Economics Letters 259 (2026) 112791

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

New critical values for likelihood ratio inference of threshold regression ☆

Ping Yu

HKU Business School, The University of Hong Kong, Pokfulam Road, Hong Kong

ARTICLE INFO

JEL classification:
C22
C24

Keywords:
Threshold regression
Likelihood ratio
Convex hull
Critical value

ABSTRACT

This paper shows that the critical values for likelihood ratio inference of the threshold point in Hansen (2000) are too large if we restrict the confidence set as an interval, which can partially explain why Hansen's confidence interval is conservative when the threshold effect is not too small. We provide appropriate critical values and show that different from conventional critical values, these new critical values are invariant to the structural change in the error and covariate distributions.

1. Introduction

Threshold regression (TR), as a parsimonious model of nonlinear relationships between a response and some covariates, is very popular in current practice of econometrics; see Hansen (2011) for an excellent review of applications in time series, cross sections and panel data. The TR model usually assumes

$$\begin{aligned} y_i &= \mathbf{x}_i'\beta_1 1\left(q_i \le \gamma\right) + \mathbf{x}_i'\beta_2 1\left(q_i > \gamma\right) + \varepsilon_i \\ &= \mathbf{x}_i'\beta_2 + \mathbf{x}_i'\delta_n 1\left(q_i \le \gamma\right) + \varepsilon_i, \end{aligned} \tag{1}$$

where y_i is the dependent variable or the response, q_i is the threshold variable which is used to split the sample, $\mathbf{x}_i = \left(1, x_i', q_i\right)' \in \mathbb{R}^k$ is the set of covariates and may include q_i as a component, ε_i is the error term and satisfies $\mathbb{E}\left[\varepsilon_i|\mathcal{F}_{i-1}\right] = 0$ with $\mathcal{F}_{i-1}$ being the sigma field generated by $\left\{x_{i-j}, q_{i-j}, \varepsilon_{i-1-j}|j \ge 0\right\}$, and the parameter of interest is $\theta = \left(\gamma, \beta'\right)'$ with $\beta = \left(\beta_1', \beta_2'\right)'$, or equivalently, $\theta = \left(\gamma, \beta_2', \delta_n'\right)'$ with $\delta_n = \beta_1 - \beta_2$ being the threshold effect in conditional mean of y_i. Note here that we use subscript n in δ_n to emphasize the dependence of $\beta_1 - \beta_2$ on n. This model is similar to the linear regression except that the regression coefficients depend on whether the threshold variable q crosses the threshold point γ.

Because $\mathbb{E}\left[\varepsilon_i|\mathcal{F}_{i-1}\right] = 0$, we can estimate θ based on least squares. Specifically, θ is estimated by minimizing the following objective function,

$$S_n(\theta) = \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\beta_1 1\left(q_i \le \gamma\right) - \mathbf{x}_i'\beta_2 1\left(q_i > \gamma\right)\right)^2.$$

Denote the least squares estimator (LSE) of θ as $\hat{\theta} = \left(\hat{\gamma}, \hat{\beta}_1', \hat{\beta}_2'\right)'$. Often, a two-step procedure is used to obtain $\hat{\theta}$. First, given γ, run least squares on the data with $q_i \le \gamma$ and $q_i > \gamma$ separately to obtain $\hat{\beta}_1(\gamma)$ and $\hat{\beta}_2(\gamma)$. Second, minimize the concentrated objective function

$$S_n(\gamma) = \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i'\hat{\beta}_1(\gamma) 1\left(q_i \le \gamma\right) - \mathbf{x}_i'\hat{\beta}_2(\gamma) 1\left(q_i > \gamma\right)\right)^2$$

to obtain $\hat{\gamma}$ and set $\hat{\beta}_\ell = \hat{\beta}_\ell(\hat{\gamma})$, $\ell = 1, 2$. The threshold effect δ_n is estimated by $\hat{\delta} = \hat{\beta}_1 - \hat{\beta}_2$. $\hat{\gamma}$, $\hat{\beta}_\ell$ and $\hat{\delta}$ are all consistent. One main concern in the literature is to develop a confidence interval (CI) for γ.

Currently, the dominant or benchmark CI for γ in the literature is the likelihood ratio (LR)-CI of Hansen (2000). Such a CI relies on the shrinking-threshold-effect asymptotics borrowed from the structural change literature such as Picard (1985) and Bai (1997). In the simulations and empirical application of Hansen (2000), the LR-CI is often too conservative, i.e., it is wide and has a coverage larger than the nominal level. Hansen (2000) attributes this phenomenon to the insufficiency in using the asymptotic distribution of the LR statistic when the threshold effect is shrinking to approximate that when the threshold effect is fixed as assumed in Chan (1993). In this paper, we offer a new perspective on the conservativeness of the LR-CI. We demonstrate that the confidence set obtained by inverting the LR statistic is not an interval. Consequently, when restricting the confidence set to a standard interval — as is conventional — the critical values proposed in Hansen (2000) tend to be too large. We address this issue by deriving new critical values that are appropriate for the conventional LR-CI. Before presenting our results, we first review the original LR-CI framework introduced in Hansen (2000).

☆ I acknowledge support from the GRF of Hong Kong Government under Grant No. 106200228.
E-mail address: pingyu@hku.hk.

https://doi.org/10.1016/j.econlet.2025.112791
Received 20 June 2025; Received in revised form 11 December 2025; Accepted 15 December 2025
Available online 17 December 2025
0165-1765/

2. Review of Hansen (2000)

To review the asymptotic distribution of $\widehat{\gamma}$ in the framework of Hansen (2000), we first replicate his Assumption 1 as Assumption D below. Let $f(q)$ be the density function of q, and γ_0 be the true value of γ,

$$M(\gamma) = \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'1(q_i \le \gamma)\right], M = \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'\right],$$
$$D(\gamma) = \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'|q_i = \gamma\right], D = D(\gamma_0),$$
$$V(\gamma) = \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'\varepsilon_i^2|q_i = \gamma\right], V = V(\gamma_0).$$

Assumption D:

(1) $(x_i, q_i, \varepsilon_i)$ is strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty}\rho_m^{1/2} < \infty$.
(2) $\mathbb{E}\left[\varepsilon_i|\mathcal{F}_{i-1}\right] = 0$.
(3) $\mathbb{E}\left[\left|\mathbf{x}_i\right|^4\right] < \infty$ and $\mathbb{E}\left[\left|\mathbf{x}_i\varepsilon_i\right|^4\right] < \infty$.
(4) For all $\gamma \in \Gamma$, $\mathbb{E}\left[\left|\mathbf{x}_i\right|^4\left|\varepsilon_i\right|^4|q_i = \gamma\right] \le C$ and $\mathbb{E}\left[\left|\mathbf{x}_i\right|^4|q_i = \gamma\right] \le C$ for some $C < \infty$, and $f(\gamma) \le \overline{f} < \infty$, where Γ is the parameter space of γ.
(5) $f(\gamma)$, $D(\gamma)$, and $V(\gamma)$ are continuous at $\gamma = \gamma_0$.
(6) $\delta_n = cn^{-\varphi}$, with $c \neq 0$ and $\varphi \in (0, 1/2)$.
(7) $c'Dc > 0$, $c'Vc > 0$, and $f = f(\gamma_0) > 0$.
(8) $M > M(\gamma) > 0$ for all $\gamma \in \Gamma$.

Hansen (2000) provides detailed discussions on these assumption after his Assumption 1, so we only briefly mention some key points. First, Assumption D6 assumes that δ_n shrinks to zero but stays out of the contiguous neighborhood of $\delta_n = 0$ (i.e., $\delta_n = cn^{-1/2}$) so that γ can still be point identified. Second, Assumption D7 excludes the continuous threshold model discussed in Chan and Tsay (1998) and Hansen (2017). Third, Assumption D8 restricts Γ to be a proper subset of the support of q_i. In practice, we often set $\Gamma = \left[\underline{\gamma}, \overline{\gamma}\right]$, where $\underline{\gamma}$ and $\overline{\gamma}$ are the lower and upper $\epsilon\%$ quantiles of $\left\{q_i\right\}_{i=1}^n$. This guarantees that each regime contains at least $\epsilon\%$ of the whole dataset for some $\epsilon > 0$ (typically, 5, 10 or 15). Note that $S_n(\gamma)$ is constant on $[q_{(i)}, q_{(i+1)})$, where $\left\{q_{(i)}\right\}_{i=1}^n$ is the sorted (ascending) version of $\left\{q_i\right\}_{i=1}^n$. This is why we need only check $\gamma \in \Gamma_n$ to search for $\widehat{\gamma}$ in practice, where $\Gamma_n = \left\{q_i | q_i \in \Gamma\right\}$. In other words, $\widehat{\gamma}$ is taken as the left endpoint of the minimizing interval of $S_n(\gamma)$. Yu (2012, 2015) suggests to take the middle point of this interval as $\widehat{\gamma}$ to improve its finite-sample performance, but under Assumption D6, taking any point in this interval as $\widehat{\gamma}$ does not affect its asymptotic properties. Intuitively, this is because the convergence rate of $\widehat{\gamma}$ is slower than n, while the distance between $q_{(i)}$ and $q_{(i+1)}$ is $O\left(n^{-1}\right)$.

Under Assumption D, Theorem 1 of Hansen (2000) implies that

$$n\widehat{f}\frac{\left(\widehat{\delta}'\widehat{D}\widehat{\delta}\right)^2}{\widehat{\delta}'\widehat{V}\widehat{\delta}}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \arg\max_{v}\left[-\frac{|v|}{2} + B(v)\right], \tag{2}$$

where note that the population normalization factor $nf\frac{\left(\delta_n'D\delta_n\right)^2}{\delta_n'V\delta_n} = O\left(n^{1-2\varphi}\right)$, which matches the convergence rate of $\widehat{\gamma}$ in Hansen's Theorem 1, $\widehat{f}$, $\widehat{D}$ and $\widehat{V}$ are consistent estimators of f, D, and V, respectively, and $B(v) = B_1(-v)1(v \le 0) + B_2(v)1(v > 0)$ with $B_1(v)$ and $B_2(v)$ being two independent standard Brownian motions on $[0, \infty)$. The CI of γ can be constructed by inverting the t statistic in testing $H_0: \gamma = \gamma_0$ vs. $H_1: \gamma \neq \gamma_0$. Specifically, the $(1-\alpha)$ t-CI is

$$\left[\widehat{\gamma} - \frac{c_{\alpha/2}^t}{n\widehat{f}}\frac{\widehat{\delta}'\widehat{V}\widehat{\delta}}{\left(\widehat{\delta}'\widehat{D}\widehat{\delta}\right)^2}, \widehat{\gamma} + \frac{c_{\alpha/2}^t}{n\widehat{f}}\frac{\widehat{\delta}'\widehat{V}\widehat{\delta}}{\left(\widehat{\delta}'\widehat{D}\widehat{\delta}\right)^2}\right],$$

where c_α^t is the upper αth quantile of the distribution of $\arg\max_v\left[-\frac{|v|}{2} + B(v)\right]$ which is developed in Bhattacharya and Brockwell (1976). In the homoskedastic case where $\mathbb{E}\left[\varepsilon^2|\mathbf{x}\right] = \mathbb{E}\left[\varepsilon^2\right] = \sigma^2$, $\frac{\widehat{\delta}'\widehat{V}\widehat{\delta}}{\left(\widehat{\delta}'\widehat{D}\widehat{\delta}\right)^2}$ can be replaced by $\frac{\widehat{\sigma}^2}{\widehat{\delta}'\widehat{D}\widehat{\delta}}$, where $\widehat{\sigma}^2$ is a consistent estimator of σ^2.

Because γ cannot be identified when $\delta_n = 0$, following Dufour (1997), Hansen (2000) suggests to use the LR-CI to improve performances.[1] The LR statistic is defined as

$$LR_n(\gamma) = \frac{S_n(\gamma) - S_n\left(\widehat{\gamma}\right)}{\widehat{\eta}^2},$$

where $\widehat{\eta}^2 = \frac{\widehat{\delta}'\widehat{V}\widehat{\delta}}{\widehat{\delta}'\widehat{D}\widehat{\delta}}$ is a consistent estimator of

$$\eta^2 = \frac{\delta_n'V\delta_n}{\delta_n'D\delta_n} = \frac{c'Vc}{c'Dc},$$

It can be shown that

$$LR_n\left(\gamma_0\right) \xrightarrow{d} \xi, \tag{3}$$

where $\xi = \max_v\left[-|v| + 2B(v)\right] = 2\cdot\max_v\left[-\frac{|v|}{2} + B(v)\right]$. When $\mathbb{E}\left[\varepsilon^2|\mathbf{x}\right] = \mathbb{E}\left[\varepsilon^2\right] = \sigma^2$, $\eta^2 = \sigma^2$, and $\widehat{\eta}^2$ can be replaced by an estimator of σ^2 such as $\widehat{\sigma}^2 := S_n\left(\widehat{\gamma}\right)/n$. As a result, the $(1-\alpha)$ LR-CI is

$$\widehat{\Gamma}(1-\alpha) = \left\{\gamma : LR_n(\gamma) \le c_\alpha\right\},$$

where c_α is the upper αth quantile of the distribution of ξ which is

$$P(\xi \le x) = \left(1 - e^{-x/2}\right)^2 =: F_\xi(x). \tag{4}$$

Compared with the t-CI, the LR-CI does not need to estimate f and D in the homoskedastic case.

Because $LR_n(\gamma)$ behaves like a drifted Brownian motion, and the Brownian motion fluctuates rapidly (due to nowhere differentiability), $\widehat{\Gamma}(1-\alpha)$ is often a union of segments rather than an interval. Consequently, a common practice is to take the convex hull of $\widehat{\Gamma}(1-\alpha)$ as the CI for γ, denoted as $\text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}$. Because $LR_n(\gamma)$ is flat on $[q_{(i)}, q_{(i+1)})$, the convex hull of $\widehat{\Gamma}(1-\alpha)$ takes the form of $[q_{(i)}, q_{(j)})$ for some $i < j$. We refer to this CI as Hansen's CI. For comparison, the t-CI is always an interval. The disjointness of LR-CI is also observed in the structural change literature, e.g., Siegmund (1986, 1988), where the convex CI is also considered. Note that $P\left(\gamma_0 \in \text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}\right) \ge P\left(\gamma_0 \in \widehat{\Gamma}(1-\alpha)\right) \to 1-\alpha$, so the critical value c_α is too large for $\text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}$. We label $P\left(\gamma_0 \in \text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}\right)$ as the interval coverage, and $P\left(\gamma_0 \in \widehat{\Gamma}(1-\alpha)\right)$ as the actual coverage.

3. Genuine critical values in Hansen's framework

From the discussions in the last section, we know that the critical value c_α is too large if we use the convex hull $\text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}$ instead of $\widehat{\Gamma}(1-\alpha)$ as the CI for γ. A natural question is what the genuine critical value should be if $\text{conv}\left\{\widehat{\Gamma}(1-\alpha)\right\}$ is employed. The following theorem answers this question.

Theorem 1. *Under Assumption D,*

$$P\left(\gamma_0 \in \mathit{conv}\left\{\gamma : LR_n(\gamma) \le x\right\}\right) \to P\left(\max\left[\xi_1, \xi_2\right] - \min\left[\xi_1, \xi_2\right] \le \frac{x}{2}\right) = 1 - e^{-\frac{x}{2}} =: p(x),$$

where ξ_1 and ξ_2 are independent and both follow $Exp(1)$, and $Exp(a)$ denotes the exponential distribution with mean $a > 0$.

[1] Another advantage of the LR-CI over the t-CI is that the former is shorter than the latter both asymptotically and in finite samples as shown in Eo and Morley (2015) in the structural change context. Technically, this is because the t-test is less powerful than the LR test.

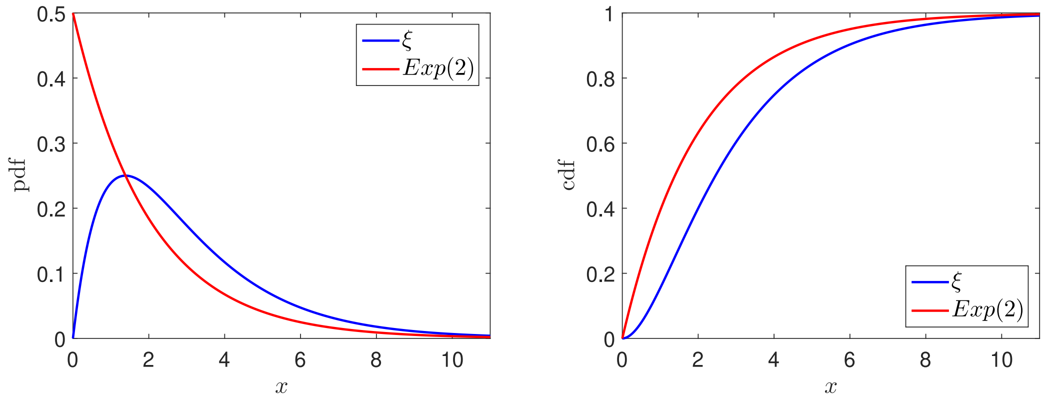

Fig. 1. Comparison between ξ and Exp(2).

Table 1
Comparison between two types of critical values.

$1-\alpha$	.80	.85	.90	.925	.95	.975	.99
$F_\xi^{-1}(1-\alpha)$	4.497	5.101	5.939	6.528	7.352	8.751	10.592
$p(c_\alpha)$	0.894	0.922	0.949	0.962	0.975	0.987	0.995
$p^{-1}(1-\alpha)$	3.219	3.794	4.605	5.181	5.991	7.378	9.210

Note that the event $\left\{\gamma_0 \in \text{conv}\left\{\gamma : LR_n(\gamma) \le x\right\}\right\}$ is equivalent to $\{LR_n(\gamma_{\_}) \le x$ for some $\gamma_{\_} \le \gamma_0$ and $LR_n(\gamma_+) \le x$ for some $\gamma_+ \ge \gamma_0\}$, or equivalently, $\{\inf_{\gamma\le\gamma_0} LR_n(\gamma) \le x$ and $\inf_{\gamma\ge\gamma_0} LR_n(\gamma) \le x\}$. It turns out that the event $\left\{\inf_{\gamma\le\gamma_0} LR_n(\gamma) \le x\right\}$ converges to $\left\{2\left[\max\left(\xi_1,\xi_2\right)-\xi_1\right] \le x\right\}$ and the event $\left\{\inf_{\gamma\ge\gamma_0} LR_n(\gamma) \le x\right\}$ converges to $\left\{2\left[\max\left(\xi_1,\xi_2\right)-\xi_2\right] \le x\right\}$. So the event $\left\{\gamma_0 \in \text{conv}\left\{\gamma : LR_n(\gamma) \le x\right\}\right\}$ converges to

$$\begin{aligned}&\left\{2\left[\max\left(\xi_1,\xi_2\right)-\xi_1\right] \le x\right\} \cap \left\{2\left[\max\left(\xi_1,\xi_2\right)-\xi_2\right] \le x\right\}\\ =\ &\left\{2\left[\max\left(\xi_1,\xi_2\right)-\min\left(\xi_1,\xi_2\right)\right] \le x\right\} = \left\{2\left|\xi_1-\xi_2\right| \le x\right\},\end{aligned}$$

while $|\xi_1-\xi_2| \sim Exp(1)$ by symmetry of ξ_1 and ξ_2 and memorylessness of exponential distributions. It follows that $2\left|\xi_1-\xi_2\right| \sim Exp(2)$, where note that $1-e^{-\frac{x}{2}}$ in Theorem 1 is the cdf of $Exp(2)$.

For comparison, in Hansen (2000),

$$\begin{aligned}P\left(LR_n\left(\gamma_0\right) \le x\right) \to P(\xi \le x) &= P\left(\max\left[\xi_1,\xi_2\right] \le \frac{x}{2}\right)\\ &\le P\left(\max\left[\xi_1,\xi_2\right]-\min\left[\xi_1,\xi_2\right] \le \frac{x}{2}\right),\end{aligned}$$

where ξ is defined in (3). Now, $p^{-1}(x) = -2\log(1-x)$ while $F_\xi^{-1}(x) = -2\log\left(1-\sqrt{x}\right)$, where F_ξ is defined in (4). Fig. 1 shows the difference between these two distributions. Because $Exp(2)$ has a much thinner right tail than ξ, its critical values are significantly smaller and $p\left(F_\xi^{-1}(1-\alpha)\right) = p\left(c_\alpha\right) > 1-\alpha$ as shown in Table 1.

One may wonder why we need not consider the difference between these two types of CIs for regular parameters. To explain the reason, consider the LR-CI for μ, where μ is the mean of a normal distribution with unit variance, i.e., we observe iid $X_i \sim N(\mu,1)$, $i=1,\dots,n$. The LR statistic is

$$L_n(\mu) = 2\log\frac{\mathcal{L}_n(\hat{\mu})}{\mathcal{L}_n(\mu)} = n\left(\overline{X}-\mu\right)^2,$$

where $\hat{\mu} = \overline{X}$ is the MLE of μ, and $\mathcal{L}_n(\mu)$ is the likelihood function for μ. As a result, $\left\{\mu : L_n(\mu) \le c_\alpha\right\} = \left[\overline{X}-n^{-1/2}c_\alpha^{1/2}, \overline{X}+n^{-1/2}c_\alpha^{1/2}\right]$, exactly the same as the t-CI, where c_α is the upper αth quantile of $\chi^2(1)$. Why do we need only to check whether $L_n\left(\mu_0\right) \le c_\alpha$ to determine whether the CI covers μ_0? This is because $L_n\left(\mu_0\right) \le c_\alpha$ is equivalent to $\mu_0 \in \left\{\mu : L_n(\mu) \le c_\alpha\right\}$ when $\left\{\mu : L_n(\mu) \le c_\alpha\right\}$ is an interval. On the contrary, for the LR-CI of γ in TR, the event $\left\{LR_n\left(\gamma_0\right) \le c_\alpha\right\}$ is smaller than $\left\{\gamma_0 \in \text{conv}\left\{\gamma : LR_n(\gamma) \le c_\alpha\right\}\right\}$. Fig. 2 illustrates this point. From the two upper graphs, we can see if $L_n\left(\mu_0\right) \le c_\alpha$, then the CI covers μ_0, and vice versa. On the contrary, from the two lower graphs, although $LR_n\left(\gamma_0\right) \le c_\alpha$ implies the CI covers γ_0, $LR_n\left(\gamma_0\right) > c_\alpha$ does not imply the convex CI excludes γ_0.

3.1. Extension

Assumption D5 assumes that $D(\gamma)$, and $V(\gamma)$ are continuous at $\gamma = \gamma_0$. This essentially assumes that the conditional distributions of the error ε_i and covariates $\mathbf{x}_i$ given $q_i = \gamma$ are continuous at γ_0. When this assumption fails, i.e., there are structural changes in the error and covariate distributions across the two regimes, Bai (1997) develops the asymptotic distribution of $\hat{\gamma}$ in the structural change model where q_i is the time index. To state the counterpart of Theorem 1 in this generalized setup of threshold regression, we first modify the notations and Assumption D as follows. Define

$$\begin{aligned}D_\pm(\gamma) &= \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'|q_i=\gamma\pm\right], D_\pm = D_\pm\left(\gamma_0\right),\\ V_\pm(\gamma) &= \mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'\varepsilon_i^2|q_i=\gamma\pm\right], V_\pm = V_\pm\left(\gamma_0\right),\end{aligned}$$

where $D_+(\gamma)$ is understood as the right limit of $\mathbb{E}\left[\mathbf{x}_i\mathbf{x}_i'|q_i=\cdot\right]$ at γ, and $D_-(\gamma)$ and $V_\pm(\gamma)$ are similarly understood.

Assumption D′:

Assumptions 1, 2, 3, 6 and 8 are the same as in Assumption D.

4. For all $\gamma \in \Gamma$, $\mathbb{E}\left[\left|\mathbf{x}_i\right|^4\left|\varepsilon_i\right|^4|q_i=\gamma\pm\right] \le C$ and $\mathbb{E}\left[\left|\mathbf{x}_i\right|^4|q_i=\gamma\pm\right] \le C$ for some $C<\infty$, and $f(\gamma) \le \overline{f} < \infty$.
5. $f(\gamma)$, $D_\pm(\gamma)$, and $V_\pm(\gamma)$ are continuous at $\gamma=\gamma_0$.
7. $c'D_\pm c > 0$, $c'V_\pm c > 0$, and $f = f\left(\gamma_0\right) > 0$.

Redefine

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

where $\hat{\eta}^2 = \frac{\hat{\delta}'\hat{V}_-\hat{\delta}}{\hat{\delta}'\hat{D}_-\hat{\delta}}$ is a consistent estimator of

$$\eta^2 = \frac{\delta_n'V_-\delta_n}{\delta_n'D_-\delta_n} = \frac{c'V_-c}{c'D_-c}.$$

If the model is homoskedastic in the left regime, i.e., $\mathbb{E}\left[\varepsilon^2|x,q\right] = \sigma_-^2$ for $q \le \gamma_0$, then $\eta^2 = \sigma_-^2$, and $\hat{\eta}^2$ can be replaced by

$$\hat{\sigma}_-^2 := S_n^-(\hat{\gamma})/n_1,$$

where $S_n^-(\hat{\gamma}) = \sum_{i=1}^n\left(y_i-\mathbf{x}_i'\hat{\beta}_1\right)^2 1\left(q_i\le\hat{\gamma}\right)$, and $n_1 = \sum_{i=1}^n 1\left(q_i\le\hat{\gamma}\right)$.

The following corollary states the asymptotic distribution of $\hat{\gamma}$, as well as the asymptotic coverages of the CI by inverting LR_n and its convex hull.

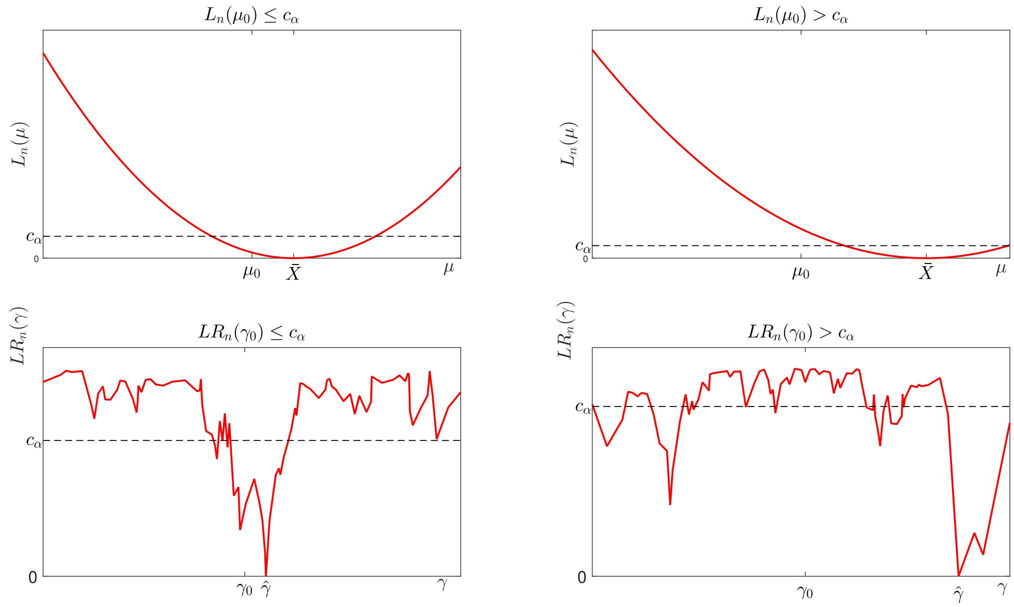

Fig. 2. Comparison of the CI construction for μ and γ.

Corollary 1. *Under Assumption D′,*

$$n\hat f \frac{\left(\hat\delta' \hat D_- \hat\delta\right)^2}{\hat\delta' \hat V_- \hat\delta}\left(\hat\gamma - \gamma_0\right) \xrightarrow{d} \arg\max_{v} \begin{cases} -\frac{1}{2}|v| + B_1(-v), & \text{if } v \le 0, \\ -\frac{1}{2}\phi v + \sqrt{\omega} B_2(v), & \text{if } v > 0. \end{cases}$$

$$P\left(\gamma_0 \in \left\{\gamma : LR_n(\gamma) \le x\right\}\right) \to P\left(\max\left[\xi_1, \xi_2(\phi,\omega)\right] \le \frac{x}{2}\right) = (1-e^{-x})(1-e^{-x\phi/\omega}),$$

and

$$P\left(\gamma_0 \in conv\left\{\gamma : LR_n(\gamma) \le x\right\}\right) \to P\left(\max\left[\xi_1, \xi_2(\phi,\omega)\right] - \min\left[\xi_1, \xi_2(\phi,\omega)\right] \le \frac{x}{2}\right) = p(x) = 1 - e^{-\frac{x}{2}},$$

where $\hat D_-$ and $\hat V_-$ are consistent estimators of D_- and V_-, respectively, $\phi = \frac{c' D_+ c}{c' D_- c}$, $\omega = \frac{c' V_+ c}{c' V_- c}$, $\xi_1 \sim Exp(1)$ and $\xi_2(\phi,\omega) \sim Exp(\omega/\phi)$ are independent.

It is interesting to notice that although the coverage of the traditional CI, $P\left(\gamma_0 \in \left\{\gamma : LR_n(\gamma) \le x\right\}\right)$, depends on the nuisance parameter ω/ϕ, the coverage of the convex CI, $P\left(\gamma_0 \in \text{conv}\left\{\gamma : LR_n(\gamma) \le x\right\}\right)$, does not. In other words, we can use the same critical values $p^{-1}(1-\alpha)$ in Table 1 regardless of the value of ω/ϕ if the convex CI is employed. In this sense, the convex CI combined with the critical values $p^{-1}(1-\alpha)$ are more natural and convenient to use in practice.

4. Simulations

In this section, we re-examine the simulation studies in Hansen (2000) to illustrate our theoretical results above, where the data are iid, $\mathbf{x}_i = (1, z_i)'$ with $z_i = q_i$ or $z_i = x_i$, $x_i \sim N(0,1)$, $q_i \sim N(2,1)$, $\varepsilon_i \sim N(0,1)$, $\beta_2 = 0$, $\delta_n = (0, \delta_{20})'$, and $\gamma_0 = 2$. We label the case with $z_i = q_i$ as DGP1 and $z_i = x_i$ as DGP2. We will concentrate on the scenarios where Hansen's CIs overcover. Specifically, we set $\delta_{20} = 0.5, 1$ (1,2) in DGP1 (DGP2) and $n = 50, 100, 500$; the replication number is set as 10,000 to improve the precision of simulation. As for why δ_{20} in DGP2 is set as double of that in DGP1, the reason is as follows. The normalization factor in (2) converges to

$$nf \frac{\left(\delta_n' D \delta_n\right)^2}{\delta_n' V \delta_n} = n \frac{\delta_n' D \delta_n}{\sqrt{2\pi}} = \frac{n\delta_{20}^2 D_{22}}{\sqrt{2\pi}},$$

where $D = \mathbb{E}\left[\begin{pmatrix} 1 \\ q \end{pmatrix}(1, q) \middle| q = 2\right] = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$ in DGP1 and $D = \mathbb{E}\left[\begin{pmatrix} 1 \\ x \end{pmatrix}(1, x)\right] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ in DGP2, and $V = D$.[2] Because D_{22} in DGP1 is four times of D_{22} in DGP2, we set δ_{20} in DGP1 as half of δ_{20} in DGP2 to ensure $\delta_{20}^2 D_{22}$ the same in the two DGPs.

The coverages and lengths of three kinds of 90% LR-CIs for γ are summarized in Table 2. From Table 2, we can draw the following conclusions. First, as expected, the actual coverages using Hansen's critical values (upper row) and the interval coverages using our new critical values (lower row) are lower than the coverages of Hansen's CIs (middle row). Second, when δ_{20} gets larger, the gap between the actual coverage and the interval coverage using Hansen's critical values gets smaller. Actually, when δ_{20} is large, $\hat\Gamma(1-\alpha)$ tends to be an interval such that the two coverages are close. Third, as expected, when Hansen's CIs overcover, our new convex CIs tend to overcover less and match the nominal level better in most cases. Fourth, when $\sqrt{n}\delta_{20}$ is small, our new convex CIs may undercover a little bit. This is because our asymptotics require $\sqrt{n}\delta_{20} \to \infty$, and a small $\sqrt{n}\delta_{20}$ value may invalidate the asymptotic arguments. Fifth, all CIs are shorter as δ_{20} gets larger and/or n gets larger; our new convex CIs (lower row) are shorter than Hansen's CIs (upper row) because smaller critical values are used.

5. Empirical application

In this section, we re-analyze the dataset of Hansen (2000) using our new critical values. This dataset was used in Durlauf and Johnson (1995) to check whether the growth pattern depends on the initial conditions. The growth theory with multiple equilibria motivates the following threshold regression model (see Box I).

For each country i, $\left(\frac{Y}{L}\right)_{i,t}$ is the real GDP per member of the population aged 15–64 in year t, $\left(\frac{I}{Y}\right)_i$ is the investment to GDP ratio, n_i is the growth rate of the working-age population, and S_i is the fraction of working-age population enrolled in secondary schools. The variables

[2] The calculation here also confirms the result of Dufour (1997) in our setup, i.e., when δ_{20} is close to zero, the length of the t-CI diverges to infinity because the normalization factor shrinks to zero.

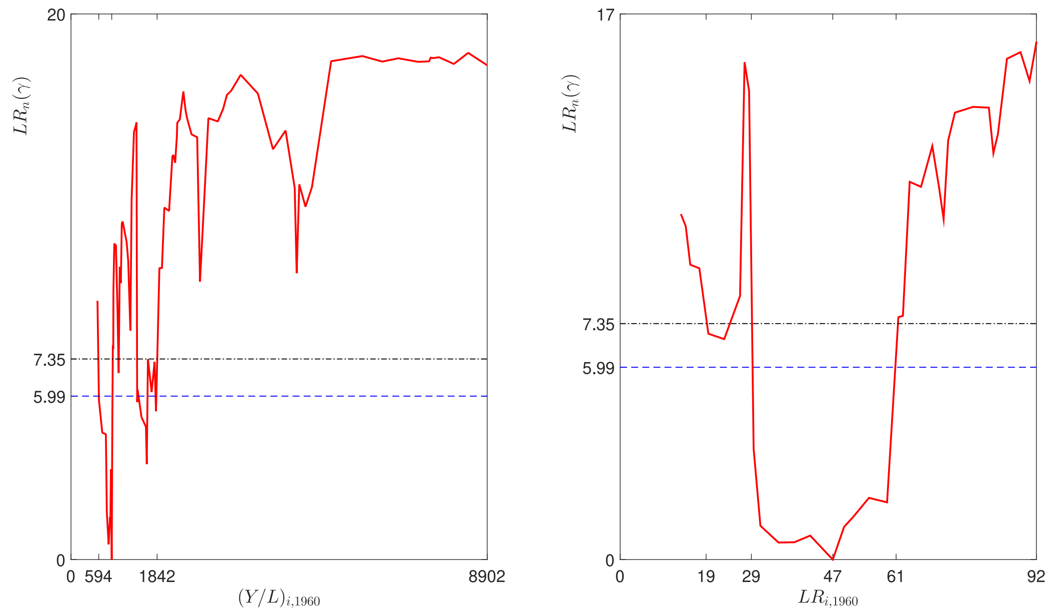

Fig. 3. Confidence interval construction for γ in two stages.

Table 2
Coverage and length of 90% LR-CI.

$n \searrow$	$\delta_{20} \longrightarrow$	$z = q$		$z = x$	
		0.5	1	1	2
Coverage	50	0.851	0.940	0.903	0.957
		0.909	0.949	0.931	0.962
		0.840	0.913	0.879	0.938
	100	0.887	0.956	0.929	0.970
		0.925	0.960	0.945	0.974
		0.868	0.930	0.905	0.949
	500	0.938	0.967	0.946	0.975
		0.955	0.970	0.958	0.979
		0.899	0.942	0.928	0.957
Length	50	1.597	0.530	1.097	0.350
		1.301	0.409	0.881	0.304
	100	1.287	0.168	0.578	0.169
		0.989	0.140	0.465	0.146
	500	0.098	0.024	0.091	0.033
		0.074	0.020	0.075	0.029

Note: For coverage, the upper and middle rows are actual and interval coverages using Hansen's critical values, and the lower row is interval coverage using new critical values. For length, the upper and lower rows are the average lengths corresponding to the CIs in the middle and lower rows of coverage, respectively.

$$\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960} = \begin{cases} \beta_{10} + \beta_{11}\ln\left(\frac{Y}{L}\right)_{i,1960} + \beta_{12}\ln\left(\frac{I}{Y}\right)_i + \beta_{13}\ln\left(n_i+g+\delta\right) + \beta_{14}\ln S_i + \varepsilon_i, & \text{if } q_i \le \gamma; \\ \beta_{20} + \beta_{21}\ln\left(\frac{Y}{L}\right)_{i,1960} + \beta_{22}\ln\left(\frac{I}{Y}\right)_i + \beta_{23}\ln\left(n_i+g+\delta\right) + \beta_{24}\ln S_i + \varepsilon_i, & \text{if } q_i > \gamma. \end{cases}$$

Box I.

not indexed by t are annual averages over the period 1960–1985. Following Durlauf and Johnson (1995), we set $g+\delta = 0.05$. As suggested in Hansen (2000), we will check two possible threshold variables; the first one is $\ln\left(\frac{Y}{L}\right)_{i,1960}$ and the second one is the adult literacy rate in 1960, $LR_{i,1960}$. Also following Hansen (2000), we will consider only the heteroskedastic-consistent procedures.

In the first stage, we use $\ln\left(\frac{Y}{L}\right)_{i,1960}$ as q_i. The middle-point LSE of γ is \$871, which generates the same sample splitting as the left-endpoint LSE \$863. We employ the critical value in Table 1 to construct the CI for γ. The 95% critical value is 5.991, much smaller than 7.352 as suggested in Hansen (2000). If we use the kernel method to estimate η^2, then both Hansen's critical value and ours generate the same CI for γ, [\$594, \$1842]. The left graph of Fig. 3 illustrates these two CIs. Obviously, this CI is quite wide since it covers 40 out of the 96 countries in the sample.

In the second stage, we use $LR_{i,1960}$ as q_i and apply threshold regression to the right regime in the first stage defined by $\ln\left(\frac{Y}{L}\right)_{i,1960} > 871$. The middle-point LSE of γ is 47%. We will still use the critical value in Table 1 to construct the CI for γ. It turns out that Hansen's CI is [19%, 61%), while ours is [29%, 61%), which is much shorter. Hansen's CI covers 19 while ours covers 14 out of the 78 countries in this subsample, so our shorter CI indicates less uncertainty in the sample splitting using $LR_{i,1960}$. The right graph of Fig. 3 illustrates these two CIs.

As commented in Hansen (2000), his CI is sufficiently wide that there is considerable uncertainty regarding the threshold value. This paper shows that his CI can be much shortened by using an appropriate critical value.

Appendix. Proofs

We here collect some notations for future reference. $a_n = n^{1-2\varphi}$. $\rightsquigarrow$ signifies weak convergence over a metric space. $u = \left(u_1', u_2'\right)'$ and v are local parameters for β and γ. $\overset{d}{=}$ denotes equality in distribution.

Proof of Theorem 1. Note that when $\gamma = \gamma_0 + a_n^{-1} v$, the true γ value is still γ_0. Re-write $LR_n(\gamma)$ as

$$LR_n(\gamma) = \frac{\left[S_n\left(\gamma, \hat{\beta}(\gamma)\right) - S_n\left(\gamma_0, \beta_0\right)\right] - \left[S_n\left(\hat{\gamma}, \hat{\beta}\right) - S_n\left(\gamma_0, \beta_0\right)\right]}{\hat{\eta}^2}.$$

Because

$$\begin{aligned} & S_n\left(\gamma_0 + a_n^{-1} v, \beta_0 + n^{-1/2} u\right) - S_n\left(\gamma_0, \beta_0\right) \\ \rightsquigarrow\ & u_1' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q \le \gamma_0)}\right] u_1 + u_2' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q > \gamma_0)}\right] u_2 - 2u_1' W_1 - 2u_2' W_2 + \mu |v| + 2\sqrt{\lambda} B(v), \end{aligned}$$

where $\mu = c' D c f$ and $\lambda = c' V c f$, we have

$$S_n\left(\gamma, \hat{\beta}(\gamma)\right) - S_n\left(\gamma_0, \beta_0\right) \rightsquigarrow -W_1' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q \le \gamma_0)}\right]^{-1} W_1 - W_2' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q > \gamma_0)}\right]^{-1} W_2 + \mu |v| + 2\sqrt{\lambda} B(v),$$

by taking minimum with respect to u, and

$$S_n\left(\hat{\gamma}, \hat{\beta}\right) - S_n\left(\gamma_0, \beta_0\right) \rightsquigarrow -W_1' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q \le \gamma_0)}\right]^{-1} W_1 - W_2' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q > \gamma_0)}\right]^{-1} W_2 + \min_v \left\{\mu |v| + 2\sqrt{\lambda} B(v)\right\}$$

by taking minimum with respect to both u and v. As result,

$$\begin{aligned} LR_n\left(\gamma_0 + a_n^{-1} v\right) \rightsquigarrow LR_\infty(v) &:= \frac{1}{\eta^2}\left[\mu |v| + 2\sqrt{\lambda} B(v) - \min_v \left\{\mu |v| + 2\sqrt{\lambda} B(v)\right\}\right] \\ &\overset{d}{=} \frac{1}{\eta^2}\left[\mu |v| + 2\sqrt{\lambda} B(v) + \sup_v \left\{-\mu |v| + 2\sqrt{\lambda} B(v)\right\}\right]. \end{aligned}$$

The interval coverage converges to

$$P\left(\inf_{v \le 0} LR_\infty(v) \le c_\alpha \text{ and } \inf_{v \ge 0} LR_\infty(v) \le c_\alpha\right),$$

where c_α is the upper αth quantile of $\sup_v \{-|v| + 2B(v)\}$. Making the change-of-variables $v = \left(\frac{\lambda}{\mu^2}\right) r$ and noting the distributional equality $B\left(a^2 r\right) = aB(r)$, we have

$$\begin{aligned} \inf_{v \le 0} LR_\infty(v) &= \frac{1}{\eta^2}\left[\inf_{r \le 0}\left\{\mu \left|\frac{\lambda}{\mu^2} r\right| + 2\sqrt{\lambda} B\left(\frac{\lambda}{\mu^2} r\right)\right\} - \inf_r \left\{\mu \left|\frac{\lambda}{\mu^2} r\right| + 2\sqrt{\lambda} B\left(\frac{\lambda}{\mu^2} r\right)\right\}\right] \\ &= \frac{\lambda}{\eta^2 \mu}\left[\inf_{r \le 0}\{|r| + 2B(r)\} - \inf_r \{|r| + 2B(r)\}\right] \\ &= \inf_{v \le 0}\{|v| + 2B(v)\} - \inf_v \{|v| + 2B(v)\}, \end{aligned}$$

where the last equality is from the fact that $\eta^2 = \lambda/\mu$; similarly,

$$\inf_{v \ge 0} LR_\infty(v) = \inf_{v \ge 0}\{|v| + 2B(v)\} - \inf_v \{|v| + 2B(v)\}.$$

As a result,

$$\begin{aligned} p\left(c_\alpha\right) &= P\left(\inf_{v \le 0}\{|v| + 2B(v)\} + \xi \le c_\alpha \text{ and } \inf_{v \ge 0}\{|v| + 2B(v)\} + \xi \le c_\alpha\right) \\ &= P\left(-2\xi_1 + 2\max\left[\xi_1, \xi_2\right] \le c_\alpha \text{ and } -2\xi_2 + 2\max\left[\xi_1, \xi_2\right] \le c_\alpha\right) \\ &= P\left(\max\left[\xi_1, \xi_2\right] - \min\left[\xi_1, \xi_2\right] \le \frac{c_\alpha}{2}\right) \\ &= P\left(\xi_1 - \xi_2 \le \frac{c_\alpha}{2} \text{ and } \xi_1 > \xi_2\right) + P\left(\xi_2 - \xi_1 \le \frac{c_\alpha}{2} \text{ and } \xi_2 > \xi_1\right) \\ &= 2\int_0^\infty \left(e^{-\xi_2} - e^{-\frac{c_\alpha}{2} - \xi_2}\right) e^{-\xi_2} d\xi_2 = 2\left(1 - e^{-\frac{c_\alpha}{2}}\right)\int_0^\infty e^{-2\xi_2} d\xi_2 = 1 - e^{-\frac{c_\alpha}{2}}, \end{aligned}$$

where

$$\xi = -\inf_{v}\{|v| + 2B(v)\} \overset{d}{=} \sup_{v}\{-|v| + 2B(v)\} = 2\max\left(\xi_1, \xi_2\right),$$

with

$$\xi_1 := \sup_{v\le 0}\left\{-\frac{|v|}{2} + B(v)\right\} \sim Exp(1) \text{ and } \xi_2 := \sup_{v\ge 0}\left\{-\frac{|v|}{2} + B(v)\right\} \sim Exp(1)$$

being independent. ■

Proof of Corollary 1. In the proof of Theorem 1, we have

$$S_n\left(\gamma, \hat{\beta}(\gamma)\right) - S_n\left(\gamma_0, \beta_0\right) \rightsquigarrow -W_1'\mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q\le\gamma_0)}\right]^{-1} W_1 - W_2'\mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q>\gamma_0)}\right]^{-1} W_2 + \begin{cases} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0,\end{cases} \tag{5}$$

and

$$S_n\left(\hat{\gamma}, \hat{\beta}\right) - S_n\left(\gamma_0, \beta_0\right) \rightsquigarrow -W_1'\mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q\le\gamma_0)}\right]^{-1} W_1 - W_2'\mathbb{E}\left[\mathbf{x}\mathbf{x}' 1_{(q>\gamma_0)}\right]^{-1} W_2 + \inf_v \begin{cases} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0,\end{cases}$$

where $\mu_\pm = c'D_\pm cf$ and $\lambda_\pm = c'V_\pm cf$. From (5),

$$a_n\left(\hat{\gamma} - \gamma_0\right) \rightsquigarrow \arg\min_v \begin{cases} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0.\end{cases}$$

Making the change-of-variables $v = \left(\frac{\lambda_-}{\mu_-^2}\right) r$ and noting the distributional equality $B_\ell\left(a^2 r\right) = aB_\ell(r)$, we have

$$\begin{aligned}
&\arg\min_v \begin{cases} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0,\end{cases}\\
&= \frac{\lambda_-}{\mu_-^2}\arg\min_r \begin{cases} \frac{\lambda_-}{\mu_-}|r| + 2\frac{\lambda_-}{\mu_-}B_1(-r), & \text{if } r\le 0,\\ \frac{\lambda_-}{\mu_-}\frac{\mu_+}{\mu_-}r + 2\frac{\lambda_-}{\mu_-}\sqrt{\frac{\lambda_+}{\lambda_-}}B_2(r), & \text{if } r>0,\end{cases}\\
&\overset{d}{=} \frac{\lambda_-}{\mu_-^2}\arg\max_v \begin{cases} -\frac{1}{2}|v| + B_1(-v), & \text{if } v\le 0,\\ -\frac{1}{2}\phi v + \sqrt{\omega}B_2(v), & \text{if } v>0.\end{cases}
\end{aligned}$$

It follows that

$$a_n\frac{\mu_-^2}{\lambda_-}\left(\hat{\gamma} - \gamma_0\right) = nf\frac{\left(\delta_n' D_-\delta_n\right)^2}{\delta_n' V_-\delta_n}\left(\hat{\gamma} - \gamma_0\right) \rightsquigarrow \arg\max_v \begin{cases} -\frac{1}{2}|v| + B_1(-v), & \text{if } v\le 0,\\ -\frac{1}{2}\phi v + \sqrt{\omega}B_2(v), & \text{if } v>0.\end{cases}$$

Because $a_n\left(\hat{\gamma} - \gamma_0\right) = O_p(1)$, and

$$n^{2\varphi}\hat{f}\frac{\left(\hat{\delta}'\hat{D}_-\hat{\delta}\right)^2}{\hat{\delta}'\hat{V}_-\hat{\delta}} \Bigg/ n^{2\varphi}f\frac{\left(\delta_n' D_-\delta_n\right)^2}{\delta_n' V_-\delta_n} = \frac{\hat{f}}{f}\left(\frac{\hat{\delta}'\hat{D}_-\hat{\delta}}{\delta_n' D_-\delta_n}\right)^2\frac{\delta_n' V_-\delta_n}{\hat{\delta}'\hat{V}_-\hat{\delta}} = 1 + o_p(1),$$

we conclude that $n\hat{f}\frac{\left(\hat{\delta}'\hat{D}_-\hat{\delta}\right)^2}{\hat{\delta}'\hat{V}_-\hat{\delta}}\left(\hat{\gamma} - \gamma_0\right)$ and $nf\frac{\left(\delta_n' D_-\delta_n\right)^2}{\delta_n' V_-\delta_n}\left(\hat{\gamma} - \gamma_0\right)$ have the same asymptotic distribution.

Next,

$$\begin{aligned}
&LR_n\left(\gamma_0 + a_n^{-1}v\right) \rightsquigarrow LR_\infty(v)\\
&:= \frac{1}{\eta^2}\left[\left\{\begin{array}{ll} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0,\end{array}\right. - \inf_v\left\{\left\{\begin{array}{ll} \mu_-|v| + 2\sqrt{\lambda_-}B_1(-v), & \text{if } v\le 0,\\ \mu_+ v + 2\sqrt{\lambda_+}B_2(v), & \text{if } v>0,\end{array}\right.\right\}\right]\\
&\overset{d}{=} \frac{1}{\eta^2}\frac{\lambda_-}{\mu_-}\left[\left\{\begin{array}{ll} |v| + 2B_1(-v), & \text{if } v\le 0,\\ \phi v + 2\sqrt{\omega}B_2(v), & \text{if } v>0,\end{array}\right. - \inf_v\left\{\begin{array}{ll} |v| + 2B_1(-v), & \text{if } v\le 0,\\ \phi v + 2\sqrt{\omega}B_2(v), & \text{if } v>0\end{array}\right.\right],
\end{aligned}$$

so

$$LR_n\left(\gamma_0\right) \rightsquigarrow LR_\infty(0) \overset{d}{=} \sup_v \begin{cases} -|v| + 2B_1(-v), & \text{if } v\le 0,\\ -\phi v + 2\sqrt{\omega}B_2(v), & \text{if } v>0,\end{cases}$$

where recall that $\eta^2 = \frac{\lambda_-}{\mu_-}$. As a result,

$$\begin{aligned}
P\left(\gamma_0 \in \left\{\gamma : LR_n(\gamma) \le x\right\}\right) &= P\left(LR_n\left(\gamma_0\right) \le x\right) \to P\left(\xi(\phi,\omega) \le x/2\right) = P\left(\xi_1 \le x/2, \xi_2(\phi,\omega) \le x/2\right)\\
&= P\left(\xi_1 \le x/2\right)P\left(\xi_2(\phi,\omega) \le x/2\right) = (1 - e^{-x/2})(1 - e^{-x\phi/2\omega}),
\end{aligned}$$

where

$$\xi(\phi,\omega) = \sup_v \begin{cases} -|v| + 2B_1(-v), & \text{if } v\le 0,\\ -\phi v + 2\sqrt{\omega}B_2(v), & \text{if } v>0,\end{cases} = \max\left(\xi_1, \xi_2(\phi,\omega)\right),$$

with

$$\xi_1 := \sup_{v\le 0}\left\{-\frac{|v|}{2}+B_1(v)\right\}\sim Exp(1) \text{ and } \xi_2(\phi,\omega) := \sup_{v\ge 0}\left\{-\frac{\phi v}{2}+\sqrt{\omega}B_2(v)\right\}\sim Exp\left(\frac{\omega}{\phi}\right)$$

being independent.

Finally, for $x>0$,

$$\begin{aligned}
&P\left(\gamma_0\in \text{conv}\left\{\gamma : LR_n^*(\gamma)\le x\right\}\right)\\
&\to P\left(\inf_{v\le 0}\left\{|v|+2B_1(v)\right\}+\xi(\phi,\omega)\le x \text{ and } \inf_{v\ge 0}\left\{\phi v+2\sqrt{\omega}B_2(v)\right\}+\xi(\phi,\omega)\le x\right)\\
&= P\left(-2\xi_1+2\max\left[\xi_1,\xi_2(\phi,\omega)\right]\right)\le x \text{ and } -2\xi_2(\phi,\omega)+2\max\left[\xi_1,\xi_2(\phi,\omega)\right]\le x\\
&= P\left(\max\left[\xi_1,\xi_2(\phi,\omega)\right]-\min\left[\xi_1,\xi_2(\phi,\omega)\right]\le x/2\right)\\
&= P\left(\xi_1-\xi_2(\phi,\omega)\le \tfrac{x}{2} \text{ and } \xi_1>\xi_2(\phi,\omega)\right)+P\left(\xi_2(\phi,\omega)-\xi_1\le \tfrac{x}{2} \text{ and } \xi_2(\phi,\omega)>\xi_1\right)\\
&= P\left(\xi_2(\phi,\omega)<\xi_1\le \xi_2(\phi,\omega)+\tfrac{x}{2}\right)+P\left(\xi_1<\xi_2(\phi,\omega)\le \xi_1+\tfrac{x}{2}\right)\\
&= \int_0^\infty\left(e^{-\xi_2}-e^{-\xi_2-\frac{x}{2}}\right)\tfrac{\phi}{\omega}e^{-\frac{\phi}{\omega}\xi_2}d\xi_2+\int_0^\infty\left(e^{-\frac{\phi}{\omega}\xi_1}-e^{-\frac{\phi}{\omega}\left(\xi_2+\frac{x}{2}\right)}\right)e^{-\xi_1}d\xi_1\\
&= \int_0^\infty\left(e^{-\frac{\omega}{\phi}z}-e^{-\frac{\omega}{\phi}z-\frac{x}{2}}\right)e^{-z}dz+\int_0^\infty\left(e^{-\frac{\phi}{\omega}z}-e^{-\frac{\phi}{\omega}\left(z+\frac{x}{2}\right)}\right)e^{-z}dz\\
&= \left(1-e^{-\frac{x}{2}}\right)\Big/\left(1+\tfrac{\omega}{\phi}\right)+\left(1-e^{-\frac{x}{2}}\right)\Big/\left(1+\tfrac{\phi}{\omega}\right)\\
&= \left(1-e^{-\frac{x}{2}}\right),
\end{aligned}$$

which is the cdf of $Exp(2)$. ■

Data availability

Data will be made available on request.

References

Bai, J., 1997. Estimation of a change point in multiple regression models. Rev. Econ. Stat. 79 (4), 551–563.

Bhattacharya, P.K., Brockwell, P.J., 1976. The minimum of an additive process with applications to signal estimation and storage theory. Z. Für Wahrscheinlichkeitstheorie Und Verwandte Geb. 37 (1), 51–75.

Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann. Statist. 520–533.

Chan, K.S., Tsay, R.S., 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika 85 (2), 413–426.

Dufour, J.M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econ.: J. Econ. Soc. 1365–1387.

Durlauf, S.N, Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. J. Appl. Econometrics 10 (4), 365–384.

Eo, Y., Morley, J., 2015. Likelihood-ratio-based confidence sets for the timing of structural breaks. Quant. Econ. 6 (2), 463–497.

Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68 (3), 575–603.

Hansen, B.E., 2011. Threshold autoregression in economics. Stat. Interface 4 (2), 123–127.

Hansen, B.E., 2017. Regression kink with an unknown threshold. J. Bus. Econom. Statist. 35 (2), 228–240.

Picard, D., 1985. Testing and estimating change-points in time series. Adv. in Appl. Probab. 17 (4), 841–867.

Siegmund, D., 1986. Boundary crossing probabilities and statistical applications. Ann. Statist. 361–404.

Siegmund, D., 1988. Confidence sets in change-point problems. Int. Stat. Review/Revue Int. de Stat. 31–48.

Yu, P., 2012. Likelihood estimation and inference in threshold regression. J. Econometrics 167 (1), 274–294.

Yu, P., 2015. Adaptive estimation of the threshold point in threshold regression. J. Econometrics 189 (1), 83–100.