}$. Then since $Y = X\theta + X_0\delta_n + e$ and X lies in the space spanned by $P_\gamma = X_\gamma^*(X_\gamma^{*\prime}X_\gamma^*)^{-1}X_\gamma^{*\prime}$,

$$S_n(\gamma) - e'e = Y'(I - P_\gamma)Y - e'e = -e'P_\gamma e + 2\delta_n'X_0'(I - P_\gamma)e + \delta_n'X_0'(I - P_\gamma)X_0\delta_n.$$

Using Assumption 1.6, Lemma A.4, and (23), we see that uniformly over $\gamma \in \Gamma$,

$$n^{-1+2\alpha}(S_n(\gamma) - e'e) = n^{-1}c'(X_0'(I - P_\gamma)X_0)c + o_p(1).$$

The projection P_γ can be written as the projection onto $[X_\gamma, Z_\gamma]$, where $Z_\gamma = X - X_\gamma$ is a matrix whose ith row is $x_i'(1 - d_i(\gamma))$. Observe that $X_\gamma'Z_\gamma = 0$, and for $\gamma \geq \gamma_0$, $X_0'X_\gamma = X_0'X_0$ and $X_0'Z_\gamma = 0$. Then using (23) we calculate that uniformly over $\gamma \in [\gamma_0, \bar{\gamma}]$

$$n^{-1}c'(X_0'(I - P_\gamma)X_0)c = c'(M_n(\gamma_0) - M_n(\gamma_0)M_n(\gamma)^{-1}M_n(\gamma_0))c$$

$$\to_p c'(M(\gamma_0) - M(\gamma_0)M(\gamma)^{-1}M(\gamma_0))c \equiv b_1(\gamma),$$

say. Note that $M(\gamma)^{-1}$ exists under Assumption 1.8 so $b_1(\gamma)$ is well defined.

A calculation based on (14) shows

$$\frac{d}{d\gamma}M(\gamma) = D(\gamma)f(\gamma), \tag{24}$$

where $M(\gamma)$ and $D(\gamma)$ are defined in (6) and (7). Using (24),

$$\frac{d}{d\gamma}b_1(\gamma) = c'M(\gamma_0)M(\gamma)^{-1}D(\gamma)f(\gamma)M(\gamma)^{-1}M(\gamma_0)c \geq 0$$

so $b_1(\gamma)$ is continuous and weakly increasing on $[\gamma_0, \bar{\gamma}]$. Additionally,

$$\frac{d}{d\gamma}b_1(\gamma_0) = c'Dfc > 0$$

by Assumption 1.7, so $b_1(\gamma)$ is uniquely minimized at γ_0 on $[\gamma_0, \bar{\gamma}]$.

Symmetrically, we can show that uniformly over $\gamma \in [\underline{\gamma}, \gamma_0]$, $n^{-1}c'(X_0'(I - P_\gamma)X_0)c \to_p b_2(\gamma)$, a weakly decreasing continuous function which is uniquely minimized at γ_0. Thus uniformly over $\gamma \in \Gamma$, $n^{-1+2\alpha}(S_n(\gamma) - e'e) \to_p b_1(\gamma)\{\gamma \geq \gamma_0\} + b_2(\gamma)\{\gamma < \gamma_0\}$, a continuous function with unique minimizer γ_0. Since $\hat{\gamma}$ minimizes $S_n(\gamma) - e'e$, it follows that $\hat{\gamma} \to_p \gamma_0$ (see, e.g., Theorem 2.1 of Newey and McFadden (1994)). *Q.E.D.*

LEMMA A.6: *$n^\alpha(\hat{\theta} - \theta) = o_p(1)$ and $n^\alpha(\hat{\delta} - \delta_n) = o_p(1)$.*

PROOF: We show the result for $\hat{\delta}$, as the derivation for $\hat{\theta}$ is similar. Let $P_X = I - X(X'X)^{-1}X'$. Then on $\gamma \in \Gamma$, using (23),

$$\frac{1}{n}X_\gamma'P_XX_\gamma = M_n(\gamma) - M_n(\gamma)M_n^{-1}M_n(\gamma) \Rightarrow M(\gamma) - M(\gamma)M^{-1}M(\gamma) = M^*(\gamma),$$

say. Observe that for $\gamma \in \Gamma$, $M^*(\gamma)$ is invertible since $M - M(\gamma)$ is invertible by Assumption 1.8. Also,

$$\frac{1}{n}X_\gamma'P_XX_0 \Rightarrow M(\gamma \wedge \gamma_0) - M(\gamma)M^{-1}M(\gamma_0) = M_0^*(\gamma),$$

say. Thus

$$n^{\alpha}(\hat{\delta}(\gamma)-\delta_n)=\left(\frac{1}{n}X_\gamma' P_X X_\gamma\right)^{-1}\left(\frac{1}{n}X_\gamma' P_X((X_0-X_\gamma)c+en^{\alpha})\right)$$

$$\Rightarrow M^*(\gamma)^{-1}(M_0^*(\gamma)-M^*(\gamma))c$$

$$=\delta(\gamma),$$

say. Note that $M_0^*(\gamma_0)=M^*(\gamma_0)$ so $\delta(\gamma_0)=0$. Since $M(\gamma)$ is continuous and $M^*(\gamma)$ is invertible, $\delta(\gamma)$ is continuous in Γ. Lemma A.5 allows us to conclude that

$$n^{\alpha}(\hat{\delta}-\delta_n)=n^{\alpha}(\hat{\delta}(\hat{\gamma})-\delta_n)\to_p \delta(\gamma_0)=0. \qquad \textit{Q.E.D.}$$

Define $G_n(\gamma)=n^{-1}\sum_{i=1}^n(c'x_i)^2|d_i(\gamma)-d_i(\gamma_0)|$ and $K_n(\gamma)=n^{-1}\sum_{i=1}^n k_i^2(\gamma_0,\gamma)$. Set $a_n=n^{1-2\alpha}$.

LEMMA A.7: *There exist constants $B>0$, $0<d<\infty$, and $0<k<\infty$, such that for all $\eta>0$ and $\varepsilon>0$, there exists a $\bar{v}<\infty$ such that for all n,*

$$P\left(\inf_{\frac{\bar{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{G_n(\gamma)}{|\gamma-\gamma_0|}<(1-\eta)d\right)\le\varepsilon, \tag{25}$$

$$P\left(\sup_{\frac{\bar{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{K_n(\gamma)}{|\gamma-\gamma_0|}>(1+\eta)k\right)\le\varepsilon.$$

PROOF: We show (25), as the proof of (26) is similar. First, note that for $\gamma\ge\gamma_0$,

$$EG_n(\gamma)=c'(M(\gamma)-M(\gamma_0))c, \tag{27}$$

so by (24), $dEG_n(\gamma)/d\gamma=c'D(\gamma)f(\gamma)c$, and the sign is reversed if $\gamma<\gamma_0$. Since $c'D(\gamma_0)f(\gamma_0)c>0$ (Assumption 1.8) and $c'D(\gamma)f(\gamma)c$ is continuous at γ_0 (Assumption 1.5), then there is a B sufficiently small such that

$$d=\min_{|\gamma-\gamma_0|\le B}c'D(\gamma)f(\gamma)c>0.$$

Since $EG_n(\gamma_0)=0$, a first-order Taylor series expansion about γ_0 yields

$$\inf_{|\gamma-\gamma_0|\le B}EG_n(\gamma)\ge d|\gamma-\gamma_0|. \tag{28}$$

Lemma A.2 (16) yields

$$E|G_n(\gamma)-EG_n(\gamma)|^2\le|c|^4E|K_n(\gamma)-EK_n(\gamma)|^2\le|c|^4n^{-1}K|\gamma-\gamma_0|. \tag{29}$$

For any η and ε, set

$$b=\frac{1-\eta/2}{1-\eta}>1 \tag{30}$$

and

$$\bar{v}=\frac{8|c|^4K}{\eta^2d^2(1-1/b)\varepsilon}. \tag{31}$$

We may assume that n is large enough so that $\bar{v}/a_n \le B$, else the inequality (25) is trivial. For $j = 1, 2, \ldots, N+1$, set $\gamma_j = \gamma_0 + \bar{v}b^{j-1}/a_n$, where N is the integer such that $\gamma_N - \gamma_0 = \bar{v}b^{N-1}/a_n \le B$ and $\gamma_{N+1} - \gamma_0 > B$. (Note that $N \ge 1$ since $\bar{v}/a_n \le B$.)

Markov's inequality, (29), and (28) yield

$$P\left(\sup_{1\le j\le N}\left|\frac{G_n(\gamma_j)}{EG_n(\gamma_j)} - 1\right| > \frac{\eta}{2}\right) \le \left(\frac{2}{\eta}\right)^2 \sum_{j=1}^{N} \frac{E|G_n(\gamma_j) - EG_n(\gamma_j)|^2}{|EG_n(\gamma_j)|^2}$$

$$\le \frac{4}{\eta^2}\sum_{j=1}^{N}\frac{|c|^4 K n^{-1}|\gamma_j - \gamma_0|}{d^2|\gamma_j - \gamma_0|^2}$$

$$= n^{-2\alpha}\frac{4|c|^4 K}{\eta^2 d^2 \bar{v}}\sum_{j=1}^{\infty}\frac{1}{b^{j-1}}$$

$$\le \frac{4|c|^4 K}{\eta^2 d^2 \bar{v}}\frac{1}{1-1/b} = \varepsilon/2$$

where the final equality is (31). Thus with probability exceeding $1 - \varepsilon/2$,

$$\left|\frac{G_n(\gamma_j)}{EG_n(\gamma_j)} - 1\right| \le \frac{\eta}{2} \tag{32}$$

for all $1 \le j \le N$.

So for any γ such that $(\bar{v}/a_n) \le \gamma - \gamma_0 \le B$, there is some $j \le N$ such that $\gamma_j < \gamma < \gamma_{j+1}$, and on the event (32),

$$\frac{G_n(\gamma)}{|\gamma - \gamma_0|} \ge \frac{G_n(\gamma_j)}{EG_n(\gamma_j)}\frac{EG_n(\gamma_j)}{|\gamma_{j+1} - \gamma_0|} \tag{33}$$

$$\ge \left(1 - \frac{\eta}{2}\right)\frac{d|\gamma_j - \gamma_0|}{|\gamma_{j+1} - \gamma_0|}$$

$$= (1 - \eta)d,$$

using (28), the construction $(\gamma_j - \gamma_0)/(\gamma_{j+1} - \gamma_0) = 1/b$, and (30). Since this event has probability exceeding $1 - \varepsilon/2$, we have established

$$P\left(\inf_{\frac{\bar{v}}{a_n} \le \gamma - \gamma_0 \le B}\frac{G_n(\gamma)}{|\gamma - \gamma_0|} < (1 - \eta)d\right) \le \varepsilon/2.$$

A symmetric argument establishes a similar inequality where the infimum is taken over $-(\bar{v}/a_n) \ge \gamma - \gamma_0 \ge -B$, establishing (25). *Q.E.D.*

LEMMA A.8: *For all $\eta > 0$ and $\varepsilon > 0$, there exists some $\bar{v} < \infty$ such that for any $B < \infty$,*

$$P\left(\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B}\frac{|J_n(\gamma) - J_n(\gamma_0)|}{\sqrt{a_n}|\gamma - \gamma_0|} > \eta\right) \le \varepsilon. \tag{34}$$

PROOF: Fix $\eta > 0$. For $j = 1, 2, \ldots$, set $\gamma_j - \gamma_0 = \bar{v} 2^{j-1}/a_n$, where $\bar{v} < \infty$ will be determined later. A straightforward calculation shows that

$$\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{|J_n(\gamma) - J_n(\gamma_0)|}{\sqrt{a_n}|\gamma - \gamma_0|} \le 2 \sup_{j>0} \frac{|J_n(\gamma_j) - J_n(\gamma_0)|}{\sqrt{a_n}|\gamma_j - \gamma_0|} + 2 \sup_{j>0} \sup_{\gamma_j \le \gamma \le \gamma_{j+1}} \frac{|J_n(\gamma) - J_n(\gamma_j)|}{\sqrt{a_n}|\gamma_j - \gamma_0|}. \tag{35}$$

Markov's inequality, the martingale difference property, and (12) imply

$$\begin{aligned} P\left(2 \sup_{j>0} \frac{|J_n(\gamma_j) - J_n(\gamma_0)|}{\sqrt{a_n}|\gamma_j - \gamma_0|} > \eta\right) &\le \frac{4}{\eta^2} \sum_{j=1}^{\infty} \frac{E|J_n(\gamma_j) - J_n(\gamma_0)|^2}{a_n|\gamma_j - \gamma_0|^2} \\ &= \frac{4}{\eta^2} \sum_{j=1}^{\infty} \frac{E(|x_i e_i|^2 |d_i(\gamma_j) - d_i(\gamma_0)|)}{a_n|\gamma_j - \gamma_0|^2} \\ &\le \frac{4}{\eta^2} \sum_{j=1}^{\infty} \frac{C_1|\gamma_j - \gamma_0|}{a_n|\gamma_j - \gamma_0|^2} \\ &= \frac{4C_1}{\eta^2} \sum_{j=1}^{\infty} \frac{1}{\bar{v} 2^{j-1}} = \frac{8C_1}{\eta^2 \bar{v}}. \end{aligned} \tag{36}$$

Set $\delta_j = \gamma_{j+1} - \gamma_j$ and $\eta_j = \sqrt{a_n}|\gamma_j - \gamma_0|\eta$. Then

$$\begin{aligned} &P\left(2 \sup_{j>0} \sup_{\gamma_j \le \gamma \le \gamma_{j+1}} \frac{|J_n(\gamma) - J_n(\gamma_j)|}{\sqrt{a_n}|\gamma_j - \gamma_0|} > \eta\right) \\ &\quad \le 2 \sum_{j=1}^{\infty} P\left(\sup_{\gamma_j \le \gamma \le \gamma_j + \delta_j} |J_n(\gamma) - J_n(\gamma_j)| > \eta_j\right). \end{aligned} \tag{37}$$

Observe that if $\bar{v} \ge 1$, then $\delta_j \ge a_n^{-1} \ge n^{-1}$. Furthermore, if $\bar{v} \ge K_2/\eta$, then $\eta_j = a_n^{-1/2} 2^{j-1} \bar{v} \eta \ge K_2 a_n^{-1/2} \ge K_2 n^{-1/2}$. Thus if $\bar{v} \ge \max[1, K_2/\eta]$ the conditions for Lemma A.3 hold, and the right-hand side of (37) is bounded by

$$2 \sum_{j=1}^{\infty} \frac{K_1 \delta_j^2}{\eta_j^4} = \sum_{j=1}^{\infty} \frac{K_1 |\gamma_{j+1} - \gamma_j|^2}{a_n^2 |\gamma_j - \gamma_0|^4 \eta^4} = \frac{8K_1}{3\eta^4 \bar{v}^2}. \tag{38}$$

(35), (36), (37), and (38) show that if $\bar{v} \ge \max[1, K_2/\eta]$,

$$P\left(\sup_{\frac{\bar{v}}{a_n} \le |\gamma - \gamma_0| \le B} \frac{|J_n(\gamma) - J_n(\gamma_0)|}{\sqrt{a_n}|\gamma - \gamma_0|} > 2\eta\right) \le \frac{8C_1}{\eta^2 \bar{v}} + \frac{8K_1}{3\eta^4 \bar{v}^2},$$

which can be made arbitrarily small by picking suitably large $\bar{v}$. *Q.E.D.*

LEMMA A.9: $a_n(\hat{\gamma} - \gamma_0) = O_p(1)$.

PROOF: Let B, d, and k be defined as in Lemma A.7. Pick $\eta > 0$ and $\kappa > 0$ small enough so that

$$(1 - \eta)d - 2(|c| + \kappa)\eta - 2(|c| + \kappa)\kappa(1 + \eta)k - (2|c| + \kappa)\kappa(1 + \eta)k > 0. \tag{39}$$

Let E_n be the joint event that $|\hat{\gamma}-\gamma_0|\le B, n^{\alpha}|\hat{\theta}-\theta_0|\le\kappa, n^{\alpha}|\hat{\delta}-\delta_0|\le\kappa$,

$$\inf_{\frac{\bar{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{G_n(\gamma)}{|\gamma-\gamma_0|}\ge(1-\eta)d, \tag{40}$$

$$\sup_{\frac{\bar{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{K_n(\gamma)}{|\gamma-\gamma_0|}\le(1+\eta)k, \tag{41}$$

and

$$\sup_{\frac{\bar{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{|J_n(\gamma)-J_n(\gamma_0)|}{\sqrt{a_n}|\gamma-\gamma_0|}<\eta. \tag{42}$$

Fix $\varepsilon>0$, and pick $\bar{v}$ and $\bar{n}$ so that $P(E_n)\ge1-\varepsilon$ for all $n\ge\bar{n}$, which is possible under Lemmas A.5–A.8.

Let $S_n^*(\gamma)=S_n(\hat{\theta},\hat{\delta},\gamma)$, where $S_n(\cdot,\cdot,\cdot)$ is the sum of squared errors function (5). Since $Y=X\theta+X_0\delta_n+e$,

$$Y-X\hat{\theta}-X_\gamma\hat{\delta}=(e-X(\hat{\theta}-\theta)-X_0(\hat{\delta}-\delta_n))-\Delta X_\gamma\hat{\delta},$$

where $\Delta X_\gamma=X_\gamma-X_0$. Hence

$$\begin{aligned}S_n^*(\gamma)-S_n^*(\gamma_0)&=\left(Y-X\hat{\theta}-X_\gamma\hat{\delta}\right)'\left(Y-X\hat{\theta}-X_\gamma\hat{\delta}\right)\\&\quad-(Y-X\hat{\theta}-X_0\hat{\delta})'(Y-X\hat{\theta}-X_0\hat{\delta})\\&=\hat{\delta}'\Delta X_\gamma'\Delta X_\gamma\hat{\delta}-2\hat{\delta}'\Delta X_\gamma'e+2\hat{\delta}'\Delta X_\gamma'\Delta X_\gamma(\hat{\theta}-\theta)\\&=\delta_n'\Delta X_\gamma'\Delta X_\gamma\delta_n-2\hat{\delta}'\Delta X_\gamma'e+2\hat{\delta}'\Delta X_\gamma'\Delta X_\gamma(\hat{\theta}-\theta)\\&\quad+(\delta_n+\hat{\delta})'\Delta X_\gamma'\Delta X_\gamma(\hat{\delta}-\delta_n).\end{aligned} \tag{43}$$

Suppose $\gamma\in[\gamma_0+\bar{v}/a_n,\gamma_0+B]$ and suppose E_n holds. Let $\hat{c}=n^{\alpha}\hat{\delta}$ so that $|\hat{c}-c|\le\kappa$. By (39)–(43),

$$\begin{aligned}\frac{S_n^*(\gamma)-S_n^*(\gamma_0)}{a_n(\gamma-\gamma_0)}&=\frac{c'\Delta X_\gamma'\Delta X_\gamma c}{n(\gamma-\gamma_0)}-\frac{2\hat{c}'\Delta X_\gamma'e}{n^{1-\alpha}(\gamma-\gamma_0)}+\frac{2\hat{c}'\Delta X_\gamma'\Delta X_\gamma n^{\alpha}(\hat{\theta}-\theta)}{n(\gamma-\gamma_0)}\\&\quad+\frac{(c+\hat{c})'\Delta X_\gamma'\Delta X_\gamma(\hat{c}-c)}{n(\gamma-\gamma_0)}\\&\ge\frac{G_n(\gamma)}{(\gamma-\gamma_0)}-2|\hat{c}|\frac{|J_n(\gamma)-J_n(\gamma_0)|}{\sqrt{a_n}(\gamma-\gamma_0)}-2|\hat{c}|\,|n^{\alpha}(\hat{\theta}-\theta)|\frac{K_n(\gamma)}{(\gamma-\gamma_0)}\\&\quad-|c+\hat{c}|\,|\hat{c}-c|\frac{K_n(\gamma)}{(\gamma-\gamma_0)}\\&\ge(1-\eta)d-2(|c|+\kappa)\eta-2(|c|+\kappa)\kappa(1+\eta)k\\&\quad-(2|c|+\kappa)\kappa(1+\eta)k\\&>0.\end{aligned}$$

We have shown that on the set E_n, if $\gamma\in[\gamma_0+\bar{v}/a_n,\gamma_0+B]$, then $S_n^*(\gamma)-S_n^*(\gamma_0)>0$. We can similarly show that if $\gamma\in[\gamma_0-B,\gamma_0-\bar{v}_\varepsilon/a_n]$ then $S_n^*(\gamma)-S_n^*(\gamma_0)>0$. Since $S_n^*(\hat{\gamma})-S_n^*(\gamma_0)\le0$,

this establishes that E_n implies $|\hat{\gamma} - \gamma_0| \leq \bar{v}/a_n$. As $P(E_n) \geq 1 - \varepsilon$ for $n \geq \bar{n}$, this shows that $P(a_n|\hat{\gamma} - \gamma_0| > \bar{v}) \leq \varepsilon$ for $n \geq \bar{n}$, as required. *Q.E.D.*

Let $G_n^*(v) = a_n G_n(\gamma_0 + v/a_n)$ and $K_n^*(v) = a_n K_n(\gamma_0 + v/a_n)$.

LEMMA A.10: *Uniformly on compact sets* Ψ,

$$G_n^*(v) \to_p \mu|v|, \tag{44}$$

and

$$K_n^*(v) \to_p |Df||v|, \tag{45}$$

where $\mu = c'Dcf$.

PROOF: We show (44). Fix $v \in \Psi$. Using (27),

$$EG_n^*(v) = a_n c'(M(\gamma_0 + v/a_n) - M(\gamma_0))c \to |v|c'Dfc = |v|\mu \tag{46}$$

as $n \to \infty$. By (29),

$$E|G_n^*(v) - EG_n^*(v)|^2 = a_n^2 E|G_n(\gamma_0 + v/a_n) - EG_n(\gamma_0 + v/a_n)|^2 \tag{47}$$

$$\leq \frac{a_n^2}{n}|c|^4 K\left|\frac{v}{a_n}\right| = |c|^4 K|v|n^{-2\alpha} \to 0.$$

Markov's inequality, (46) and (47) show that $G_n^*(v) \to_p \mu|v|$.

Suppose $\Psi = [0, \bar{v}]$. Since $G_n^*(v)$ is monotonically increasing on Ψ and the limit function is continuous, the convergence is uniform over Ψ. To see this, set $G(v) = \mu v$. Pick any $\varepsilon > 0$, then set $J = \bar{v}\mu/\varepsilon$ and for $j = 0, 1, \ldots, J$, set $v_j = \bar{v}\mu j/J$. Then pick n large enough so that $\max_{j \leq J}|G_n^*(v_j) - G(v_j)| \leq \varepsilon$ with probability greater than $1 - \varepsilon$, which is possible by pointwise consistency. For any $j \geq 1$, take any $v \in (v_{j-1}, v_j)$. Both $G_n^*(v)$ and $G(v)$ lie in the interval $[G(v_{j-1}) - \varepsilon, G(v_j) + \varepsilon]$ (with probability greater than $1 - \varepsilon$), which has length bounded by 3ε. Since v is arbitrary, $|G_n^*(v) - G(v)| \leq 3\varepsilon$ uniformly over Ψ.

An identical argument yields uniformity over sets of the form $[-\bar{v}, 0]$, and thus for arbitrary compact sets Ψ. *Q.E.D.*

Let $R_n(v) = \sqrt{a_n}(J_n(\gamma_0 + v/a_n) - J_n(\gamma_0))$.

LEMMA A.11: *On any compact set* Ψ,

$$R_n(v) \Rightarrow B(v)$$

where $B(v)$ *is a vector Brownian motion with covariance matrix* $E(B(1)B(1)') = Vf$.

PROOF: Our proof proceeds by establishing the convergence of the finite dimensional distributions of $R_n(v)$ to those of $B(v)$ and then showing that $R_n(v)$ is tight.

Fix $v \in \Psi$. Define $d_i^*(v) = d_i(\gamma_0 + v/a_n) - d_i(\gamma_0)$ and $u_{ni}(v) = \sqrt{a_n}x_i e_i d_i^*(v)$ so that $R_n(v) = n^{-1/2}\sum_{i=1}^n u_{ni}(v)$, and let $V_n(v) = n^{-1}\sum_{i=1}^n u_{ni}(v)u_{ni}(v)'$. Under Assumption 1.2, $\{u_{ni}(v), F_i\}$ is a martingale difference array (MDA). By the MDA central limit theorem (for example, Theorem 24.3 of Davidson (1994, p. 383)) sufficient conditions for $R_n(v) \to_d N(0, |v|Vf)$ are

$$V_n(v) \to_p |v|Vf \tag{48}$$

and

$$n^{-1/2}\max_{1\le i\le n}|u_{ni}(\upsilon)|\to_p 0. \tag{49}$$

First, a calculation based on (14) shows $dE(x_ix_i'e_i^2d_i(\gamma))/d\gamma = V(\gamma)f(\gamma)$, where $V(\gamma)$ is defined in (8). Thus by the continuity of $V(\gamma)f(\gamma)$ at γ_0 (Assumption 1.5), if $\upsilon > 0$,

$$\begin{aligned} EV_n(\upsilon) &= a_nE(x_ix_i'e_i^2|d_i^*(\upsilon)|) \\ &= a_nE\left(x_ix_i'e_i^2d_i\left(\gamma_0+\frac{\upsilon}{a_n}\right)\right)-E(x_ix_i'e_i^2d_i(\gamma_0)) \\ &\to \upsilon Vf, \end{aligned} \tag{50}$$

with the sign reversed if $\upsilon < 0$. By (15),

$$\begin{aligned} E|V_n(\upsilon)-EV_n(\upsilon)|^2 &= \frac{a_n^2}{n}E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n(|x_ie_i|^2d_i^*(\upsilon)-E(|x_ie_i|^2|d_i^*(\upsilon)|))\right|^2 \\ &\le \frac{a_n^2}{n}K\frac{\upsilon}{a_n} = n^{-2\alpha}K\upsilon\to 0 \end{aligned} \tag{51}$$

as $n\to\infty$, which with (50) combines to yield (48). By (12),

$$\begin{aligned} E\left|n^{-1/2}\max_{1\le i\le n}|u_{ni}(\upsilon)|\right|^4 &\le \frac{1}{n}E|u_{ni}(\upsilon)|^4 \\ &= \frac{a_n^2}{n}E(|x_ie_i|^4|d_i^*(\upsilon)|) \\ &\le \frac{a_n^2}{n}C_1\frac{|\upsilon|}{a_n} \\ &= n^{-2\alpha}C_1|\upsilon|\to 0, \end{aligned}$$

which establishes (49) by Markov's inequality. We conclude that $R_n(\upsilon)\to_d N(0,|\upsilon|Vf)$. This argument can be extended to include any finite collection $[\upsilon_1,\dots,\upsilon_k]$ to yield the convergence of the finite dimensional distributions of $R_n(\upsilon)$ to those of $B(\upsilon)$.

We now show tightness. Fix $\varepsilon > 0$ and $\eta > 0$. Set $\delta = \varepsilon\eta^4/K_1$ and $\bar{n} = (\max[\delta^{-1/2}, K_2/\eta])^{1/\alpha}$, where K_1 and K_2 are defined in Lemma A.3. Set $\gamma_1 = \gamma_0 + \upsilon_1/a_n$. By Lemma A.3, for $n\ge\bar{n}$,

$$\begin{aligned} P\left(\sup_{\upsilon_1\le\upsilon\le\upsilon_1+\delta}|R_n(\upsilon)-R_n(\upsilon_1)|>\eta\right) &= P\left(\sup_{\gamma_1\le\gamma\le\gamma_1+\delta/a_n}|J_n(\gamma)-J_n(\gamma_1)|>\frac{\eta}{a_n^{1/2}}\right) \\ &\le \frac{K_1\left(\frac{\delta}{a_n}\right)^2}{a_n^{-2}\eta^4}\le\delta\varepsilon. \end{aligned}$$

(The conditions for Lemma A.3 are met since $\delta/a_n\ge n^{-1}$ when $n^\alpha\ge\delta^{-1/2}$, and $\eta/a_n^{1/2}\ge K_2/n^{1/2}$ when $n^\alpha\ge K_2/\eta$, and these hold for $n\ge\bar{n}$.) As discussed in the proof of Lemma A.4, this shows that $R_n(\gamma)$ is tight, so $R_n(\upsilon)\Rightarrow B(\upsilon)$. *Q.E.D.*

LEMMA A.12: $\sqrt{n}(\underline{\hat{\theta}} - \hat{\underline{\theta}}(\gamma_0)) \to_p 0$, *and* $\sqrt{n}(\hat{\underline{\theta}}(\gamma_0) - \underline{\theta}) \to_d Z \sim N(0, V_\theta)$.

PROOF: Let $J_n = n^{-1/2} X'e$, and observe that Lemma A.4 implies

$$\frac{1}{\sqrt{n}} X_\gamma^{*\prime} e = \begin{pmatrix} J_n \\ J_n(\gamma) \end{pmatrix} \Rightarrow \begin{pmatrix} J \\ J(\gamma) \end{pmatrix},$$

a Gaussian process with continuous sample paths. Thus on any compact set Ψ,

$$J_n^*(v) = \begin{pmatrix} J_n \\ J_n\left(\gamma_0 + \dfrac{v}{a_n}\right) \end{pmatrix} \Rightarrow \begin{pmatrix} J \\ J(\gamma_0) \end{pmatrix}$$

and (23) implies

$$M_n^*(v) = \begin{pmatrix} M_n & M_n\left(\gamma_0 + \dfrac{v}{a_n}\right) \\ M_n\left(\gamma_0 + \dfrac{v}{a_n}\right) & M_n\left(\gamma_0 + \dfrac{v}{a_n}\right) \end{pmatrix} \to_p \begin{pmatrix} M & M(\gamma_0) \\ M(\gamma_0) & M(\gamma_0) \end{pmatrix}$$

uniformly on Ψ. Also

$$A_n(v) = \frac{a_n}{n} \Delta X'_{\gamma_0 + v/a_n} \Delta X_{\gamma_0 + v/a_n} \le K_n^*(v) \to_p 0$$

uniformly on Ψ by Lemma A.10, and hence

$$\sqrt{n}\left(\underline{\hat{\theta}}\left(\gamma_0 + \frac{v}{a_n}\right) - \underline{\theta}\right) = M_n^*(v)^{-1}\left(J_n^*(v) - \begin{pmatrix} I \\ I \end{pmatrix} A_n(v) c a_n^{-1/2}\right) \tag{52}$$

$$\Rightarrow \begin{pmatrix} M & M(\gamma_0) \\ M(\gamma_0) & M(\gamma_0) \end{pmatrix}^{-1} \begin{pmatrix} J \\ J(\gamma_0) \end{pmatrix} = Z,$$

which establishes $\sqrt{n}(\underline{\hat{\theta}}(\gamma_0) - \underline{\theta}) \to_d Z$ as stated.

Pick $\varepsilon > 0$ and $\delta > 0$. Lemma A.9 and (52) show that there is a $\bar{v} < \infty$ and $\bar{n} < \infty$ so that for all $n \ge \bar{n}$, $P(a_n|\hat{\gamma} - \gamma_0| > \bar{v}) \le \varepsilon$ and

$$P\left(\sup_{-\bar{v} \le v \le \bar{v}} \sqrt{n} \left| \underline{\hat{\theta}}\left(\gamma_0 + \frac{v}{a_n}\right) - \hat{\underline{\theta}}(\gamma_0) \right| > \delta \right) \le \varepsilon.$$

Hence, for $n \ge \bar{n}$,

$$P(\sqrt{n}|\underline{\hat{\theta}} - \hat{\underline{\theta}}(\gamma_0)| > \delta) \le P\left(\sup_{-\bar{v} \le v \le \bar{v}} \sqrt{n} \left| \underline{\hat{\theta}}\left(\gamma_0 + \frac{v}{a_n}\right) - \hat{\underline{\theta}}(\gamma_0) \right| > \delta \right)$$

$$+ P(a_n|\hat{\gamma} - \gamma_0| > \bar{v}) \le 2\varepsilon,$$

establishing that $\sqrt{n}|\underline{\hat{\theta}} - \hat{\underline{\theta}}(\gamma_0)| \to_p 0$ as stated. *Q.E.D.*

Let $Q_n(v) = S_n(\hat{\theta}, \hat{\delta}, \gamma_0) - S_n(\hat{\theta}, \hat{\delta}, \gamma_0 + v/a_n)$.

LEMMA A.13: *On any compact set* Ψ, $Q_n(v) \Rightarrow Q(v) = -\mu|v| + 2\sqrt{\lambda} W(v)$, *where* $\lambda = c'Vcf$.

PROOF: From (43) we find $Q_n(v) = -G_n^*(v) + 2c'R_n(v) + L_n(v)$, where

$$|L_n(v)| \le 2\sqrt{n}|\hat{\delta} - \delta_n| |R_n(v)| + (2n^\alpha |\hat{c}| |\hat{\theta} - \theta| + |c + \hat{c}| |\hat{c} - c|) |K_n^*(v)|$$

$$\Rightarrow 0$$

since (uniformly on Ψ) $K_n^*(v) = O_p(1)$, $R_n(v) = O_p(1)$, $\sqrt{n}\,|\hat{\theta} - \theta| = O_p(1)$, and $\sqrt{n}\,|\hat{\delta} - \delta_n| = O_p(1)$, by Lemmas A.10, A.11, and A.12. Applying again Lemmas A.10 and A.11, $Q_n(v) \Rightarrow -\mu|v| + 2c'B(v)$. The process $c'B(v)$ is a Brownian motion with variance $\lambda = c'Vcf$, so can be written as $\sqrt{\lambda}\,W(v)$, where $W(v)$ is a standard Brownian motion. *Q.E.D.*

PROOF OF THEOREM 1: By Lemma A.9, $a_n(\hat{\gamma} - \gamma_0) = \operatorname{argmax}_v Q_n(v) = O_p(1)$, and by Lemma A.13, $Q_n(v) \Rightarrow Q(v)$. The limit functional $Q(v)$ is continuous, has a unique maximum, and $\lim_{|v|\to\infty} Q(v) = -\infty$ almost surely (which is true since $\lim_{v\to\infty} W(v)/v = 0$ almost surely). It therefore satisfies the conditions of Theorem 2.7 of Kim and Pollard (1990) which implies

$$a_n(\hat{\gamma} - \gamma_0) \to^d \operatorname*{argmax}_{v\in R} Q(v).$$

Making the change-of-variables $v = (\lambda/\mu^2)r$, noting the distributional equality $W(a^2r) \equiv aW(r)$, and setting $\omega = \lambda/\mu^2$, we can re-write the asymptotic distribution as

$$\begin{aligned}\operatorname*{argmax}_{-\infty<v<\infty} [-\mu|v| + 2\sqrt{\lambda}\,W(v)] &= \frac{\lambda}{\mu^2} \operatorname*{argmax}_{-\infty<r<\infty} \left[-\frac{\lambda}{\mu}|r| + 2\sqrt{\lambda}\,W\left(\frac{\lambda}{\mu^2}r\right)\right] \\ &\equiv \omega \operatorname*{argmax}_{-\infty<r<\infty} \left[-\frac{\lambda}{\mu}|r| + 2\frac{\lambda}{\mu}W(r)\right] \\ &= \omega \operatorname*{argmax}_{-\infty<r<\infty} \left[-\frac{|r|}{2} + W(r)\right]. \qquad \textit{Q.E.D.}\end{aligned}$$

PROOF OF THEOREM 2: Lemma A.12 and (23) show that

$$\begin{aligned}\hat{\sigma} LR_n(\gamma_0) - Q_n(\hat{v}) &= (S_n(\underline{\hat{\theta}}(\gamma_0), \gamma_0) - S_n(\underline{\hat{\Theta}}, \hat{\gamma})) - (S_n(\underline{\hat{\theta}}, \gamma_0) - S_n(\underline{\hat{\theta}}, \hat{\gamma})) \\ &= S_n(\underline{\hat{\theta}}(\gamma_0), \gamma_0) - S_n(\underline{\hat{\theta}}, \gamma_0) \\ &= (\underline{\hat{\theta}}(\gamma_0) - \hat{\theta})' X_\gamma^{*\prime} X_\gamma^* (\underline{\hat{\theta}}(\gamma_0) - \hat{\theta}) \to_p 0.\end{aligned}$$

Now applying Lemma A.13 and the continuous mapping theorem,

$$LR_n(\gamma_0) = \frac{Q_n(\bar{v})}{\hat{\sigma}^2} + o_p(1) = \frac{\sup_v Q_n(v)}{\hat{\sigma}^2} + o_p(1) \to_d \frac{\sup_v Q(v)}{\sigma^2}.$$

This limiting distribution equals

$$\begin{aligned}\frac{1}{\sigma^2}\sup_v [-\mu|v| + 2\sqrt{\lambda}\,W(v)] &= \frac{1}{\sigma^2}\sup_r \left[-\mu\left|\frac{\lambda}{\mu^2}r\right| + 2\sqrt{\lambda}\,W\left(\frac{\lambda}{\mu^2}r\right)\right] \\ &\equiv \frac{\lambda}{\sigma^2\mu}\sup_r [-|r| + 2W(r)] = \eta^2\xi\end{aligned}$$

by the change-of-variables $v = (\lambda/\mu^2)r$, the distributional equality by $W(a^2r) \equiv aW(r)$, and the fact $\eta^2 = \lambda/(\sigma^2\mu)$.

To find the distribution function of ξ, note that $\xi = 2\max[\xi_1, \xi_2]$, where $\xi_1 = \sup_{s\le 0}[W(s) - \frac{1}{2}|s|]$ and $\xi_2 = \sup_{0\le s}[W(s) - \frac{1}{2}|s|]$. ξ_1 and ξ_2 are iid exponential random variables with distribution function $P(\xi_1 \le x) = 1 - e^{-x}$ (see Bhattacharya and Brockwell (1976)). Thus

$$P(\xi \le x) = P(2\max[\xi_1, \xi_2] \le x) = P(\xi_1 \le x/2)P(\xi_2 \le x/2) = (1 - e^{-x/2})^2.$$

PROOF OF THEOREM 3: Note that by the invariance property of the likelihood ratio test, $LR_n(\gamma_0)$ is invariant to reparameterizations, including those of the form $\gamma \to \gamma^* = F_n(\gamma)$. Since the threshold variable q_i only enters the model through the indicator variables $\{q_i \le \gamma\}$, by picking $F_n(x)$ to be the

empirical distribution function of the q_i, we see that

$$\{q_i \le \gamma\} = \{F_n(q_i) \le F_n(\gamma)\} = \left\{\frac{j}{n} \le \gamma^*\right\}$$

for some $1 \le j \le n$. Without loss of generality, we therefore will assume that $q_i = i/n$ for the remainder of the proof. Let j_0 be the largest integer such that $j_0/n < \gamma_0$. Without loss of generality, we can also set $\sigma^2 = 1$.

If we set $\alpha = 0$, the proof of Lemma A.13 shows that uniformly in υ

$$Q_n(\gamma_0 + \upsilon/n) = -G_n^*(\upsilon) + 2\delta' R_n(\upsilon) + o_p(1) \tag{53}$$

where for $\upsilon > 0$

$$\begin{aligned} G_n^*(\upsilon) &= \sum_{i=1}^{n} (\delta' x_i)^2 \{\gamma_0 < q_i \le \gamma_0 + \upsilon/n\} \\ &= \sum_{i=1}^{n} (\delta' x_i)^2 \left\{\frac{j_0}{n} < q_i \le \frac{j_0 + \upsilon}{n}\right\} \end{aligned}$$

and

$$R_n(\nu) = \sum_{i=1}^{n} x_i e_i \left\{\frac{j_0}{n} < q_i \le \frac{j_0 + \nu}{n}\right\}.$$

While Lemma A.13 assumed that $\alpha > 0$, it can be shown that (53) continues to hold when $\alpha = 0$.

Note that the processes $G_n^*(\upsilon)$ and $\delta_n' R_n(\upsilon)$ are step functions with steps at integer-valued υ. Let N^+ denote the set of positive integers and $D_n(\upsilon)$ be any continuous, strictly increasing function such that $G_n^*(k) = D_n(k)$ for $k \in N^+$. Let $N_n(\upsilon)$ be a mean-zero Gaussian process with covariance kernel

$$E(N_n(\upsilon_1) N_n(\upsilon_2)) = D_n(\upsilon_1 \wedge \upsilon_2).$$

Since $\delta' R_n(\upsilon)$ is a mean-zero Gaussian process with covariance kernel $\sigma^2 G_n^*(\upsilon_1 \wedge \upsilon_2)$, the restriction of $\delta' R_n(\upsilon)$ to the positive integers has the same distribution as $N_n(\upsilon)$.

Since $D_n(\upsilon)$ is strictly increasing, there exists a function $\upsilon = g(s)$ such that $D_n(g(s)) = s$. Note that $N_n(g(s)) = W(s)$ is a standard Brownian motion on R^+. Let $G^+ = \{s\colon g(s) \in N^+\}$.

It follows from (53) that

$$\begin{aligned} \max_{\gamma > \gamma_0} \frac{Q_n(\gamma)}{\hat\sigma^2} &= \max_{k \in N^+} [-G_n^*(k) + 2\delta' R_n(k)] + o_p(1) \\ &\equiv \max_{k \in N^+} [-D_n(k) + 2N_n(k)] + o_p(1) \\ &= \max_{s \in G^+} [-D_n(g(s)) + 2N_n(g(s))] + o_p(1) \\ &= \max_{s \in G^+} [-s + 2W(s)] + o_p(1) \\ &\le \max_{s \in R^+} [-s + 2W(s)] + o_p(1). \end{aligned}$$

We conclude that

$$\begin{aligned} LR_n(\gamma_0) &= \max\left[\max_{\gamma > \gamma_0} \frac{Q_n(\gamma)}{\hat\sigma^2}, \max_{\gamma < \gamma_0} \frac{Q_n(\gamma)}{\hat\sigma^2}\right] \\ &\le \max\left[\max_{s \ge 0} [-|s| + 2W(s)], \max_{s \le 0} [-|s| + 2W(s)]\right] + o_p(1) \\ &\to^d \max_{-\infty < s < \infty} [-|s| + 2W(s)] = \xi. \end{aligned}$$

This shows that $P(LR_n(\gamma_0) \ge x) \le P(\xi \ge x) + o(1)$, as stated. *Q.E.D.*

REFERENCES

BAI, J. (1997): "Estimation of a Change Point in Multiple Regression Models," *Review of Economics and Statistics*, 79, 551–563.

BHATTACHARYA, P. K, and P. J. Brockwell (1976): "The Minimum of an Additive Process with Applications to Signal Estimation and Storage Theory," *Z. Wahrschein. Verw. Gebiete*, 37, 51–75.

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley.

BREIMAN, L., J. L. FRIEDMAN, R. A. OLSHEN, AND C. J. STONE (1984): *Classification and Regression Trees*. Belmont, CA: Wadsworth.

CHAN, K. S. (1989): "A Note on the Ergodicity of a Markov Chain," *Advances in Applied Probability*, 21, 702–704.

——— (1993): "Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model," *The Annals of Statistics*, 21, 520–533.

CHAN, K. S., AND R. S. TSAY (1998): "Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model," *Biometrika*, 85, 413–426.

CHU, C. S., AND H. WHITE (1992): "A Direct Test for Changing Trend," *Journal of Business and Economic Statistics*, 10, 289–299.

DAVIDSON, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians*. Oxford: Oxford University Press.

DUFOUR, J. M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural Variables and Dynamic Models," *Econometrica*, 65, 1365–1387.

DÜMBGEN, L. (1991): "The Asymptotic Behavior of Some Nonparametric Change Point Estimators," *The Annals of Statistics*, 19, 1471–1495.

DURLAUF, S. N., AND P. A. JOHNSON (1995): "Multiple Regimes and Cross-Country Growth Behavior," *Journal of Applied Econometrics*, 10, 365–384.

HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.

HANSEN, B. E. (1992): "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

——— (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

——— (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, forthcoming.

HARDLE, W., AND O. LINTON (1994): "Applied Nonparametric Methods," in *Handbook of Econometrics, Vol. IV*, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier Science, pp. 2295–2339.

IBRAGIMOV, I. A. (1975): "A Note on the Central Limit Theorem for Dependent Random Variables," *Theory of Probability and Its Applications*, 20, 135–141.

KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Annals of Statistics*, 18, 191–219.

NEWEY, W. K., AND D. L. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics, Vol. IV*, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, pp. 2113–2245.

PELIGRAD, M. (1982): "Invariance Principles for Mixing Sequences of Random Variables," *The Annals of Probability*, 10, 968–981.

PICARD, D. (1985): "Testing and Estimating Change-Points in Time Series," *Advances in Applied Probability*, 17, 841–867.

POTTER, S. M. (1995): "A Nonlinear Approach to U.S. GNP," *Journal of Applied Econometrics*, 2, 109–125.

TONG, H. (1983): *Threshold Models in Nonlinear Time Series Analysis: Lecture Notes in Statistics*, 21. Berlin: Springer.

——— (1990): *Non-Linear Time Series: A Dynamical System Approach*. Oxford: Oxford University Press.

YAO, Y.-C. (1987): "Approximating the Distribution of the ML Estimate of the Changepoint in a Sequence of Independent R.V.'s," *The Annals of Statistics*, 3, 1321–1328.