An Article Submitted to

Studies in Nonlinear Dynamics & Econometrics

Manuscript 1322

A Threshold Model of Real U.S. GDP and the Problem of Constructing Confidence Intervals in TAR Models

WALTER ENDERS* BARRY L. FALK† PIERRE SIKLOS‡

*UNIVERSITY OF ALABAMA, wenders@cba.ua.edu
†Iowa State University, bfalk@iastate.edu
‡Wilfred Laurier University, psiklos@wlu.ca

A Threshold Model of Real U.S. GDP and the Problem of Constructing Confidence Intervals in TAR Models *

WALTER ENDERS, BARRY L. FALK, and PIERRE SIKLOS

Abstract

We estimate real US GDP growth as a threshold autoregressive process, and construct confidence intervals for the parameter estimates. However, there are various approaches that can be used in constructing the confidence intervals. We construct confidence intervals for the slope coefficients and the threshold using asymptotic results and bootstrap methods, finding that the results for the different methods have very different economic implications. We perform a Monte Carlo experiment to evaluate the various methods. Surprisingly, the confidence intervals are wide enough to cast doubt on the assertion that the time-series responses of GDP to negative growth rates are different than the responses to positive growth rates.

*Walter Enders Department of Economics, Finance and Legal Studies University of Alabama Tuscalooosa, AL 35487 wenders@cba.ua.edu 1-205-348-8972 Barry Falk Department of Economics Iowa State University Ames, IA 50011 bfalk@iastate.edu 1-515-294-5875 Pierre L. Siklos Department of Economics Wilfrid Laurier University Waterloo, ON N2L 3C5 1-519-884-1970

1. Introduction

Threshold autoregressive (TAR) models are popular, providing a straightforward, economically appealing, and econometrically tractable nonlinear extension of the linear autoregressive (AR) model. TAR models are particularly suited for time-series processes that are subject to periodic shifts due to regime changes. Examples of early applications include Burgess (1992), Cao and Tsay (1992), Enders and Granger (1998), Galbraith (1996), Hansen (1997), Krager and Kugler (1993), Potter (1995), and Rothman (1991).

Although there have been a number of important developments in the asymptotic theory for estimation and inference in TAR models [e.g., Hansen (1997, 2000), Chan and Tsay (1998), and Gonzalo and Wolf (2005)], there has been little research concerning the finite sample properties of these procedures. There are a number of ways to perform inference in TAR models and we explore the small-sample properties of some of these methods using Monte Carlo experiments. One complicating factor is the need to know if the process is continuous at the threshold. The issue is important as a comparison of Hansen (1997) and Chan and Tsay (1998) indicates that the distributions relevant for inference in a continuous (C-TAR) model are different from those in a discontinuous (D-TAR) model. Whereas economic analysis may predict the existence of different regimes, it may not be clear whether a C-TAR or a D-TAR model is most appropriate. Enders and Siklos (2004) show that it is the combined values of the intercepts, threshold and autoregressive coefficients that determine whether the model is continuous at the threshold. Although a C-TAR model can be viewed as a restricted version of a D-TAR model, their work shows that testing this restriction is problematic since conventional test statistics are not asymptotically pivotal (i.e. the asymptotic distribution depends on nuisance parameters).

To address some of these issues, we apply Monte Carlo methods to study the small sample coverage properties of confidence intervals for the slope coefficients and the threshold parameter in the class of first-order, stationary, two-regime threshold autoregressive models. The procedures we consider include approaches suggested by asymptotic theory and bootstrap methods. We show that these confidence intervals have poor coverage in a variety of conditions. As a result, the appropriate way to conduct inference in TAR models in small samples is unclear, particularly when the threshold is unknown, and our results cast doubt on some standard methods. Economic theory should provide guidance for choosing among D-TAR or C-TAR alternatives since there seems to be no other obvious way of choosing one over the other.

We consider the implications of our results for the behavior of real US GDP growth, one of the most widely studied time series. The consensus opinion seems to be that the growth rate of real US GDP is a nonlinear process, perhaps of the threshold autoregressive variety. Several papers, such as Kapetanios (2003), Peel and Speight (1998) and Potter (1995), indicate that GDP behaves very differently in periods of high growth than in periods of low growth. However, we show that different methods of constructing confidence intervals lead to different implications concerning the way that GDP behaves in expansions versus contractions. The confidence intervals for the alternative persistence parameters overlap and the location of the threshold value is unclear. As a result, concluding that there are different degrees of persistence in positive versus negative growth regimes may be problematic, throwing into doubt some widely held beliefs concerning the properties of US real GDP growth.

2. The TAR Model

The simplest D-TAR model can be formulated as follows:

$$y_t = (\alpha_0 + \alpha_1 y_{t-1})I_t + (\beta_0 + \beta_1 y_{t-1})(1-I_t) + \varepsilon_t\,, \quad t = 1, \ldots, T \tag{2.1}$$

where I_t is the indicator function defined in terms of the threshold parameter τ as

$$I_t = \begin{Bmatrix} 1 & if & y_{t-d} > \tau \\ 0 & if & y_{t-d} \le \tau \end{Bmatrix},$$

such that d is the delay parameter, and the ε_t's are *i.i.d.* $(0, \sigma^2)$ random variables.

We assume that $\sigma^2 < \infty$ and that the autoregressive parameters in (2.1) satisfy stationarity conditions. Sufficient conditions for stationarity are: $0 < \alpha_1 < 1$ and $0 < \beta_1 < 1$. Petrucelli and Woolford (1984) provide further discussion of conditions for stationarity and ergodicity.

We will also consider the inference problem for the following C-TAR model that is a constrained version of (2.1):

$$y_t = \tau + \alpha_1(y_{t-1} - \tau)I_t + \beta_1(y_{t-1} - \tau)(1-I_t) + \varepsilon_t\,, \quad t = 1, \ldots, T \tag{2.2}$$

where τ denotes the thresholds parameter, α_1 and β_1 are slope coefficients that satisfy stationarity conditions, and I_t and ε_t are stochastic processes defined as in (2.1).

Model (2.1) has the two attractors: $\alpha_0/(1 - \alpha_1)$ and $\beta_0/(1 - \beta_1)$. Model (2.2) implies that y_t has a unique long-run equilibrium, which is equal to the threshold parameter. The short-run dynamics of the C-TAR model depend on whether the system is above or below the long-run equilibrium. In some applications, this version of the model may be a more "natural" representation of the TAR model than (2.1). Note that, in contrast to version (2.1), version (2.2) implies that y_t is continuous in the neighborhood of the threshold.

Extensions of these models allowing for unit roots and higher order autoregressive terms have been considered in the theoretical and applied literature. Note that if $\alpha_1 = \beta_1$ (and in (2.1), $\alpha_0 = \beta_0$) the D-TAR and C-TAR models collapse to an AR(1) model. However, testing this restriction is complicated by the failure of τ to be identified under the null hypothesis.[1]

The ordinary least squares estimator of α_1 and β_1 is asymptotically efficient when the threshold parameter is known, or when it is unknown but is replaced by a consistent estimator. In particular, conditioning on τ, or a consistent estimator of τ, the least squares t-statistics associated with α_0, α_1, β_0, and β_1 converge in distribution to standard normal random variables. As described in Enders (2004), a grid search over all potential thresholds and delay parameters yields a consistent estimate of the threshold. Consequently, the standard textbook approach to confidence interval construction provides intervals with asymptotically correct coverage.

However, even in the special case of the linear AR(1) model, the OLS t-statistics for the slope coefficient will not be approximately normal or even approximately pivotal in small samples, particularly for values close to the unit root boundary. Monte Carlo simulations by Hansen (1999) illustrate the poor finite sample performance of normal confidence intervals, bootstrap-t, and bootstrap-percentile confidence intervals for the AR(1) model. We will apply Monte Carlo simulations to evaluate the finite sample performance of these intervals for the slope coefficients and for the threshold parameter τ in the D-TAR and C-TAR models.

Confidence interval construction for the threshold parameter τ has been considered by Hansen (1997) for model (2.1), by Chan and Tsay (1998) for model (2.2), and Gonzalo and Wolf (2005) for both models. These procedures involve inverting a likelihood ratio statistic constructed from a model estimated using a consistent estimate of the threshold. Hansen's procedure is based on the limiting distribution of the likelihood ratio statistic for model (2.1), a nonstandard distribution derived by Hansen (1996). Chan and Tsay (1998) show that the limiting distribution of the likelihood ratio statistic is chi-square for model (2.2). The difference in the asymptotic behavior of the likelihood ratio statistic according to whether model (2.1) or model (2.2) is assumed is what motivates our consideration of both models. Gonzalo and Wolf (2005) use sub-sampling to generate confidence intervals for an unknown threshold for both the D-TAR and

[1] For further discussion of this issue, see Andrews and Ploberger (1994) and Hansen (1996).

C-TAR models. They also propose a test for continuity but coverage probabilities are good only for relatively large samples (e.g., $T = 500$).

We apply Monte Carlo simulations to evaluate the finite sample performance of these procedures as well as intervals constructed from bootstrapped distributions of these (asymptotically pivotal) statistics. Specifically, we will examine the finite sample properties of the following types of confidence intervals for the slope parameters:

- The so-called 'normal approximation' uses intervals constructed from the *t*-statistics for α_1 and β_1 obtained from a standard *t* distribution.
- Bootstrap-percentile confidence intervals. For example, a 90% confidence interval for α_1 (β_1) can be constructed from the lowest 5% and highest 95% of the ordered bootstrapped estimates of α_1 (β_1).
- Bootstrap-*t* confidence intervals (which assume that the *t*-statistics are approximately pivotal though not necessarily student-*t*).[2] For example, a 90% confidence interval for α_1 (β_1) can be constructed from inverting the lowest 5% and highest 95% of the ordered bootstrapped *t*-statistics for the null hypothesis $\alpha_1 = 0$ ($\beta_1 = 0$).

Confidence intervals for the threshold parameter itself will be constructed from inversion of the likelihood ratio statistic using its asymptotic and bootstrapped distributions. In addition, the bootstrap percentile methods will be applied to construct these intervals.

3. Confidence Intervals

For each parameterization of the TAR model, 1000 realizations of $y_1, \ldots, y_T$ were generated for $T = 236$.[3] The threshold parameter τ was always set to 0, the initial

---

[2] Bootstrap confidence intervals are discussed in detail in Efron and Tibshirani (1993). The grid-bootstrap approach described in Hansen (1999) and the bias-corrected bootstrap approach described in Efron and Tibshirani (1993) were designed to improve the performance of the bootstrap-*t* and bootstrap-percentile methods, respectively. We did not consider these approaches for practical reasons. Our Monte Carlo experiments were very time-consuming when the threshold parameter was treated as unknown because of the nonlinear estimation problem that had to be solved for each bootstrap sample.

[3] Note that 236 equals the number of observations in our GDP data set. In an earlier version of this paper, we reported results using $T = 100$. The results of these simulations are available from the authors on request. Also note that in simulating TAR processes, it is possible that the constructed series never crosses the true threshold. This turned out to be especially true for values of β_1 equal to 0.9 and 0.95. Unless there are two observations on each side of the threshold it is impossible to fit a TAR model to the data. In practice, researchers searching for an unknown threshold typically discard the largest and smallest 10 or 15 percent of the ordered data from their search. If one of our

value y_0 was set to the unconditional mean of the process, and the ε_t's were drawn from the standard normal distribution. Each series was generated for T + 100 data points and the initial 100 observations were discarded. For each of the realized series, we used the standard grid-search method described in Enders (2004) to find a consistent estimate of the threshold.[4] The t-statistics and student-t distributions were used to construct confidence intervals with nominal coverage equal to 0.75, 0.9, 0.95, 0.975, and 0.99 for each of the two slope coefficients. Next, for each of these y_t series, the estimated slope coefficients and estimated value of τ were used to construct 1000 bootstrap samples in order to construct the bootstrap-percentile and bootstrap-t intervals. Hence, there are 1000 bootstrap samples for each of the 1000 generated y_t series. Actual coverage percentages were computed as the proportion of instances in which the true slope coefficients fell into each type of constructed interval. Note that for each realization of the y_t process, the bootstrap samples used to construct the bootstrap-percentile and bootstrap-t intervals were generated using the estimated threshold rather than the true threshold and that the threshold parameter was re-estimated (along with the intercept and slope coefficients) for each bootstrap sample.

These simulations were very time-consuming because of the need to search for τ in each of the bootstrap samples. Therefore, we used a relatively small set of parameter combinations for the data-generating process. Specifically, the threshold parameter τ was set to zero, the slope coefficient α_1 was set to 0.3, and the slope coefficient β_1 was sequentially selected from {0.6, 0.9, 0.95}. For the D-TAR model (2.1) we set the intercepts α_0 and β_0 equal to 0 and 0.9, and for the C-TAR model we set $\alpha_0 = \beta_0 = 0$.

3.1 Confidence intervals for the slope coefficients in the D-TAR model

The simulated coverage probabilities for the slope coefficients using the three methods of interval estimation are presented in Table 1 for the D-TAR model. Consider, for example, the D-TAR model with $T = 236$, $\alpha_0 = 0$ $\alpha_1 = 0.3$, $\tau = 0$, $\beta_0 = 0.9$ and $\beta_1 = 0.6$. The nominal 90% confidence interval constructed using the normal approximation (i.e., $\hat{\beta}_1 \pm s.e.(\hat{\beta}_1)*1.64$) included the true value of α_1 in 87.1% of the trials and included the true value of β_1 in 78.1% of the trials. As such, these confidence intervals are "too narrow" in that the simulated coverage is smaller than the nominal coverage. Notice that for these same parameter values,

---

simulated series did not contain at least three points on each side of the threshold, it was discarded and replaced with another simulated series. We applied this rule throughout this study, including the bootstrap simulations.

[4] This is the estimation procedure used in Chan (1993) and Hansen (1997). The meaningful candidates for the threshold are the observed values of the data series.

TABLE 1: Simulated Coverage Probabilities for the Slope Coefficients in the D-TAR Model

$\alpha_1 = 0.3,\ \tau = 0,\ \alpha_0 = 0,\ \beta_0 = 0.9,\ T = 236$

	Nominal Coverage	Normal Approximation		Bootstrap Percentile		Bootstrap-*t*	
		α_1	β_1	α_1	β_1	α_1	β_1
$\beta_1 = 0.6$	75%	71.4	60.5	79.0	69.4	73.8	69.1
	90%	87.1	78.1	94.3	87.2	88.1	81.8
	95%	92.2	84.6	97.9	93.2	93.6	85.9
	97.5%	95.7	89.4	98.9	96.2	95.8	88.7
	99%	97.6	92.9	99.6	98.0	97.5	91.1
$\beta_1 = 0.9$	75%	71.9	65.5	74.2	59.9	73.3	74.3
	90%	85.9	82.6	91.9	82.7	89.3	89.0
	95%	92.0	89.4	96.3	90.9	93.8	94.3
	97.5%	95.7	93.6	98.3	95.6	96.2	97.5
	99%	97.1	96.9	99.5	98.3	98.0	99.4
$\beta_1 = 0.95$	75%	68.4	64.7	75.7	53.4	77.2	74.3
	90%	83.6	81.7	92.6	79.8	90.3	89.1
	95%	89.9	89.6	96.7	90.1	95.3	94.3
	97.5%	93.9	93.6	98.9	95.1	97.3	97.6
	99%	97.1	97.3	99.7	98.2	98.4	98.6

the bootstrapped *t*-statistic yielded confidence intervals closer to the nominal values than the normal approximation (88.1 for α_1 and 81.8 for β_1). Among the key points to note in Table 1 are:

- The confidence intervals, constructed using the normal approximation, are always too narrow in that their simulated coverage is less than their nominal coverage. Hence, the use of the normal approximation (i.e., the 'usual' *t*-test) to test the null hypothesis $\alpha_1 = 0$ or $\beta_1 = 0$ is likely to result in too few rejections.
- The percentile method yields confidence intervals for α_1 that are very close to their nominal values. Those for β_1 are generally too narrow, although they are better than those generated from the normal approximation. Notice that the coverage properties for β_1 clearly deteriorate as the magnitude of β_1 increases, and improve as the nominal size of the confidence interval increases.

- Of the three methods, the bootstrap-*t* generally has the best coverage. Although tending to produce intervals that undercover, the interval coverage rates for α_1 are almost always within one-to-two percent of the desired rate. Regarding β_1, the interval coverage is within one-percent of the desired rate when $\beta_1 = 0.9$ and 0.95.

We conclude that for the D-TAR model confidence intervals constructed from the normal approximation were the least satisfactory while the bootstrap-*t* intervals were usually the most satisfactory.

3.2 Confidence intervals for the slope coefficients in the C-TAR model

The simulated coverage probabilities for the slope coefficients of the C-TAR model are presented in Table 2. As in the D-TAR model, the intervals constructed from the normal approximation perform the worst. For all cases considered, the normal approximation yields simulated coverage percentages for both α_1 and β_1 that are very low when compared to the nominal percentages. Relative to the D-TAR model, the performance of the normal approximation actually deteriorates for the C-TAR model. The bootstrap percentile intervals for α_1 work very well with actual coverage rates almost always within one-percent of the nominal coverage rates. The bootstrap percentile intervals for β_1 work reasonably for small β_1 (i.e., $\beta_1 = 0.6$) but very poorly for large β_1 (i.e., $\beta_1 = 0.9$ and 0.95). The bootstrap-*t* intervals for β_1 work very well with actual coverage rates almost always within one-percent of the nominal coverage rates. While the bootstrap-*t* intervals for α_1 are not quite as good as those generated using the bootstrap percentile method, they are reasonable. As such, in applied work the bootstrap-*t* seems to be the best choice among the three methods. Alternatively, it may be best to use a combination of the two bootstrap methods, using the percentile method for the smallest slope coefficient and the bootstrap-*t* for the largest slope coefficient.

3.3 Confidence intervals for the threshold parameter

Hansen (1997) derived the (non-standard) asymptotic distribution of the least-squares estimator of the threshold parameter and the likelihood ratio statistic for inference concerning the threshold parameter in the D-TAR model. Critical values for these distributions are tabulated in Hansen (1997). Chan and Tsay (1998) studied the asymptotic distribution of the least squares estimator of the threshold parameter and the likelihood ratio statistic in the C-TAR model, showing that in

TABLE 2: Simulated Coverage Probabilities for the Slope Coefficients in the C-TAR Model

$\alpha_1 = 0.3$, $\tau = 0$, $T = 236$

	Nominal Coverage	Normal Approximation		Bootstrap Percentile		Bootstrap-*t*	
		α_1	β_1	α_1	β_1	α_1	β_1
$\beta_1 = 0.6$	75%	56.2	62.2	74.9	73.9	76.6	77.0
	90%	73.8	79.3	89.7	86.3	92.5	91.0
	95%	81.3	87.0	94.8	92.7	95.5	95.6
	97.5%	86.7	91.7	97.1	96.2	97.3	97.7
	99%	91.4	94.6	98.9	98.5	99.0	98.7
$\beta_1 = 0.9$	75%	50.5	69.7	72.7	67.5	72.9	74.0
	90%	65.1	86.2	89.3	82.1	86.5	89.6
	95%	73.8	92.3	95.0	87.9	91.8	95.2
	97.5%	80.1	95.9	98.1	91.3	95.0	97.4
	99%	85.5	98.0	99.4	95.6	97.0	98.7
$\beta_1 = 0.95$	75%	49.0	69.6	77.9	65.0	70.3	71.4
	90%	64.6	86.8	91.2	76.7	86.1	87.5
	95%	73.0	92.9	95.4	83.0	91.5	93.8
	97.5%	77.9	95.7	98.0	85.8	93.2	97.4
	99%	82.6	97.8	99.6	90.2	94.8	98.5

this case the asymptotic distribution of the likelihood ratio statistic is a chi-square with one degree of freedom.

In this section, we evaluate the finite sample coverage properties of confidence intervals for τ constructed using three different procedures. The first procedure is to invert the likelihood ratio statistic using the asymptotic critical values. That is, the likelihood ratio statistic to test the null hypothesis that $\tau = \tau_0$ is

$$LR(\tau_0) = \frac{SSR(\tau_0) - SSR(\hat{\tau})}{SSR(\hat{\tau})}$$

where: $SSR(\hat{\tau})$ is the sum of squared residuals from the regression model (2.1) or (2.2) using a grid-search procedure to estimate τ and $SSR(\tau_0)$ is the sum of squared residuals from the regression model (2.1) or (2.2) fixing τ at τ_0. The δ-percent confidence interval for τ found by inversion of the likelihood ratio statistic is $\Gamma(\delta) = \{\tau : LR(\tau) \leq C(\delta)\}$ where $C(\delta)$ is the δ-level critical value from the asymptotic distribution of $LR(\tau)$. Following Hansen (1997), we use the convexified region

$\Gamma^*(\delta) = [\tau_1\ \tau_2]$, where τ_1 and τ_2 are the minimum and maximum elements of $\Gamma(\delta)$, respectively.

The second procedure is identical to the first except that it uses the bootstrapped distribution of $LR(\tau)$ to determine the critical value $C(\delta)$. The third procedure is the bootstrap percentile distribution constructed as the values of τ falling within the $(1-\delta)/2$ and $1-\delta+(1-\delta)/2$ percentiles of the bootstrap distribution of $\hat{\tau}$.

The results for a nominal 90% confidence interval are presented in Table 3 for the D-TAR and C-TAR models. Values for other percentages were found to be ordered similarly and are not reported here. These results were obtained as part of the Monte Carlo experiments used to construct the confidence intervals for the slope coefficients in the TAR models. For the bootstrap procedures, the estimated threshold from each simulated series was used to generate the bootstrap samples and re-estimated for each bootstrap sample so as to simulate the bootstrap distributions of the least-squares estimator of τ and the likelihood ratio statistic $LR(\tau)$.[5]

According to Table 3, none of the three procedures performs satisfactorily for the D-TAR model. All three methods over-cover in the sense that the confidence intervals are too wide. Surprisingly, the bootstrapped likelihood ratio method has the worst performance—the confidence intervals were so wide they have 100% coverage for $\beta_1 = 0.6$ and $\beta_1 = 0.90$ and 99.8% coverage for $\beta_1 = 0.95$. The poor performance of the bootstrap-*LR* procedure is somewhat surprising since the likelihood ratio statistic is asymptotically pivotal. The bootstrap-percentile procedure provides good coverage when $\beta_1 = 0.6$, (simulated coverage is 91.7%) but not for $\beta_1 = 0.9$ or $\beta_1 = 0.95$. The normal approximation provides coverage rates greater than 96% in all three cases.

In contrast, the normal approximation works best among the three procedures applied to the C-TAR model. The simulated coverage is reasonably close to 90% for all three values of β_1. The confidence intervals from the two bootstrapped procedures are far too wide for the C-TAR model. The percentile method works worse than the bootstrap-LR method when β_1 is large. Hence, the normal approximation works poorly for the slope coefficients but works reasonably well for the threshold parameter (especially in the C-TAR model).

3.4 Estimating the C-TAR model as a D-TAR process

[5] Note that the bootstrapped values of the likelihood-ratio statistic can be negative since the estimated threshold for any given bootstrap sample can generate a smaller sum of squared residuals than the sum of squared residuals obtained from the estimated threshold fit to the original sample.

In many applications it is not clear whether the true data generating process is continuous or discontinuous at the threshold. Since a C-TAR is a restricted form of a D-TAR model, it might seem plausible to estimate a D-TAR model in the form of (2.1) and then test whether the restriction implied by (2.2) [i.e., $\tau = (\beta_0 - \alpha_0)/(\alpha_1 - \beta_1)$] is binding. However, this strategy is not feasible. Enders and Siklos

TABLE 3: Simulated Coverage Probabilities for the Threshold Parameter
Nominal Coverage = 90-percent

$\alpha_0 = 0, \beta_0 = 0.9, \alpha_1 = 0.3, \tau = 0, T = 236$

	Asymptotic Approx.	**BS – Percentile**	**BS – LR**
$\beta_1 =$		**Coverage in the D-TAR Model**	
0.6	98.0	91.7	100
0.9	96.4	96.3	100
0.95	96.2	98.1	99.8
$\beta_1 =$		**Coverage in the C-TAR Model**	
0.6	93.1	100	100
0.9	91.6	100	99.8
0.95	87.7	100	93.9

(2004) demonstrate that an F-statistic for the null hypothesis $\tau = (\beta_0 - \alpha_0)/(\alpha_1 - \beta_1)$ is not asymptotically pivotal. An important issue, then, is to analyze the consequences of estimating the wrong functional form. We focus our attention on the case of estimating a C-TAR model in the functional form of (2.1) since the C-TAR process is nested within a D-TAR model. In contrast, a D-TAR model estimated as a C-TAR process results in a misspecification error.

We generated 1000 C-TAR series using the parameter set and methodology described above. However, unlike the results described in Sections 3.2 and 3.3, we estimated each simulated series as a D-TAR process and calculated the coverage properties of each method of constructing confidence intervals. The results for the slope coefficients are reported in Table 4. Notice that the simulated coverage of the normal approximation is always far too low. For example, for the case of $\beta_1 = 0.6$, the calculated coverage using a nominal 90% confidence interval was only 69.5% for α_1 and 67.2% for β_1. The intervals for the percentile method were too wide for $\beta_1 = 0.6$ but were generally too narrow for $\beta_1 = 0.9$ and $\beta_1 = 0.95$. The intervals for the bootstrap-t method were always too narrow.

A comparison of the results in Tables 2 and 4 indicates the cost of estimating the over-parameterized D-TAR model when the true data generating process (DGP) is a C-TAR model. Clearly, the simulated coverage values for the normal approximation and bootstrap-*t* methods shown in Table 4 are even narrower than those shown in Table 2. The percentage differences for the normal approximation are small. For example, for a nominal 90% confidence interval, when β_1 = 0.6, the coverage for α_1 shown in Table 2 is 73.8% and the coverage shown in Table 4 is 69.5%. In percentage terms, the losses using the bootstrap-*t* are far larger when a D-TAR model is used to estimate a C-TAR process. As illustrated by a nominal 90% confidence interval for α_1, the simulated coverage shown in Table 2 is 92.5%, whereas the simulated coverage shown in Table 4 is 71.7%. The results for the percentile method are tricky to interpret since some of the confidence intervals are too narrow and others are too wide.

TABLE 4: Slope Coefficient Coverage for a C-TAR Model Estimated as a D-TAR

$\alpha_1 = 0.3$, $\tau = 0$, $T = 236$

	Nominal Coverage	Normal Approximation		Bootstrap Percentile		Bootstrap-*t*	
		α_1	β_1	α_1	β_1	α_1	β_1
$\beta_1 = 0.6$	75%	48.7	48.5	86.3	88.4	56.5	55.1
	90%	69.5	67.2	98.2	98.0	71.7	73.0
	95%	78.7	79.1	99.5	99.4	80.3	82.2
	97.5%	86.3	86.6	99.8	99.9	87.1	87.3
	99%	92.1	92.3	100.0	100.0	92.3	90.9
$\beta_1 = 0.9$	75%	47.6	49.7	70.4	67.3	53.0	69.3
	90%	61.5	70.3	92.3	90.4	70.8	84.6
	95%	71.8	79.9	96.3	96.7	77.5	88.9
	97.5%	78.5	86.7	97.9	98.6	81.7	91.4
	99%	84.8	93.2	98.6	99.9	86.0	94.5
$\beta_1 = 0.95$	75%	43.8	52.8	60.2	55.4	49.9	67.1
	90%	58.2	69.0	87.0	83.4	66.7	83.0
	95%	66.2	78.1	93.8	93.7	73.0	89.7
	97.5%	71.7	86.2	96.5	98.1	77.6	92.0
	99%	76.7	92.9	98.3	99.6	81.1	94.7

If we use a nominal 90% confidence interval, the coverage properties for the threshold parameter are

Coverage of Threshold Parameter

	Asymptotic Approx.	BS – Percentile	BS – LR
$\beta_1 =$			
0.6	93.9	91.7	100.0
0.9	91.6	87.5	99.7
0.95	85.3	86.5	95.2

Notice that the confidence intervals for the bootstrapped likelihood ratio statistic are always too wide. The asymptotic approximation and the percentile methods work similarly--sometimes the intervals are too wide and sometimes they are too narrow. In comparing these results to those shown in Table 3, it is interesting that the coverage properties of the percentile method actually improve when the C-TAR process is estimated as a D-TAR process.

Overall, the losses from estimating the D-TAR model when the actual DGP is a C-TAR process can be small. The most serious loss involves the bootstrap-*t* method for the slope coefficients. Nevertheless, if there is little knowledge of the actual form of the DGP, it seems preferable to estimate the D-TAR model than a possibly misspecified C-TAR model.

4. Confidence Intervals for TAR Estimates of US GDP

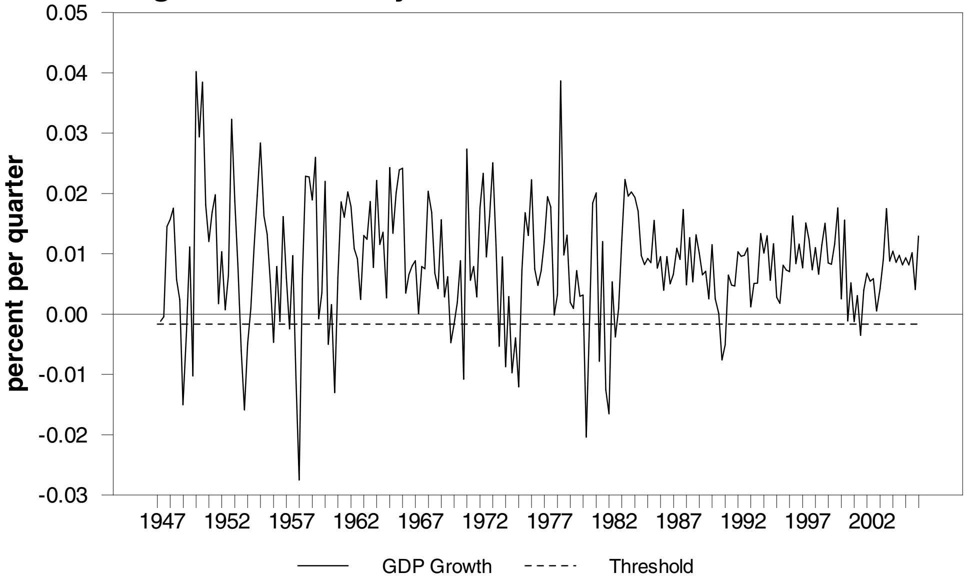
The aim of this section is to compare the various methods for constructing confidence intervals for the threshold and slope parameters of the real US GDP series. The time path of the logarithmic change in real US GDP (y_t) over the 1947:Q1 to 2006:Q1 sample period is shown as the solid line in Figure 1. It is quite possible that this series represents a litmus test for nonlinear time series modeling. For example, Potter (1995) modeled and fit the logarithmic difference of real US GNP (not GDP) through 1990:Q4 to a threshold autoregressive model under the assumption that the threshold is known and equal to zero.[6]

4.1 Model selection

We followed Hansen's (1999) procedure to test for linearity and, if linearity is rejected, perform a test to select the appropriate number of regimes for the TAR model of GDP. The first step in the process is to fit GDP to an AR(p) model.

[6] In addition, Potter (1995) estimates the model for each regime separately to allow for heteroskedasticity across regimes, which is straightforward when the threshold is assumed to be known.

Figure 1: Quarterly Growth Rate of Real U.S. GDP



Choosing among lag lengths 1, …, 8, the AIC selected a lag length of 4, yielding the following estimated AR(4) model of GDP.[7] Below, and in what follows, *t*-statistics are reported in parenthesis, *aic* is the sample value of the AIC and *rss* is the sum of squared residuals.

$$y_t = \underset{(6.25)}{0.006} + \underset{(4.51)}{0.298}y_{t-1} + \underset{(2.03)}{0.139}y_{t-2} - \underset{(-1.23)}{0.085}y_{t-3} - \underset{(-1.64)}{0.108}y_{t-4} + \varepsilon_t \qquad (4.1)$$

$$aic = -903.13,\ rss = 0.0195$$

Next, we fit a two-regime TAR model to GDP by minimizing the sum of squared residuals with respect to the intercept, slope, threshold, and delay parameters, maintaining the lag length of four. The estimated threshold was constrained to require that at least 10-percent of the observations fall above and below the threshold. This estimator selected a delay parameter of two (as does Potter 1995) and produced the following estimated TAR model:

[7] The Schwartz criteria selected a lag length of 1. In this case, the null of linearity is not rejected against the TAR alternative. Since our paper is concerned with interval estimation in settings where TAR effects are present, we followed the path implied by the AIC.

$$y_t = I_t\,[0.006 + 0.320 y_{t\text{-}1} + 0.137 y_{t\text{-}2} - 0.083 y_{t\text{-}3} - 0.067 y_{t\text{-}4}\,] \qquad (4.2)$$
$$\quad (4.64) \quad (4.38) \quad (1.56) \quad (-1.15) \;-(0.097)$$

$$+ (1 - I_t)[-0.003 + 0.208 y_{t\text{-}1} - 0.909 y_{t\text{-}2} - 0.156 y_{t\text{-}3} - 0.506 y_{t\text{-}4}\,] + \varepsilon_t$$
$$(-1.01) \quad (1.53) \quad (-3.03) \quad (-0.86) \quad (-2.79)$$

$$I_t = \begin{cases} 1 \;\; if \;\; y_{t-2} \geq -0.00167 \\ 0 \;\; if \;\; y_{t-2} < -0.00167 \end{cases} \quad ; \qquad aic = -918.01,\; rss = 0.0178$$

To test the null of linearity against the two-regime TAR alternative, we used the test statistic

$$F_{12} = T(\frac{S_1 - S_2}{S_2})$$

where S_1 is the sum of squared residuals from the estimated linear autoregression and S_2 is the sum of squared residuals from the estimated two-regime TAR. Following Hansen (1999), we used the bootstrap (with 1000 bootstrap samples) to estimate the percentiles of the asymptotic null distribution of F_{12}. The value of the F_{12} statistic turned out to be 22.54 and the resulting p-value was 0.026. Therefore, we rejected the null of linearity against the alternative of a two-regime TAR model.

We also tested the null of a two-regime TAR model against the three-regime alternative, using the test statistic

$$F_{23} = T(\frac{S_2 - S_3}{S_3})$$

where S_3 is the sum of squared residuals from the estimated three-regime TAR. The value of F_{23} was 15.6 and the bootstrapped p-value was 0.216. Therefore, we did not reject the null of a two-regime TAR model.

On the basis of these tests, we conclude that the two-regime threshold model is the appropriate choice within the class of TAR models for GDP. Figure 1, the time series graph of quarterly real GDP growth rates, includes a dashed horizontal line at the estimated threshold of –0.00167. Of the 236 observations, 30 fell below the threshold and 206 fell above the threshold. Thus, approximately 12-percent of the observations fell into the ‘low-growth’ regime. Interestingly, this is roughly the same proportion of quarters over the sample period that are NBER-dated recession periods.[8] It is also interesting to note that the value of the threshold is near zero and that the sum of the lag coefficients is positive (0.307) in

[8] The US economy has been in recession for 104 months between 1947Q1-2006Q1, according to the NBER (www.nber.org). This represents almost 15% of the sample.

the 'high-growth' regime but strongly negative (–1.363) in the low-growth regime. These results are roughly in line with those reported by Potter (1995, Table II). The point estimates of the AR coefficients in the low-growth regime violate the stationarity condition. However, the interval estimates we present below suggest that these coefficients are measured very imprecisely, which is not surprising since there are only 30 observations in the low-growth regime.

The next issue is to construct confidence intervals for the TAR parameters.

4.2 Confidence intervals for the threshold

Figure 2 shows the value of the *LR*-statistic as a function of the potential threshold values where:

$$LR(\tau_0) = \frac{SSR(\tau_0) - SSR(\hat{\tau})}{SSR(\hat{\tau})} \tag{4.3}$$

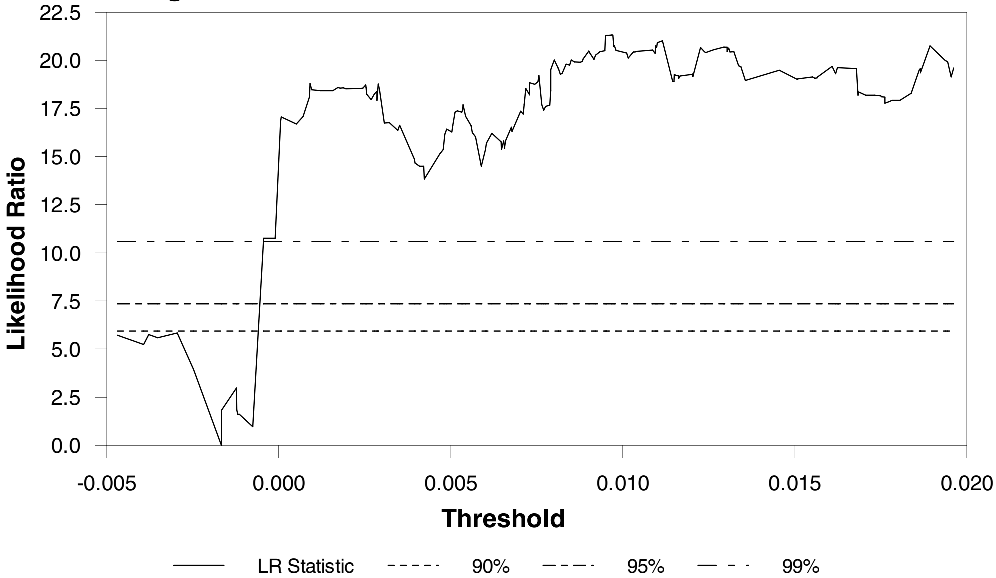
In (4.3), $SSR(\tau_0)$ is the sum of squared residuals from fitting the growth rate of real GDP to a four-order, two-regime TAR model with delay parameter equal to two and threshold equal to τ_0 and $SSR(\hat{\tau})$ is the sum of squared residuals from (4), i.e., $\hat{\tau}$ = –0.00167 and $SSR(\hat{\tau})$ = 0.0178. Hansen (1997) shows that when evaluated at the true value of τ, $LR(\tau)$ is asymptotically chi-square with one degree of freedom. The 90%, 95% and 99% critical values for $LR(\tau_0)$ are drawn as horizontal dashed lines in Figure 2. It is clear from the figure that the change in *LR* is quite pronounced around the estimated threshold τ = –0.00167. The jump in the function $LR(\tau_0)$ is so sharp that the 90%, 95% and 99% confidence intervals, which are found by inverting the likelihood ratio statistic, are precisely that same. (Recall that the estimated threshold is only identified up to the observed values of the dependent variable.) Specifically, the confidence intervals implied from the asymptotic approximations are:

Asymptotic Confidence Intervals for τ

	Low	**High**
90%	–0.00469	–0.00076
95%	–0.00469	–0.00076
99%	–0.00469	–0.00076

Since there is only one clear trough in the figure, there is fairly strong evidence of a single threshold; in a three-regime model there should be two distinct threshold values. In other words, Figure 2 is consistent with the formal test results we presented above regarding the appropriate number of regimes.

Figure 2: Confidence Intervals for the Threshold



We bootstrapped equation (4.2) 2500 times, generating the 2500 bootstrap estimates of the threshold. Retaining only the middle 90%, 95% and 99% of the ordered threshold estimates yielded the bootstrap percentile confidence intervals shown below. Note that these intervals are far larger than the confidence intervals reported above and always span $\tau = 0$. In fact, the 99% confidence interval constructed using the percentile method spans nearly all of the data set.

Bootstrap Percentile Confidence Intervals for τ

	Low	**High**
90%	–0.00289	0.00097
95%	–0.00348	0.00443
99%	–0.00486	0.01577

Finally, since $LR(\tau_0)$ is an asymptotically pivotal statistic, we also bootstrapped the 90%, 95% and 99% critical values of its distribution and used these to construct the confidence intervals for τ. The 90% (and, therefore, 95% and 99%) critical values were so large that all of the data points that were candidate threshold values fell into each of these intervals. Consequently, the threshold confidence intervals implied by this procedure went from the 10-th percentile of the data (-0.00469) to the 90-th percentile of the data (0.01961). This is consistent with our simulation indicating that the actual coverage of the 90%

bootstrap-LR intervals for the threshold in the D-TAR model were 100% for each parameter configuration we considered.

Bootstrap-LR Confidence Intervals for τ

	Low	**High**
90%	–0.00469	0.01961
95%	–0.00469	0.01961
99%	–0.00469	0.01961

Our conclusion is that the evidence to support the claim that the threshold for real US GDP growth is negative is not very compelling. Although we cannot reject the null hypothesis of threshold behavior, the problem is to produce a reliable confidence interval for the threshold parameter. When we construct confidence intervals using an asymptotic approximation we can rule out the possibility of a positive threshold. However, the bootstrap methodology does not support the assertion that the time-series properties of negative growth rates behave differently from positive growth rates.

4.3 Confidence intervals for the slope parameters

Although there are four lags in the model, we focus on the two first-order slope coefficients and the sum of the slope coefficients within each regime as these sums are a measure of within-regime persistence. The two first-order slope coefficients are 0.320 and 0.208 with *t*-statistics of 4.38 and 1.53, respectively, and standard errors of 0.0729 and 0.1352, respectively. Since we use a consistent estimate of the threshold, asymptotically valid confidence intervals for these slope coefficients can be constructed using the percentiles of the normal distribution. The 90%, 95% and 99% confidence intervals for the two slope coefficients (called α_1 and β_1) using the normal approximation are reported in the top-left portion of Table 5. For example, the 95% confidence intervals for α_1 and β_1 run from 0.1908 to 0.4727 and from –0.0574 to 0.4725, respectively.

The 2500 bootstrap replications of equation (4.2) also gave us 2500 bootstrap estimates of the two first-order slope coefficients. Retaining only the middle 90%, 95% and 99% of the ordered slope coefficients yielded the percentile confidence intervals reported in the top-middle portion of Table 5. For each bootstrapped series, we also constructed the bootstrap *t*-statistic for the null hypothesis $\alpha_1 = 0.320$ and $\beta_1 = 0.208$. This bootstrap *t*-statistic allows us to 'back-out' the confidence intervals reported in the top-right portion of Table 5.

As shown in Table 5, the confidence intervals for the slope coefficient α_1 are roughly the same, both in location and length, across the three procedures. All three include only positive values of α_1. Closer inspection shows that the intervals for α_1 constructed from the normal approximation are slightly shifted to the right relative to those constructed from the bootstrap-percentile and the intervals for α_1 constructed from the bootstrap-*t* are slightly shifted to the right relative to those constructed from the normal approximation. The confidence intervals for the slope coefficient β_1 show less uniformity across methods. However, for any given method and coverage rate, the interval for β_1 is larger than the interval for α_1. In fact, $\beta_1 \leq 0$ is in the 95% confidence interval constructed from each of the three methods. In addition, note that the percentile method yielded 95% and 99% confidence intervals for β_1 that fully contain the confidence intervals constructed from the two other methods. Also, each confidence interval for β_1 constructed from the normal approximation is contained within the corresponding interval constructed from the bootstrap-*t*. Thus, the percentile method appears to be the most conservative method, and the normal approximation appears to be the least

TABLE 5: Confidence Intervals for the Slope and the Persistence Coefficients in the D-TAR Model of GDP

	Normal Approx.		**BS-Percentile**		**Bootstrap-*t***	
Slope Coefficients	**lower bound**	**upper bound**	**lower bound**	**upper bound**	**lower bound**	**upper bound**
α_1						
90%	0.2136	0.4500	0.2070	0.4388	0.2200	0.4614
95%	0.1908	0.4727	0.1788	0.4656	0.1919	0.4865
99%	0.1470	0.5165	0.1060	0.5238	0.1408	0.5539
β_1						
90%	0.0147	0.4298	0.1005	0.4912	-0.0320	0.4433
95%	-0.0574	0.4725	-0.1827	0.5479	-0.0828	0.4987
99%	-0.1398	0.5549	-0.3962	0.7245	-0.1795	0.6042
Persistence	**lower bound**	**upper bound**	**lower bound**	**upper bound**	**lower bound**	**upper bound**
$y_{t-2} \geq \tau$						
90%	0.1183	0.4974	0.0760	0.4708	0.1248	0.5280
95%	0.0819	0.5339	0.0323	0.5093	0.0862	0.5689
99%	0.0115	0.6042	-0.2643	0.5902	0.0124	0.6381
$y_{t-2} < \tau$						
90%	-1.9791	-0.7478	-2.0471	-0.5655	-2.1810	-0.6691
95%	-2.0974	-0.6294	-2.2018	-0.0472	-3.2648	-0.5373
99%	-2.3259	-0.4010	-2.5048	0.3608	-6.6580	-0.3275

conservative method, to obtain the confidence intervals for β_1. The percentile method seems to exacerbate the effect of poorly estimated coefficients.

Perhaps, the more important results concern those pertaining to the sum of the lagged coefficients in each regime since this sum is an indication of the degree of persistence within a regime. The confidence intervals for the persistence parameter in the high-growth regime are roughly the same, both in location and length, across the three procedures. All three include only positive values of this parameter, with the exception of the 99% interval constructed using the bootstrap-percentile procedure. Closer inspection shows that the intervals for the high-growth persistence parameter constructed from the normal approximation are slightly shifted to the right relative to those constructed from the bootstrap-percentile procedure and the intervals for the high-growth persistence parameter constructed from the bootstrap-*t* procedure are slightly shifted to the right relative to those constructed from the normal approximation. Notice that the confidence intervals for persistence in the high-growth regime are far tighter than those for the corresponding intervals in the low-growth regime. For example, a 95% confidence interval for the persistence parameter using the normal approximation runs from 0.0819 to 0.5339 for the high-growth regime, and from –2.0974 to –0.6294 for the low-growth regime. This is not very surprising given the relatively small number of observations that define the low-growth regime. All of the low-growth intervals contain only negative values of the persistence parameter, except for the 99% interval constructed using the bootstrap-percentile procedure.

4.4 The C-TAR model

Since there is no *a priori* way of knowing whether real GDP growth is a C-TAR or a D-TAR process, we also estimated y_t as a continuous threshold process in the form of (2.2). A lag length of two and a delay parameter of two provided the best fitting C-TAR model, resulting in:

$$y_t = \tau + I_t[\underset{(3.97)}{0.358}(y_{t-1} - \tau) + \underset{(2.52)}{0.213}(y_{t-2} - \tau)] \quad (4.3)$$

$$+ (1-I_t)[\underset{(2.54)}{0.242}(y_{t-1}-\tau)\underset{(-0.831)}{-0.089}(y_{t-2}-\tau)] + \varepsilon_t$$

$$\tau = 0.00571; \quad I_t = \begin{cases} 1 \text{ if } & y_{t-2} \geq \tau \\ 0 \text{ if } & y_{t-2} < \tau \end{cases} \quad ; \; aic = -908.37 \; rss = 0.0199$$

In comparing this C-TAR model to the D-TAR model in (4.2), notice that the *aic* selects the D-TAR model even though the C-TAR model is far more parsimonious. The difference, however, is not very large. The first-order autoregressive coefficients in the high and low-growth regimes are quite similar across the fitted C-TAR and D-TAR models. However, the estimated threshold is positive in the fitted C-TAR model. Figure 3 shows the value of the *LR*-statistic constructed from (4.3) as a function of the potential threshold values.

Figure 3: Confidence Intervals for the GDP C-TAR Model

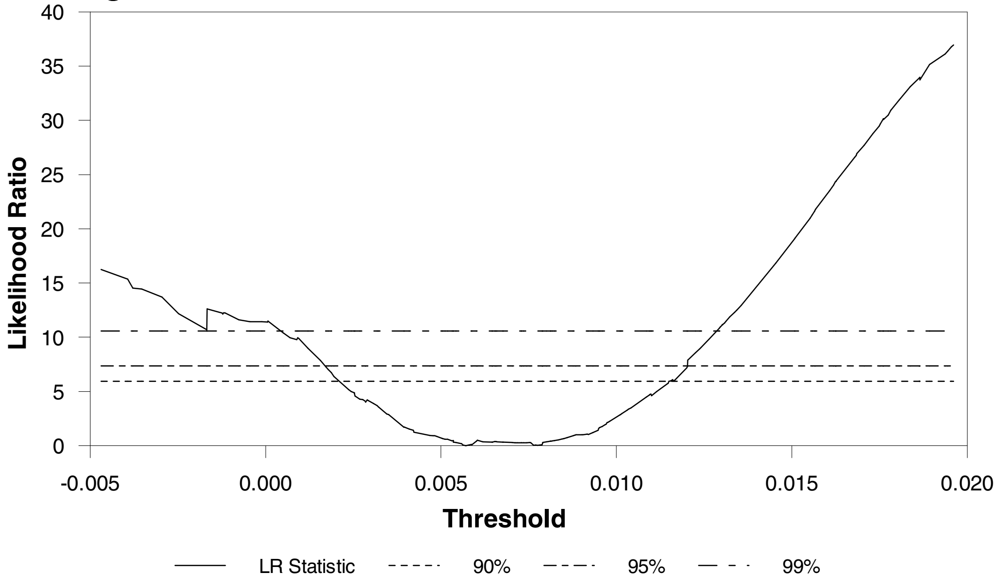


Table 6 provides the confidence intervals for the slope parameter in the high and low-growth regimes constructed from the normal approximation, the bootstrap-percentile and bootstrap-*t* procedures. The confidence intervals for each of the slope coefficients, α_1 and β_1, are roughly the same, both in location and length, across the three procedures. All include only positive values of α_1 and, except for the 99% intervals constructed using the methods, only positive values of β_1. Closer inspection shows that every interval for α_1 and β_1 constructed from the normal approximation is contained within the corresponding intervals constructed from the two bootstrap procedures. Every interval for α_1 constructed from the bootstrap percentile method is slightly shifted to the left relative to the corresponding interval constructed from the bootstrap-*t* method. Every interval for β_1 constructed from the bootstrap percentile method contains the corresponding interval constructed from the bootstrap-*t* method.

Table 6 also provides the confidence intervals for the persistence parameter in the high and low-growth regimes constructed from the normal approximation, the bootstrap-percentile and bootstrap-*t* procedures. These intervals behave in most respects like the intervals for the persistence parameters in the D-TAR model. The confidence intervals for the persistence parameter in the high-growth regime are roughly the same, both in location and length, across the three procedures. All three include only positive values of this parameter, with the exception of the 99% interval constructed using the bootstrap-percentile procedure. The intervals for the high-growth persistence parameter constructed from the normal approximation are generally slightly shifted to the right relative to those constructed from the bootstrap-percentile procedure and the intervals for the high-growth persistence parameter constructed from the bootstrap-*t* procedure are generally slightly shifted to the right relative to those constructed from the normal approximation.

TABLE 6: Confidence Intervals for the Slope and the Persistence Coefficients in the C-TAR Model of GDP

	Normal Approx.		BS-Percentile		Bootstrap-*t*	
Slope Coefficients	**lower bound**	**upper bound**	**lower bound**	**upper bound**	**lower bound**	**upper bound**
α_1						
90%	0.21022	0.50658	0.19422	0.51334	0.19486	0.53114
95%	0.18174	0.53507	0.15628	0.53740	0.16646	0.57212
99%	0.12676	0.59005	0.08998	0.58954	0.11999	0.63253
β_1						
90%	0.08896	0.40053	0.04527	0.45727	0.05735	0.42427
95%	0.05902	0.43048	0.00012	0.51110	0.01486	0.45757
99%	0.00121	0.48828	-0.09890	0.6083	-0.0504	0.51218
Persistence	**lower bound**	**upper bound**	**lower bound**	**upper bound**	**lower bound**	**upper bound**
$y_{t-2} \geq \tau$						
90%	0.43015	0.71293	0.14308	0.53408	0.60507	0.85847
95%	0.40297	0.74011	0.07845	0.57148	0.57159	0.89065
99%	0.35051	0.79257	-0.01209	0.62456	0.51839	0.94490
$y_{t-2} < \tau$						
90%	-0.07613	0.35835	-0.00261	0.42540	-0.29427	0.30599
95%	-0.11789	0.40011	-0.05940	0.46373	-0.37596	0.36792
99%	-0.19850	0.48072	-0.22295	0.53570	-0.56542	0.50131

The confidence intervals for the persistence parameter in the high-growth regime constructed using the bootstrap percentile method tend to be much larger than the corresponding intervals constructed using the normal approximation which, in turn, tend to be slightly larger than the intervals constructed using the bootstrap-*t* method. The confidence intervals for the persistence parameter in the high-growth regime are far tighter than those for the corresponding intervals in the low-growth regime. All of the low-growth intervals contain zero. The bootstrap-*t* intervals for the persistence parameter in the low-growth regime are much larger than the corresponding intervals constructed using the normal approximation which, in turn, tend be slightly larger than the intervals constructed using the bootstrap-percentile methods. The intervals for the high-growth persistence parameter constructed from the normal approximation are generally shifted to the right relative to those constructed from the bootstrap-*t* procedure and the intervals for α_1 constructed from the bootstrap-percentile procedure are generally shifted to the right relative to those constructed from the normal approximation.

As shown in Table 7, the 90%, 95% and 99% confidence intervals for τ using the asymptotic approximation are fairly tight. However, those formed using the percentile method are even tighter, being fully contained within those formed using the asymptotic approximation. For both methods, the confidence intervals for τ are such that they rule out the plausibility of a negative threshold. As in the D-TAR model, the confidence intervals formed from the bootstrap-*LR* method are essentially non-informative in that they span the entire range of potential thresholds.[9]

Table 7: Confidence Intervals for the Threshold in the C-TAR Model of GDP

	Normal Approximation		**BS-Percentile**		**Bootstrap-LR**	
	lower bound	**upper bound**	**lower bound**	**upper bound**	**lower bound**	**upper bound**
90%	0.00237	0.01148	0.00350	0.00941	–0.00469	0.01961
95%	0.00172	0.01203	0.00283	0.01013	–0.00469	0.01961
99%	0.00005	0.01264	0.00114	0.01201	–0.00469	0.01961

[9] We do not report confidence intervals for the autoregressive coefficients since α_1 is so similar to β_1.

5. Summary and Conclusions

Monte Carlo methods were applied to study the finite-sample performance of standard regression approaches to confidence interval construction in threshold autoregressive models. More specifically, intervals based upon asymptotic approximations and bootstrap methods were generated for the coefficients in the stationary, first-order, threshold autoregressive model. Interval coverage probabilities were used to measure the quality of the various procedures. When the true threshold is unknown, and is estimated along with the slope parameters, none of the procedures provide intervals for the slope coefficients with good coverage properties over the full range of parameters considered. Standard-*t* intervals performed especially poorly in this case.

Confidence intervals for the threshold parameter itself were constructed by inversion of the asymptotic distribution of the likelihood ratio statistic, by inversion of the bootstrap distribution of the likelihood ratio statistic, and by the bootstrap-percentile method. None of the procedures perform satisfactorily across the full range of parameter values. Interestingly, the bootstrap-*LR* procedure generated overly large confidence intervals with nearly 100% actual coverage for each parameterization of the D-TAR. This suggests that the bootstrap-*LR* procedure may not be very useful in D-TAR models. However, the bootstrap-*LR* procedure worked reasonably well, and better than the other procedures, for constructing confidence intervals for the slope coefficients.

We applied these procedures to obtain confidence intervals for the slope and threshold coefficients in a D-TAR model of real GDP growth rates, a model similar to the one estimated by Potter (1995) where he assumed that the threshold growth rate is zero. The confidence intervals for the threshold constructed using asymptotic theory always excluded zero and other nonnegative numbers. However, bootstrapped confidence intervals included zero, positive, and negative values for the threshold. The intervals from the bootstrapped likelihood ratio were so large that all candidate thresholds are included in these intervals. This is in line with our simulation results. Intervals for the slope coefficients appear to be more stable across the three procedures considered.

One message of the paper is that when the threshold parameter is unknown, asymptotic and bootstrap approximations of finite sample distributions do not lead to satisfactory confidence intervals for slope or threshold parameters in stationary TAR models. Since inference in a TAR model is problematic, caution must be exercised in applications attempting to conduct inference in threshold models.

REFERENCES

Andrews, D.W.K. and W. Ploberger (1994): "Optimal tests when a nuisance parameter is present only under the alternative," *Econometrica*, 64, 1383-1414.

Burgess, S.M. (1992): "Nonlinear dynamics in a structural model of employment," *Journal of Applied Econometrics*, 7, S101-S118.

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

Chan, K.S. and R.S.Tsay (1998): "Limiting properties of the least squares estimator of a continuous threshold autoregressive model," *Biometrika*, 45, 413-426.

Cao, C. Q. and Tsay, R.S. (1992): "Nonlinear time-series analysis of stock volatilities," *Journal of Applied Econometrics*, 7, S165-S185.

Efron, B. and R.J. Tibshirani (1993): *An Introduction to the Bootstrap*. New York: Chapman and Hall.

Enders, W. (2004): *Applied Econometric Time Series*, Hoboken, N.J.: John Wiley and Sons.

Enders, W. and P.L. Siklos (2004): "D-TAR versus C-TAR models? Modeling the dynamics of inflation", in R. Becker and S. Hurn, eds., *Advances in Economics and Econometrics,* Northampton, MA: Edward Elgar, 205 – 225.

Enders, W. and C.W.J. Granger (1998): "Unit-root tests and asymmetric adjustment with an example using the term structure of interest rates," *Journal of Business and Economic Statistics*, 16, 304-311.

Galbraith, J.W. (1996): "Credit rationing and threshold effects in the relation between money and output," *Journal of Applied Econometrics*, 11, 419-429.

Gonzalo, J. and M. Wolf (2005): "Subsampling Inference in Threshold Autoregressive Models," *Journal of Econometrics*, 127, 201-224.

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

Hansen, B.E. (1997): "Inference in TAR models," *Studies in Nonlinear Dynamics and Econometrics*, 1, 119-131.

Hansen, B.E. (1999): "The grid bootstrap and the autoregressive model," *The Review of Economics and Statistics*, 81, 594-607.

Hansen, B.E. (2000): "Sample Splitting and Threshold Estimation," *Econometrica*, 68, 575-603.

Kapetanios, G. (2003): "Threshold models for trended time series." *Empirical Economics*, 28, 687-707.

Krager, H. and P. Kugler (1993): "Non-linearities in foreign exchange markets: a different perspective," *Journal of International Money and Finance*, 12, 195-208.

Peel, D. A. and A.E. Speight (1998): "Threshold nonlinearities in output: some international evidence," *Applied Economics*, 30, 323 – 333.

Petrucelli, J. and S. Woolford (1984): "A threshold AR(1) model," *Journal of Applied Probability*, 21, 270-286.

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

Rothman, P. (1991): "Further evidence on the asymmetric behavior of unemployment rates over the business cycle," *Journal of Macroeconomics*, 13, 291-298.