



THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2014 - 04

**Improving Likelihood-Ratio-Based Confidence
Intervals for Threshold Parameters in Finite
Samples**

**Luigi Donayre, Yunjong Eo
and James Morley**

March 2014

Improving Likelihood-Ratio-Based Confidence Intervals for Threshold Parameters in Finite Samples

Luiggi Donayre ^{*} Yunjong Eo [†] James Morley [‡]

This Draft: March 25, 2014

Abstract

We propose an improved method for constructing likelihood-ratio-based confidence intervals for threshold parameters in threshold regressions. Related methods have been extensively developed in the literature and are asymptotically valid. However, their performance in finite samples is not satisfactory. We suggest two modifications to the standard inverted likelihood ratio approach. First, we consider a middle point adjustment for the boundaries of confidence intervals. Second, we propose an interpolation approach for evaluating the likelihood ratio profile at non-observable threshold values. Our extensive Monte Carlo simulations suggest that our proposed confidence intervals outperform existing methods, including bootstrap approaches, by attaining very accurate coverage rates with relatively short lengths in finite samples.

Keywords: Threshold regression; Finite-sample inference; Inverted likelihood ratio
JEL classification: C13, C20

^{*}Department of Economics, University of Minnesota - Duluth, 1318 Kirby Dr. Duluth, MN 55812, USA; Email: adonayre@d.umn.edu

[†]School of Economics, University of Sydney, Sydney 2006, Australia; Tel: +61 2 9351 3075; Fax: + 61 2 9351 4341; Email: yunjong.eo@sydney.edu.au

[‡]School of Economics, University of New South Wales, Sydney 2052, Australia; Tel: +61 2 9385 3366; Fax: +61 2 9313 6337; Email: james.morley@unsw.edu.au

1 Introduction

Threshold regression models and their various extensions have become standard for the specification of nonlinear relationships between economic variables (see, for example, Potter, 1995; Pesaran and Potter, 1997; Hansen, 1997; Balke, 2000; Koop and Potter, 2004; Enders et al., 2007; Auerbach and Gorodnichenko, 2012, among many others.)¹ In threshold regression models, the regression coefficients are not stable over the entire sample which, as a consequence, is split into subsamples conditional on values of a threshold variable. The true threshold parameters that drive change in coefficients are unknown and we need to estimate them and construct confidence interval for their values. In this paper, we study how to improve the finite-sample performance of confidence intervals for threshold parameters.

In the literature, Chan (1993) shows that a threshold parameter estimate converges to a functional of a compound Poisson process in the context of threshold autoregressions (TAR). However, the distribution depends on nuisance parameters. To circumvent their existence, Hansen (2000) develops an asymptotic theory for a threshold parameter under the assumption that the threshold effect (i.e., the difference between the regression parameters in different regimes) shrinks as the sample size increases. He establishes an asymptotic distribution for a likelihood ratio (LR) test of the threshold parameter and constructs a confidence interval by inverting the LR test for different hypothesized threshold values. More recently, Gonzalo and Wolf (2005) employ the subsampling method to make inferences for TAR processes. They construct confidence intervals that are asymptotically valid even when the (dis)continuity of the TAR process is unknown. However, their coverage probabilities are accurate only for relatively large samples (*e.g.*, $n = 500$). Meanwhile, Seo and Linton (2007) consider a more general threshold regression that may include continuous or discrete variables, or a linear combination of such variables. They propose a smoothed least squares estimator for this more general threshold model after replacing the indicator function with an integrated kernel and develop an asymptotic theory for this estimator without assuming a shrinking threshold effect. They also show how to construct asymptotic and bootstrap confidence intervals in their specification and compare their finite-sample performance with

¹For a comprehensive review of threshold applications in Economics, see Hansen (2011), Tong (2011), and Gonzalo and Pitarakis (2013).

the inverted likelihood ratio (ILR) approach in Hansen (2000). Their approach outperforms Hansen’s approach for large values of the threshold parameter. However, their coverage rates are far from nominal levels for small threshold effects and are sensitive to the choice of the bandwidth for the kernel.

While there have been important developments in the asymptotic theory for inference in threshold regression models (Chan, 1993; Hansen, 1996; Chan and Tsay, 1998; Hansen, 2000; Gonzalo and Wolf, 2005; Seo and Linton, 2007; Yu, 2012), only a few studies have focused on the finite-sample performance of confidence intervals for the threshold parameter. For example, Enders et al. (2007) study the small sample coverage properties of confidence intervals for the threshold parameter by comparing asymptotic and bootstrap methods using Monte Carlo experiments. They find that none of the procedures performs satisfactorily across a range of parameter values.

In this paper, we conduct extensive Monte Carlo experiments for three data-generating processes (DGP) previously considered in the literature. Our results confirm the poor performance of the ILR approach in finite samples for some DGPs, even for a large sample of $n = 1,000$. Specifically, we find that the ILR approach performs worse when the threshold effect is larger. Note that this finding is in striking contrast to the asymptotic results from Theorem 3 in Hansen (2000), which states that his asymptotic distribution provides an upper bound to the LR statistic if the threshold effect is fixed or large under the auxiliary assumptions of *iid* Gaussian errors and the independence of the errors from the regressors (assumption which are satisfied in all of our Monte Carlo simulations). In this sense, the ILR confidence intervals, according to Hansen (2000), are asymptotically conservative.² However, we argue that the LR profile evaluated at the discrete observations econometricians observe for the threshold variable is a bad approximation to the asymptotic LR profile in finite samples. When the threshold effect is large or fixed, the empirical LR profile is too sharp and a sequence of LR tests for the finite threshold observations are rejected too often, leading to the inclusion of too few threshold observations in the ILR confidence intervals. These few observations in the confidence intervals do not provide enough information to approximate

²In a similar context of structural break models, ILR confidence intervals for structural break dates proposed by Eo and Morley (2014) attain coverage rates greater than the nominal level and shorter confidence sets as the magnitude of the structural break gets larger.

the LR profile and to correctly make inferences about the true threshold parameter. This finding motivates our proposed approach.

We modify Hansen’s (1997; 2000) ILR approach to constructing confidence intervals for the threshold parameter in two directions: (i) a middle point adjustment for the boundaries of confidence intervals; and (ii) an interpolation approach for likelihood profile at non-observable threshold values. First, the middle point adjustment can be motivated as follows: Given a *finite* number of observations of the threshold variable, any point in the interval of the threshold variable which includes the maximum likelihood estimate maximizes the likelihood function. By convention, the left endpoint is usually taken as the estimator. However, Yu (2012) shows that the middle point of such an interval is asymptotically more efficient. We extend this idea to determining the boundaries of confidence intervals and shift their upper and lower bounds from the left endpoints to the middle points, respectively. Second, we approximate the LR profile for the threshold parameter values that we do not observe from the data by using an interpolation approach. In this way, we consider threshold values that could potentially be included in the confidence interval if they were observed. We take into account each modification by itself to determine whether both are required to attain coverage rates close to a given nominal level. We then employ both modifications together to construct our proposed confidence intervals.

To evaluate the performance of our proposed approach, we compare it to a variety of asymptotic and bootstrap methods by considering three different designs of Monte Carlo simulations previously examined in the literature.³ The results suggest that our proposed approach performs best overall in terms of producing accurate coverage rates and relatively short confidence intervals, even in small samples. Other approaches to constructing confidence intervals either massively undercover the true threshold parameter for large threshold effects or overcover with confidence intervals that are almost twice as long as for our approach. Therefore, when considering both coverage rates and average lengths, our proposed approach performs better than existing asymptotic and bootstrap methods.

The remainder of the paper is organized as follows. Section 2 describes the threshold

³Because asymptotic confidence intervals will not necessarily exhibit good coverage rates in small sample settings or environments with fixed or large threshold effects, we also consider two different bootstrap methods to constructing confidence intervals.

regression model. In Section 3, we present the likelihood-ratio-based approach to constructing asymptotic confidence intervals, as well as our proposed refinements. We evaluate the performance of our approach by means of Monte Carlo experiments in Section 4. Concluding remarks are presented in Section 5.

2 Threshold Regressions

We consider a general class of threshold regressions. Following Hansen (2000), regression parameters, or a subset of them, switch between two regimes according to the following specification:

$$y_i = \theta'_1 x_i + e_i, \quad \text{if } q_i \leq \gamma \quad (1)$$

$$y_i = \theta'_2 x_i + e_i, \quad \text{if } q_i > \gamma \quad (2)$$

for $i = 1, \dots, n$, where $x_i \in \mathbb{R}^k$ is a vector of regressors; the threshold variable q_i , which may be an element of x_i , splits the sample into two regimes; γ is the unknown threshold parameter; y_i is generated by either (1) or (2) depending on the value of q_i relative to γ ; and the random variable e_i is a regression error.

For expositional purposes, the threshold regression model (1)-(2) can be rewritten in a single-equation form:

$$y_i = \theta' x_i + \delta'_n x_i(\gamma) + e_i \quad (3)$$

where $\theta = \theta_2$, $\delta_n = (\theta_1 - \theta_2)$, $x_i(\gamma) = x_i d_i(\gamma)$, $d_i(\gamma) = \mathbf{1}\{q_i \leq \gamma\}$, and $\mathbf{1}\{\cdot\}$ is the indicator function.

The regression parameters $(\theta, \delta_n, \gamma)$ are then estimated by least squares as in Hansen (2000). The sum of squared errors is

$$S_n(\theta, \delta, \gamma) = (y - X\theta - X_\gamma\delta)'(y - X\theta - X_\gamma\delta) \quad (4)$$

where $y = (y_1, \dots, y_n)'$, X and X_γ are stacking matrices of the vectors x'_i and $x'_i(\gamma)$ in equation (3) respectively. Thus, we minimize the nonlinear function (4) jointly with respect

to θ , δ_n , and γ .⁴ As it is standard in the literature, the estimate γ is restricted to be in a bounded set $\Gamma = [\underline{\gamma}, \bar{\gamma}]$ to avoid end-of-sample distortions.⁵

The model follows the assumptions described in Hansen (2000). We briefly present the main assumptions and refer readers to Hansen (2000) for details. Let $f(q)$ be the density function of q_i and γ_0 denote the true value of γ . Then, the assumptions in this paper are stated as follows:

Assumptions

A1. The observed sample $\{y_i, q_i, e_i\}_{i=1}^n$ is strictly stationary, ergodic and ρ -mixing, with ρ -mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$. Also, $E(e_i | \mathcal{F}_{i-1}) = 0$.

A2. $\delta_n = cn^{-\kappa}$, for some $c \neq 0$ and $0 < \kappa < \frac{1}{2}$.

A3. Let $M(\gamma) = E(x_i x_i' \{q_i \leq \gamma\})$, $D(\gamma) = E(x_i x_i' | q_i = \gamma)$, and $V(\gamma) = E(x_i x_i' e_i^2 | q_i = \gamma)$. Also, $D = D(\gamma_0)$, $V = V(\gamma_0)$, $f = f(\gamma_0)$, and $M = E(x_i x_i')$. Then, $f(\gamma)$, $D(\gamma)$, and $V(\gamma)$ are continuous at $\gamma = \gamma_0$ and $M > M(\gamma) > 0$ for all $\gamma \in \Gamma$. Also, $c'Dc > 0$, $c'Vc > 0$, and $f > 0$.

A4. $E|x_i|^4 < \infty$ and $E|x_i e_i|^4 < \infty$. Also, for all $\gamma \in \Gamma$, $E(|x_i|^4 e_i^4 | q_i = \gamma) \leq C$ and $E(|x_i|^4 | q_i = \gamma) \leq C$ for some $C < \infty$ and $f(\gamma) \leq f < \infty$.

Because we are interested in constructing a confidence interval for the threshold parameter γ and want to make the minimization of (4) easy, we consider a concentrated estimator for γ . Conditional on γ , (3) is linear in θ and δ . The conditional estimators $\theta(\gamma)$ and $\delta(\gamma)$ can be found by regressing y on $X_\gamma^* = [X \ X_\gamma]$. Then, the sum of squared errors function for γ is given by

$$S_n(\gamma) = S_n(\theta(\gamma), \delta(\gamma), \gamma) = y'y - y'X_\gamma^*(X_\gamma^{*'}X_\gamma^*)^{-1}X_\gamma^{*'}y. \quad (5)$$

⁴Under the assumption that e_i is iid $N(0, \sigma^2)$, this least estimator is equivalent to the maximum likelihood estimator.

⁵In practice, the first and last 15% of threshold observations are trimmed. We follow this practice in our Monte Carlo simulations.

The estimate of γ can be obtained in two steps. First, we calculate the sum of squared errors over the restricted set Γ_n defined as

$$\Gamma_n = \Gamma \cap \{q_i\}_{i=1}^n \quad (6)$$

and second, we find the value of γ which minimizes (5):

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma_n} S_n(\gamma). \quad (7)$$

3 Confidence Intervals for Threshold Parameters

3.1 Benchmark ILR Approach

Following Hansen (2000), we construct a $1 - \alpha$ level confidence interval for γ by inverting an α -level likelihood ratio test of the hypothesis $H_0 : \gamma = \gamma_0$. Hansen shows that the log likelihood ratio statistic under the auxiliary assumption that $e_i \sim iidN(0, \sigma^2)$ is given by

$$LR_n(\gamma) = n \frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})} \quad (8)$$

with $S_n(\gamma)$ defined as in equation (5), and under Assumptions A1-A4

$$LR_n(\gamma_0) \xrightarrow{d} \eta^2 \xi \quad (9)$$

where $\xi = \max_{s \in \mathbb{R}} [2W(s) - |s|]$, $W(s)$ is a two-sided Brownian motion on \mathbb{R} , $\eta^2 = (c'Vc)/(\sigma^2 c'Dc)$, and the distribution function of ξ is $P(\xi \leq x) = (1 - \exp(-x/2))^2$. It is well known that the distribution of the LR statistic in (9) is non-standard but free of nuisance parameters under the assumption of conditional homoskedasticity (i.e., $\eta^2 = 1$, see Hansen (2000) for more details).

Hansen's (2000) analysis allows for the construction of asymptotically valid confidence intervals for the threshold parameter based on inverting the LR statistic. In this way, a $1 - \alpha$ confidence interval for the threshold parameter consists of all the possible values of $\gamma \in \Gamma_n$ for which the null hypothesis would not be rejected at the α level. Then, his confidence

interval is given by

$$C_d = \{\gamma : LR_n(\gamma) \leq CV_{1-\alpha}, \gamma \in \Gamma_n\} \quad (10)$$

where $CV_{1-\alpha}$ is the critical value calculated by direct inversion of the distribution function of ξ with the estimate of η^2 .

Note that the confidence interval in (10) may be disjoint. However, we can construct a convexified version of *confidence interval* by connecting all disjoint segments from the confidence interval in (10). We set this convexified confidence interval as the *benchmark* in the paper. To describe the different approaches to constructing confidence intervals, we illustrate a hypothetical LR profile in Figure 1. Let $q(j)$ denote the j -th ordered threshold value among all $q_i \in \Gamma_n$. Suppose the l -th threshold value $q(l)$ and the u -th threshold value $q(u)$ are the boundaries of the ILR confidence interval, defined as the minimum and maximum values in the ILR confidence set (10), respectively:

$$q(l) = \min \{q_i : LR_n(q_i) \leq CV_{1-\alpha}, q_i \in \Gamma_n\} \quad (11)$$

$$q(u) = \max \{q_i : LR_n(q_i) \leq CV_{1-\alpha}, q_i \in \Gamma_n\} \quad (12)$$

These can be easily seen in Figure 1. Then, the $1 - \alpha$ benchmark ILR confidence interval is given by

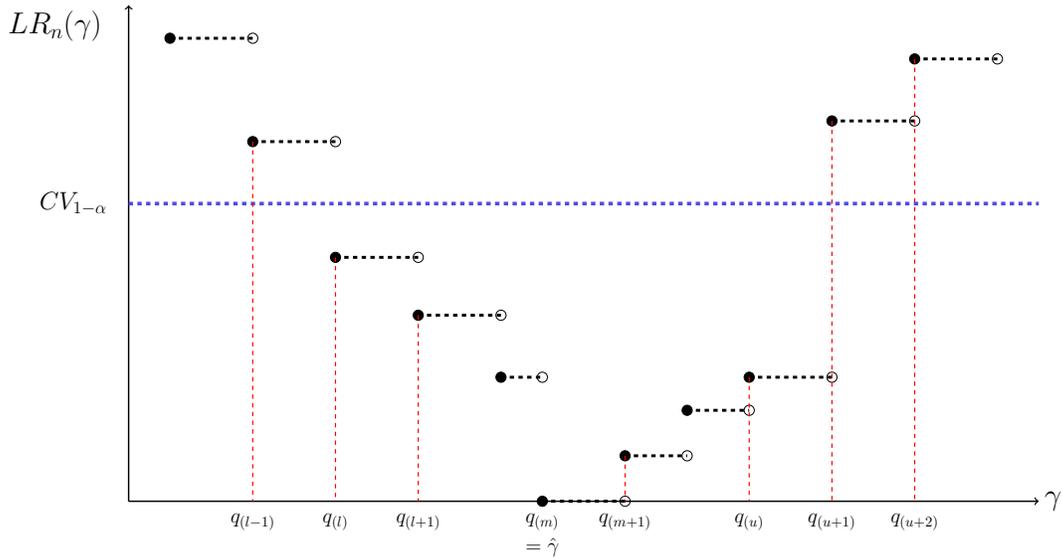
$$C_b = \{\gamma : q(l) \leq \gamma \leq q(u)\} \quad (13)$$

where $q(l)$ and $q(u)$ are defined in (11) and (12), respectively.

Theorem 3 in Hansen (2000) shows that the asymptotic distribution associated with (10) is an upper bound on the asymptotic distribution in which δ_n is fixed (i.e., $\kappa = 0$ in Assumption A2) under the additional assumptions that the errors e_i are iid $N(0, \sigma^2)$ and strictly independent of the regressors x_i as well as the threshold variable q_i . In this case, the expected coverage rate of the benchmark ILR confidence interval (13) should be, asymptotically, greater than $1 - \alpha$ since the confidence interval is constructed by completing the disjoint segments in the confidence set given by (10).

However, as shown in our Monte Carlo simulations in the next section and previously found in the literature, the empirical coverage rate of the benchmark confidence interval is

Figure 1: Illustrated Example of Log-Likelihood Ratio Profile for the Threshold Parameter



Note: A hypothetical LR profile is depicted. Given a finite number of observations of the threshold variable, the likelihood ratio is evaluated discretely. Thus, for all $q_i \in [q(j), q(j+1))$, they have the same likelihood ratio value $LR_n(q_i) = LR_n(q(j))$. It is drawn by a dotted black line with an interval. The left endpoint of the interval $q(j)$ is denoted by a solid point and the right endpoint $q(j+1)$ is denoted by a hollow point. A critical value $CV_{1-\alpha}$ is indicated by a blue dashed line.

far lower than the desired $1 - \alpha$ level (e.g., Enders et al., 2007; Gonzalo and Wolf, 2005). This discrepancy motivates us to propose a modified version of the likelihood-ratio-based confidence interval.

3.2 Modified ILR Approaches

In this section, we propose a new approach to constructing a confidence interval for the threshold parameter. We believe that the benchmark ILR approach attains unsatisfactory coverage rates because likelihood ratio evaluations based on $\{q_i\}_{i=1}^n$ are poor approximations to likelihood ratios from the true DGP.⁶ To overcome this issue, we modify the ILR approach in two directions: (i) a middle point LR evaluation; and (ii) an interpolation approach for the LR values.

⁶We can confirm this conjecture in the following section from Monte Carlo simulations.

3.2.1 Middle point ILR approach

We describe here how the middle point evaluation works and why this modification is important for improving coverage rates of confidence intervals. Let $q(m)$ be the estimate of the threshold parameter defined in (7) (i.e. $\hat{\gamma} = q(m)$), which is also shown in Figure 1. Although $f(\gamma = \gamma_0)$ is continuous in the DGP (see Assumption A.3), we only have discrete observations of the threshold variable. Therefore, $q(m)$ is the left endpoint estimator in the sense that any threshold value $\gamma \in [q(m), q(m+1))$ can produce the same split subsamples and the same LR value, given the specification in equations (1)-(2). Yu (2012) shows that the middle point estimator $\frac{1}{2}(q(m) + q(m+1))$ is more efficient than the left endpoint estimator, even though both are asymptotically consistent estimators.⁷ However, in finite samples, they are distinctly different, so that $q(m) \neq \frac{1}{2}(q(m) + q(m+1))$.

This idea can also be applied to determining the boundaries of the confidence interval.⁸ Therefore, we can propose a modified approach to constructing a confidence interval for the threshold parameter labeled as the middle point ILR confidence interval (C_m):

$$C_m = \left\{ \gamma : \frac{1}{2}(q(l) + q(l+1)) \leq \gamma \leq \frac{1}{2}(q(u) + q(u+1)) \right\} \quad (14)$$

where $q(l)$ and $q(u)$ are defined as before, $q(l+1) = \min \{q_i : LR_n(q_i) \leq CV_{1-\alpha}, q_i > q(l), q_i \in \Gamma_n\}$, and $q(u+1) = \min \{q_i : LR_n(q_i) > CV_{1-\alpha}, q_i > q(u), q_i \in \Gamma_n\}$. For a visual representation, refer to Figure 1.

⁷We can also confirm this by taking into account an alternative specification. Consider the example in Figure 1. Before we introduce the alternative specification, we denote the estimate of the threshold parameter from the threshold model (1)-(2), which is the left endpoint estimator, by $\hat{\gamma}_{LE} = q(m)$ and the middle point estimator by $\hat{\gamma}_{ME} = \frac{1}{2}(q(m) + q(m+1))$. Suppose we rewrite the threshold model (1)-(2) slightly differently by changing inequality signs for the conditions of the threshold parameter from $q_i \leq \gamma$ to $q_i < \gamma$ and from $q_i > \gamma$ to $q_i \geq \gamma$, respectively:

$$y_i = \theta'_1 x_i + e_i, \quad \text{if } q_i < \gamma \quad (1')$$

$$y_i = \theta'_2 x_i + e_i, \quad \text{if } q_i \geq \gamma. \quad (2')$$

The estimate of the threshold parameter from the rewritten threshold model (1')-(2'), which is the right endpoint estimator, would be $\hat{\gamma}_{RE} = q(m+1)$. Note that as the sample size increases, $\lim_{n \rightarrow \infty} \hat{\gamma}_{LE} = \lim_{n \rightarrow \infty} \hat{\gamma}_{RE} = \lim_{n \rightarrow \infty} \hat{\gamma}_{ME} = \gamma_0$ and the middle point estimator is the same for either (1)-(2) or (1')-(2'). Thus, this alternative specification would not make any difference under Assumption A3 that $f(\gamma)$ is continuous at γ_0 . However, in finite samples they are all different (i.e. $q(m) \neq q(m+1) \neq \frac{1}{2}(q(m) + q(m+1))$).

⁸Yu (2012) suggested a Bayesian approach to constructing a confidence interval. However, we focus on a frequentist approach only in this paper.

3.2.2 Interpolation ILR approach

An additional direction in which we modify the ILR approach considers an interpolation for calculating likelihood values associated with non-observable thresholds. While discussing this second modification, we put aside the first modification described by the middle point approach. However, we will also consider both modifications together in the next section.

The motivation for this second modification stems from the fact that we cannot evaluate the likelihood functions for points that we do not observe (i.e., any points of $\gamma \notin \Gamma_n$) because Γ_n is a collection of discrete observations in the parameter space of Γ in finite samples. For example, we cannot determine the likelihood ratio values for threshold variables between $q(u)$ and $q(u + 1)$ and between $q(l - 1)$ and $q(l)$. However, it is likely that there are some threshold parameter values $\hat{\gamma} \in (q(u), q(u + 1))$ such that $LR_n(\hat{\gamma}) \leq CR_{1-\alpha}$. Similarly, it is possible that there are some threshold parameter values $\hat{\gamma} \in (q(l - 1), q(l))$ such that $LR_n(\hat{\gamma}) \leq CR_{1-\alpha}$ where $q(l - 1) = \max \{q_i : LR_n(q_i) > CV_{1-\alpha}, q_i < q(l), q_i \in \Gamma_n\}$. If they are not included in the confidence interval, it may exclude the true threshold value and its coverage rate could be far lower than $1 - \alpha$. Thus, if there are such γ 's, they should be included in the confidence interval.

To remedy this potential problem, we consider an interpolation method using observable LR values and the critical value $CV_{1-\alpha}$ so that we can include possible threshold values between $q(u)$ and $q(u + 1)$ and between $q(l - 1)$ and $q(l)$ for which the null hypothesis would not be rejected, even if we do not have observations for them. The interpolation ILR confidence interval is given by:

$$C_{int} = \{\gamma : w_l q(l) + (1 - w_l)q(l - 1) \leq \gamma \leq w_u q(u) + (1 - w_u)q(u + 1)\} \quad (15)$$

where

$$w_l = \frac{LR_n(q(l - 1)) - CV_{1-\alpha}}{LR_n(q(l - 1)) - LR_n(q(l))}, \quad (16)$$

$$w_u = \frac{LR_n(q(u + 1)) - CV_{1-\alpha}}{LR_n(q(u + 1)) - LR_n(q(u))}, \quad (17)$$

so that $w_l \in (0, 1)$ and $w_u \in (0, 1)$.

Note that, as $LR_n(q(l-1)) \rightarrow CV_{1-\alpha}$, $w_l \rightarrow 0$ and the interpolation ILR confidence interval would include more values between $q(l-1)$ and $q(l)$ than the benchmark confidence interval. Similarly, as $LR_n(q(u+1)) \rightarrow CV_{1-\alpha}$, $w_u \rightarrow 0$ and the interpolation ILR confidence interval would include more values between $q(u)$ and $q(u+1)$ than the benchmark confidence interval.

We can further adopt a more conservative approach and set the weights $w_u = 0$ and $w_l = 0$, regardless of LR values around the boundaries. Hence, we can define a conservative ILR confidence interval as

$$C_c = \{\gamma : q(l-1) \leq \gamma \leq q(u+1)\}. \quad (18)$$

By construction, the conservative confidence interval C_c in (18) is always longer than either the benchmark confidence interval (13) or the interpolation confidence interval (15). Also, note that the interpolation confidence interval (15) is always longer than the benchmark confidence interval (13).

3.2.3 Proposed approach to constructing confidence intervals

Up until now, we have considered the middle point approach and the interpolation approach separately. Building on these refinements, we propose a new approach considers both modifications together. This new confidence interval is given by

$$C_{new} = \{\gamma : q_L^* \leq \gamma \leq q_U^*\} \quad (19)$$

where

$$q_L^* = w_l \times \frac{1}{2}(q(l) + q(l+1)) + (1 - w_l) \times \frac{1}{2}(q(l-1) + q(l))$$

$$q_U^* = w_u \times \frac{1}{2}(q(u) + q(u+1)) + (1 - w_u) \times \frac{1}{2}(q(u+1) + q(u+2))$$

with $q(u+2) = \min\{q_i : LR_n(q_i) > CV_{1-\alpha}, q_i > q(u+1), q_i \in \Gamma_n\}$. The weights w_l and w_u are defined as in (16) and (17). In our proposed confidence intervals, the lower boundary of the confidence interval is determined by the weighted average of two middle points $\frac{1}{2}(q(l) + q(l+1))$ and $\frac{1}{2}(q(l-1) + q(l))$ instead of the weighted average of $q(l)$ and $q(l-1)$. Similarly, the

upper boundary is determined by the weighted average of two middle points $\frac{1}{2}(q(u)+q(u+1))$ and $\frac{1}{2}(q(u+1)+q(u+2))$ instead of the weighted average of $q(u)$ and $q(u+1)$.

Notice that if $LR(q(l)) \rightarrow CV_{1-\alpha}$, then $w_l \rightarrow 1$ for the case of the lower boundary. Meanwhile, if $LR(q(l-1)) \rightarrow CV_{1-\alpha}$, then $w_l \rightarrow 0$ for the case of the upper boundary of the confidence interval. Thus, the middle point ILR confidence interval in (14) is a special case of our new proposed approach when $w_l = 1$ and $w_u = 1$.

4 Monte Carlo Experiments

In this section, we conduct extensive Monte Carlo simulations to evaluate the finite-sample performance of the competing methods for constructing confidence intervals of the threshold parameter. We compare our new proposed approach (ILR_{new}), as well as the middle point approach (ILR_m) and the interpolation approach (ILR_{int}), to the following existing methods in the literature: the disjoint ILR confidence interval and the benchmark ILR confidence interval (ILR_d and ILR_b) in Hansen (1997, 2000), the bootstrap percentile confidence interval (BP), and the bootstrap ILR confidence interval ($BILR$). In addition, we also consider the conservative ILR approach (ILR_c), as described above. By taking into account only one modification of the benchmark ILR approach at a time (either the middle point modification only or the interpolation modification only), we are able to evaluate the effect of each modification and determine whether both are required to improve the performance of the ILR approach. To be concise, we omit the details of these well-known methods and encourage interested readers to refer to Hansen (1997, 2000) and MacKinnon (2006) for specifics.

To evaluate and contrast the overall performance of all methods, we examine the empirical coverage rates and the average lengths of confidence intervals for the threshold parameter. The coverage rate is computed as the frequency of Monte Carlo simulations for which the constructed intervals contain the true threshold parameter. Its accuracy is determined by comparing it to the nominal confidence level $1 - \alpha$. In all experiments, we construct 95% confidence intervals. The average length of the confidence interval is defined as the difference between the upper and the lower boundaries of the confidence interval averaged across Monte Carlo simulations. For ease of comparison, the average lengths for all approaches are

normalized by the length of our proposed $ILLR_{new}$ approach in all specifications.

In the case of bootstrap methods, we consider $B = 199$ bootstrap samples. To construct the bootstrap samples, we take a parametric approach and use the estimated coefficients and threshold parameters instead of the parameters from the DGP and errors from a normal distribution with variance $\hat{\sigma}^2$. In all cases, we calculate confidence intervals for sample sizes set to $n = 50, 100, 250, 500$ and $1,000$ and consider 1,000 Monte Carlo replications for each experiment.⁹

For ease of comparison, we summarize the competing approaches to constructing confidence intervals for the threshold parameter in our Monte Carlo analysis in Table 1. To

Table 1: Competing Approaches to Constructing Confidence Intervals

	Description	Confidence Interval
$ILLR_d$	Hansen's disjoint ILR	$\{\gamma : LR_n(\gamma) \leq CV_{1-\alpha}\}$
$ILLR_b$	Hansen's benchmark ILR	$\{\gamma : q(l) \leq \gamma \leq q(u)\}$
$ILLR_c$	Inverted LR conservative ILR	$\{\gamma : q(l-1) \leq \gamma \leq q(u+1)\}$
$ILLR_m$	Middle Point ILR	$\{\gamma : \frac{1}{2}(q(l) + q(l+1)) \leq \gamma \leq \frac{1}{2}(q(u) + q(u+1))\}$
$ILLR_{int}$	Interpolation ILR	$\{\gamma : w_l q(l) + (1-w_l)q(l-1) \leq \gamma \leq w_u q(u) + (1-w_u)q(u+1)\}$
$ILLR_{new}$	New Proposed Approach	$\left\{ \gamma : \begin{array}{l} w_l \times \frac{1}{2}(q(l) + q(l+1)) + \\ (1-w_l) \times \frac{1}{2}(q(l-1) + q(l)) \end{array} \leq \gamma \leq \begin{array}{l} w_u \times \frac{1}{2}(q(u) + q(u+1)) + \\ (1-w_u) \times \frac{1}{2}(q(u+1) + q(u+2)) \end{array} \right\}$
BP	Bootstrap percentile	$\{\gamma : q_{1-\alpha/2}^* \leq \gamma \leq q_{\alpha/2}^*\}$
$BILR$	Bootstrap ILR	$\{\gamma : LR_n(\gamma) \leq CV_{1-\alpha}^*\}$

Note: The percentiles $q_{1-\alpha/2}^*$ and $q_{\alpha/2}^*$ for the BP approach and the critical value $CV_{1-\alpha}^*$ for the BILR approach are calculated for each simulation by bootstrapping. See MacKinnon (2006) for more details.

evaluate the performance of our proposed approach in different settings, we consider three different DGPs previously examined in the literature by Tong (1990), Hansen (2000) and Chan and Tsay (1998). The DGP and results for each experiment are described in the following subsections.

4.1 Monte Carlo Experiment 1: Tong's (1990) DGP

We generate data according to the following self-exciting TAR (SETAR) model:

$$y_i = \begin{cases} \alpha_0 + \sum_{j=1}^p \alpha_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} \leq \gamma \\ \beta_0 + \sum_{j=1}^p \beta_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} > \gamma \end{cases} \quad (20)$$

⁹Each series was generated for $(n + 200)$ observations after the first 200 observations were discarded to avoid any distortions from initial values.

where $\epsilon_i \sim N(0, 1)$ for $i = 1, \dots, n$. To reduce the computational burden, we focus on the simplest case where $p = d = 1$ and set $\alpha_0 = 0.7$, $\alpha_1 = -0.5$, $\beta_0 = -1.8$, $\beta_1 = 0.7$ and $\gamma = 0$, which is also the DGP studied by Tong (1990) and Gonzalo and Wolf (2005).

Table 2 shows empirical coverage rates and average lengths for the DGP described in (20) and different sample sizes. The results show that most approaches perform poorly in the sense that they cannot produce coverage rates close to the nominal level for all sample sizes. Exceptions are given by the $ILLR_{new}$ approach, the BP approach, and the $ILLR_c$ approach. For example, the $ILLR_c$ approach massively overcovers, as it exhibits coverage rates near 100%. This can be explained by the excessively long average lengths associated with it. The confidence intervals constructed using the $ILLR_c$ approach are, by far, the longest from all asymptotic methods. Although our proposed $ILLR_{new}$ approach (92% to 96%) and the BP approach (94% to 96%) produce quite accurate coverage rates across different sample sizes, the lengths of our $ILLR_{new}$ proposed approach are substantially shorter than those from the BP approach, which are 70% to 80% longer, for all sample sizes. Furthermore, the relative length of the BP approach, compared to our $ILLR_{new}$ approach, increases with the sample size. Thus, the accurate coverage of the BP approach is attained at the expense of a much longer length and, consequently, less precise information about the threshold parameter. At the same time, undercoverage from the $ILLR_m$ and $ILLR_{int}$ approaches implies that both modifications are necessary for better coverage accuracy of the general $ILLR_b$ approach. Another interesting finding is that the $BILR$ approach massively undercovers (35% to 45%) and exhibits average lengths that are excessively short. It is important to note that the coverage rates of all these methods do not appear to increase with the sample size.

4.2 Monte Carlo Experiment 2: Hansen’s (2000) DGP

In the second experiment, we use two different DGPs considered by Hansen (2000). Specifically, the DGP is described by equation (3) with *iid* data and $x_i = (1 \ z_i)'$, $e_i \sim N(0, 1)$, $q_i \sim N(2, 1)$ and $\gamma = 2$. Following Hansen (2000), two different regressors z_i are considered: (i) $z_i \sim iidN(0, 1)$ and (ii) $z_i = q_i$. Partitioning $\delta_n = (\delta_1 \ \delta_2)'$, we set $\delta_1 = 0$ and assess coverage rates and normalized average lengths allowing δ_2 to vary. Specifically, δ_2 is set to

0.25, 0.50, 1.00, 1.50, and 2.00.¹⁰ The sample size n also varies as before.¹¹

The results are reported in panels (a) and (b) of Table 3 for the cases with $z_i \sim iid N(0, 1)$ and $z_i = q_i$, respectively. From panel (a), when the threshold effect is small ($\delta_2 = 0.25, 0.50$), we find that coverage rates for all asymptotic approaches slightly overcover for all sample sizes, with the exception of the $ILLR_d$ approach. Meanwhile, the $BILLR$ approach slightly undercovers for small samples, but coverage rates tend to increase with the sample size n . Similarly, the BP approach exhibits low coverage rates for small samples but tends to converge to the nominal level of 95% when $\delta_2 = 0.50$. However, their coverage rates are quite low when $\delta_2 = 0.25$. In general terms, normalized average lengths for all approaches are not very different in small samples ($n \leq 100$) for $\delta_2 = 0.25$. As the sample size n increases, the normalized average length of the $ILLR_d$ approach decreases significantly, while the normalized average length of the BP approach increases, becoming 33% longer than our proposed approach when $n = 1,000$. This same pattern also occurs for the case of $\delta_2 = 0.5$ and the length differences for $ILLR_d$ and BP , relative to the length of our proposed approach, are only accentuated.

As the threshold effect δ_2 becomes larger, the overcoverage of the $ILLR_c$ approach increases to the point of reaching almost 100% for $\delta_2 = 2$. This is explained by their increasing normalized average lengths relative to our proposed $ILLR_{new}$ approach, which becomes almost twice as long for $\delta_2 = 2$. By contrast, the coverage rates of all other asymptotic approaches, as well as the $BILLR$ approach, fall with the threshold effect. The $BILLR$ confidence intervals perform particularly poorly as their coverage rates fall close to zero for very large threshold effects ($\delta_2 = 2$). Intuitively, the decrease in coverage probabilities is explained by the fact that, with a larger threshold effect, the identification of the threshold parameter is very

¹⁰Both setups in Monte Carlo simulations are also used in the supplemental appendix to Seo and Linton (2007). They found that their smoothed least squares estimator performs well for large threshold effects. However, the coverage rates are far below the nominal level for small threshold effects. See Seo and Linton (2006) for more details. They note that their confidence intervals are asymptotically longer than those in Hansen (2000) because their estimator converges at a slower rate. We do not consider their approach in this paper because our new proposed approach can attain coverage rates close to the nominal level and the expected length of our new proposed approach is asymptotically the same as in Hansen (2000).

¹¹Using observations of the threshold variable for each Monte Carlo sample, we find somewhat different results to those in Table II in Hansen (2000). We are only able to replicate his results when evaluating the likelihood ratio at the true threshold value from the DGP, which would not be considered in practice given a finite-sample of data drawn from a continuous distribution.

precise. As a consequence, the confidence intervals become very narrow and include very few points, in some cases even just one threshold value.¹² This can be clearly seen by comparing normalized average lengths for all asymptotic approaches and the *BILR* approach in panel (a) of Table 2. For example, for $\delta_2 = 2$, the average length of the benchmark *ILLR_b* approach is, roughly, 25%-30% that of our proposed *ILLR_{new}* approach. To get a sense of how narrow such confidence intervals become with large threshold effects, the average lengths for our *ILLR_{new}* confidence intervals prior to normalization are 0.093, 0.044, 0.018, 0.009 and 0.004 for $n = 50, 100, 250, 500$ and $1,000$, respectively.

Notably, the coverage rates of our proposed *ILLR_{new}* approach are very accurate even for threshold effects as large as $\delta_2 = 1$. For the exceptional cases of threshold effects that are unusually high, like $\delta_2 = 1.5$ or $\delta_2 = 2$, our constructed confidence intervals undercover. In spite of this, the coverage rates of our *ILLR_{new}* approach are always higher than those of the *BILR* approach and all other asymptotic confidence intervals (with the exception of *ILLR_c*), especially for large threshold effects. This is explained by the fact that, even when the estimation of the threshold becomes very precise for large threshold effects, our confidence intervals still contain enough information to include the true threshold parameter. Meanwhile, the coverage probabilities from the *BP* confidence intervals are very close to the nominal level of 95% for large values of δ_2 (although they undercover for small threshold effects, as noted previously). Intuitively, their coverage rate is closer to 95% because the confidence intervals are significantly longer than our proposed confidence intervals. For example, for δ_2 , they are, on average, more than double the length at 133% longer.

The results for the case with $z_i = q_i$, reported in panel (b) of Table 2, are not much different, qualitatively, from those reported in panel (a). In general, the *ILLR_c* approach always overcovers, especially for large values of δ_2 , in part, because the confidence intervals are longer than all other asymptotic confidence intervals by construction. The *BILR* approach, on the other hand, generates confidence intervals with accurate coverage rates for small threshold effects, but with very low coverage rates for large threshold effects, since the confidence intervals become prohibitively narrow in that case. The confidence intervals from

¹²Hansen's (2000) benchmark confidence intervals include less than three points over 70% of the time in Monte Carlo simulations when $\delta_2 = 2$. In that case, even our linear interpolation modification cannot help attain coverage rates close to the nominal level of 95%.

all other asymptotic approaches exhibit accurate coverage probabilities for small threshold effects, but they fall below 95% as δ_2 increases, given the small number of threshold values included in them, as explained above. Notably, our $ILLR_{new}$ approach produces accurate coverage probabilities, even for $\delta_2 = 2$. Finally, the BP approach produces accurate confidence intervals for large threshold effects, but they undercover for $\delta_2 = 0.25$, even when $n = 1,000$ and also for $\delta_2 = 0.5$ and $\delta_2 = 1$ when samples are small. As before, the BP approach produces confidence intervals that are always relatively longer than those from our $ILLR_{new}$ approach and their normalized average lengths increase with δ_2 .

4.3 Monte Carlo Experiment 3: Chan and Tsay’s (1998) DGP

In the third experiment, we consider the continuous SETAR process given by DGP (20) with $p = d = 1$ and set $\alpha_0 = 0.52$, $\alpha_1 = 0.6$, $\beta_0 = 1.48$, $\beta_1 = -0.6$, $\gamma = 0.8$, and $\epsilon_i \sim N(0, 1)$ for $i = 1, \dots, n$, which corresponds to the DGP studied by Chan and Tsay (1998) and Gonzalo and Wolf (2005).¹³

The results are reported in Table 4. The $BILR$ approach and all asymptotic approaches (with the exception of the $ILLR_d$ approach) exhibit empirical coverage rates that are very close to the 95% nominal level, even in small samples. Similarly, normalized average lengths for all these approaches are not very different, although the $ILLR_c$ approach exhibits slightly longer normalized average lengths in small samples and the $BILR$ approach, for large samples. Meanwhile, the $ILLR_d$ approach slightly undercovers, even for sample sizes as large as $n = 1,000$, consistent with the findings from the previous two Monte Carlo experiments. This is explained by the considerably shorter normalized average lengths associated with the $ILLR_d$ approach, which decrease with n .

The BP approach is massively outperformed by our proposed $ILLR_{new}$ approach and all other methods. While the normalized average length of the BP confidence intervals is consistently higher relative to our proposed $ILLR_{new}$ approach, as before, they undercover the true threshold parameter by a large margin. Furthermore, the coverage rate remains

¹³In the original DGP, Chan and Tsay (1998) consider heteroscedastic errors: ϵ_i for the first regime and $2\epsilon_i$ for the second regime. We set the innovations terms to have the same size of variance in both regimes, as this is a necessary assumption to develop the asymptotic distribution in Hansen (2000).

substantially below the 95% nominal level even in very large sample sizes (*e.g.*, $n = 1,000$).

4.4 Summary of Monte Carlo Results

Our Monte Carlo simulation results show that the benchmark ILR approach to constructing confidence intervals undercovers the true threshold parameter value in finite samples for some DGPs, even for a large sample of $n = 1,000$. Specifically, we find that the ILR approach performs worse when the threshold effect is larger. This finding is in contrast to the asymptotic results in the literature. However, our proposed $ILLR_{new}$ approach produces confidence intervals with empirical coverage rates that are almost always very close to the nominal 95% level, while keeping a relatively short average length. To the extent that confidence intervals with short length are desirable as long as they also exhibit high coverage accuracy, we infer from these experiments that our proposed $ILLR_{new}$ approach generates accurate confidence intervals with the appropriate length. That is, considering both coverage rates and average lengths, our proposed approach performs better than all other asymptotic and bootstrap confidence intervals.

5 Concluding Remarks

We have proposed two modifications for the ILR approach to constructing confidence intervals for the threshold parameter in threshold regressions (Hansen, 2000). First, we consider a middle point adjustment for the boundaries of confidence intervals. Second, we consider an interpolation approach for the likelihood profile evaluated at non-observable threshold values. Our extensive Monte Carlo simulations suggest that our proposed modified approach produces confidence intervals that exhibit very accurate coverage rates with reasonably short lengths in finite samples when compared to other existing asymptotic and bootstrap methods. Also, we find that both modifications are necessary to improve coverage accuracy of the ILR approach.

References

- Auerbach, Alan J. and Yuriy Gorodnichenko (2012), “Measuring the output responses to fiscal policy.” *American Economic Journal: Economic Policy*, 4(2), 1–27.
- Balke, Nathan S. (2000), “Credit and economic activity: Credit regimes and nonlinear propagation of shocks.” *The Review of Economics and Statistics*, 82, 344–349.
- Chan, Kan S. (1993), “Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model.” *Annals of Statistics*, 21, 520–533.
- Chan, Kan S. and Ruey S. Tsay (1998), “Limiting properties of the least squares estimator of a continuous threshold autorregressive model.” *Biometrika*, 45, 413–426.
- Enders, Walter, Barry Falk, and Pierre Siklos (2007), “A threshold model of u.s. real gdp and the problem of constructing confidence intervals in tar models.” *Studies in Nonlinear Dynamics and Econometrics*, 11(3), Article 4.
- Eo, Yunjong and James Morley (2014), “Likelihood-ratio-based confidence sets for the timing of structural breaks.” *manuscript*.
- Gonzalo, Jesús and Jean-Yves Pitarakis (2013), “Estimation and inference in threshold type regime switching models.” In *Handbook of Research Methods and Applications in Empirical Macroeconomics* (Nigar Hashimzade and Michael A. Thornton, eds.), 189–204, Cheltenham: Edward Elgar Publishing Limited.
- Gonzalo, Jesús and Michael Wolf (2005), “Subsampling inference in threshold autoregressive models.” *Journal of Econometrics*, 127(2), 201–224.
- Hansen, Bruce E. (1996), “Inference when a nuisance parameter is not identified under the null hypothesis.” *Econometrica*, 64, 413–430.
- Hansen, Bruce E. (1997), “Inference in tar models.” *Studies in Nonlinear Dynamics and Econometrics*, 2(1), 1–14.

- Hansen, Bruce E. (2000), "Sample splitting and threshold estimation." *Econometrica*, 68, 575–603.
- Hansen, Bruce E. (2011), "Threshold autoregression in economics." *Statistics and Its Interface*, 4, 123–127.
- Koop, Gary and Simon Potter (2004), "Dynamic asymmetries in u.s. unemployment." *Journal of Business and Economic Statistics*, 17(3), 298–312.
- MacKinnon, James (2006), "Bootstrap methods in econometrics." *Economic Record*, 82(S1), S2–S18.
- Pesaran, Hashem and Simon Potter (1997), "A floor and ceiling model of u.s. output." *Journal of Economic Dynamics and Control*, 21(4-5), 661–695.
- Potter, Simon (1995), "A nonlinear approach to u.s. gnp." *Journal of Applied Econometrics*, 10, 109–25.
- Seo, Myung Hwan and Oliver Linton (2006), "Appendum to a smoothed least squares estimator for threshold regression models: Simulations." *manuscript*.
- Seo, Myung Hwan and Oliver Linton (2007), "A smoothed least squares estimator for threshold regression models." *Journal of Econometrics*, 141(2), 704–735.
- Tong, Howell (1990), *Non-linear Time Series. A Dynamical System Approach*. Oxford University Press, New York.
- Tong, Howell (2011), "Threshold models in time series analysis - 30 years on." *Statistics and Its Interface*, 4, 107–136.
- Yu, Ping (2012), "Likelihood estimation and inference in threshold regression." *Journal of Econometrics*, 167(1), 274–294.

Appendix: Tables

Table 2: Coverage and average length of 95% confidence intervals: Tong (1990)

$n =$	Coverage Rate					Average Length				
	50	100	250	500	1,000	50	100	250	500	1,000
ILL_d	0.61	0.55	0.56	0.58	0.59	0.61	0.57	0.60	0.60	0.60
ILL	0.66	0.59	0.59	0.61	0.61	0.85	0.73	0.68	0.67	0.66
ILL_c	0.99	0.99	0.99	1.00	0.99	1.16	1.25	1.30	1.29	1.30
ILL_m	0.62	0.58	0.58	0.59	0.59	0.83	0.72	0.67	0.67	0.67
ILL_{int}	0.88	0.86	0.85	0.87	0.88	1.01	1.01	1.01	1.00	1.00
ILL_{new}	0.93	0.93	0.93	0.92	0.96	1.00	1.00	1.00	1.00	1.00
BP	0.96	0.95	0.94	0.95	0.95	1.61	1.69	1.78	1.75	1.81
$BILL$	0.45	0.35	0.35	0.37	0.36	0.56	0.40	0.37	0.37	0.37

Note: The average lengths are normalized by the length of our proposed approach (ILL_{new}) for each sample size. The average lengths for ILL_{new} are 0.650, 0.212, 0.071, 0.035, 0.017 for sample sizes 50, 100, 250, 500, and 1,000 respectively. In our Monte Carlo simulations, we consider a SETAR model with a DGP:

$$y_i = \begin{cases} 0.70 - 0.50y_{i-1} + \epsilon_t, & \text{if } y_{i-1} \leq 0 \\ -1.80 + 0.70y_{i-1} + \epsilon_t, & \text{if } y_{i-1} > 0 \end{cases} \quad (21)$$

where $\epsilon_t \sim N(0, 1)$ for $i = 1, \dots, n$.

Table 3: Coverage and average length of 95% confidence intervals: Hansen (2000)

(a) $x_i = q_i$

$n =$	Coverage Rate					Average Length				
	50	100	250	500	1,000	50	100	250	500	1,000
$\delta_2 = 0.25$										
<i>ILL_d</i>	0.92	0.89	0.94	0.92	0.95	0.91	0.90	0.83	0.72	0.65
<i>ILL</i>	0.97	0.95	0.97	0.96	0.97	1.01	1.00	1.00	1.00	0.99
<i>ILL_c</i>	0.98	0.96	0.98	0.97	0.97	1.03	1.01	1.01	1.01	1.01
<i>ILL_m</i>	0.97	0.95	0.97	0.96	0.97	0.99	0.99	1.00	1.00	0.99
<i>ILL_{int}</i>	0.98	0.95	0.97	0.96	0.97	1.02	1.01	1.00	1.00	1.00
<i>ILL_{new}</i>	0.98	0.96	0.97	0.97	0.97	1.00	1.00	1.00	1.00	1.00
<i>BP</i>	0.86	0.66	0.75	0.89	0.94	1.09	1.05	1.06	1.16	1.33
<i>BILR</i>	0.92	0.92	0.96	0.96	0.96	0.99	0.99	1.01	1.02	1.01
$\delta_2 = 0.50$										
<i>ILL_d</i>	0.90	0.91	0.93	0.93	0.94	0.84	0.78	0.68	0.72	0.75
<i>ILL</i>	0.94	0.94	0.95	0.95	0.97	0.99	0.99	0.97	0.95	0.95
<i>ILL_c</i>	0.97	0.97	0.98	0.98	0.99	1.03	1.02	1.03	1.05	1.06
<i>ILL_m</i>	0.94	0.94	0.95	0.96	0.96	0.98	0.98	0.97	0.96	0.95
<i>ILL_{int}</i>	0.96	0.96	0.97	0.97	0.98	1.01	1.01	1.00	1.00	1.00
<i>ILL_{new}</i>	0.97	0.96	0.97	0.97	0.98	1.00	1.00	1.00	1.00	1.00
<i>BP</i>	0.90	0.88	0.95	0.95	0.96	1.14	1.16	1.44	1.54	1.44
<i>BILR</i>	0.88	0.92	0.94	0.92	0.93	0.95	0.95	0.79	0.83	0.83
$\delta_2 = 1.00$										
<i>ILL_d</i>	0.71	0.70	0.76	0.73	0.74	0.65	0.61	0.70	0.69	0.70
<i>ILL</i>	0.74	0.75	0.79	0.75	0.77	0.90	0.84	0.81	0.80	0.79
<i>ILL_c</i>	0.98	0.99	0.99	0.99	0.99	1.10	1.14	1.17	1.21	1.20
<i>ILL_m</i>	0.77	0.75	0.77	0.76	0.78	0.90	0.85	0.81	0.80	0.79
<i>ILL_{int}</i>	0.91	0.92	0.94	0.93	0.94	1.00	1.00	1.00	1.00	1.00
<i>ILL_{new}</i>	0.95	0.95	0.97	0.96	0.96	1.00	1.00	1.00	1.00	1.00
<i>BP</i>	0.96	0.96	0.94	0.96	0.94	1.53	1.63	1.54	1.64	1.63
<i>BILR</i>	0.58	0.58	0.61	0.55	0.58	0.74	0.61	0.58	0.55	0.54
$\delta_2 = 1.50$										
<i>ILL_d</i>	0.38	0.34	0.37	0.39	0.41	0.49	0.46	0.47	0.48	0.47
<i>ILL</i>	0.40	0.37	0.39	0.41	0.43	0.62	0.51	0.53	0.52	0.51
<i>ILL_c</i>	0.99	0.99	1.00	0.99	1.00	1.35	1.43	1.43	1.45	1.42
<i>ILL_m</i>	0.43	0.40	0.41	0.39	0.39	0.65	0.53	0.53	0.51	0.52
<i>ILL_{int}</i>	0.75	0.76	0.74	0.75	0.80	0.99	0.99	0.99	1.00	0.99
<i>ILL_{new}</i>	0.87	0.88	0.88	0.88	0.89	1.00	1.00	1.00	1.00	1.00
<i>BP</i>	0.95	0.95	0.96	0.95	0.95	1.94	1.95	1.91	1.97	1.90
<i>BILR</i>	0.21	0.15	0.17	0.17	0.18	0.34	0.22	0.21	0.20	0.20
$\delta_2 = 2.00$										
<i>ILL_d</i>	0.16	0.13	0.12	0.13	0.16	0.28	0.26	0.23	0.26	0.27
<i>ILL</i>	0.16	0.14	0.13	0.14	0.16	0.32	0.27	0.25	0.28	0.27
<i>ILL_c</i>	1.00	1.00	0.99	1.00	1.00	1.82	1.93	1.93	1.93	1.90
<i>ILL_m</i>	0.18	0.16	0.14	0.14	0.14	0.33	0.29	0.27	0.27	0.27
<i>ILL_{int}</i>	0.54	0.53	0.53	0.56	0.57	0.97	0.99	0.99	1.00	1.00
<i>ILL_{new}</i>	0.74	0.70	0.71	0.73	0.72	1.00	1.00	1.00	1.00	1.00
<i>BP</i>	0.96	0.94	0.94	0.95	0.94	2.33	2.36	2.28	2.37	2.33
<i>BILR</i>	0.03	0.02	0.01	0.01	0.01	0.07	0.03	0.02	0.01	0.02

Note: The average lengths are normalized by the length of our proposed approach (*ILL_{new}*) for each value of δ_2 sample size. The average lengths for *ILL_{new}* are 0.0933, 0.0438, 0.0180, 0.0086, 0.0044 for $\delta_2 = 2$ and sample sizes 50, 100, 250, 500, and 1,000 respectively. Other average lengths are available upon request. In our Monte Carlo simulations, we consider a threshold model with a DGP:

$$y_i = \begin{cases} 1 + x_i + \epsilon_t, & \text{if } q_i \leq 2 \\ 1 + (1 + \delta_2)x_i + \epsilon_t, & \text{if } q_i > 2 \end{cases}$$

where $x_i = q_i \sim N(2, 1)$ and $\epsilon_t \sim N(0, 1)$ for $i = 1, \dots, n$.

(b) $x_i \sim N(0, 1)$

$n =$	Coverage Rate					Average Length				
	50	100	250	500	1,000	50	100	250	500	1,000
$\delta_2 = 0.25$										
$ILLR_d$	0.95	0.94	0.93	0.92	0.93	0.96	0.95	0.90	0.84	0.76
$ILLR$	0.98	0.98	0.96	0.96	0.96	1.01	1.00	1.00	1.00	1.00
$ILLR_c$	0.98	0.98	0.96	0.96	0.96	1.02	1.01	1.01	1.00	1.00
$ILLR_m$	0.98	0.97	0.96	0.96	0.96	0.99	1.00	1.00	1.00	1.00
$ILLR_{int}$	0.98	0.98	0.96	0.96	0.96	1.02	1.01	1.00	1.00	1.00
$ILLR_{new}$	0.98	0.98	0.96	0.96	0.96	1.00	1.00	1.00	1.00	1.00
BP	0.87	0.58	0.61	0.76	0.89	1.10	1.05	1.03	1.09	1.21
$BILR$	0.96	0.96	0.96	0.96	0.96	1.00	1.00	1.01	1.02	1.04
$\delta_2 = 0.50$										
$ILLR_d$	0.94	0.92	0.94	0.94	0.96	0.93	0.89	0.78	0.75	0.78
$ILLR$	0.97	0.96	0.97	0.97	0.98	1.00	1.00	0.99	0.99	0.99
$ILLR_c$	0.98	0.97	0.98	0.97	0.99	1.03	1.02	1.01	1.01	1.01
$ILLR_m$	0.97	0.96	0.97	0.96	0.98	0.98	0.99	0.99	0.99	0.99
$ILLR_{int}$	0.98	0.96	0.97	0.97	0.98	1.02	1.01	1.00	1.00	1.00
$ILLR_{new}$	0.98	0.96	0.98	0.97	0.99	1.00	1.00	1.00	1.00	1.00
BP	0.90	0.76	0.89	0.95	0.95	1.14	1.12	1.23	1.43	1.44
$BILR$	0.94	0.94	0.95	0.95	0.97	0.99	0.99	0.98	0.98	0.94
$\delta_2 = 1.00$										
$ILLR_d$	0.90	0.91	0.92	0.93	0.93	0.82	0.78	0.80	0.80	0.82
$ILLR$	0.93	0.93	0.94	0.94	0.94	0.97	0.96	0.96	0.96	0.96
$ILLR_c$	0.99	0.97	0.98	0.99	0.98	1.05	1.04	1.05	1.05	1.06
$ILLR_m$	0.93	0.93	0.94	0.94	0.94	0.96	0.96	0.96	0.95	0.96
$ILLR_{int}$	0.96	0.95	0.96	0.97	0.97	1.01	1.00	1.00	1.00	1.00
$ILLR_{new}$	0.97	0.96	0.97	0.98	0.97	1.00	1.00	1.00	1.00	1.00
BP	0.92	0.92	0.95	0.95	0.95	1.30	1.39	1.49	1.48	1.51
$BILR$	0.90	0.89	0.91	0.91	0.90	0.90	0.88	0.83	0.82	0.79
$\delta_2 = 1.50$										
$ILLR_d$	0.84	0.87	0.87	0.86	0.84	0.79	0.79	0.80	0.81	0.83
$ILLR$	0.87	0.89	0.88	0.89	0.86	0.92	0.92	0.91	0.91	0.91
$ILLR_c$	0.98	0.98	0.99	0.99	0.99	1.09	1.10	1.10	1.10	1.10
$ILLR_m$	0.86	0.88	0.89	0.87	0.87	0.92	0.92	0.91	0.91	0.91
$ILLR_{int}$	0.94	0.94	0.94	0.95	0.94	1.00	1.00	1.00	1.00	1.00
$ILLR_{new}$	0.96	0.96	0.97	0.97	0.96	1.00	1.00	1.00	1.00	1.00
BP	0.95	0.95	0.95	0.94	0.96	1.51	1.50	1.57	1.56	1.59
$BILR$	0.80	0.81	0.82	0.79	0.78	0.79	0.75	0.73	0.71	0.73
$\delta_2 = 2.00$										
$ILLR_d$	0.77	0.78	0.76	0.77	0.77	0.76	0.80	0.78	0.80	0.79
$ILLR$	0.79	0.80	0.78	0.79	0.79	0.88	0.87	0.87	0.87	0.85
$ILLR_c$	0.98	0.99	0.99	0.99	0.99	1.16	1.17	1.17	1.18	1.18
$ILLR_m$	0.78	0.77	0.78	0.79	0.78	0.87	0.87	0.87	0.86	0.85
$ILLR_{int}$	0.89	0.90	0.89	0.90	0.91	1.00	1.00	1.00	1.01	1.00
$ILLR_{new}$	0.93	0.92	0.93	0.94	0.94	1.00	1.00	1.00	1.00	1.00
BP	0.95	0.94	0.96	0.95	0.95	1.70	1.67	1.70	1.74	1.80
$BILR$	0.68	0.69	0.68	0.67	0.69	0.66	0.65	0.64	0.64	0.63

Note: The average lengths are normalized by the length of our proposed approach ($ILLR_{new}$) for each value of δ_2 and sample size. The average lengths for $ILLR_{new}$ are 0.3742, 0.1863, 0.0721, 0.0351, 0.0169 for $\delta_2 = 2$ and sample sizes 50, 100, 250, 500, and 1,000 respectively. Other average lengths are available upon request. In our Monte Carlo simulations, we consider a threshold model with a DGP:

$$y_i = \begin{cases} 1 + x_i + \epsilon_t, & \text{if } q_i \leq 2 \\ 1 + (1 + \delta_2)x_i + \epsilon_t, & \text{if } q_i > 2 \end{cases}$$

where $q_i \sim N(2, 1)$, $x_i \sim N(0, 1)$, and $\epsilon_t \sim N(0, 1)$ for $i = 1, \dots, n$.

Table 4: Coverage and average length of 95% confidence intervals: Chan and Tsay (1998)

$n =$	Coverage Rate					Average Length				
	50	100	250	500	1,000	50	100	250	500	1,000
ILR_d	0.89	0.91	0.91	0.90	0.92	0.92	0.88	0.85	0.85	0.84
ILR	0.94	0.96	0.96	0.95	0.96	1.00	1.00	0.99	0.99	1.00
ILR_c	0.95	0.97	0.96	0.95	0.96	1.03	1.02	1.01	1.01	1.00
ILR_m	0.94	0.97	0.96	0.95	0.96	0.98	0.99	0.99	1.00	1.00
ILR_{int}	0.94	0.97	0.96	0.95	0.96	1.02	1.01	1.00	1.00	1.00
ILR_{new}	0.95	0.97	0.96	0.95	0.96	1.00	1.00	1.00	1.00	1.00
BP	0.49	0.44	0.46	0.45	0.51	1.07	1.03	1.03	1.05	1.07
$BILR$	0.93	0.95	0.96	0.95	0.96	0.99	1.00	1.02	1.02	1.03

Note: The average lengths are normalized by the length of our proposed approach (ILR_{new}) for each sample size. The average lengths for ILR_{new} are 1.802, 1.739, 1.385, 1.069, and 0.816 for sample sizes 50, 100, 250, 500, and 1,000 respectively. In our Monte Carlo simulations, we consider a SETAR model with a DGP:

$$y_i = \begin{cases} 0.52 + 0.60y_{i-1} + \epsilon_t, & \text{if } y_{i-1} \leq 0.80 \\ 1.48 - 0.60y_{i-1} + \epsilon_t, & \text{if } y_{i-1} > 0.80 \end{cases} \quad (22)$$

where $\epsilon_t \sim N(0, 1)$ for $i = 1, \dots, n$.