Sample Splitting and Threshold Estimation

Bruce E. Hansen*
Boston College†

April 1996
Revised: May 1998

Abstract

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 non-parametric function estimation. For example, the threshold autoregressive model (TAR) is popular in the non-linear 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. The previous literature has demonstrated 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 non-standard. 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).

*This research was supported by a grant from the National Science Foundation and a Alfred P. Sloan Foundation Research Fellowship. Thanks go to two referees, Robert de Jong and James MacKinnon for insightful comments and Mehmet Caner and Metin Celebi for excellent research assistance.

†Department of Economics, Boston College, Chestnut Hill MA 02167-3806. bruce.hansen@bc.edu. Homepage: http://www2.bc.edu/~hansenb/

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 sub-samples. Sometimes the sub-samples are selected on categorical variables, such as gender, but in other cases the sub-samples 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.

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 or "regimes". The random variable e_i is a regression error.

Formal threshold models arise in the econometric literature. One example is the Threshold Autoregressive (TAR) model of Tong (1983), 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 non-linearities 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 $q_i = i$. The proposed solution is to let the "threshold effect" $\theta_2 - \theta_1$ decrease with sample size. Under this assumption, it has been found (see Picard (1985) and Bai (1997)) that the asymptotic distribution of the changepoint estimate is non-standard yet free of nuisance parameters (other than a scale effect). We make a similar technical assumption, letting the difference $\theta_2 - \theta_1$ decrease with sample size. Interestingly, we find 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 the sampling distribution of 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 $\theta_2 - \theta_1$ decreases with sample size, and that this asymptotic distribution is an upper bound on the asymptotic distribution for the case that $\theta_2 - \theta_1$ 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 which 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 assume 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 $d_i(\gamma) = \{q_i \leq \gamma\}$ where $\{\cdot\}$ is the indicator function, set $x_i(\gamma) = x_i d_i(\gamma)$, so that (1)-(2) equal

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

where $\theta = \theta_2$ and $\delta = \theta_1 - \theta_2$. 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

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

The regression parameters are (θ, δ, γ), 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. Then by definition the LS estimators $\hat{\theta}, \hat{\delta}, \hat{\gamma}$ jointly maximize $S_n(\theta, \delta, \gamma)$. 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, yielding the conditional OLS estimators $\hat{\theta}(\gamma)$ and $\hat{\delta}(\gamma)$ by regression of Y on $X_\gamma^* = [X \quad X_\gamma]$. The concentrated sum of squared errors function is

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

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

$$\hat{\gamma} = \underset{q_i \in \Gamma}{\operatorname{argmin}}\, S_n(q_i)$$

which requires at most 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 $N \leq n$, let $q_{(j)}$ denote the (j/N)'th quantile of the sample $\{q_1, ..., q_n\}$, and let $\Gamma_N = \Gamma \cap \{q_{(1)}, ..., q_{(N)}\}$. Then $\hat{\gamma}_N = \text{argmin}_{\gamma \in \Gamma_N} S_n(\gamma)$ 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 (1992b, 1997) 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 $R_n(\gamma) = \sum_{i=1}^{n} x_i e_i \{q_i \leq \gamma\}$ 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

To allow for time-series data, we employ the weak dependence concept of absolute regularity (β-mixing). The absolute regular mixing coefficient $\beta(\mathcal{A}, \mathcal{B})$ between σ-fields $\mathcal{A}$ and $\mathcal{B}$ is

$$\beta(\mathcal{A}, \mathcal{B}) = \frac{1}{2} \sup \sum_{(i,j) \in (I,J)} \left| P(A_i \cap B_j) - P(A_i) P(B_j) \right|,$$

where $A_i \subset A$, $B_j \subset B$, and the supremum is taken over all the finite partitions $(A_i)_{i \in I}$ and $(B_j)_{j \in J}$ respectively $\mathcal{A}$ and $\mathcal{B}$ measurable. Absolute regularity was first defined by

Volkonskii and Rozanov (1959), and is stronger than strong mixing yet weaker than uniform mixing. Pham and Tran (1985) have shown that a wide class of linear processes with iid innovations (such as ARMA processes) are absolutely regular when the innovation has a bounded, continuous density. The β-mixing coefficients for the strictly stationary sequence (x_i, q_i, e_i) are given by $\beta_j = \beta(F_0, F_j)$, where $F_i = \sigma(x_{j+1}, q_{j+1}, e_j : j \leq i)$. Our theory also allows for cross-section data, for independent sequences are trivially absolutely regular with coefficients $\beta_m = 0$ for $m \geq 1$.

Define the conditional moment functionals

$$D(\gamma) = E\left(x_i x_i' \mid q_i = \gamma\right), \tag{6}$$

$$V(\gamma) = E\left(x_i x_i' e_i^2 \mid q_i = \gamma\right), \tag{7}$$

and

$$V_2(\gamma) = E\left(|x_i|^4 e_i^4 \mid q_i = \gamma\right).$$

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

Assumption 1 *For some* $\phi > 6$,

1. (x_i, q_i, e_i) *is strictly stationary with* β*-mixing coefficients* $\beta_m = O(m^{-\phi})$;
2. $E\left(e_i \mid F_{i-1}\right) = 0$;
3. $E\left|x_i\right|^4 < \infty$ *and* $E\left|e_i\right|^4 < \infty$;
4. $f(\gamma)$, $D(\gamma)$, $V(\gamma)$ *and* $V_2(\gamma)$ *are continuous at* $\gamma = \gamma_0$;
5. $P\left(q_i \in \Gamma\right) < 1$;
6. $\delta = \delta_n = cn^{-\alpha}$ *with* $c \neq 0$ *and* $\frac{1}{\phi} < \alpha < \frac{1}{2} - \frac{1}{\phi}$;
7. $c'Dc > 0$, $c'Vc > 0$, *and* $f > 0$.

Assumption 1.1 excludes time trends, integrated processes, and long memory processes. The parameter ϕ controls the degree of serial dependence in the data. If the mixing coefficients β_m decay exponentially (or if the data are independent) then we can set $\phi = \infty$.

Assumption 1.2 imposes that (1)-(2) is a correct specification of the conditional mean. Assumption 1.4 requires the threshold variable to have a continuous distribution, and essentially requires the conditional variance $E\left(e_i^2 \mid q_i = \gamma\right)$ to be continuous at γ_0, excluding regime-dependent heteroskedasticity. Assumption 1.5 requires that Γ be restricted to a subset of the support of q_i. This is a technical condition which simplifies our consistency proof.

Assumption 1.7 is a full-rank condition needed to have non-degenerate asymptotic distributions. While the restriction $c'Dc > 0$ might appear innocuous, it excludes the interesting special case of a "continuous threshold" model, which is (1)-(2) with $x_i = (1 \quad q_i)'$ and $\delta'\gamma_0^* = 0$ where $\gamma_0^* = (1 \quad \gamma_0)$. In this case the conditional mean takes the form of a continuous linear spline. From definition (6) we can calculate that $c'Dc = E\left(x_i x_i' \mid q_i = \gamma_0\right) = c'\gamma_0^*\gamma_0^{*\prime}c = 0$. A recent paper which explores the asymptotic distribution of the least squares estimates in this model is Chan and Tsay (1998).

Assumption 1.6 is the most unusual condition. 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. The parameter α controls the rate at which δ_n decreases to zero, i.e., how small we are forcing δ to be. Smaller values of α are thus less restrictive. The assumption restricts $\alpha > \phi^{-1}$, which implies a trade-off with the serial correlation in the data. For cross section data (or $\beta-$mixing processes with exponential decay) the restriction simplifies to $\alpha > 0$, which is quite mild. The reason for Assumption 1.6 is that Chan (1993) found that with δ fixed, $n(\hat{\gamma} - \gamma_0)$ converged to an asymptotic distribution which 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.

3.2 Asymptotic Distribution

Let $W_1(\nu)$ and $W_2(\nu)$ be two independent standard Brownian motions on $[0, \infty)$. A two-sided Brownian motion $W(\nu)$ on the real line is defined as

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

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, 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 $E(x_i x_i' \mid q_i = \gamma_0)$ 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 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

$$E\left(e_i^2 \mid q_i\right) = \sigma^2, \tag{8}$$

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.

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

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

and

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

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

If homoskedasticity (8) 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 non-standard, but free of nuisance parameters under (8). 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 $c_\xi(\beta) = -2\ln\left(1 - \sqrt{\beta}\right)$. Thus a test of $H_0 : \gamma = \gamma_0$ rejects at the asymptotic level of α if $LR_n(\gamma_0)$ exceeds $c_\xi(1-\alpha)$. Selected critical values are reported in Table 1.

Table 1: Asymptotic Critical Values

	.80	.85	.90	.925	.95	.975	.99
$P(\xi \leq x)$	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 parameters in order to use this theory in applications. Let $r_{1i} = (\delta_n' x_i)^2 (e_i^2/\sigma^2)$ and $r_{2i} = (\delta_n' x_i)^2$. Then

$$\eta^2 = \frac{E(r_{1i} \mid q_i = \gamma_0)}{E(r_{2i} \mid q_i = \gamma_0)} \tag{9}$$

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

A simple estimator of the ratio (9) 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

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

where $K_h(u) = h^{-1}K(u/h)$ for some bandwidth h and kernel $K(u)$, such as the Epanechnikov $K(u) = \frac{3}{4}(1 - u^2)\{|u| \leq 1\}$. 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 and indirectly a function 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 $\delta = 0$. These concerns have encouraged us to explore the construction of confidence regions based on the likelihood ratio statistic $LR_n(\gamma)$.

Let β denote the desired asymptotic confidence level (e.g. $\beta = .95$), and let $c_\xi(\beta)$ be the β-level critical value for ξ Table 1. Set

$$\hat{\Gamma} = \left\{\gamma : LR_n(\gamma) \leq c_\xi(\beta)\right\}.$$

Theorem 2 shows that $P(\gamma_0 \in \hat{\Gamma}) \to \beta$ as $n \to \infty$ under the homoskedasticity assumption (8). Thus $\hat{\Gamma}$ is an asymptotic β-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 $c_\xi(\beta)$. (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 $S_n(\hat{\gamma}) + \hat{\sigma}^2 c_\xi(\beta)$.

If the homoskedasticity condition (8) 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

$$\hat{\Gamma}^* = \left\{\gamma : LR_n^*(\gamma) \leq c_\xi(\beta)\right\}.$$

Since $\hat{\eta}^2$ is consistent for η^2, $P(\gamma_0 \in \hat{\Gamma}^*) \to \beta$ as $n \to \infty$ whether or not (8) holds, so $\hat{\Gamma}^*$ is a heteroskedasticity-robust asymptotic β-level confidence region for γ.

The region $\hat{\Gamma}$ is an asymptotic β-level confidence interval under the assumption that $\delta_n \to 0$ as $n \to \infty$, which suggests that the actual coverage of the interval may differ from β for large values of δ. We now consider the case of $\alpha = 0$, which implies that δ 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*

$$P\left(LR_n(\gamma_0) \geq x\right) \leq P\left(\xi \geq x\right) + o(1).$$

Table 2: Confidence Interval Converage 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

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 is relatively large. Unfortunately, we do not know if Theorem 3 generalizes to the case of non-normal errors or regressors which are not strictly exogenous.. The proof of Theorem 3 relies up 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 $x_i = (1\ z_i)'$, $e_i \sim N(0,1)$, and $q_i \sim N(2,1)$. The regressor z_i was either iid $N(0,1)$ or $z_i = q_i$. The likelihood ratio statistic is invariant to θ. Partitioning $\delta = (\delta_1\ \delta_2)'$ we set $\delta_1 = 0$ and $\gamma = 2$, and assessed the coverage probability of the confidence interval $\hat{\Gamma}$ 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 2 reports the coverage probabilities for nominal 10% 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 $\underline{\theta}= (\theta, \delta)$, Lemma A.12 in the Appendix shows that

$$\sqrt{n}\left(\hat{\underline{\theta}}(\gamma) - \underline{\theta}_0\right) = N(0, \Psi) + O_p(n^{\alpha-1/2}) \tag{10}$$

uniformly in a $n^{1-2\alpha}$–neighborhood of γ_0, where Ψ is the standard asymptotic covariance matrix if $\gamma = \gamma_0$ were fixed. Since $\hat{\gamma}$ is $n^{1-2\alpha}$-consistent and $\alpha < 1/2$, 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 β-level confidence region for $\underline{\theta}$ constructed under the assumption that γ is known. (10) shows that $P\left(\underline{\theta} \in \hat{\Theta}(\hat{\gamma})\right) \to \beta$ 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 some $\eta < 1$, let $\hat{\Gamma}(\eta)$ denote the confidence interval for γ with asymptotic coverage η. For each $\gamma \in \hat{\Gamma}(\eta)$ construct the pointwise confidence region $\hat{\Theta}(\gamma)$ and then set

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

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

This procedure is assessed using a simple Monte Carlo simulation. In Table 3 we report coverage rates of a nominal 95% confidence interval ($\beta = .95$) 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 $\eta = 0, .5$, .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 η lead to similar results for large δ_2. For small δ_2 and n, the best choice may be $\eta = .8$, although this may produce somewhat conservative confidence intervals for small δ_2.

In summary, while the naive choice $\hat{\Theta}(\hat{\gamma})$ 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 $\hat{\Theta}_\eta$, and the choice $\eta = .8$ appears to work reasonably well in simulations.

Table 3: Confidence Interval Converage 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

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.

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

$$\begin{aligned} \ln (Y/L)_{i,1985} - \ln (Y/L)_{i,1960} &= \zeta + \beta \ln (Y/L)_{i,1960} + \pi_1 \ln (I/Y)_i \\ &\quad + \pi_2 \ln (n_i + g + \delta) + \pi_3 \ln (SCHOOL)_i + e_i, \end{aligned} \tag{11}$$

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 (11) for four regimes selected via a regression tree using two possible threshold variables which 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 the bootstrap, using the regressors from the right-hand-side of (11) and the bootstrap dependent variable generated 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 procedure produces asymptotically correct p-values. Using 1000 bootstrap replications, the p-value for the threshold model using initial per capita 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 which 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 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 sub-sample. 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 sub-sample. 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 sub-sample. We could try to further split these two sub-samples, but none of the bootstrap test statistics were significant at the 10% level.

Our point estimates our 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 large sample distribution of the likelihood ratio statistic appears to be non-pivotal. This would be an interesting subject for future research.

Appendix: Mathematical Proofs

Let " $\Rightarrow$ " denote weak convergence with respect to the uniform metric. When the limit is non-random this specializes to uniform convergence in probability.

Lemma A. 1 *Uniformly over* $\gamma \in \Gamma$:

$$\frac{1}{n} X_\gamma' X_\gamma \Rightarrow M(\gamma) = E\left(x_i x_i' \{q_i \leq \gamma\}\right) \tag{12}$$

and

$$\frac{1}{\sqrt{n}} X_\gamma' e \Rightarrow X(\gamma), \tag{13}$$

where $X(\gamma)$ *is a mean-zero Gaussian process.*

Proof: Pick ψ so that $0 < \psi < (\phi - 4)/2$ (which is possible since $\phi > 4$). Set

$$r = \frac{2\phi}{2 + \phi + 2\psi} \tag{14}$$

and

$$\tau = \frac{1}{r} - \frac{1}{2} = \frac{1+\psi}{\phi}. \tag{15}$$

Note that $1 < r < 2$. Then

$$\sum_{m=1}^{\infty} \beta_m^{\tau} \leq \sum_{m=1}^{\infty} \left(C m^{-\phi}\right)^{\tau} = C^{\tau} \sum_{m=1}^{\infty} m^{-1-\psi} < \infty. \tag{16}$$

Theorem 3 of Hansen (1996) shows (12) and (13) follow from (16) and Assumptions 1.1-1.3. □

It will be helpful later in our analysis to observe that since $\alpha > 2/\phi$ (Assumption 1.6) and definition (14), it follows that

$$2\left(\frac{1}{r} - \frac{1}{2}\right) - \alpha < 2\left(\frac{1}{r} - \frac{1}{2} - \frac{1}{\phi}\right) = -2\frac{\psi}{\phi} < 0. \tag{17}$$

Lemma A. 2 $\hat{\gamma} \rightarrow_p \gamma_0$.

Proof: Define $Q_n(\gamma) = S_n(\gamma_0) - S_n(\gamma)$ and note that $\hat{\gamma} = \text{argmax}_{\gamma \in \Gamma} Q_n(\gamma)$. Assumption 1.5 guarantees that for n sufficiently large, the conditional OLS estimators $\hat{\theta}(\gamma)$ and $\hat{\delta}(\gamma)$ can be defined for all $\gamma \in \Gamma$ with probability one. Let $P_\gamma = X_\gamma^* \left(X_\gamma^{*\prime} X_\gamma^*\right)^{-1} X_\gamma^{*\prime}$, $P_0 = P_{\gamma_0}$, and $X_0 = X_{\gamma_0}$. Set $\Delta X_\gamma = X_\gamma - X_0$. Since X lies in the space spanned by P_γ, and $X_0 = X_\gamma - \Delta X_\gamma$,

$$\begin{aligned} S_n(\gamma) &= Y'(I - P_\gamma)Y \\ &= e'(I - P_\gamma)e + 2\delta_n' X_0'(I - P_\gamma)e + \delta_n' X_0'(I - P_\gamma)X_0\delta_n \\ &= e'(I - P_\gamma)e - 2\delta_n' \Delta X_\gamma'(I - P_\gamma)e + \delta_n' \Delta X_\gamma'(I - P_\gamma)\Delta X_\gamma \delta_n. \end{aligned}$$

Thus

$$Q_n(\gamma) = e'(P_\gamma - P_0)e + 2n^{-\alpha} c' \Delta X_\gamma'(I - P_\gamma)e - n^{-2\alpha} c' \Delta X_\gamma'(I - P_\gamma)\Delta X_\gamma c.$$

Using (12) and (13), and the fact that $X_\gamma' X = X_\gamma' X_\gamma$, we find that uniformly over $\gamma \in \Gamma$

$$\begin{aligned} n^{-1+2\alpha} Q_n(\gamma) &= n^{-1+2\alpha} e'(P_\gamma - P_0)e - 2n^{\alpha-1} c' X_0'(I - P_\gamma)e - n^{-1} c' X_0'(I - P_\gamma)X_0 c \\ &\Rightarrow -c' \left[\Delta M(\gamma) - \Delta M(\gamma) M^{-1} \Delta M(\gamma)\right] c \\ &\equiv b(\gamma) \end{aligned}$$

where $M = E\left(x_i x_i'\right)$ and

$$\Delta M(\gamma) = M(\gamma) - M(\gamma_0) = \int_{\gamma_0}^{\gamma} D(q) f(q) dq, \tag{18}$$

which can be found from (6). Note that $b(\gamma)$ is a continuous non-positive real function which achieves its unique maximum of 0 at γ_0. Since $\hat{\gamma}$ maximizes $n^{-1+2\alpha} Q_n(\gamma)$, and the latter converges in probability to the continuous function $b(\gamma)$ uniformly over Γ, and $b(\gamma)$ has a unique maximum at γ_0, it is well known that $\hat{\gamma} \rightarrow_p \gamma_0$ (see, e.g., Theorem 2.1 of Newey and McFadden, 1994). □

Lemma A. 3 *$n^\alpha \left(\hat{\theta} - \theta_0\right) = o_p(1)$ and $n^\alpha \left(\hat{\delta} - \delta_0\right) = o_p(1)$.*

Proof: For $\gamma \in \Gamma$, by (12) and (13),

$$\begin{aligned} n^\alpha \begin{pmatrix} \hat{\theta}(\gamma) - \theta_0 \\ \hat{\delta}(\gamma) - \delta_0 \end{pmatrix} &= \left(\frac{1}{n} X_\gamma^{*\prime} X_\gamma^*\right)^{-1} \left(\frac{1}{n} X_\gamma^{*\prime} e - \frac{1}{n} X_\gamma^{*\prime} \Delta X_\gamma c_0\right) \\ &\Rightarrow -\begin{pmatrix} M & M(\gamma) \\ M(\gamma) & M(\gamma) \end{pmatrix}^{-1} \begin{pmatrix} \Delta M(\gamma) \\ \Delta M(\gamma) \end{pmatrix} c_0 = \theta(\gamma), \end{aligned}$$

say. Since $\theta(\gamma)$ is continuous near γ_0, $\theta(\gamma_0) = 0$, and $\hat{\gamma} \rightarrow_p \gamma_0$ (Lemma A.2) then

$$n^\alpha \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\delta} - \delta_0 \end{pmatrix} = n^\alpha \begin{pmatrix} \hat{\theta}(\hat{\gamma}) - \theta_0 \\ \hat{\delta}(\hat{\gamma}) - \delta_0 \end{pmatrix} \rightarrow_p \theta(\gamma_0) = 0.$$

□

Let $\Delta_i(\gamma) = d_i(\gamma) - d_i(\gamma_0)$ and $\Delta_i(\gamma', \gamma'') = d_i(\gamma') - d_i(\gamma'')$. Define

$$g_i(\gamma) = (c' x_i)^2 |\Delta_i(\gamma)|$$

$$k_i(\gamma) = |x_i|^2 |\Delta_i(\gamma)|$$

and

$$h_i(\gamma', \gamma'') = |x_i e_i| |\Delta_i(\gamma', \gamma'')| .$$

Lemma A. 4 *For some $B > 0$, there exist some $d' > 0$, $d'' < \infty$, $k' > 0$, $k'' < \infty$, $H < \infty$, and $\overline{V} < \infty$ such that for all $|\gamma - \gamma_0| \leq B$,*

1. *$Eg_i(\gamma) \geq d' |\gamma - \gamma_0|$ and $Eg_i(\gamma) \leq d'' |\gamma - \gamma_0|$;*
2. *$Ek_i(\gamma) \geq k' |\gamma - \gamma_0|$ and $Ek_i(\gamma) \leq k'' |\gamma - \gamma_0|$;*
3. *$Eh_i(\gamma', \gamma'') \leq H |\gamma'' - \gamma'|$;*
4. *$Eh_i^2(\gamma', \gamma'') \leq \overline{V} |\gamma'' - \gamma'|$.*

Proof: Follows from Assumption 1 under standard Taylor series expansions. □

Let $G_n(\gamma) = n^{-1} \sum_{i=1}^n g_i(\gamma)$, $K_n(\gamma) = n^{-1} \sum_{i=1}^n k_i(\gamma)$, $H_n(\gamma', \gamma'') = n^{-1} \sum_{i=1}^n h_i(\gamma', \gamma'')$, and $H_{2n}(\gamma', \gamma'') = n^{-1} \sum_{i=1}^n h_i^2(\gamma', \gamma'')$.

Lemma A. 5 *There is a $K < \infty$ such that for all $|\gamma - \gamma_0| \leq B$,*

$$\|G_n(\gamma) - EG_n(\gamma)\|_r \leq K n^{1/r-1} |\gamma - \gamma_0|^{1/2}, \tag{19}$$

$$\|K_n(\gamma) - EK_n(\gamma)\|_r \leq K n^{1/r-1} |\gamma - \gamma_0|^{1/2}, \tag{20}$$

$$\|H_n(\gamma', \gamma'') - EH_n(\gamma', \gamma'')\|_r \leq K n^{1/r-1} |\gamma' - \gamma''|^{1/2}, \tag{21}$$

$$\|H_{2n}(\gamma', \gamma'') - EH_{2n}(\gamma', \gamma'')\|_r \leq K n^{1/r-1} |\gamma' - \gamma''|^{1/2}, \tag{22}$$

where r is defined in (14).

Proof: We show (19). Let r and τ be defined as in (14) and (15). By McLeish's (1975) α-mixing inequality, $\|E\left(g_i(\gamma) - Eg_i(\gamma) \mid F_{i-m}\right)\|_r \leq 6\alpha_m^\tau \|g_i(\gamma) - Eg_i(\gamma)\|_2$, where α_m are the strong mixing coefficients for x_i. (16) implies that $\sum_{m=1}^\infty \alpha_m^\tau < \infty$. Since $r < 2$, Lemma 2 of Hansen (1992a) implies that there is a $K_1 < \infty$ such that

$$\|G_n(\gamma) - EG_n(\gamma)\|_r \leq n^{1/r-1} K_1 \|g_i(\gamma)\|_2 \leq n^{1/r-1} K_1 \left(d'' |\gamma - \gamma_0|\right)^{1/2},$$

where the second inequality is Lemma A.4 (1). This implies (19) with $K = K_1\sqrt{d''}$. □

Set $a_n = n^{1-2\alpha}$.

Lemma A. 6 *For all $\eta > 0$ and $\varepsilon > 0$, there exists some $\overline{v} < \infty$ such that*

$$P\left(\sup_{\frac{\overline{v}}{a_n} \leq |\gamma - \gamma_0| \leq B} \left|\frac{G_n(\gamma)}{EG_n(\gamma)} - 1\right| > \eta\right) \leq \varepsilon, \tag{23}$$

and

$$P\left(\sup_{\frac{\overline{v}}{a_n} \leq |\gamma - \gamma_0| \leq B} \left|\frac{K_n(\gamma)}{EK_n(\gamma)} - 1\right| > \eta\right) \leq \varepsilon, \tag{24}$$

Proof: We show (23). Fix η and ε. Pick $b > 1$ so that

$$\left(1 + \frac{\eta}{2}\right)(1 + d''B(b-1)) \leq 1 + \eta \tag{25}$$

and

$$\left(1 - \frac{\eta}{2}\right)(1 - d'B(b-1)) \geq 1 - \eta$$

where B, d' and d'' are defined in Lemma A.4. Set

$$\overline{v} \geq \left(\frac{2K}{\eta d'}\right)^2 \frac{1}{\left[\left(1 - b^{-r/2}\right)\varepsilon\right]^{2/r}}. \tag{26}$$

and $\gamma_j = \gamma_0 + \overline{v} b^j / a_n$, $j \geq 0$.

By Markov's inequality, Lemma A.4(1) and Lemma A.5 (19),

$$\begin{aligned} P\left(\sup_{j \geq 0}\left|\frac{G_n(\gamma_j)}{EG_n(\gamma_j)} - 1\right| > \frac{\eta}{2}\right) &\leq \left(\frac{2}{\eta}\right)^r \sum_{j=0}^\infty \left[\frac{\|G_n(\gamma_j) - EG_n(\gamma_j)\|_r}{EG_n(\gamma_j)}\right]^r \\ &\leq \left(\frac{2}{\eta}\right)^r \sum_{j=0}^\infty \left[\frac{K n^{1/r-1} |\gamma_j - \gamma_0|^{1/2}}{|\gamma_j - \gamma_0| \, d'}\right]^r \end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{2K}{\eta d'}\right)^r n^{1-r} \sum_{j=0}^{\infty} \left(\frac{a_n}{\overline{v} b^j}\right)^{r/2} \\
&\leq \left(\frac{2K}{\eta d' \overline{v}^{1/2}}\right)^r \frac{1}{1-b^{-r/2}} \\
&\leq \varepsilon.
\end{aligned}
$$

The third inequality holds since (17) implies that $n^{1-r} a_n^{r/2} \leq 1$. The final inequality is (26).

Thus with probability exceeding $1 - \varepsilon$,

$$
\left| \frac{G_n(\gamma_j)}{EG_n(\gamma_j)} - 1 \right| \leq \frac{\eta}{2} \tag{27}
$$

for all $j \geq 0$. Suppose this event holds. Note that for $|\gamma_{j+1} - \gamma_0| \leq B$,

$$
|\gamma_{j+1} - \gamma_j| = \gamma_j (b-1) \leq B(b-1).
$$

Then again using Lemma A.4(1),

$$
\frac{EG_n(\gamma_{j+1})}{EG_n(\gamma_j)} \leq 1 + d'' |\gamma_{j+1} - \gamma_j| \leq 1 + d'' B(b-1). \tag{28}
$$

So for $\gamma_j < \gamma < \gamma_{j+1}$, using (27), (28), and (25),

$$
\begin{aligned}
\frac{G_n(\gamma)}{EG_n(\gamma)} &\leq \frac{G_n(\gamma_{j+1})}{EG_n(\gamma_j)} = \frac{G_n(\gamma_{j+1})}{EG_n(\gamma_{j+1})} \frac{EG_n(\gamma_{j+1})}{EG_n(\gamma_j)} \\
&\leq \left(1 + \frac{\eta}{2}\right) (1 + d'' B(b-1)) \leq 1 + \eta.
\end{aligned}
$$

Similarly, we can show that $G_n(\gamma)/EG_n(\gamma) \geq 1 - \eta$. Hence, with probability exceeding $1 - \varepsilon$, $|G_n(\gamma_j)/EG_n(\gamma) - 1| \leq \eta$ for all $\overline{v}/a_n \leq |\gamma - \gamma_0| \leq B$. □

Let

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

Lemma A. 7 *For all $\eta > 0$ and $\varepsilon > 0$, there exists some $\overline{v} < \infty$ such that for n sufficiently large,*

$$
P\left(\sup_{\frac{\overline{v}}{a_n} \leq |\gamma - \gamma_0| \leq B} \frac{|R_n(\gamma)|}{\sqrt{a_n} |\gamma - \gamma_0|} > \eta \right) \leq \varepsilon.
$$

Proof: Fix $\eta > 0$ and $\varepsilon > 0$. Set $b = 1 + \eta/H$ and for $j \geq 0$, $\gamma_j - \gamma_0 = \overline{\upsilon} b^j / a_n$, where $\overline{\upsilon}$ will be determined later. Let $m = n^\alpha$ and for $k = 1, ..., m$, set $\gamma_{jk} = \gamma_j + (\gamma_{j+1} - \gamma_j)k/m$. Observe that by Lemma A.4(3),

$$\sqrt{n} E H_n(\gamma_{jk}, \gamma_{jk+1}) \leq H\sqrt{n} \left|\gamma_{jk+1} - \gamma_{jk}\right| = H(b-1)\sqrt{n}\left(\gamma_j - \gamma_0\right)/m = \eta\left(\gamma_j - \gamma_0\right)\sqrt{a_n},$$

so

$$\begin{aligned}
&\sup_{\gamma_j \leq \gamma \leq \gamma_{j+1}} \left|R_n(\gamma) - R_n(\gamma_j)\right| - \max_{1 \leq k \leq m} \left|R_n(\gamma_{jk}) - R_n(\gamma_j)\right| \\
&\leq \max_{1 \leq k \leq m} \sup_{\gamma_{jk} \leq \gamma \leq \gamma_{jk+1}} \left|R_n(\gamma) - R_n(\gamma_{jk})\right| \\
&\leq \max_{1 \leq k \leq m} \sqrt{n} \left|H_n(\gamma_{jk}, \gamma_{jk+1})\right| \\
&\leq \max_{1 \leq k \leq m} \sqrt{n} \left|H_n(\gamma_{jk}, \gamma_{jk+1}) - E H_n(\gamma_{jk}, \gamma_{jk+1})\right| + \eta\sqrt{a_n}\left(\gamma_j - \gamma_0\right),
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\frac{\overline{\upsilon}}{a_n} \leq |\gamma - \gamma_0| \leq B} \frac{|R_n(\gamma)|}{\sqrt{a_n}\,|\gamma - \gamma_0|} &\leq \max_{j \geq 0} \frac{|R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} + \max_{j \geq 0} \sup_{\gamma_j \leq \gamma \leq \gamma_{j+1}} \frac{|R_n(\gamma) - R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} \\
&\leq \max_{j \geq 0} \frac{|R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} + \max_{j \geq 0} \max_{1 \leq k \leq m} \frac{|R_n(\gamma_{jk}) - R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} \\
&\quad + \max_{j \geq 0} \max_{1 \leq k \leq m} \frac{\sqrt{n}\,|H_n(\gamma_{jk}, \gamma_{jk+1}) - E H_n(\gamma_{jk}, \gamma_{jk+1})|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} + \eta.
\end{aligned}$$

Hence

$$\begin{aligned}
P\left(\sup_{\frac{\overline{\upsilon}}{a_n} \leq |\gamma - \gamma_0| \leq B} \frac{|R_n(\gamma)|}{\sqrt{a_n}\,|\gamma - \gamma_0|} > 4\eta\right) &\leq \sum_{j=0}^{\infty} P\left(\frac{|R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} > \eta\right) \\
&+ \sum_{j=0}^{\infty} P\left(\max_{1 \leq k \leq m} \frac{|R_n(\gamma_{jk}) - R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} > \eta\right) \\
&+ \sum_{j=0}^{\infty} P\left(\max_{1 \leq k \leq m} \frac{\sqrt{n}\,|H_n(\gamma_{jk}, \gamma_{jk+1}) - E H_n(\gamma_{jk}, \gamma_{jk+1})|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} > \eta\right). \qquad (29)
\end{aligned}$$

We now bound each of the elements on the right-hand-side of (29).

First, by Markov's inequality, the martingale difference property and Lemma A.4(4),

$$\begin{aligned}
\sum_{j=0}^{\infty} P\left(\frac{|R_n(\gamma_j)|}{\sqrt{a_n}\,|\gamma_j - \gamma_0|} > \eta\right) &\leq \frac{1}{\eta^2} \sum_{j=0}^{\infty} \frac{E\left|R_n(\gamma_j)\right|^2}{a_n\left|\gamma_j - \gamma_0\right|^2} \\
&= \frac{1}{\eta^2} \sum_{j=0}^{\infty} \frac{E h_i^2(\gamma_0, \gamma_j)}{a_n\left|\gamma_j - \gamma_0\right|^2}
\end{aligned}$$

$$\begin{aligned}
&\le \frac{1}{\eta^2}\sum_{j=0}^{\infty}\frac{\overline{V}\left|\gamma_j-\gamma_0\right|}{a_n\left|\gamma_j-\gamma_0\right|^2} \\
&= \frac{\overline{V}}{\overline{\upsilon}\eta^2}\frac{1}{1-1/b}. \qquad (30)
\end{aligned}$$

Second, observe that by Burkholder's inequality (see, e.g., Hall and Heyde (1980, p. 23), Minkowski's inequality, Lemma A.4(3), and Lemma A.5 (22),

$$\begin{aligned}
E\left|R_n(\gamma_{jk+1})-R_n(\gamma_{jk})\right|^{2r} &= E\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}x_i e_i\left(\Delta_i(\gamma_{jk+1})-\Delta_i(\gamma_{jk})\right)\right|^{2r} \\
&\le C'E\left|H_{2n}(\gamma_{jk},\gamma_{jk+1})\right|^r \\
&\le C'\left[EH_{2n}(\gamma_{jk},\gamma_{jk+1})+\left\|H_{2n}(\gamma_{jk},\gamma_{jk+1})-EH_{2n}(\gamma_{jk},\gamma_{jk+1})\right\|_r\right]^r \\
&\le C'\left[\overline{V}\left|\gamma_{jk+1}-\gamma_{jk}\right|+Kn^{1/r-1}\left|\gamma_{jk+1}-\gamma_{jk}\right|^{1/2}\right]^r \\
&\le C\left|\gamma_{jk+1}-\gamma_{jk}\right|^r. \qquad (31)
\end{aligned}$$

where C' is from Burkholder's inequality and $C = C'\left(\overline{V}+K\right)^r$, and the final inequality in (31) follows from

$$\left|\gamma_{jk+1}-\gamma_{jk}\right|^{-1/2}=\left|\frac{\gamma_{j+1}-\gamma_j}{m}\right|^{-1/2}\le\frac{n^{1/2-\alpha/2}}{\sqrt{\overline{\upsilon}(b-1)}}\le n^{1-1/r},$$

which holds for n sufficiently large since $\alpha/2 > 1/r - 1/2$, as shown in (17). The bound (31) and Theorem 12.2 of Billingsley (1968, p.94) implies that there is a $K' < \infty$ such that

$$P\left(\max_{1\le k\le m}\left|R_n(\gamma_{jk})-R_n(\gamma_j)\right|>\rho\right)\le\frac{K'\left(\gamma_{jm}-\gamma_j\right)^r}{\rho^{2r}}=\frac{K'\left(\gamma_{j+1}-\gamma_j\right)^r}{\rho^{2r}}.$$

Hence

$$\begin{aligned}
\sum_{j=0}^{\infty}P\left(\max_{1\le k\le m}\frac{\left|R_n(\gamma_{jk})-R_n(\gamma_j)\right|}{\sqrt{a_n}\left|\gamma_j-\gamma_0\right|}>\eta\right) &\le \sum_{j=0}^{\infty}\frac{K'\left(\gamma_{j+1}-\gamma_j\right)^r}{\left(\sqrt{a_n}\left|\gamma_j-\gamma_0\right|\eta\right)^{2r}} \\
&= \frac{K'}{(1-b^{-r})}\left(\frac{b-1}{\eta^2\overline{\upsilon}}\right)^r. \qquad (32)
\end{aligned}$$

Finally, by Markov's inequality, Lemma 3.2 of Pollard (1990, p. 10), Lemma A.5 (21), and $m < n$,

$$\sum_{j=0}^{\infty}P\left(\max_{1\le k\le m}\frac{\sqrt{n}\left|H_n(\gamma_{jk},\gamma_{jk+1})-EH_n(\gamma_{jk},\gamma_{jk+1})\right|}{\sqrt{a_n}\left|\gamma_j-\gamma_0\right|}>\eta\right)$$

$$
\begin{aligned}
&\le \left(\frac{\sqrt{n/a_n}}{\eta}\right)^r \sum_{j=0}^{\infty} E \max_{1\le k\le m} \frac{|H_n(\gamma_{jk},\gamma_{jk+1}) - EH_n(\gamma_{jk},\gamma_{jk+1})|^r}{|\gamma_j-\gamma_0|^r} \\
&\le \left(\frac{n^\alpha\sqrt{2+\log m}}{\eta}\right)^r \sum_{j=0}^{\infty} \max_{1\le k\le m} \frac{\|H_n(\gamma_{jk},\gamma_{jk+1}) - EH_n(\gamma_{jk},\gamma_{jk+1})\|_r^r}{|\gamma_j-\gamma_0|^r} \\
&\le \left(\frac{n^\alpha\sqrt{2+\log m}}{\eta}\right)^r \sum_{j=0}^{\infty} \max_{1\le k\le m} \left(\frac{Kn^{1/r-1}\,|\gamma_{jk+1}-\gamma_{jk}|^{1/2}}{|\gamma_j-\gamma_0|}\right)^r \\
&= \left(\frac{K\sqrt{2+\log n}(b-1)^{1/2}n^{1/r-1/2-\alpha/2}}{\eta\overline{v}^{1/2}}\right)^r \frac{1}{1-b^{-r/2}} \\
&\le \left(\frac{K(b-1)^{1/2}}{\eta\overline{v}^{1/2}}\right)^r \frac{1}{1-b^{-r/2}} \qquad (33)
\end{aligned}
$$

for n sufficiently large since $\alpha/2 > 1/r - 1/2$, as shown in (17).

The sum of the right-hand-sides of (30), (32) and (33) can be made less than ε by picking suitably large $\overline{v}$. Together with (29) this establishes the result. □

Lemma A. 8 $a_n(\hat{\gamma}-\gamma_0) = O_p(1)$.

Proof: Fix $\varepsilon > 0$. Pick $\eta > 0$ and $\kappa > 0$ small enough so that

$$
(1-\eta)\,d' - 2\left(|c|+\kappa\right)\eta - 2\left(|c|+\kappa\right)(1+\eta)\,k''\kappa - \left(\left(2\,|c|+\kappa\right)k''(1+\eta)\kappa\right) > 0. \qquad (34)
$$

Let $\overline{v}$ be large enough so that Lemmas A.6 and A.7 hold. Let E_n be the joint event that $|\hat{\gamma}-\gamma_0| \le B$, $n^\alpha\left|\hat{\theta}-\theta_0\right| \le \kappa$, $n^\alpha\left|\hat{\delta}-\delta_0\right| \le \kappa$,

$$
P\left(\sup_{\frac{\overline{v}}{a_n}\le|\gamma-\gamma_0|\le B}\left|\frac{G_n(\gamma)}{EG_n(\gamma)}-1\right| > \eta\right) \le \varepsilon, \qquad (35)
$$

$$
P\left(\sup_{\frac{\overline{v}}{a_n}\le|\gamma-\gamma_0|\le B}\left|\frac{K_n(\gamma)}{EK_n(\gamma)}-1\right| > \eta\right) \le \varepsilon, \qquad (36)
$$

and

$$
P\left(\sup_{\frac{\overline{v}}{a_n}\le|\gamma-\gamma_0|\le B}\frac{|R_n(\gamma)|}{\sqrt{a_n}\,|\gamma-\gamma_0|} > \eta\right) \le \varepsilon. \qquad (37)
$$

Let n be large enough so that Lemma A.7 holds and $P(E_n) \geq 1-\varepsilon$, which is possible under Lemmas A.2, A.3, A.6 and A.7.

Take any $\gamma \in [\,\gamma_0 + \overline{v}/a_n, \gamma_0 + B]$. Suppose E_n holds. (35) and Lemma A.4 (1) imply

$$\frac{c'\Delta X_\gamma' \Delta X_\gamma c}{n\,(\gamma-\gamma_0)} = \frac{G_n(\gamma)}{(\gamma-\gamma_0)} = \frac{G_n(\gamma)}{EG_n(\gamma)}\frac{EG_n(\gamma)}{(\gamma-\gamma_0)} \geq (1-\eta)\,d'. \tag{38}$$

(36) and Lemma A.4 (2) show that

$$\frac{\left|\Delta X_\gamma' \Delta X_\gamma\right|}{n\,(\gamma-\gamma_0)} = \frac{K_n(\gamma)}{(\gamma-\gamma_0)} = \frac{K_n(\gamma)}{EK_n(\gamma)}\frac{EK_n(\gamma)}{(\gamma-\gamma_0)} \leq (1+\eta)\,k''. \tag{39}$$

(37) implies

$$\frac{\left|\Delta X_\gamma' e\right|}{n^{1-\alpha}\,(\gamma-\gamma_0)} = \frac{|R_n'(\gamma)|}{\sqrt{a_n}\,(\gamma-\gamma_0)} \leq \eta. \tag{40}$$

Let $S_n^*(\gamma) = S_n(\hat{\theta},\hat{\delta},\gamma)$, where $S_n(\theta,\delta,\gamma)$ is the sum of squared errors function (5). Since $Y = X\theta_0 + X_{\gamma_0}\delta_0 + e$,

$$Y - X\hat{\theta} - X_\gamma\hat{\delta} = \left(e - X\left(\hat{\theta}-\theta_0\right) - X_{\gamma_0}\left(\hat{\delta}-\delta_0\right)\right) - \Delta X_\gamma\hat{\delta},$$

we find that

$$\begin{aligned}
S_n^*(\gamma) - S_n^*(\gamma_0) &= \left(Y - X\hat{\theta} - X_\gamma\hat{\delta}\right)'\left(Y - X\hat{\theta} - X_\gamma\hat{\delta}\right) - \left(Y - X\hat{\theta} - X_{\gamma_0}\hat{\delta}\right)'\left(Y - X\hat{\theta} - X_{\gamma_0}\hat{\delta}\right) \\
&= \hat{\delta}'\Delta X_\gamma'\Delta X_\gamma\hat{\delta} - 2\hat{\delta}'\Delta X_\gamma' e + 2\hat{\delta}'\Delta X_\gamma'\Delta X_\gamma\left(\hat{\theta}-\theta_0\right) \\
&= \delta_0'\Delta X_\gamma'\Delta X_\gamma\delta_0 - 2\hat{\delta}'\Delta X_\gamma' e + 2\hat{\delta}'\Delta X_\gamma'\Delta X_\gamma\left(\hat{\theta}-\theta_0\right) \\
&\quad + \left(\delta_0+\hat{\delta}\right)'\Delta X_\gamma'\Delta X_\gamma\left(\hat{\delta}-\delta_0\right).
\end{aligned} \tag{41}$$

Let $\hat{c} = n^\alpha\hat{\delta}$ so that $|\hat{c}-c| \leq \kappa$. By (38), (39), (40), and (34),

$$\begin{aligned}
\frac{S_n^*(\gamma) - S_n^*(\gamma_0)}{a_n\,(\gamma-\gamma_0)} &= \frac{c'\Delta X_\gamma'\Delta X_\gamma c}{n\,(\gamma-\gamma_0)} - \frac{2\hat{c}'\Delta X_\gamma' e}{n^{1-\alpha}\,(\gamma-\gamma_0)} + \frac{2\hat{c}'\Delta X_\gamma'\Delta X_\gamma n^\alpha\left(\hat{\theta}-\theta_0\right)}{n\,(\gamma-\gamma_0)} \\
&\quad + \frac{(c+\hat{c})'\,\Delta X_\gamma'\Delta X_\gamma\,(\hat{c}-c)}{n\,(\gamma-\gamma_0)} \\
&\geq (1-\eta)\,d' - 2\,(|c|+\kappa)\,\eta - 2\,(|c|+\kappa)\,(1+\eta)\,k''\kappa - ((2\,|c|+\kappa)\,k''(1+\eta)\kappa) \\
&> 0.
\end{aligned}$$

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

$S_n^*(\hat{\gamma}) - S_n^*(\gamma_0) \leq 0$, this establishes that E_n implies $|\hat{\gamma} - \gamma_0| \leq \overline{\upsilon}/a_n$. Since $P(E_n) \geq 1 - \varepsilon$, the proof is complete. $\square$

Let Ψ denote any compact subset of R. Let $G_n^*(\nu) = a_n G_n(\gamma_0 + \nu/a_n)$ and $K_n^*(\nu) = a_n K_n(\gamma_0 + \nu/a_n)$

Lemma A. 9 *Uniformly in* $\nu \in \Psi$,

$$G_n^*(\nu) \Rightarrow d\,|\nu|\,, \tag{42}$$

and

$$K_n^*(\nu) \Rightarrow |Df|\,|\nu|\,, \tag{43}$$

where $d = c'Dcf$.

Proof: We show (42). Fix $\nu \in \Psi$. Taking expectations,

$$\begin{aligned} EG_n^*(\nu) &= a_n c' E\left(x_i x_i' \left|\Delta_i(\gamma_0 + \nu/a_n)\right|\right) c \\ &= a_n \left|c' \left(M(\gamma_0 + \nu/a_n) - M(\gamma_0)\right) c\right| \\ &\to |\nu|\, c' M'(\gamma_0) c \\ &= |\nu|\, d \end{aligned} \tag{44}$$

as $n \to \infty$. The final equality comes from $M'(\gamma_0) = Df$, which can be seen from (18) and the definition $d = c'Dfc$. By Lemma A.5 (19) and (17),

$$\left\|G_n^*(\nu) - EG_n^*(\nu)\right\|_r \leq a_n K n^{1/r-1} \left|\frac{\nu}{a_n}\right|^{1/2} \to 0. \tag{45}$$

Markov's inequality, (44) and (45) show that $G_n^*(\nu) \to_p d\,|\nu|\,.$

Since $G_n^*(\nu)$ is monotonically increasing in $|\nu|$ and the limit function is continuous, the convergence is uniform over Ψ. To see this, set $G(\nu) = d\,|\nu|\,.$ Pick any $\varepsilon > 0$, then pick J and $(\nu_1, ..., \nu_J)$ so that $|G(\nu_j) - G(\nu_{j-1})| \leq \varepsilon$ for all j, which is possible since $G(\nu)$ is continuous and Ψ is compact. Then pick n large enough so that $\max_{j \leq J} |G_n^*(\nu_j) - G(\nu_j)| \leq \varepsilon$ with probability greater than $1 - \varepsilon$, which is possible by pointwise consistency. For any j, take any $\nu \in (\nu_{j-1}, \nu_j)$. Both $G_n^*(\nu)$ and $G(\nu)$ lie in the interval $[G(\nu_{j-1}) - \varepsilon, G(\nu_j) + \varepsilon]$, (with probability greater than $1 - \varepsilon$) which has length bounded by 3ε. Since ν is arbitrary, $|G_n^*(\nu) - G(\nu)| \leq 3\varepsilon$ uniformly over Ψ. $\square$

Let $R_n^*(\nu) = \sqrt{a_n} R_n (\gamma_0 + \nu/a_n)$.

Lemma A. 10 *Uniformly in* $\nu \in \Psi$,

$$R_n^*(\nu) \Rightarrow B(\nu)$$

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 $R_n^*(\nu)$ to those of $B(v)$ and then showing that $R_n^*(\nu)$ is stochastically equicontinuous.

Fix $\nu \in \Psi$. Define $u_{ni}(\nu) = \sqrt{a_n/n} x_i e_i \Delta_i(\gamma_0 + \nu/a_n)$ so that $R_n^*(\nu) = \sum_{i=1}^n u_{ni}(\nu)$, and define $V_n(\nu) = \sum_{i=1}^n u_{ni}(\nu) u_{ni}(\nu)'$. Under Assumption 1.2, $\{u_{ni}(\nu), F_i\}$ 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 $R_n^*(\nu) \to_d N(0, |\nu| Vf)$ are

$$V_n(\nu) \to_p |\nu| Vf \tag{46}$$

and

$$\max_{1\le i\le n} |u_{ni}(\nu)| \to_p 0. \tag{47}$$

Set

$$\Omega(\gamma) = E\left(x_i x_i' e_i^2 \{q_i \le \gamma\}\right)$$

so that from (7) we can find

$$\Omega'(\gamma) = V(\gamma) f(\gamma). \tag{48}$$

Then

$$\begin{aligned} EV_n(\nu) &= a_n E\left(x_i x_i' e_i^2 |\Delta_i(\nu)|\right) \\ &= a_n |\Omega(\gamma_0 + \nu/a_n) - \Omega(\gamma_0)| \\ &\to |\nu| \Omega'(\gamma_0) \\ &= |\nu| Vf \end{aligned} \tag{49}$$

by (48). Note that $|V_n(\nu)| = a_n H_{2n}(\gamma_0, \gamma_0 + \nu/a_n)$, so by Lemma A.5 (22), and (17),

$$\|V_n(\nu) - EV_n(\nu)\|_r \le a_n K n^{1/r-1} \left|\frac{\nu}{a_n}\right|^{1/2} \to 0 \tag{50}$$

as $n \to \infty$, which with (49) combines to yield (46). Note also that using Lemma A.4 (4),

$$E\left|u_{ni}(\nu)\right|^2 = \frac{1}{n}E\left|V_n(\nu)\right| = \frac{a_n}{n}EH_{2n}(\gamma_0, \gamma_0 + \nu/a_n) \le \frac{\overline{V}\left|\nu\right|}{n}, \tag{51}$$

so combined with Lemma 3.2 of Pollard (1990, p. 10), we find

$$\begin{aligned} \left\| \max_{1\le i\le n} \left|u_{ni}(\nu)\right| \right\|_2 &\le \left((2+\log n) \max_{1\le i\le n} E\left|u_{ni}(\nu)\right|^2\right)^{1/2} \\ &\le (2+\log n)\frac{\overline{V}\left|\nu\right|}{n} \to 0, \end{aligned}$$

which establishes (47) by Markov's inequality. We conclude that $R_n^*(\nu) \to_d N(0, |\nu| Vf)$. This argument can be extended to include any finite collection $[\nu_1, ..., \nu_k]$ to yield the convergence of the finite dimensional distributions of $R_n^*(\nu)$ to those of $B(\nu)$.

It remains to show that the sequence $R_n'(\nu)$ is stochastically equicontinuous on Ψ. We appeal to Theorem 1 of Doukhan, Massart and Rio (1994). First, note that $R_n^*(\nu)$ is $L_{2r}-$bounded. Indeed, by Burkholder's inequality, (51), and (50),

$$\begin{aligned} E\left|R_n^*(\nu)\right|^{2r} &\le C'E\left|V_n(\nu)\right|^r \\ &\le C'\left(\left\|V_n(\nu) - EV_n(\nu)\right\|_r + E\left|V_n(\nu)\right|\right)^r \\ &\le C'\left(K\nu^{1/2} + \overline{V}\left|\nu\right|\right)^r < \infty. \end{aligned}$$

Second, the dimension of L_2-entropy with bracketing over the function class $\{R_n^*(\nu) : \nu \in \Psi\}$ is logarithmic. Indeed, since Ψ is a bounded subset of R, if $|\nu_2 - \nu_1| \le \rho/\left(|V|\, f\right)$, then by the martingale difference property,

$$\begin{aligned} \left\|R_n^*(\nu_2) - R_n^*(\nu_1)\right\|_2^2 &= E\left|\sqrt{\frac{a_n}{n}}\sum_{i=1}^n x_i e_i\left(\Delta_i(\gamma_0+\nu_2) - \Delta_i(\gamma_0+\nu_1)\right)\right|^2 \\ &= a_n\left|\Omega(\gamma_0 + \nu_2/a_n) - \Omega(\gamma_0+\nu_1/a_n)\right| \\ &\to |V|\, f\left|\nu_2 - \nu_1\right| \\ &\le \rho, \end{aligned}$$

which means that the L_2 entropy is of order $o\left(\left|\log \rho\right|\right)$. Third, from the definition of r (14) and ψ, we calculate that $\phi\left(1 - 1/r\right) = \phi/2 - 1 - \psi = \psi^* > 1$. Thus

$$\beta_m^{1-1/r} \le C^{1-1/r} m^{-\phi\left(1-\frac{1}{r}\right)} = C^{1-1/r} m^{-\psi^*}. \tag{52}$$

As discussed by Doukhan, Massart and Rio (1994) in the last paragraph of their application 5, $L_{2r}-$boundedness, logarithmic L_2-entropy, and (52) are sufficient for stochastic equicontinuity. This completes the proof. □

Let $Q_n^*(\nu) = (a_n/n)\left(S_n^*(\gamma_0) - S_n^*(\gamma_0 + \nu/a_n)\right)$, where $S_n^*(\gamma) = S_n(\hat{\theta}, \hat{\delta}, \gamma)$.

Lemma A. 11 *Uniformly in $\nu \in \Psi$, $Q_n^*(\nu) \Rightarrow Q(\nu) = -d\,|\nu| + 2\sqrt{\lambda}W(\nu)$, where $\lambda = c'Vcf$.*

Proof: From (41) we find $Q_n^*(\nu) = -G_n^*(\nu) + 2c'R_n^*(\nu) + L_n(\nu)$, where

$$\begin{aligned} |L_n(\nu)| &\leq 2\left|\hat{\delta} - \delta\right| \sqrt{\frac{a_n}{n}}\,|R_n^*(\nu)| + 2\left|\hat{\delta}\right| \left|\hat{\theta} - \theta_0\right| K_n(\nu) + \left|\delta_0 + \hat{\delta}\right| \left|\hat{\delta} - \delta_0\right| K_n(\nu) \\ &\Rightarrow 0 \end{aligned}$$

since $\left|\hat{\theta} - \theta_0\right| \to_p 0$, $\left|\hat{\delta} - \delta_0\right| \to_p 0$, $K_n(\nu) = O_p(1)$ and $R_n^*(\nu) = O_p(1)$, by Lemmas A.3, A.9 and A.10. Applying again Lemmas A.9 and A.10, $Q_n^*(\nu) \Rightarrow -d\,|\nu| + 2c'B(\nu)$. The process $c'B(\nu)$ is a Brownian motion with variance $\lambda = c'Vcf$, so can be written as $\sqrt{\lambda}W(\nu)$, where $W(v)$ is a standard Brownian motion. □

Proof of Theorem 1: Let $\hat{\nu} = a_n(\hat{\gamma} - \gamma_0)$, or $\hat{\gamma} = \gamma_0 + \hat{\nu}/a_n$. Using this reparameterization, $\hat{\nu} = \operatorname{argmax}_{\nu \in \Psi_n} Q_n^*(\nu)$, where $\Psi_n = [a_n(\underline{\gamma} - \gamma_0), a_n(\overline{\gamma} - \gamma_0)]$. Lemma A.8 implies that for any $\varepsilon > 0$, there exists some $\overline{\nu}_\epsilon < \infty$ such that setting $\Psi_\varepsilon = [-\overline{\nu}_\epsilon, \overline{\nu}_\epsilon]$,

$$P\left(a_n(\hat{\gamma} - \gamma_0) = \operatorname*{argmax}_{\nu \in \Psi_\varepsilon} Q_n^*(\nu)\right) \geq 1 - \varepsilon.$$

That is, if we define

$$\hat{\nu}_\varepsilon = \operatorname*{argmax}_{\nu \in \Psi_\varepsilon} Q_n^*(\nu), \tag{53}$$

then $P(\hat{\nu}_\varepsilon = \hat{\nu}) \geq 1 - \varepsilon$.

To find the asymptotic distribution of $\hat{\nu}_\varepsilon$ from (53) and Lemma A.11, one might be tempted to apply the continuous mapping theorem. The "argmax" functional, however, is not continuous. Kim and Pollard (1990, Theorem 2.7) have provided an appropriate alternative. They show that under Lemma A.11, since $Q(v)$ is continuous, has a unique maximum, and $\lim_{|v| \to \infty} Q(v) = -\infty$ almost surely (which is true since $\lim_{v \to \infty} W(\nu)/\nu = 0$

almost surely), then $\hat{\nu}_\varepsilon \to^d \text{argmax}_{\nu\in\Psi_\varepsilon} Q(\nu)$. Since ε is arbitrary we find the asymptotic distribution of the threshold estimator:

$$a_n(\hat{\gamma}-\gamma_0) = \hat{\nu} \to^d \underset{\nu\in R}{\text{argmax}}\, Q(\nu).$$

The final step is to simplify the asymptotic distribution. Making the change-of-variables $\nu = (\lambda/d^2)\, r$, and noting the distributional equality $W(a^2 r) \equiv aW(r)$ we can re-write the asymptotic distribution as

$$\begin{aligned}
\underset{\nu\in R}{\text{argmax}}\, Q(\nu) &= \underset{-\infty<\nu<\infty}{\text{argmax}} \left[-d\,|\nu| + 2\sqrt{\lambda}W(\nu)\right] \\
&= \frac{\lambda}{d^2} \underset{-\infty<r<\infty}{\text{argmax}} \left[-\frac{\lambda}{d}\,|r| + 2\sqrt{\lambda}W\left(\frac{\lambda}{d^2}r\right)\right] \\
&\equiv \omega \underset{-\infty<r<\infty}{\text{argmax}} \left[-\frac{\lambda}{d}\,|r| + 2\frac{\lambda}{d}W(r)\right] \\
&= \omega \underset{-\infty<r<\infty}{\text{argmax}} \left[-\frac{|r|}{2} + W(r)\right].
\end{aligned}$$

since $\omega = \lambda/d^2$. □

Let $\underline{\theta} = (\theta, \delta)$.

Lemma A. 12 *Uniformly in a $n^{1-2\alpha}-$neighborhood of γ_0,*

$$\sqrt{n}\left(\hat{\underline{\theta}}(\gamma) - \underline{\theta}_0\right) = Z + O_p(a_n^{-1/2}),$$

where $Z \sim N(0,\Psi)$, and Ψ is the standard asymptotic covariance matrix if $\gamma = \gamma_0$ were fixed.

Proof: Lemma A.9 (43), $\frac{a_n}{n}\Delta X_\gamma' \Delta X_\gamma = O_p(1)$. Thus

$$\begin{aligned}
\sqrt{n}\left(\hat{\underline{\theta}}(\gamma) - \underline{\theta}_0\right) &= \left(\frac{1}{n}X_\gamma^{*\prime}X_\gamma^*\right)^{-1}\left(\frac{1}{\sqrt{n}}X_\gamma^{*\prime}e - \begin{pmatrix} I \\ I \end{pmatrix}\left(\frac{a_n}{n}\Delta X_\gamma' \Delta X_\gamma\right) c a_n^{-1/2}\right) \\
&= \left(\frac{1}{n}X_\gamma^{*\prime}X_\gamma^*\right)^{-1}\left(\frac{1}{\sqrt{n}}X_\gamma^{*\prime}e\right) + O_p(a_n^{-1/2}),
\end{aligned}$$

and

$$\left(\frac{1}{n}X_\gamma^{*\prime}X_\gamma^*\right)^{-1}\left(\frac{1}{n}X_\gamma^{*\prime}e\right) \Rightarrow -\begin{pmatrix} M & M(\gamma_0) \\ M(\gamma_0) & M(\gamma_0) \end{pmatrix}^{-1}\begin{pmatrix} X \\ X(\gamma_0) \end{pmatrix} = Z.$$

□

Proof of Theorem 2: From Lemma A.12,

$$\sqrt{n}\left(\hat{\underline{\theta}}(\gamma_0) - \hat{\underline{\theta}}\right) = \sqrt{n}\left(\hat{\underline{\theta}}(\gamma_0) - \underline{\theta}\right) - \sqrt{n}\left(\hat{\underline{\theta}}(\hat{\gamma}) - \underline{\theta}_0\right) \Rightarrow Z - Z = 0.$$

It is also easy to calculate that for $\underline{\theta}$ in a $\sqrt{n}$ neighborhood of $\underline{\theta}_0$,

$$\left|\frac{\partial}{\partial \underline{\theta}} S_n(\underline{\theta}, \gamma_0)\right| = O_p(n^{1/2}).$$

Hence for θ^* on a line segment joining $\hat{\underline{\theta}}$ and $\hat{\underline{\theta}}(\gamma_0)$,

$$\left|LR_n(\gamma_0) - \frac{Q_n^*(\hat{\nu})}{\hat{\sigma}^2}\right| \leq \left|\frac{\partial}{\partial \underline{\theta}} S_n(\theta^*, \gamma_0)\right| \left|\hat{\underline{\theta}}(\gamma_0) - \hat{\underline{\theta}}\right| = o_p(1).$$

Now applying Lemma A.11 and the continuous mapping theorem

$$\begin{aligned}
LR_n(\gamma_0) &= \frac{Q_n^*(\hat{\nu})}{\hat{\sigma}^2} + o_p(1) \\
&= \frac{\sup_{\nu} Q_n^*(\nu)}{\hat{\sigma}^2} + o_p(1) \\
&\to {}_d \frac{\sup_{\nu} Q(\nu)}{\sigma^2} \\
&= \frac{1}{\sigma^2} \sup_{\nu} \left[-d\left|\nu\right| + 2\sqrt{\lambda} W(\nu)\right] \\
&= \frac{1}{\sigma^2} \sup_{r} \left[-d\left|\frac{\lambda}{d^2} r\right| + 2\sqrt{\lambda} W(\frac{\lambda}{d^2} r)\right] \\
&\equiv \frac{\lambda}{\sigma^2 d} \sup_{r} \left[-\left|r\right| + 2W(r)\right] \\
&= \eta^2 \xi
\end{aligned}$$

the second equality following from the change-of-variables $\nu = \left(\lambda/d^2\right) r$, the second-to-last equality by $W(a^2 r) \equiv aW(r)$, and the final equality by the fact that $\eta^2 = \lambda/(\sigma^2 d)$.

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

$$\begin{aligned}
P\left(\xi \leq x\right) &= P\left(2\max\left[\xi_1, \xi_2\right] \leq x\right) \\
&= P\left(\xi_1 \leq x/2\right) P\left(\xi_2 \leq x/2\right) \\
&= \left(1 - e^{-x/2}\right)^2.
\end{aligned}$$

□

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 $\{q_i \leq \gamma\}$, by picking $F_n(x)$ to be the empirical distribution function of the $q_{i,}$ we see that

$$\{q_i \leq \gamma\} = \{F_n(q_i) \leq F_n(\gamma)\} = \{\frac{j}{n} \leq \gamma^*\}$$

for some $1 \leq j \leq n$. 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.11 shows that shows that uniformly in ν

$$Q_n(\gamma_0 + \nu/n) = -G_n^*(\nu) + 2\delta' R_n^*(\nu) + o_p(1) \tag{54}$$

where for $\nu > 0$

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

and

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

While Lemma A.11 assumed that $\alpha > 0$, it can be shown that (54) continues to hold when $\alpha = 0$.

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

$$E\left(N_n(\nu_1)N_n(\nu_2)\right) = D_n(\nu_1 \wedge \nu_2).$$

Since $\delta' R_n^*(\nu)$ is a mean-zero Gaussian process with covariance kernel $\sigma^2 G_n^*(\nu_1 \wedge \nu_2)$, the restriction of $\delta' R_n^*(\nu)$ to the positive integers has the same distribution as $N_n(\nu)$.

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

It follows from (54) that

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

We conclude that

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

This shows that $P\left(LR_n(\gamma_0) \geq x\right) \leq P\left(\xi \geq x\right) + o(1)$, as stated. □

References

[1] Bai, J. (1997): "Estimation of a change point in multiple regression models," *Review of Economics and Statistics*, 551-563.

[2] 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.

[3] Chan, K.S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *The Annals of Statistics*, 21, 520-533..

[4] 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.

[5] Chu, C.S. and H. White (1992): "A direct test for changing trend," *Journal of Business and Economic Statistics*, 10, 289-299.

[6] Davidson, J. (1994): *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford: Oxford University Press.

[7] Doukhan, P., P. Massart and E. Rio (1994): "Invariance principles for absolutely regular empirical processes," *Annales de l'Institut H. Poincare,* forthcoming.

[8] Dufour, J.M. (1997): "Some impossibility theorems in econometrics with applications to structural variables and dynamic models," *Econometrica*, 65, 1365-1387.

[9] Dumbgen, L. (1991): "The asymptotic behavior of some nonparametric change point estimators," *The Annals of Statistics*, 19, 1471-1495.

[10] Durlauf, S.N. and P.A. Johnson (1995): "Multiple regimes and cross-country growth behavior," *Journal of Applied Econometrics*, 10, 365-384.

[11] Hall, P. and C.C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

[12] Hansen, B.E. (1991): "Strong laws for dependent heterogeneous processes," *Econometric Theory,* 7, 213-221.

[13] Hansen, B.E. (1992a): "Erratum: Strong laws for dependent heterogeneous processes," *Econometric Theory,* 8, 421-422.

[14] Hansen, B.E. (1992b): "Tests for parameter instability in regressions with I(1) processes," *Journal of Business and Economic Statistics,* 10, 321-335.

[15] Hansen, B.E. (1996): "Inference when a nuisance parameter is not identified under the null hypothesis," *Econometrica*, 64, 413-430.

[16] Hansen, B.E. (1997): "Testing for structural change in conditional models," Boston College working paper No. 310.

[17] Hardle, W. and O. Linton (1994): "Applied Nonparametric Methods," in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics, Vol. IV*, pp. 2295-2339. Amsterdam: Elsevier Science.

[18] McLeish, D.L. (1975): "A maximal inequality and dependent strong laws," *Annals of Probability,* 3, 829-839.

[19] Pham, T.D. and L.T. Tran (1985): "Some mixing properties of time series models," *Stochastic Processes and Their Applications*, 19, 297-303.

[20] Picard, D. (1985): "Testing and estimating change-points in time series," *Advances in Applied Probability*, 17, 841-867.

[21] Pollard, D. (1990): *Empirical Processes: Theory and Applications. CBMS Conference Series in Probability and Statistics*, Vol. 2, Institute of Mathematical Statistics.

[22] Potter, S.M. (1995): "A nonlinear approach to U.S. GNP," *Journal of Applied Econometrics*, 2, 109-125.

[23] Silverman (1986): *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

[24] Tong, H. (1983): *Threshold Models in Non-linear Time Series Analysis. Lecture Notes in Statistics*, 21, Berlin: Springer.

[25] Volkonskii, V.A. and Y.A. Rozanov (1959): "Some limit theorems for random functions, Part I," *Theory of Probability and its Application*, 4, 178-197.

[26] Yao, Y.-C. (1987): "Approximating the distribution of the ML estimate of the change-point in a sequence of independent r.v.'s," *The Annals of Statistics*, 3, 1321-1328.

Figure 1: Asymptotic Density of Threshold Estimator

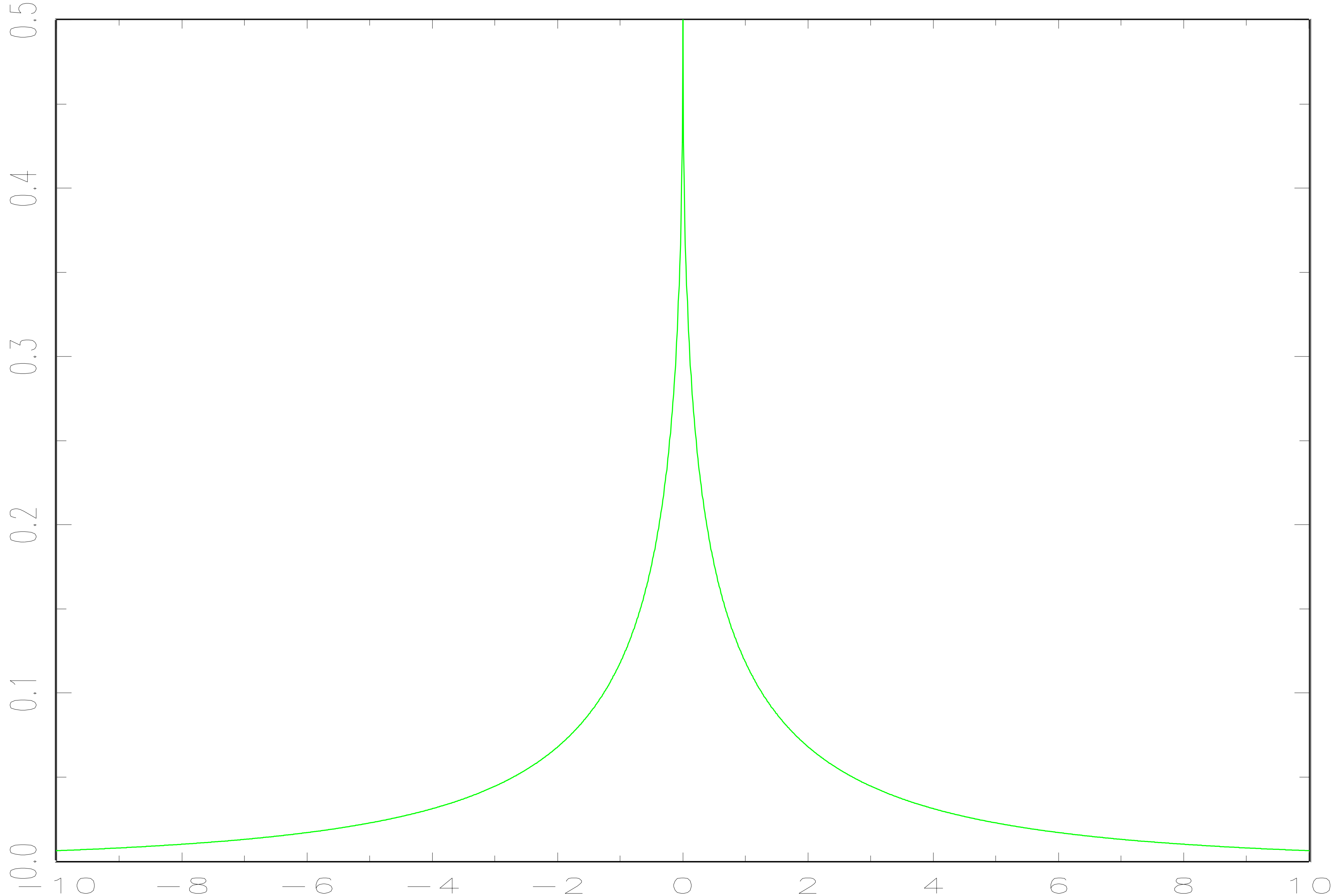

Figure 2: First Sample Split
Confidence Interval Construction for Threshold

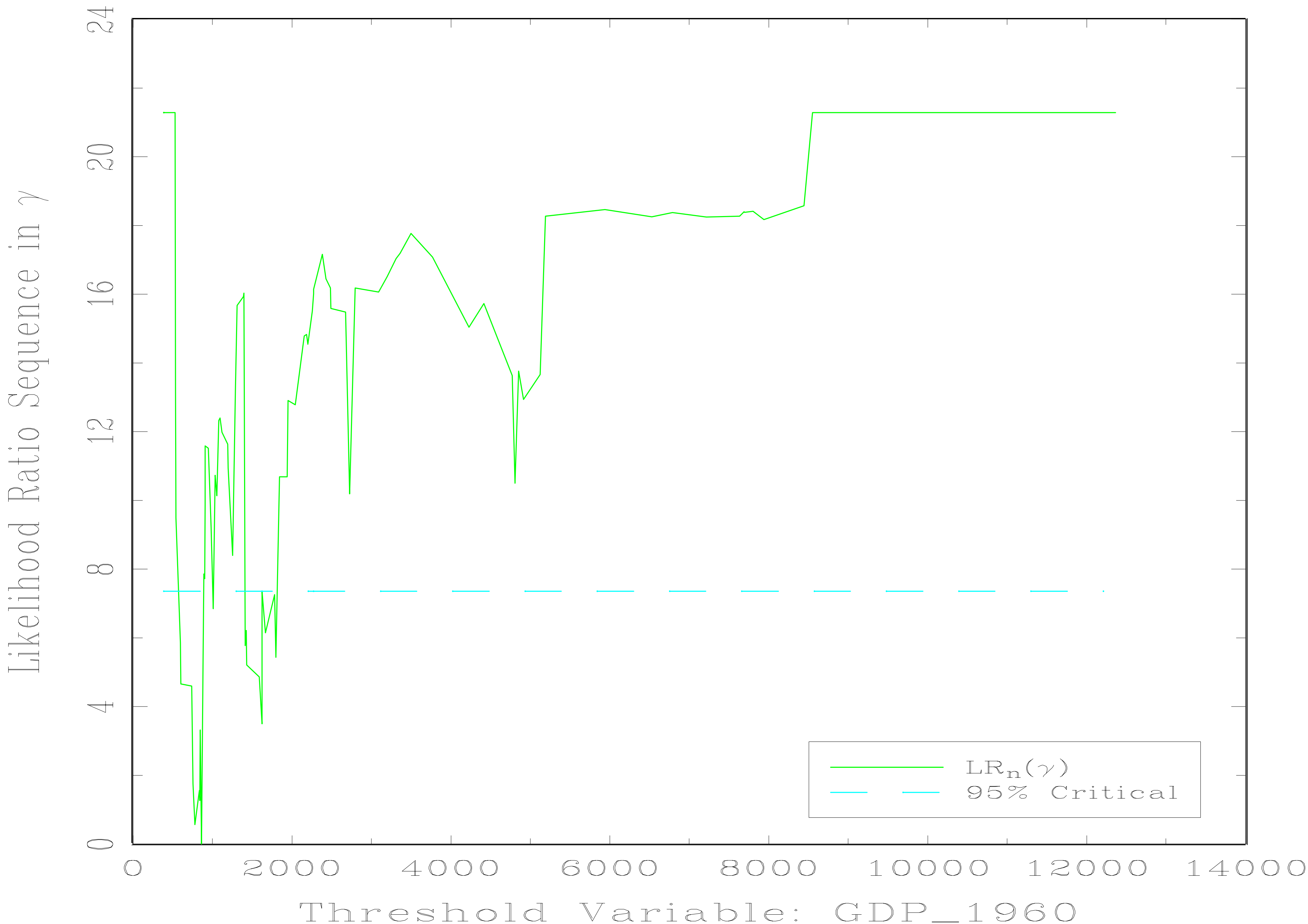

Figure 3: Second Sample Split
Confidence Interval Construction for Threshold