THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2014 - 04

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

Luiggi Donayre, Yunjong Eo and James Morley

July 2016

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

Luiggi Donayre * Yunjong Eo † James Morley ‡

Abstract

Within the context of threshold regressions, we show that asymptotically-valid likelihood-ratio-based confidence intervals for threshold parameters perform poorly in finite samples when the threshold effect is large. A large threshold effect leads to a poor approximation of the profile likelihood in finite samples such that the conventional approach to constructing confidence intervals excludes the true threshold parameter value too often, resulting in low coverage rates.

We propose a modification to the standard likelihood-ratio-based confidence interval that has coverage rates at least as high as the nominal level, while still being informative in the sense of including relatively few observations of the threshold variable

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

* *Corresponding Author*: 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 (Potter, 1995; Balke, 2000; Koop and Potter, 2004; Gonzalo and Pitarakis, 2013, among many others.)[1]. Although 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; Yu, 2012), it is well-known that the finite-sample performance of confidence intervals for the threshold parameter is poor in some settings (Enders, Falk and Siklos, 2007).

In this paper, we show that the performance of asymptotically valid likelihood-ratio-based confidence intervals (CIs), as proposed by Hansen (1997, 2000), may be inadequate when the threshold effect is large. In particular, a problem regarding the construction of CIs that are based on the inverted likelihood ratio (ILR) relates to a step function approximation of the likelihood function at threshold values that are not observed in the sample. Given this approximation, the CIs may exclude the true threshold parameter, resulting in low coverage rates, even in large samples.

We propose two possible modifications of Hansen's ILR approach to address the step function approximation: (i) a grid-search approach based on equally-spaced grid points and (ii) a conservative approach that extends the CIs to the closest observations excluded by the standard ILR approach. We then conduct Monte Carlo simulations to evaluate the performance of the original ILR approach and the proposed modifications, using two different data-generating processes (DGPs) previously considered in the literature. For each approach, we evaluate the coverage rate, average length and average number of threshold values included in the CIs.

Our results suggest that the standard ILR approach massively undercovers the true threshold parameter when the threshold effect is large, even for sample sizes as large as $n = 1,000$. This poor performance is explained by the 'sharp' likelihood profile associated with a large threshold effect, which results in too few possible threshold values being included in the CIs. The refined grid-search improves the performance by including some of the non-observed, but possible threshold values, although the coverage rates are still far below the nominal level in the most cases. Meanwhile, the conservative approach has coverage rates at least as high as the nominal level, while still being informative in the sense of including relatively few observations of the threshold variable.

Based on these results, we recommend researchers use the conservative approach when constructing CIs for threshold parameters in practical applications.

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

2 Threshold Regressions

We consider a general class of threshold regressions. Following Hansen (2000), regression parameters switch between two regimes according to

$$y_i = \theta_1' x_i + e_i, \quad if\ q_i \leq \gamma \tag{1}$$

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

for $i = 1, \ldots, n$, where $x_i \in \mathbb{R}^k$ is a vector of regressors; the threshold variable q_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 e_i is a regression error. For expositional purposes, the threshold regression model (1) - (2) can be rewritten in a single-equation form:

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

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.[2]

An estimate of γ can be obtained through concentration. Conditional on γ, (3) is linear in θ and δ. The conditional estimators $\theta(\gamma)$ and $\delta(\gamma)$ can be found by regressing $y = (y_1, \ldots, y_n)'$ on $X_\gamma^* = [X \quad X_\gamma]$, where X and X_γ are stacking matrices of the vectors x_i' and $x_i(\gamma)'$ in equation (3), respectively. As is standard in the literature, γ is restricted to be in a bounded set $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ to avoid small-sample distortions. In practice, $\underline{\gamma}$ and $\overline{\gamma}$ correspond to the first and last 15% of the vector of ordered threshold observations, respectively, which are trimmed.

We consider two sets of possible threshold values over which to search for the optimal threshold estimate and construct the confidence interval. In the first case, which is standard in the literature, the grid-search procedure occurs over $\Gamma_s = \Gamma \cap \{q_i\}_{i=1}^n$, so that all elements in Γ_s are simply all observed values of q between $\underline{\gamma}$ and $\overline{\gamma}$. In the second case, the refined grid-search occurs over $\Gamma_r = \Gamma \cap q^r$ where the elements in q^r are given by $[\underline{\gamma}, \underline{\gamma} + \zeta, \underline{\gamma} + 2\zeta, \ldots \overline{\gamma}]$, for $\zeta = \frac{(\overline{\gamma} - \underline{\gamma})}{0.7n}$. In this way, the refined grid-search captures non-observed, but possible threshold values from the threshold variable q.[3]

The sum of squared errors function for γ is given by

$$S_n(\gamma) = S_n(\theta(\gamma), \delta(\gamma), \gamma) = y'y - y' X_\gamma^* (X_\gamma^{*'} X_\gamma)^{-1} X_\gamma^{*'} y. \tag{4}$$

and the estimate of γ is given by the value that minimizes (4):

$$\hat{\gamma} = \underset{\gamma \in \Gamma_g}{\arg\min} S_n(\gamma). \tag{5}$$

for $g = s, r$ so that $\Gamma_g = \Gamma_s$ for the standard grid-search procedure and $\Gamma_g = \Gamma_r$ for the refined grid-search

[2] Assumptions made in this paper are equivalent to those in Hansen (2000) and we omit these for brevity.

[3] We trim the first and last 15% of the threshold observations for both grid-search procedures in our Monte Carlo simulations.

procedure.

3 Confidence Intervals for Threshold Parameters

We evaluate the performance of two approaches to constructing confidence intervals for each of the grid-search procedures described above. The two approaches are described in the following subsections.

3.1 Benchmark ILR Approach

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

$$LR_n(\gamma) = n\frac{S_n(\gamma) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})} \tag{6}$$

with $S_n(\gamma)$ defined as in equation (4). It is well known that the distribution of the LR statistic in (6) is non-standard.

The $1-\alpha$ ILR confidence set for the threshold parameter consists of all the possible values of $\gamma \in \Gamma_g$, $g = s, r$ for which the null hypothesis would not be rejected at the α level:

$$C_d = \{\gamma : LR_n(\gamma) \leq CV_{1-\alpha}, \gamma \in \Gamma_g\} \tag{7}$$

where $CV_{1-\alpha}$ is the critical value derived by Hansen (2000) and $g = s, r$. Note that the confidence set in (7) may be disjoint. However, we can construct a convexified confidence interval by connecting all disjoint segments, which we set as the *benchmark* confidence interval in this paper.

To illustrate the two different approaches to constructing confidence intervals, we display a hypothetical LR profile in Figure 1. Let $q(j)$ denote the j-th ordered possible threshold value among all $q_i \in \Gamma_g$. Suppose the l-th possible threshold value $q(l)$ and the u-th possible threshold value $q(u)$ are the boundaries of the ILR confidence interval, defined as the minimum and maximum values in the ILR confidence set (7), respectively:

$$q(l) = \min\{q_i : LR_n(q_i) \leq CV_{1-\alpha}, q_i \in \Gamma_s\} \tag{8}$$

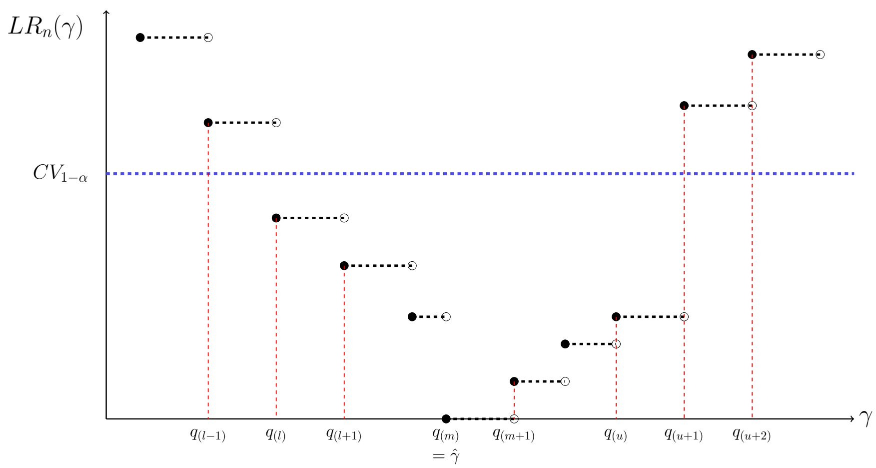
$$q(u) = \max\{q_i : LR_n(q_i) \leq CV_{1-\alpha}, q_i \in \Gamma_s\} \tag{9}$$

Then, the $1-\alpha$ benchmark ILR confidence interval is given by

$$C_b = \{\gamma : q(l) \leq \gamma \leq q(u)\} \tag{10}$$

where $q(l)$ and $q(u)$ are defined in (8) and (9), respectively. See Figure 1.

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 $q_i \in \big[q(j), q(j+1)\big)$, there is the same likelihood ratio value $LR_n(q_i) = LR_n\big(q(j)\big)$, denoted by a dashed line. 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.

Theoretically, because the confidence interval is constructed by completing the disjoint segments in (7), the coverage rate of the benchmark (10) is expected to be greater than $1-\alpha$, at least asymptotically in the case of iid Gaussian errors (see Hansen (2000)). However, the empirical coverage rate can be far lower (Gonzalo and Wolf, 2005; Enders, Falk and Siklos, 2007). This discrepancy motivates us to propose a conservative version of the likelihood-ratio-based confidence interval in (10).

3.2 Conservative ILR Approach

The motivation for the conservative modification of the standard ILR approach stems from the fact that we use a step function approximation of the likelihood function for possible values of the threshold that we do not observe (i.e., any points of $\gamma \notin \Gamma_s$) because Γ_s is a collection of discrete observations in the parameter space of Γ in finite samples. Specifically, we cannot determine the actual 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 $\acute{\gamma} \in \big(q(u), q(u+1)\big)$ such that $LR_n(\acute{\gamma}) \leq CR_{1-\alpha}$.[4] If these values are not included in the confidence interval, it may exclude the true threshold value and its coverage rate could be

[4]Similarly, it is possible that there are some threshold parameter values $\grave{\gamma} \in \big(q(l-1), q(l)\big)$ such that $LR_n(\grave{\gamma}) \leq CR_{1-\alpha}$ where $q(l-1) = \max\{q_i : LR_n(q_i) > CV_{1-\alpha}, q_i < q(l), q_i \in \Gamma_s\}$.

far lower than $1 - \alpha$.

Indeed, the benchmark ILR approach attains unsatisfactory coverage rates because LR evaluations based on $\{q_i\}_{i=1}^{n}$ are poor approximations to the LR from the true DGP. In particular, we argue that the LR profile evaluated at the discrete values of the observed threshold variable in the sample (a step function) is a bad approximation to the asymptotic LR profile in finite samples.[5] To overcome this issue, we modify the ILR approach by means of a conservative approach.

Intuitively, the conservative approach accounts for non-observed, but possible thresholds whose LR values are lower than the critical value by extending the benchmark ILR confidence interval to include the possible threshold value less than, but closest to $q(l)$ in (8) and the possible threshold value larger than, but closest to $q(u)$ in (9). Formally,

$$q(l-1) = \max\{q_i : q_i \in \Gamma_s, q_i < q(l)\} \tag{11}$$

$$q(u+1) = \min\{q_i : q_i \in \Gamma_s, q_i > q(u)\} \tag{12}$$

for $q(l)$ and $q(u)$ defined in (8) and (9), respectively. Based on Figure 1, thus, we can define the conservative confidence interval as follows:

$$C_c = \{\gamma : q(l-1) \leq \gamma \leq q(u+1)\} \tag{13}$$

where $q(l-1)$ and $q(u+1)$ are defined in (11) and (12), respectively. Therefore, the conservative confidence interval (13) includes all non-observable threshold values between $q(l-1)$ and $q(l)$ and between $q(u)$ and $q(u+1)$. Notice that, by construction, the conservative confidence interval C_c in (13) is always longer than the benchmark ILR confidence interval.

4 Monte Carlo Experiments

To evaluate and contrast the finite sample performance of the different CIs, we examine the empirical coverage rates, the average lengths, and the average number of threshold observations contained in the CIs by means of Monte Carlo simulations. 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.

[5] When the threshold effect is large or fixed, the empirical LR profile is too 'sharp' and a sequence of LR tests for the possible threshold values are rejected too often, leading to the inclusion of too few sample observations of the threshold variable being included in the ILR confidence intervals. Too few observations in the confidence intervals means that there is not enough information to approximate the LR profile and to correctly make inferences about the true threshold parameter. We confirm this conjecture in our Monte Carlo simulations.

Similarly, the average number of threshold observations is defined as the number of threshold observations that the confidence interval contains averaged across Monte Carlo simulations. For ease of comparison, the average lengths for all approaches are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample, $\overline{\gamma} - \underline{\gamma}$, while the average number of threshold observations is expressed as a percentage of the sample size. The sample sizes are set to $n = 50$, 100, 250, 500 and $1,000$ and we consider 1,000 Monte Carlo replications for each experiment.[6]

We consider two different DGPs previously examined in the literature by Tong (1990) and Hansen (2000) to evaluate the performance of the proposed approaches in different settings.[7]

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

In the first experiment, we generate data according to the following self-exciting TAR (SETAR) model:

$$y_i = \begin{cases} \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} \leq \gamma \\ \beta_0 + \sum_{j=1}^{p} \beta_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} > \gamma \end{cases} \tag{14}$$

where $\epsilon_i \sim N(0, 1)$ for $i = 1, \ldots, n$. 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 1 shows empirical coverage rates, average lengths and average number of threshold observations associated with the DGP described in (14) and different sample sizes. The benchmark and conservative approaches computed using the standard grich-search are ILR_b and ILR_c, respectively. Those computed using the refined grid-search are ILR_b^r and ILR_c^r, respectively. The results show that the benchmark approach, ILR_b, performs poorly in the sense that the coverage rates range from 0.558 to 0.653, far below the nominal level for any sample size. The refined grid-search procedure helps by accounting for non-observable threshold values, thus increasing the coverage rates of the ILR_b^r approach. This is evidenced by the larger percentage of threshold observations included in the ILR_b^r approach relative to the ILR_b approach for each sample size. However, the improvement is only marginal and their coverage rates vary from 0.646 to 0.672. Meanwhile, the conservative approach for both grid-search procedures, ILR_c and ILR_c^r, produce coverage rates that are higher than nominal levels, overcovering the true threshold parameter, from 0.987 to 0.996.

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) $z_i = q_i$ and (ii) $z_i \sim iidN(0, 1)$. Partitioning

[6] Each series was generated for $(n + 200)$ observation with the first 200 observations discarded to avoid any distortions from initial values.

[7] We also considered a third DGP examined by Chan and Tsay (1998). All approaches performed well in terms of coverage rates and average lengths. Results are available from the authors upon request.

$\delta_n = (\delta_1 \;\; \delta_2)'$, 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. The sample size n varies as before and each approach is labeled in the same fashion as in Table 1.

The results are reported in panels (a) and (b) of Table 2 for the cases with $z_i = q_i$ and $z_i \sim iid \;\; N(0,1)$, respectively. In all cases, the refined grid-search approach generates confidence intervals with slightly higher coverage rates relative to the standard grid-search approach, but the increase is only marginal. Therefore, our discussion below focuses on the distinction between the benchmark, ILR_b and conservative ILR_c approaches, since the performances of ILR_b^r and ILR_c^r are similar to those, respectively.

From panel (a), when the threshold effect is small ($\delta_2 = 0.25, 0.50$), all approaches slightly overcover for most sample sizes, with the exception of the ILR_b approach which slightly undercovers for $n \leq 100$. In general terms, the normalized average lengths and average number of threshold observations for all approaches are not very different in small samples ($n \leq 100$) for $\delta_2 = 0.25$. As the threshold effect δ_2 becomes larger, the ILR_b approach produces the coverage rates far below the nominal level. For example, when $\delta_2 = 2.00$ the coverage rates of the ILR_b approach range from 0.136 to 0.161 while the ILR_c approach always produces the coverage rates greater than the nominal level, 0.996 to 1.000.

The coverage rate of the ILR_c approach increases across sample sizes to the point of reaching almost 100% for $\delta_2 = 2.00$. This is explained by the increasing normalized average lengths relative to the ILR_b approach, which becomes even 7 or 8 times as long for $\delta_2 = 2.00$. Not only is the benchmark confidence interval substantially shorter, but it also includes few elements. While the average number of threshold observations monotonically decreases with the sample size, the conservative approach includes more than twice as many threshold observations for $\delta_2 = 2.00$ (as percentage of the sample size). However, note that in absolute terms the length difference is minor. For example, when $\delta_2 = 2.00$ and $n = 250$, the lengths for ILR_b and ILR_c (normalized by the length of the middle 70% parameter space) are 0.002 and 0.016 while the coverage rates are 0.139 and 0.998, respectively.

By contrast, the coverage rates of the benchmark approach fall with the threshold effect for all n. Intuitively, the identification of the threshold parameter is very precise as the threshold effect δ_2 increases. Hence, the confidence intervals become very narrow and include very few points. In some cases, even just one threshold value as inferred from the average number of threshold observations.[8] This is consistent with the ILR_b approach exhibiting average lengths that are 7 times shorter than the conservative confidence intervals, as discussed above.

Overall, these results suggest that, even when the estimation of the threshold becomes very precise for large threshold effects, the conservative confidence intervals still contain enough possible values to include the true threshold parameter. The results for the case with $z_i \sim iid \;\; N(0,1)$, reported in panel (b) of Table 2, are not much different, qualitatively, from those reported in panel (a). In general, the confidence intervals

[8] For example, the reported 2.79% average number of observations for $n = 50$ and $\delta_2 = 2.00$ corresponds to an average of $0.0279 \times 50 = 1.40$ threshold observations included in Hansen's (2000) benchmark confidence intervals, while the 0.13% reported for $n = 1,000$ and $\delta_2 = 2.00$ corresponds to an average of $0.13 \times 1,000 = 1.30$ threshold observations.

from the benchmark and conservative approaches exhibit accurate coverage probabilities for small threshold effects, but the coverage rates of the benchmark approach fall below 95%. Meanwhile, the ILR_c approach overcovers, especially for large values of δ_2 for the same reasons as explained above.

Finally, it is important to note that our results are different from those reported in Table II of Hansen (2000). We are only able to replicate those results when evaluating the LR at the true threshold value in the DGPs, which is unknown in practice. By contrast, for our results here, we evaluate the LR at the values of the threshold variable observed in the sample when constructing the confidence intervals.

5 Concluding Remarks

Using Monte Carlo simulations, we have shown that asymptotically valid likelihood-ratio-based confidence intervals may perform poorly, even for large samples, when the threshold effect is large. The coverage rates of the benchmark confidence interval derived in Hansen (2000) are substantially below nominal levels. Intuitively, a large threshold effect results in precise estimation of the threshold parameter, which leads to the inclusion of too few possible threshold values in the confidence intervals. A conservative modification of Hansen's benchmark approach yields coverage rates that are higher than 95%, while still being informative in the sense of including relatively few observations of the threshold variable.

References

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.

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.* , ed. Nigar Hashimzade and Michael A. Thornton, 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.

Potter, Simon. 1995. “A Nonlinear Approach to U.S. GNP.” *Journal of Applied Econometrics*, 10: 109–25.

Tong, Howell. 1990. *Non-linear Time Series. A Dynamical System Approach.* New York:Oxford University Press.

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

	Coverage Rate					Average Length					Av. # of thresholds				
$n =$	50	100	250	500	1,000	50	100	250	500	1,000	50	100	250	500	1,000
ILR_b	0.653	0.612	0.612	0.580	0.558	0.167	0.045	0.014	0.006	0.003	10.29	3.31	1.89	0.55	0.28
ILR_c	0.987	0.992	0.996	0.994	0.990	0.227	0.077	0.027	0.013	0.007	14.18	5.33	3.23	0.95	0.48
ILR_b^r	0.672	0.658	0.683	0.646	0.656	0.185	0.051	0.017	0.008	0.004	12.84	4.26	2.56	0.75	0.38
ILR_c^r	0.986	0.996	0.995	0.997	0.992	0.240	0.080	0.028	0.013	0.007	16.69	6.25	3.89	1.13	0.58

Note: The average lengths are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample size. The non-normalized average lengths for ILR_b are 0.556, 0.153, 0.047, 0.022, 0.011 for sample sizes 50, 100, 250, 500, and 1,000 respectively. The average number of threshold observations is expressed as a percentage of the sample size. In our Monte Carlo simulations, we consider a SETAR model with a DGP:

$$y_i = \begin{cases} 0.70 - 0.50 y_{i-1} + \epsilon_i, & \text{if } y_{i-1} \leq 0 \\ -1.80 + 0.70 y_{i-1} + \epsilon_i, & \text{if } y_{i-1} > 0 \end{cases} \tag{15}$$

where $\epsilon_i \sim N(0,1)$ for $i = 1, \ldots, n$.

Table 2: Coverage and average length of 95% confidence intervals: Hansen (2000)

(a) $z_i = q_i$

	Coverage Rate					Average Length					Av. # of thresholds				
$n =$	50	100	250	500	1,000	50	100	250	500	1,000	50	100	250	500	1,000
								$\delta_2 = 0.25$							
ILR_b	0.943	0.952	0.960	0.970	0.976	0.885	0.859	0.675	0.420	0.174	59.42	55.95	40.54	23.22	9.27
ILR_c	0.952	0.962	0.966	0.976	0.980	0.897	0.867	0.681	0.424	0.176	60.24	56.44	40.97	23.44	9.46
ILR_b^r	0.955	0.960	0.958	0.974	0.976	0.901	0.869	0.685	0.426	0.175	63.45	57.99	41.36	23.20	8.75
ILR_c^r	0.961	0.968	0.968	0.980	0.978	0.911	0.876	0.690	0.431	0.178	64.15	58.45	41.77	23.52	8.94
								$\delta_2 = 0.50$							
ILR_b	0.907	0.937	0.949	0.957	0.958	0.701	0.528	0.193	0.053	0.023	45.88	31.69	10.74	3.33	1.59
ILR_c	0.954	0.970	0.978	0.976	0.984	0.728	0.547	0.203	0.058	0.025	47.74	33.06	11.50	3.73	1.79
ILR_b^r	0.922	0.929	0.948	0.946	0.957	0.728	0.546	0.200	0.054	0.023	49.97	33.57	10.80	3.16	1.51
ILR_c^r	0.958	0.972	0.982	0.982	0.985	0.751	0.565	0.211	0.060	0.026	51.65	34.88	11.56	3.56	1.71
								$\delta_2 = 1.00$							
ILR_b	0.763	0.745	0.765	0.778	0.774	0.265	0.082	0.022	0.011	0.005	17.24	5.58	1.76	0.86	0.42
ILR_c	0.977	0.981	0.986	0.993	0.992	0.315	0.110	0.032	0.016	0.008	20.79	7.56	2.56	1.26	0.62
ILR_b^r	0.790	0.759	0.752	0.765	0.787	0.286	0.092	0.023	0.011	0.006	19.38	6.40	1.92	0.96	0.48
ILR_c^r	0.982	0.986	0.987	0.991	0.994	0.335	0.120	0.034	0.017	0.008	22.86	8.37	2.72	1.36	0.68
								$\delta_2 = 1.50$							
ILR_b	0.403	0.400	0.391	0.396	0.385	0.061	0.021	0.007	0.003	0.002	5.35	2.17	0.82	0.41	0.20
ILR_c	0.993	0.991	0.991	0.997	0.999	0.123	0.053	0.020	0.010	0.005	9.32	4.17	1.62	0.81	0.40
ILR_b^r	0.592	0.583	0.554	0.536	0.516	0.080	0.032	0.011	0.005	0.003	7.07	3.17	1.14	0.55	0.28
ILR_c^r	0.996	0.996	0.994	1.000	0.996	0.135	0.060	0.022	0.011	0.005	11.03	5.17	1.94	0.95	0.48
								$\delta_2 = 2.00$							
ILR_b	0.150	0.161	0.139	0.144	0.136	0.014	0.007	0.002	0.001	0.001	2.79	1.36	0.52	0.26	0.13
ILR_c	0.997	1.000	0.998	0.998	0.996	0.082	0.043	0.016	0.008	0.004	6.79	3.36	1.32	0.66	0.33
ILR_b^r	0.459	0.466	0.448	0.422	0.400	0.039	0.019	0.007	0.004	0.002	4.71	2.34	0.91	0.45	0.22
ILR_c^r	0.996	0.997	0.999	0.996	0.998	0.094	0.048	0.019	0.009	0.005	8.70	4.34	1.71	0.85	0.42

Note: The average lengths are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample size. The non-normalized average lengths for ILR_b are 0.029, 0.014, 0.005, 0.002, 0.001 for $\delta_2 = 2.00$ and sample sizes 50, 100, 250, 500, and 1,000 respectively. Other average lengths are available upon request. The average number of threshold observations is expressed as a percentage of the sample size. In our Monte Carlo simulations, we consider a threshold model with a DGP:

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

where $z_i = q_i \sim N(2,1)$ and $\epsilon_i \sim N(0,1)$ for $i = 1, \ldots, n$.

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

	Coverage Rate					Average Length					Av. # of thresholds				
$n=$	50	100	250	500	1,000	50	100	250	500	1,000	50	100	250	500	1,000
								$\delta_2 = 0.25$							
ILR_b	0.953	0.950	0.973	0.969	0.961	0.913	0.893	0.838	0.675	0.440	63.56	59.83	54.98	41.87	25.90
ILR_c	0.958	0.957	0.973	0.970	0.964	0.923	0.900	0.842	0.679	0.442	64.22	60.27	55.26	42.13	26.08
ILR_b^r	0.964	0.955	0.973	0.967	0.965	0.925	0.899	0.844	0.679	0.443	66.50	61.44	55.20	41.01	24.41
ILR_c^r	0.968	0.960	0.974	0.969	0.967	0.933	0.905	0.847	0.682	0.445	67.05	61.84	55.47	41.26	24.59
								$\delta_2 = 0.50$							
ILR_b	0.954	0.945	0.964	0.965	0.979	0.820	0.728	0.459	0.226	0.098	56.71	47.82	28.07	14.22	6.44
ILR_c	0.964	0.956	0.975	0.974	0.983	0.840	0.743	0.469	0.231	0.101	58.01	48.83	28.77	14.62	6.64
ILR_b^r	0.954	0.951	0.966	0.971	0.974	0.834	0.737	0.463	0.228	0.099	59.70	48.76	27.19	13.06	5.72
ILR_c^r	0.968	0.956	0.974	0.976	0.986	0.851	0.750	0.473	0.234	0.102	60.89	49.73	27.88	13.45	5.92
								$\delta_2 = 1.00$							
ILR_b	0.925	0.934	0.943	0.947	0.944	0.545	0.316	0.101	0.048	0.025	37.57	20.60	7.22	3.50	1.79
ILR_c	0.964	0.971	0.973	0.983	0.983	0.585	0.343	0.111	0.053	0.027	40.48	22.50	8.02	3.90	1.99
ILR_b^r	0.928	0.932	0.932	0.927	0.943	0.554	0.320	0.102	0.048	0.025	38.83	20.37	6.74	3.24	1.66
ILR_c^r	0.968	0.970	0.975	0.986	0.988	0.593	0.347	0.113	0.054	0.027	41.65	22.27	7.54	3.64	1.86
								$\delta_2 = 1.50$							
ILR_b	0.863	0.861	0.866	0.893	0.865	0.277	0.132	0.049	0.023	0.012	19.67	9.65	3.74	1.80	0.91
ILR_c	0.978	0.986	0.988	0.992	0.989	0.331	0.158	0.059	0.028	0.014	23.50	11.64	4.54	2.20	1.11
ILR_b^r	0.887	0.858	0.874	0.883	0.867	0.285	0.133	0.049	0.024	0.012	20.46	9.58	3.65	1.76	0.89
ILR_c^r	0.981	0.990	0.991	0.993	0.990	0.338	0.161	0.061	0.029	0.015	24.26	11.57	4.45	2.16	1.09
								$\delta_2 = 2.00$							
ILR_b	0.801	0.788	0.795	0.800	0.810	0.169	0.079	0.030	0.015	0.007	13.17	6.19	2.45	1.23	0.60
ILR_c	0.984	0.986	0.986	0.990	0.989	0.223	0.107	0.041	0.021	0.010	17.14	8.19	3.25	1.63	0.80
ILR_b^r	0.801	0.802	0.796	0.764	0.790	0.179	0.083	0.032	0.016	0.008	14.12	6.52	2.55	1.26	0.61
ILR_c^r	0.987	0.998	0.991	0.990	0.994	0.234	0.111	0.043	0.021	0.010	18.08	8.52	3.35	1.66	0.81

Note: The average lengths are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample size. The non-normalized average lengths for ILR_b are 0.346, 0.160, 0.063, 0.032, 0.015 for $\delta_2 = 2.00$ and sample sizes 50, 100, 250, 500, and 1,000 respectively. Other average lengths are available upon request. The average number of threshold observations is expressed as a percentage of the sample size. In our Monte Carlo simulations, we consider a threshold model with a DGP:

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

where $q_i \sim N(2,1)$, $z_i \sim N(0,1)$, and $\epsilon_i \sim N(0,1)$ for $i = 1, \ldots, n$.