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 §

July 10, 2016

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; Confidence Interval; Finite-sample inference

JEL classification: C13, C20

* All errors are our own.

† *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 specify that regression functions can be divided into several regimes based on the value of an observed variable, called a threshold variable, related to threshold parameters. Threshold regression models and their various extensions have become standard for the specification of nonlinear relationships between economic variables (see Potter, 1995; Balke, 2000; Koop and Potter, 2004; Gonzalo and Pitarakis, 2013, among many others.)[1]. There have been important developments in the asymptotic theory for inference in threshold regression models (see Chan, 1993; Hansen, 1996; Chan and Tsay, 1998; Hansen, 2000). However, Enders, Falk and Siklos (2007) show that when the threshold parameter is unknown, asymptotic and bootstrap approximations of finite sample distributions do not result in satisfactory confidence intervals (CIs) for slope or threshold parameters in stationary threshold autoregressive models.

In this paper, we are particularly interested in the finite sample performance of asymptotically-valid likelihood-ratio-based CIs for the threshold parameter proposed by Hansen (1997, 2000). Using Monte Carlo experiments, we show that the performance of the CIs becomes particularly problematic in finite samples when the threshold effect is relatively large. This finding is puzzling because the coverage rates of CIs are expected to converge to a nominal level when the threshold effect increases (i.e. there is more precise information about the true threshold value).

We argue that when the threshold effect is large, the approximation of the profile likelihood becomes poor and leads to lower coverage rates of the CIs. As noted above, we would expect large threshold effects to help the CIs achieve accurate coverage rates relative to a nominal level given the benefits in terms of econometric identification. However, the large threshold effects also lead to a poor approximation to the profile likelihood for the threshold parameter. Thus, a large threshold effect has two conflicting impacts and the performance of the CIs depends on which impact is bigger. When the magnitude of the threshold effect is particularly large, the poor approximation dominates the benefit from the more precise information and the standard CIs perform poorly.

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

Why does the large threshold effect make the approximation so poor? To construct the CIs, Hansen (2000) inverts the likelihood-ratio test for the threshold parameter by evaluating the profile likelihood at observed threshold values and includes the threshold values for which the likelihood-ratio test cannot be rejected. The asymptotic theory for the likelihood-ratio test is developed under the assumption that the threshold variable is distributed with a *continuous* distribution. However, in finite samples, the threshold variables are observed discretely and the profile likelihood for the test is constructed using a step function approximation for the threshold values that are not observed in the sample. When the threshold effect is small (i.e. there is less information about the true threshold value), the likelihood-ratio tests for the threshold parameter are rarely rejected and the CIs includes many threshold values. Thus, the step function would approximate the likelihood function effectively when constructing the CIs. However, when the threshold effect is large, the likelihood-ratio tests for the threshold parameters are rejected too often and the CIs include few threshold observations. With few observations, the step function then becomes a poor approximation of the likelihood and the CIs may exclude the true threshold parameter, resulting in low coverage rates, even in large samples.

We consider two possible modifications to Hansen's inverted likelihood-ratio (ILR) approach in order to address the step function approximation: (i) an equally-spaced grid-search approach; 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). 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' profile likelihood associated with a large threshold effect, which results in too few possible threshold values being included in the CIs. Thus, the large threshold effect leads to a poor approximation of the profile likelihood in finite samples. The refined grid-search improves the performance by including some of the non-observed, but possible threshold values, but the coverage rates are still far below the nominal level in 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.

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. Then, the grid-search procedure occurs over $\Gamma_n = \Gamma \cap \{q_i\}_{i=1}^n$, so that all elements

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

in Γ_n are simply all observed values of q_i between $\underline{\gamma}$ and $\overline{\gamma}$.

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):

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

3 Confidence Intervals for Threshold Parameters

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 $\gamma \in \Gamma_n$ for which the null hypothesis would not be rejected at the α level:

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

where $CV_{1-\alpha}$ is the critical value derived by Hansen (2000). 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 benchmark approach 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 $q_i \in \Gamma_n$. 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:

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

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

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.

Theoretically, because the confidence interval is constructed by completing the disjoint segments in (7), the coverage rate of the benchmark interval (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 when the threshold effect is particulary large. We will show this in our Monte Carlo experiments in Section 4. 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 to 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 $\gamma \notin \{q_i\}_{i=1}^n$) because Γ_n is a collection of discrete observations in the parameter space of Γ in finite samples. Specifically, the threshold values between $q(u)$ and $q(u+1)$ and between $q(l-1)$ and $q(l)$ are excluded in the benchmark confidence interval. However, it is likely that there are some threshold parameter values $\acute{\gamma} \in \big(q(u), q(u+1)\big)$ such that $LR_n(\acute{\gamma}) \leq CR_{1-\alpha}$.[3] If these values are not included in the confidence interval, it may exclude the true threshold value and its coverage

[3] 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 $q(l-1) = \max\{q_i : LR_n(q_i) > CV_{1-\alpha}, q_i < q(l), q_i \in \Gamma_n\}$.

rate could be far lower than $1-\alpha$.

Indeed, the benchmark ILR approach attains unsatisfactory coverage rates when the threshold effect is large. This large threshold effect leads to the 'sharp' empirical LR profile. This implies that 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 benchmark ILR confidence intervals. Then, LR evaluations based on $\{q_i\}_{i=1}^n$ are poor approximations to the profile likelihood for threshold parameter γ so that the threshold parameter spaces between $q(u)$ and $q(u+1)$ and between $q(l-1)$ and $q(l)$ become relatively large. The large spaces between $q(u)$ and $q(u+1)$ and between $q(l-1)$ and $q(l)$ would lead to low coverage rates.[4] 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 threshold values whose LR values are lower than the critical value by extending the benchmark ILR confidence interval to include the possible threshold value smaller than, but closest to $q(l)$ in (8) and the possible threshold value larger than, but closest to $q(u)$ in (9) in a conservative way. Formally,

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

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

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

[4]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.

confidence interval C_c in (13) is always longer than the benchmark ILR confidence interval.

3.3 Refined Grid-Search

In addition to the benchmark and conservative approaches, we consider the refined grid-search over the equally-spaced grid $\Gamma_r = \Gamma \cap q^r$ where the elements in q^r are given by $q^r = \{\underline{\gamma}, \underline{\gamma} + \zeta, \underline{\gamma} + 2\zeta, \ldots, \overline{\gamma}\}$ and the size of the grid step is given by $\zeta = \left(\overline{\gamma} - \underline{\gamma}\right)/(0.7n)$.[5] In this way, the refined grid-search captures non-observed, but possible threshold values from the threshold variable q_i. For the refined grid-search, we use the same benchmark and conservative approaches but conduct the likelihood-ratio tests over the equally-spaced gridpoints in Γ_r rather than Γ_n.

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. 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.

We consider two different DGPs to evaluate the performance of the proposed approaches in different settings and 1,000 Monte Carlo replications for each experiment.

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

4.1 Monte Carlo Experiment 1

In the first experiment, we generate data according to

$$y_i = \begin{cases} \alpha_0 + \alpha_1 x_i + \epsilon_i, & \text{if } q_i \leq \gamma \\ \beta_0 + \beta_1 x_i + \epsilon_i, & \text{if } q_i > \gamma \end{cases} \tag{14}$$

where $\alpha_0 = 1$, $\alpha_1 = 1$, $\beta_0 = 1$, $\epsilon_i \sim i.i.d.N(0,1)$ for $i = 1, \ldots, n$. The threshold variable follows $q_i \sim N(3,1)$ and $x_i = q_i$. The true threshold parameter is given by $\gamma_0 = 3$. To see whether the magnitude of the threshold effect affects the performance of the CIs, the slope coefficient β_1 is set to 1.25, 1.50, and 2.00. The threshold effect can be calculated as $\delta = \beta_1 - \alpha_1$. The sample sizes are set to $n = 50$, 100, 250, 500 and $1,000$.

The results are reported in Table 1. The benchmark and conservative approaches using the standard grid-search are ILR_b and ILR_c, respectively. Those using the refined grid-search are ILR_b^r and ILR_c^r, 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.

When the threshold effect is small ($\beta_1 = 1.25$), all approaches slightly overcover for most sample sizes, with the exception of the ILR_b approach which slightly undercovers for $n = 50$. As the threshold effect becomes larger, the ILR_b approach produces coverage rates far below the nominal level. For example, when $\beta_1 = 2.00$ the coverage rates of the ILR_b approach range from 0.37 to 0.41 while the ILR_c approach always produces coverage rates greater than the nominal level, e.g. 0.99 to 1.00. Intuitively, the identification of the threshold parameter is very precise as the threshold effect increases. Hence, the confidence intervals become very narrow and include very few points. This relatively small number of average threshold points results in a poor approximation to the profile likelihood for the threshold parameter γ. Our interpretation on this undercoverage for the ILR_b approach is evidenced by the average threshold points across Monte Carlo simulations included in the CIs: ranging from 7.2 to 15.2 for $\beta_1 = 1.5$ and from 1.97 to 2.56 for $\beta_1 = 2.0$ depending on the sample size. Note that the average number of threshold points ranges from 27.6 to 50.7 when the threshold effect is small ($\beta_1 = 1.25$) so that this large number of the threshold points help approximate the profile likelihood.

In addition, the conservative approach can achieve significantly more accurate coverage rates at the cost of a trivial increase in the normalized average length of the CIs. This increase ranges from 0.3 to 6.1 percentage points.

Overall, the results of this Monte Carlo experiment suggest that the conservative approach can achieve more accurate coverage rates with a relatively marginal increase in the average length in comparison to the benchmark approach.

4.2 Monte Carlo Experiment 2

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

$$y_i = \begin{cases} \alpha_0 + \sum_{j=1}^{p} \alpha_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} \leq \gamma \\ \beta_0 + \sum_{j=1}^{p} \beta_j y_{i-j} + \epsilon_i, & \text{if } y_{i-d} > \gamma \end{cases} \tag{15}$$

To reduce the computational burden, we focus on the simplest case where $p = d = 1$ and set $\alpha_0 = 0$, $\alpha_1 = 0.3$, $\beta_0 = 0.9$, $\beta_1 = 0.6$ and $\gamma_0 = 0$. Because the DGP follows a SETAR model, it is not easy to measure the magnitude of the threshold effect. Thus, we vary the error variance according to $\epsilon_i \sim i.i.d.N(0, \sigma^2)$ for $i = 1, \ldots, n$ and set $\sigma = 0.3, 0.5, 1.0$. A small error variance implies a high signal-to-noise ratio and this specification generates a big threshold effect. The DGP with a unit error variance was studied by Enders, Falk and Siklos (2007), but we consider the various variance sizes to examine the impact of the magnitude of the threshold effect on the performance of the CIs. The sample size is $n = 236$, which is the same as in Enders, Falk and Siklos (2007).

Table 2 presents empirical coverage rates, average lengths and average number of threshold observations across different variance sizes. The results show that the benchmark approach, ILR_b, performs poorly when the threshold effect is large (i.e. $\sigma = 0.3, 0.5$) in the sense that the coverage rates are 0.41 and 0.84, which are far below the nominal level. The refined grid-search procedure helps by accounting for non-observable threshold values, but the improvement is only marginal, resulting in the coverage rates of 0.55 and 0.84, respectively. Meanwhile, the conservative approach ILR_c produces coverage rates that are higher than the nominal level, overcovering the true threshold parameter, from 0.97 to 0.99 at the trivial cost of marginally longer confidence intervals. Note that the normalized lengths of the CIs based on the ILR_c approach are about 1.2 to 1.5 percentage points longer than those for the ILR_b approach.

We find the poor performance of the benchmark approach occurs because the few threshold variable observations included in the CIs produce a poor approximation to the profile likelihood, as argued in the previous section. The average number of the threshold observations in the CIs is 36 when the threshold effect is small ($\sigma = 1$). However, that number falls significantly (about 2 to 6 observations) when the threshold effect becomes large ($\sigma = 0.3, 0.5$).

5 Counterfactual Experiment

The coverage rates are determined by the frequency of Monte Carlo simulations for which the likelihood-ratio test is not rejected at the true threshold parameter value. In the previous sections, we argue that the true threshold value is likely to exist either in $\left(q_{(l-1)}, q_{(l)}\right)$ or in $\left(q_{(u)}, q_{(u+1)}\right)$ when the threshold effect is large, given the poor approximation to the profile likelihood. Any threshold value either in $\left(q_{(l-1)}, q_{(l)}\right)$ or in $\left(q_{(u)}, q_{(u+1)}\right)$ leads to rejecting the test and this results in the poor performance of the CIs. Based on this argument, we have proposed the use of the conservative approach by extending the benchmark confidence interval to include all threshold values in $\left[q_{(l-1)}, q_{(u+1)}\right]$.

To examine whether our argument is valid, we conduct a counterfactual experiment. We repeat the first Monte Carlo experiment with $n = 250$ and $\beta_1 = 1.10, 1.15, ..., 2.00$, but consider two different cases: (i) the true threshold parameter $\gamma_0 \in \{q_i\}_{i=1}^n$; and, (ii) $\gamma_0 \notin \{q_i\}_{i=1}^n$. Thus, in case (i), we force the true threshold value to become observable when generating the threshold variable in the simulation. This setting is a counterfactual experiment because the true threshold value would be included in the data set of the threshold variable with probability 0 if the threshold variable is assumed to follow a continuous distribution as studied in the literature. In this case, the true threshold value must be equal to one of threshold variable observations. Therefore, we can conduct the likelihood-ratio test at the true threshold value against the threshold estimate in each simulation without using the step function approximation:

$$LR_n(\gamma^0) = n\frac{S_n(\gamma^0) - S_n(\hat{\gamma})}{S_n(\hat{\gamma})} \tag{16}$$

where $\gamma_0 \in \{q_i\}_{i=1}^n$. Note that the threshold estimate is not necessarily equal to the true threshold value because the threshold estimate is determined by the threshold value which minimizes the SSR.

If our argument is correct, the simulation setting in case (i) would result in a rejecting frequency of 5% or less at the true threshold value. Hence, the coverage rates would be equal to or greater than 95%.[6] Case (ii) is the same as the Monte Carlo experiment setting in Section 4.1. We construct confidence intervals using the benchmark approach.

Figure 2 plots the coverage rates for the two different cases against the magnitude of the threshold effect. The results show that the coverage rates are equal to or greater than 95% when the true threshold value is included (case (i) so that $\gamma_0 \in \{q_i\}_{i=1}^n$). However, in case (ii), the coverage rates are close to 95% when threshold effect is relatively small but they decrease when the magnitude is significantly large, as shown in Section 4.1.

6 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 particularly large. The coverage rates of the benchmark confidence interval derived in Hansen (2000) are substantially below nominal levels. We have proposed a conservative modification to Hansen's benchmark approach and this modification yields coverage rates that are equal to or higher than 95%, while still being informative in the sense of marginally longer confidence intervals.

[6]Note that the likelihood-ratio test is conservative when the threshold effect is large (see Theorem 3 in Hansen 2000).

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.

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. 2011. "Threshold Models in Time Series Analysis - 30 years on." *Statistics and Its Interface*, 4: 107–136.

Appendix: Tables

Table 1: Monte Carlo Experiment 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
								$\beta_1 = 1.25$							
ILR_b	0.94	0.96	0.96	0.96	0.96	0.825	0.706	0.364	0.137	0.044	27.62	44.51	50.72	37.90	29.35
ILR_c	0.95	0.97	0.97	0.97	0.97	0.843	0.720	0.373	0.142	0.047	28.18	45.46	52.39	39.85	31.35
ILR_b^r	0.94	0.96	0.96	0.96	0.96	0.847	0.722	0.370	0.140	0.044	29.66	46.63	50.76	35.93	26.74
ILR_c^r	0.96	0.98	0.97	0.97	0.97	0.860	0.734	0.379	0.146	0.047	30.15	47.51	52.41	37.88	28.74
								$\beta_1 = 1.50$							
ILR_b	0.86	0.88	0.91	0.90	0.92	0.475	0.206	0.045	0.020	0.009	15.22	12.34	8.02	7.41	7.20
ILR_c	0.97	0.98	0.99	0.98	0.98	0.516	0.231	0.055	0.025	0.012	16.62	14.21	10.02	9.41	9.20
ILR_b^r	0.88	0.88	0.88	0.89	0.87	0.503	0.218	0.047	0.021	0.010	16.96	13.36	8.14	7.59	7.35
ILR_c^r	0.98	0.98	0.99	0.99	0.99	0.541	0.245	0.059	0.026	0.013	18.30	15.22	10.13	9.59	9.35
								$\beta_1 = 2.00$							
ILR_b	0.41	0.39	0.41	0.37	0.39	0.053	0.018	0.008	0.003	0.002	2.56	2.12	2.07	1.97	2.00
ILR_c	0.99	0.99	1.00	0.99	1.00	0.114	0.050	0.020	0.010	0.005	4.54	4.12	4.07	3.97	4.00
ILR_b^r	0.57	0.54	0.55	0.53	0.52	0.075	0.027	0.011	0.005	0.003	3.47	2.88	2.90	2.77	2.79
ILR_c^r	1.00	1.00	1.00	1.00	1.00	0.130	0.056	0.022	0.011	0.005	5.46	4.88	4.90	4.76	4.79

For the first experiment, we consider a threshold model with the following DGP:

$$y_i = \begin{cases} 1 + x_i + \epsilon_i, & \text{if } q_i \leq 3 \\ 1 + \beta_1 x_i + \epsilon_i, & \text{if } q_i > 3 \end{cases}$$

where $x_i = q_i \sim N(3,1)$ and $\epsilon_i \sim i.i.d.N(0,1)$ for $i = 1, \ldots, n$. The average lengths are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample size, while the average number of threshold observations is expressed as a percentage of the sample size.

Table 2: Monte Carlo Experiment 2

	Coverage Rate			Average Length			Av. # of thresholds		
$\sigma =$	1.0	0.5	0.3	1.0	0.5	0.3	1.0	0.5	0.3
ILR_b	0.95	0.84	0.41	0.284	0.038	0.008	36.00	5.71	2.04
ILR_c	0.97	0.99	0.99	0.294	0.050	0.023	37.78	7.71	4.04
ILR_b^r	0.95	0.84	0.55	0.288	0.040	0.012	37.13	6.93	2.99
ILR_c^r	0.98	0.99	1.00	0.299	0.052	0.024	38.90	8.93	4.99

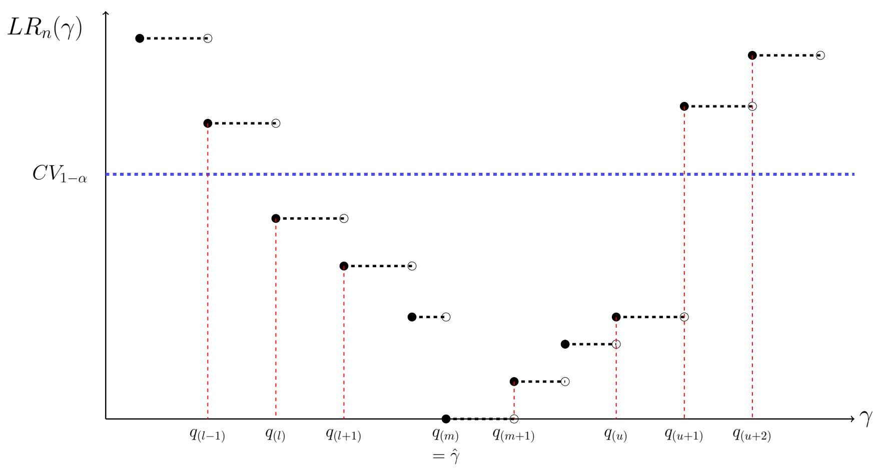
Note: We consider a SETAR model and set $\sigma = 1.0, 0.5, 0.3$ for the error variance. The DGP is given by

$$y_i = \begin{cases} 0.9 + 0.6y_{i-1} + \epsilon_i, & \text{if } y_{i-1} \leq 0 \\ 0.0 + 0.3y_{i-1} + \epsilon_i, & \text{if } y_{i-1} > 0 \end{cases} \tag{17}$$

where $\epsilon_i \sim i.i.d.N(0, \sigma^2)$ for $i = 1, \ldots, n$ and $n = 236$. The average lengths are normalized by the length of the bounded parameter space $\Gamma = [\underline{\gamma}, \overline{\gamma}]$ for each sample size, while the average number of threshold observations is expressed as a percentage of the sample size.

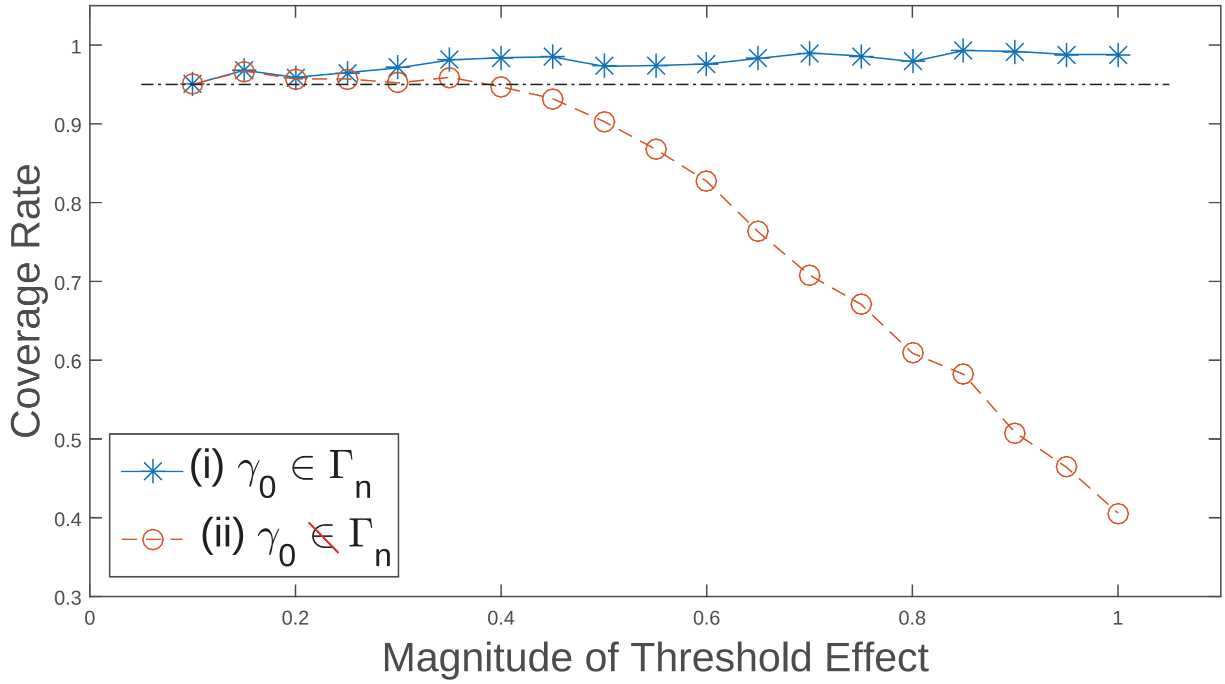
Appendix: Figures

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. The critical value $CV_{1-\alpha}$ is indicated by a blue dashed line.

Figure 2: Counterfactual Experiment



Note: The true threshold parameter is γ_0 and the magnitude of the threshold effect is measured by $\delta = \beta_1 - \alpha_1$.