Available online at www.sciencedirect.com
SCIENCE DIRECT®

ELSEVIER

Journal of Econometrics 127 (2005) 201–224

JOURNAL OF Econometrics

www.elsevier.com/locate/jeconom

Subsampling inference in threshold autoregressive models

Jesús Gonzalo[a,1], Michael Wolf[b,*,2]

[a]*Department of Economics, Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Spain*
[b]*Department of Economics and Business, Universitat Pompeu Fabra, Ramon Trias Fargas 25–27, 08005 Barcelona, Spain*

Received 1 May 2003
Available online 6 October 2004

Abstract

This paper discusses inference in self-exciting threshold autoregressive (SETAR) models. Of main interest is inference for the threshold parameter. It is well-known that the asymptotics of the corresponding estimator depend upon whether the SETAR model is continuous or not. In the continuous case, the limiting distribution is normal and standard inference is possible. In the discontinuous case, the limiting distribution is non-normal and it is not known how to estimate it consistently. We show that valid inference can be drawn by the use of the subsampling method. Moreover, the method can even be extended to situations where the (dis)continuity of the model is unknown. In this case, the inference for the regression parameters of the model also becomes difficult and subsampling can be used again. In addition, we consider an hypothesis test for the continuity of a SETAR model. A simulation study examines small sample performance and an application illustrates how the proposed methodology works in practice.

JEL: C14; C15; C22; C32

Keywords: Confidence intervals; Continuity; Regime shifts; Subsampling; Threshold autoregressive models

*Corresponding author.
E-mail addresses: jesus.gonzalo@uc3m.es (J. Gonzalo), michael.wolf@upf.edu (M. Wolf).
[1]Research supported by the Spanish Secretary of Education grants SEC01-0890 and SEJ2004-04101ECON.
[2]Research supported by the Spanish Ministry of Science and Technology and FEDER, grant BMF2003-03324, and by the Barcelona Economics Program of CREA.

doi:10.1016/j.jeconom.2004.08.004

1. Introduction

Over the last two decades, there has been an increasing interest in non-linear time series analysis; for example, see Tong (1990) as a general reference. One of the most popular non-linear time series models is the self-exciting threshold autoregressive (SETAR) model or sometimes just called the threshold autoregressive (TAR) model. A two-regime SETAR model is defined as

$$X_t = \begin{cases} \phi_{10} + \phi_{11} X_{t-1} + \cdots + \phi_{1p} X_{t-p} + \sigma_1 \varepsilon_t & \text{if } X_{t-d} \leqslant r, \\ \phi_{20} + \phi_{21} X_{t-1} + \cdots + \phi_{2p} X_{t-p} + \sigma_2 \varepsilon_t & \text{if } X_{t-d} > r. \end{cases} \tag{1}$$

Here, $d \leqslant p$ is a positive integer referred to as the threshold lag, r is the threshold, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) variables with mean zero and unit variance; also, ε_t is assumed to be independent of the past $X_{t-1}, X_{t-2}, \ldots$. The positive constants σ_1 and σ_2 allow the innovations to have different standard deviations in the two regimes. Throughout the paper, it will be assumed that $\{X_t\}$ is stationary ergodic, having finite second moments, and that the stationary distribution of $(X_1, X_2, \ldots, X_p)'$ admits a density positive everywhere.

Heuristically speaking, X_t is generated by one of two distinct autoregressive models according to the level of X_{t-d}. This model can be generalized to have more than two distinct regimes and/or to depend on the levels of more than one lagged variable. SETAR models are popular because they can exhibit many non-linear phenomena such as limit cycles, chaos, harmonic distortion, jump phenomena, and time irreversibility. They can be used as a general, parsimonious strategy for modeling non-linear economic time series. For a number of applications, see Tong (1990), Tiao and Tsay (1994), Potter (1995), and Chan and Tsay (1998), among others.

It is important to distinguish between discontinuous and continuous SETAR models. Let $\Phi_i = (\phi_{i0}, \phi_{i1}, \ldots, \phi_{ip})'$ be the autoregressive coefficient vector of model (1) in regime i. Then the model is said to have a discontinuous autoregressive function if there exists $Z_* = (1, z_{p-1}, \ldots, z_0)'$, where $z_{p-d} = r$, such that $(\Phi_1 - \Phi_2)' Z_* \neq 0$. In this case, the threshold r constitutes the jump point of the autoregressive function. Otherwise, that is, if $(\Phi_1 - \Phi_2)' Z_* = 0$ for all Z_* satisfying the above condition, the model has a continuous autoregressive function. It is easy to see that the latter case is equivalent to the requirement that $\phi_{1j} = \phi_{2j}$ for $1 \leqslant j \neq d \leqslant p$ and that $\phi_{10} + r\phi_{1d} = \phi_{20} + r\phi_{2d}$. Therefore, in the continuous case, the SETAR model can be written as

$$X_t = \phi_0 + \sum_{j=1, j \neq d}^{p} \phi_j X_{t-j} + \begin{cases} \phi_{d-}(X_{t-d} - r) + \sigma_1 \varepsilon_t & \text{if } X_{t-d} \leqslant r, \\ \phi_{d+}(X_{t-d} - r) + \sigma_2 \varepsilon_t & \text{if } X_{t-d} > r, \end{cases} \tag{2}$$

where $\phi_0 = \phi_{10} + r\phi_{1d}$, $\phi_{d-} = \phi_{1d}$, $\phi_{d+} = \phi_{2d}$ and $\phi_j = \phi_{1j}$ for $j \neq d$. The importance of distinguishing between discontinuous and the continuous SETAR models stems from the fact that the asymptotics of the (conditional) least squares estimator of the parameter $\vartheta = (\Phi_1', \Phi_2', r, d)'$ are different in the two cases. While $\hat{\Phi}_{i,n}$ always converges to a normal distribution with mean zero at rate square root of n, with n being the sample size, the asymptotic covariance matrix depends upon whether the model is continuous or not. For discontinuous models, $\hat{r}_n$ converges to a non-standard distribution at rate n and is asymptotically independent of $\hat{\Phi}_{i,n}$. But for continuous models, $\hat{r}_n$ converges to a normal distribution at rate square root of n and is asymptotically correlated with $\hat{\Phi}_{i,n}$. See Chan (1993) and Chan and Tsay (1998) for the results concerning the discontinuous and the continuous case, respectively. It should be pointed out that Chan and Tsay (1998) base the estimation of ϑ on the restricted model (2), thereby enforcing the estimated model to be continuous.

A main goal of this paper is to construct asymptotically valid confidence intervals for the threshold parameter r. In principle, the inference problem can be considered solved when it is known that the SETAR model is continuous. In this case, Chan and Tsay (1998) show that $n^{1/2}(\hat{r}_n - r)$ converges weakly to a normal distribution with mean zero and a variance that can be estimated consistently. On the other hand, the discontinuous case remains without a satisfactory solution. While Chan (1993) demonstrates that $n(\hat{r}_n - r)$ converges weakly to a non-degenerate distribution, the limiting distribution depends, in a very complicated way, on the underlying probability mechanism and apparently cannot be estimated consistently. It is not known whether a bootstrap approach would work. Under more restrictive conditions, such as i.i.d. normal innovations and the threshold effect vanishing asymptotically, the method of Hansen (2000) can be employed; see Section 4. In case it is unknown whether the SETAR model is continuous or not, an additional complication arises; this case has not been studied so far.

As will be demonstrated, one can solve the inference problem for the threshold parameter r by the use of the subsampling method dating back to Politis and Romano (1994); for a broader reference, see Politis, Romano, and Wolf (1999), abbreviated by PRW (1999) in the sequel. We will first discuss the case when the (dis)continuity of the SETAR model is known and then focus on the general case when it is unknown. Moreover, the subsampling method can also be used to make inference for regression parameters ϕ_{ij}. This is especially interesting in the general case, since the form of the limiting variance of $\hat{\phi}_{ij,n}$ depends upon whether the SETAR model is continuous or not and hence cannot be estimated consistently by standard methods unless the (dis)continuity of the model is known.

A problem that has not been discussed in the literature yet is the construction of a hypothesis test for the continuity of a SETAR model. As will be shown, the subsampling method can be employed to this end as well.

The remainder of the paper is organized as follows. In Section 2, we provide some key facts of the subsampling method to make the exposition self-complete. In Section 3, we discuss how to use subsampling to compute confidence intervals for SETAR model parameters. In Section 4, we compare our method to that of Hansen (2000). In Section 5, we present a hypothesis test for the continuity of the SETAR model. In Section 6, we discuss the choice of the block size, which is an important model parameter of the subsampling method. In Section 7, we conduct some simulation studies to examine finite-sample properties. In Section 8, we provide an empirical application to unemployment data. In Section 9, we provide a discussion. The mathematical details are postponed to the Appendix.

2. Subsampling in a nutshell

In this section, the subsampling method for dependent data is briefly reviewed. We consider the construction of confidence intervals for real-valued parameters and the construction of hypothesis test for general null hypotheses.

2.1. Confidence intervals for a parameter

Consider the case of a time series $\{X_1, X_2, X_3, \ldots\}$ governed by a probability law P. The goal is to construct asymptotically valid confidence intervals for a real-valued parameter $\theta = \theta(P)$ on the basis of observing the finite segment $X_1, \ldots, X_n$. For brevity we only consider two-sided symmetric confidence intervals; one-sided confidence intervals and two-sided equal-tailed intervals are treated similarly. The existence of an estimator $\hat{\theta}_n = \hat{\theta}(X_1, \ldots, X_n)$ is assumed. The basis of constructing confidence intervals for θ is the estimation of the two-sided sampling distribution of $\hat{\theta}_n$, properly normalized. To this end let

$$J_n(x, P) = Prob_P\{\tau_n|\hat{\theta}_n - \theta| \leqslant x\},$$

where $\{\tau_n\}$ is a normalizing sequence. We shall assume here that $\tau_n = n^\beta$ for some positive real number β.

The subsampling approximation to $J_n(x, P)$ is defined by

$$L_{n,b}(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b|\hat{\theta}_{b,a} - \hat{\theta}_n| \leqslant x\},$$

where the integer $1 < b < n$ is referred to as the block size, $\hat{\theta}_{b,a} = \hat{\theta}(X_a, \ldots, X_{a+b-1})$ is the estimator of θ computed on the block (or subsample) of data $\{X_a, \ldots, X_{a+b-1}\}$. The quantiles of the subsampling distribution $L_{n,b}$ can then be used to construct asymptotically valid confidence intervals for θ. To be more specific, let $c_{n,b}(1-\alpha)$ be an $(1-\alpha)$ quantile of $L_{n,b}$. The symmetric subsampling interval is then given as

$$I_{\text{sym}} = [\hat{\theta}_n \pm \tau_n^{-1} c_{n,b}(1-\alpha)]. \tag{3}$$

This interval can be shown to have the right coverage probability asymptotically under very weak conditions. Specifically, the following corollary is a special case of Corollary 3.2.1 of PRW (1999).

Corollary 2.1. *Assume that* (i) *$J_n(P)$ converges weakly to a continuous limiting distribution*; (ii) *the sequence $\{X_t\}$ is strong mixing*; *and* (iii) $b \to \infty$, $b/n \to 0$, $\tau_b \to \infty$ *and* $\tau_b/\tau_n \to 0$ *as* $n \to \infty$.

Then the confidence interval I_{sym} of (3) *has asymptotic coverage probability of* $1-\alpha$.

To use this construction, one has to know the rate of convergence τ_n. For our application of the threshold parameter r, this would be $n^{1/2}$ for a continuous SETAR model and n for a discontinuous SETAR model. Therefore, in the general case, when the (dis)continuity of the model is unknown, the standard subsampling method is not applicable. One can get around this problem by using subsampling in conjunction with an estimated rate of convergence. Assume an estimator of the rate denoted by $\hat{\tau}_n$ is available. Then one simply uses the standard method with τ_n replaced by $\hat{\tau}_n$. Let

$$\hat{L}_{n,b}(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\hat{\tau}_b |\hat{\theta}_{b,a} - \hat{\theta}_n| \leqslant x\}.$$

Denoting an $(1-\alpha)$ quantile of $\hat{L}_{n,b}$ by $\hat{c}_{n,b}(1-\alpha)$, the symmetric subsampling interval based on the estimated rate of convergence is then given as

$$\hat{I}_{\text{sym}} = [\hat{\theta}_n \pm \hat{\tau}_n^{-1} \hat{c}_{n,b}(1-\alpha)]. \tag{4}$$

In typical applications, the (unknown) rate is of the form $\tau_n = n^\beta$ with β unknown. The problem of rate estimation is then reduced to estimating the power β. Given an estimator $\hat{\beta}$, the estimated rate becomes $\hat{\tau}_n = n^{\hat{\beta}}$. If $\hat{\beta}$ converges sufficiently fast to β, interval (4) can be shown to have the right coverage probability asymptotically under very weak conditions. Specifically, the following corollary is a special case of Theorem 8.3.2 of PRW (1999).

Corollary 2.2. *Assume that* (i) $J_n(P)$ *converges weakly to a continuous limiting distribution*; (ii) *the sequence* $\{X_t\}$ *is strong mixing*; (iii) $b \to \infty$, $b/n \to 0$, $\tau_b \to \infty$ *and* $\tau_b/\tau_n \to 0$ *as* $n \to \infty$; *and* (iv) $\tau_n = n^\beta$ *and* $\hat{\beta} = \beta + o_P((\log n)^{-1})$.

Then the confidence interval $\hat{I}_{\text{sym}}$ *of* (4) *has asymptotic coverage probability of* $1-\alpha$.

2.2. General hypothesis tests

More generally, assume that the unknown law P is assumed to belong to a certain class of laws $\mathscr{P}$. The null hypothesis H_0 asserts $P \in \mathscr{P}_0$, and the alternative hypothesis H_1 is $P \in \mathscr{P}_1$, where $\mathscr{P}_i \subset \mathscr{P}$and $\mathscr{P}_0 \cup \mathscr{P}_1 = \mathscr{P}$. The goal is to construct an asymptotically valid test based on a given test statistic,

$$W_n = \tau_n w_n = \tau_n w_n(X_1, \ldots, X_n),$$

where τ_n is a normalizing sequence. Let

$$G_n(x, P) = Prob_P\{\tau_n w_n \leqslant x\}.$$

It will be assumed that $G_n(\cdot, P)$ converges in distribution, at least for $P \in \mathscr{P}_0$. Of course, this would imply (as long as $\tau_n \to \infty$) that $w_n \to 0$ in probability for $P \in \mathscr{P}_0$. Naturally, w_n should somehow be designed to distinguish between the competing hypotheses. The method we describe assumes w_n is constructed to satisfy the following: $w_n \to w(P)$ in probability, where $w(P)$ is a constant which satisfies $w(P) = 0$ if $P \in \mathscr{P}_0$ and $w(P) > 0$ if $P \in \mathscr{P}_1$.

To describe the test construction, let $w_{b,a}$ be equal to the statistic w_b evaluated at the block of data $\{X_a, \ldots, X_{a+b-1}\}$. The sampling distribution of W_n is then approximated by

$$\hat{G}_{n,b}(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b w_{b,a} \leqslant x\}.$$

Using this estimated sampling distribution, the critical value for the test is obtained as the $1-\alpha$ quantile of $\hat{G}_{n,b}(\cdot)$; specifically, define

$$g_{n,b}(1-\alpha) = \inf\{x : \hat{G}_{n,b}(x) \geqslant 1-\alpha\}.$$

Finally, the nominal level α test rejects H_0 if and only if

$$W_n > g_{n,b}(1-\alpha). \tag{5}$$

This construction can be shown to lead to a test that has asymptotically the right size and is consistent under very weak conditions. Specifically, the following corollary is a special case of Theorem 3.5.1 of PRW (1999).

Corollary 2.3. *Assume that* (i) *for* $P \in \mathscr{P}_0$, $G_n(P)$ *converges weakly to a continuous limit distribution*; (ii) *the sequence* $\{X_t\}$ *is strong mixing*; (iii) $b \to \infty$, $b/n \to 0$, $\tau_b \to \infty$ *and* $\tau_b/\tau_n \to 0$ *as* $n \to \infty$; *and* (iv) $w(P) = 0$ *if* $P \in \mathscr{P}_0$ *and* $w(P) > 0$ *if* $P \in \mathscr{P}_1$.

Then probability (5) *tends to* α *when* $P \in \mathscr{P}_0$ *and it tends to* 1 *when* $P \in \mathscr{P}_1$.

Remark 2.1. Alternatively, one could compute a subsampling P-value given as

$$PV_{n,b} = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{\tau_b w_{b,a} \geqslant W_n\}.$$

In this case, the nominal level α test rejects H_0 if and only if

$$PV_{n,b} < \alpha.$$

3. Confidence intervals for SETAR model parameters

This section describes how to use subsampling to construct confidence intervals for SETAR model parameters. The (joint) estimation of the parameter vector ϑ is carried out by the method of conditional least squares (CLS); see Chan (1993) and Chan and Tsay (1998). Note that in empirical applications, the lag parameter d is sometimes assumed to be known and is then not estimated from the data; an example is the application of Chan and Tsay (1998).

The theoretical results to follow refer to regularity conditions of Chan (1993) and of Chan and Tsay (1998), respectively. For self-completeness, we briefly restate these conditions:

Regularity Conditions 3.1 (*Chan (1993), Theorem 2*). *Let* P^l *denote the* l*-step transition probability of the Markov chain* $\{\boldsymbol{Z}_t\}$, *where* $\boldsymbol{Z}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})'$. *Assume the following*:

C1 $\{\boldsymbol{Z}_t\}$ *admits a unique invariant measure* $\pi(\cdot)$ *such that there exist constants* $K < \infty$ *and* $\rho < 1$ *so that for all* $z \in \mathbb{R}^p$ *and for all* $l \in \mathbb{N}$, $||P^l(z,\cdot) - \pi(\cdot)|| \leqslant K(1 + |z|)\rho^l$, *where* $||\cdot||$ *and* $|\cdot|$ *denote the total variation norm and the Euclidean norm, respectively*.

C2 ε_t *is absolutely continuous with a uniformly continuous and positive probability density function. Furthermore,* $E(\varepsilon_t^4) < \infty$.

C3 $\{X_t\}$ *is stationary with* $E(X_t^4) < \infty$.

Regularity Conditions 3.2 (Chan and Tsay (1998), *Theorem 2.2*). *Assume that* $\{X_t\}$ *generated by* (2) *is a* β*-mixing stationary process with a geometrically decaying mixing rate. Assume that* $\phi_{d-} \neq \phi_{d+}$, *and, for some* $q > 2$, $E(|X_t|^q) < \infty$. *Assume further that the stationary probability density function of* X_t *is positive everywhere and is bounded over a neighborhood of the true threshold parameter* r.

We would like to point out that these are the regularity conditions used to derive the limiting distribution of CLS estimators in the discontinuous and the continuous case, respectively. We do not need to strengthen these conditions in order to prove the validity of the subsampling inference methods, even in the general case when the (dis)continuity of the model is unknown.

3.1. Confidence intervals for the threshold parameter r

First, consider the continuous case. Chan and Tsay (1998), basing the estimation on model (2), construct normal theory confidence intervals for r. A simulation study in their paper shows that this method tends to undercover quite a bit. As an alternative, the subsampling method can be used.

Theorem 3.1. *Base the estimation of* r *on estimating model* (2). *Assume the model is continuous and assume Regularity Conditions* 3.2. *Let* $\theta = r$ *and* $\tau_n = n^{1/2}$. *Further, assume that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$.

Then the confidence interval (3) *has asymptotic coverage probability* $1 - \alpha$.

Next, consider the discontinuous case. Chan (1993) proves the strong consistency and the limiting distribution of $\hat{r}_n$. However, the distribution is non-standard and depends, in a very complicated way, on the underlying probability mechanism. Indeed, $n(\hat{r}_n - r)$ converges weakly to a random variable M_-, where $[M_-, M_+)$ is the unique random interval over which a compound Poisson process attains its global minimum. Even though the underlying probability mechanism arguably can be estimated consistently, it is not clear how one could go from there to consistently estimate the distribution of M_- as a basis for asymptotic inference for r. The subsampling method can be used to construct valid confidence intervals.

Theorem 3.2. *Base the estimation of* r *on estimating model* (1). *Assume the model is discontinuous and assume Regularity Conditions* 3.1. *Let* $\theta = r$ *and* $\tau_n = n$. *Further, assume that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$.

Then the confidence interval (3) *has asymptotic coverage probability* $1 - \alpha$.

Finally, consider the general case. We apply subsampling in conjunction with an estimated rate of convergence. It is known that $\tau_n = n^\beta$, where β is equal to either 0.5 (if the model is continuous) or to 1 (if the model is discontinuous). The following theorem demonstrates that the asymptotic validity of subsampling confidence intervals is not affected as long as the estimator of β converges sufficiently fast.

Theorem 3.3. *Base the estimation of* r *on estimating model* (1). *If the model is discontinuous, assume Regularity Conditions* 3.1. *If the model is continuous, assume Regularity Conditions* 3.2. *Let* $\theta = r$ *and* $\hat{\tau}_n = n^{\hat{\beta}}$, *where* $\hat{\beta} = \beta + o_P((\log n)^{-1})$. *Further, assume that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$.

Then the confidence interval (4) *has asymptotic coverage probability* $1 - \alpha$.

Remark 3.1. The key ingredients of the theorem are that in both the continuous and the discontinuous cases $\tau_n(\hat{r}_n - r)$ has a proper limiting distribution, that the rate τ_n is allowed to depend on the case, and that it can be estimated consistently satisfying a certain regularity condition. In the discontinuous case, the convergence of $n(\hat{r}_n - r)$ to a proper, albeit non-standard limiting distribution is proved in Chan (1993). While Chan and Tsay (1998) discuss continuous SETAR models, their results cannot be used for our theorem because they consider a restricted fit based on model (2). What is needed instead is the asymptotic distribution of $\hat{r}_n$ when the model is continuous but the unrestricted model (1) is estimated. A corresponding result is stated as Theorem A.1 in the Appendix.

The applicability of the suggested method now hinges on an estimator of β, the power of n in the rate of convergence. Indeed, subsampling can be applied to this end as well. The basic idea is the following. Since $\tau_n|\hat{r}_n - r|$ converges to a non-degenerate distribution, loosely speaking, $|\hat{r}_n - r|$ converges to the point mass zero at rate τ_n. Therefore, by comparing a number of subsampling distributions, based on distinct block sizes $b_1 \ldots b_I$, which estimate the sampling distribution of the un-scaled statistic $|\hat{r}_n - r|$, one can consistently estimate the rate τ_n. In the interest of space, we can only present the formula of the resulting estimator; for a detailed discussion, the reader is referred to PRW (1999, Section 8.2). Define

$$K_{n,b}(x) = \frac{1}{n-b+1} \sum_{a=1}^{n-b+1} 1\{|\hat{r}_{b,a} - \hat{r}_n| \leqslant x\},$$

and denote by $K_{n,b}^{-1}(t)$ a t-quantile of $K_{n,b}$. Now, let $b_i = \lfloor n^{\gamma_i} \rfloor$, for constants $0 < \gamma_1 < \cdots < \gamma_I$, let t_j, for $j = 1, \ldots, J$, be some points in $(0.5, 1)$, and let

$$y_{i,j} = \log(K_{n,b_i}^{-1}(t_j)).$$

The following estimator of β then satisfies $\hat{\beta}_{I,J} = \beta + o_P((\log n)^{-1})$:

$$\hat{\beta}_{I,J} = -\frac{\sum_{i=1}^{I}(y_{i,\cdot} - \bar{y})(\log b_i - \overline{\log})}{\sum_{i=1}^{I}(\log b_i - \overline{\log})^2}, \tag{6}$$

where

$$y_{i,\cdot} = J^{-1}\sum_{j=1}^{J} y_{i,j}, \quad \bar{y} = (IJ)^{-1}\sum_{i=1}^{I}\sum_{j=1}^{J} y_{i,j}, \quad \text{and} \quad \overline{\log} = I^{-1}\sum_{i=1}^{I}\log(b_i).$$

3.2. Confidence intervals for regression parameters ϕ_{ij}

It is also of interest to make inference for the regression parameters ϕ_{ij}. On grounds of consistency, the problem can be considered solved when it is known whether the SETAR model is continuous or not. In both cases, $n^{1/2}(\hat{\phi}_{ij,n} - \phi_{ij})$ converges to normal distribution with mean zero and a variance that can be consistently estimated. The result for the continuous case is given by Chan and Tsay (1998) and the one for the discontinuous case by Chan (1993). It should be mentioned, though, that the method of Chan (1993) tends to undercover in finite samples because it does not take the estimation uncertainty about $\hat{r}_n$ into account (e.g., Hansen, 2000). When the (dis)continuity of the model is unknown, standard inference is rendered infeasible, since the form of the limiting variance is different in the two cases. Instead, the subsampling method can be used. Given that the rate of convergence of $\hat{\phi}_{ij,n}$ does not depend on the continuity of the SETAR model, the complication of the rate estimation does not occur.

The following theorem shows that when the estimation is based on model (1) subsampling confidence intervals for $\phi_{i,j}$ will always have asymptotically correct coverage probability. The validity of this approach when the true model is continuous again hinges on Theorem A.1.

Theorem 3.4. *Base the estimation of ϕ_{ij} on estimating model* (1). *If the model is discontinuous, assume Regularity Conditions* 3.1. *If the model is continuous, assume Regularity Conditions* 3.2. *Let $\theta = \phi_{ij}$ and $\tau_n = n^{1/2}$. Further, assume that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$.*

Then the confidence interval (3) *has asymptotic coverage probability* $1 - \alpha$.

4. Comparison with a related method

To construct confidence intervals for SETAR parameters in discontinuous models, the approach of Hansen (2000) could also be adopted. Note that Hansen's framework is richer, since it allows general regression models where the predictor variables do not have to be lagged variables of the response. In what follows, we will discuss how his method is applied to SETAR models as a special case. To circumvent the non-standard and difficult asymptotics of $\hat{r}_n$ in discontinuous models, Hansen (2000) assumes that the "threshold effect", that is, the difference between the two regression coefficient parameters, shrinks to zero as the sample size increases:

$$\Phi_1 - \Phi_2 = \Delta n^{-\alpha} \text{ with } \varDelta \neq 0 \quad \text{and} \quad 0 \leqslant \alpha < 0.5.$$

Under the assumption that $\sigma_1 = \sigma_2$, Hansen (2000) is able to construct confidence intervals for r by inverting a likelihood ratio test for r. The ensuing intervals are asymptotically correct when $\alpha > 0$. In the case of a fixed, non-vanishing threshold effect (that is, when $\alpha = 0$), the intervals are shown to be asymptotically conservative under the additional assumption that the innovations are Gaussian.

The method has a number of problems. First, it is doubtful that it can be extended to non-Gaussian innovations because the proof relies heavily on the Gaussian innovation structure. Second, it is assumed that $\sigma_1 = \sigma_2$, so that the innovation terms are required to have the same variance in the two regimes. Third, the confidence intervals are conservative when $\alpha = 0$ and the simulations in Hansen (2000) show that unless $\Phi_1 - \Phi_2$ is close to zero and n is small, the intervals overcover significantly. Fourth, the method cannot be extended to continuous SETAR models, so it would not work in the general case.

We can also compare inference for regression parameters ϕ_{ij}. Hansen (2000) shows that, as in Chan (1993) in his model also, $n^{1/2}(\hat{\phi}_{ij,n} - \phi_{ij})$ converges to a normal distribution with mean zero and that the limiting variance is the same as when r is known. He argues correctly that 'in finite samples, this procedure seems likely to under-represent the true sampling uncertainty, since it is not the case that $\hat{r}_n = r$ in any given sample'. Therefore, he suggests a Bonferroni-type bound in the following way. First, one constructs a $1 - \varrho$ level confidence interval for r. Next, for each $r^\dagger$ contained in that interval, one constructs a $1 - \alpha$ level confidence interval for ϕ_{ij}, acting as if $r^\dagger$ were the true parameter. Finally, one takes the union over $r^\dagger$ of all the $1 - \alpha$ intervals for ϕ_{ij}. The question is how to choose the model parameter ϱ; note that the choice $\varrho = 1$ would correspond to treating $\hat{r}_n$ as the true parameter, that is, the approach of Chan (1993). Based on some simulations, Hansen (2000) suggests to use $\varrho = 0.2$. Obviously, this is an ad hoc method whose asymptotic properties are not clear. On the other hand, the subsampling inference for ϕ_{ij} yields confidence intervals with asymptotically correct coverage probability. It does not under-represent the true sampling uncertainty, since r is also estimated from the subsamples.

5. A test for continuity

An important issue that has not been explored in the literature is to test whether a SETAR model is continuous or not. Chan and Tsay (1998) apply both continuous and discontinuous models to quarterly U.S. unemployment rates and note that the two estimated models are close to each other, 'which is indicative of using a continuous model'. But they are not able to test whether this hypothesis may be violated. We will now describe how the general subsampling hypothesis testing approach of Section 2.2 can be adopted to this end. As was noted earlier, a necessary and sufficient condition for a SETAR model to be continuous is that $\phi_{1j} = \phi_{2j}$ for $1 \leqslant j \neq d \leqslant p$ and that $\phi_{10} + r\phi_{1d} = \phi_{20} + r\phi_{2d}$. Obviously, this is equivalent to $h(\vartheta) = 0$, with

$$h(\vartheta) = |\phi_{10} + r\phi_{1d} - \phi_{20} - r\phi_{2d}| + \sum_{1 \leqslant j \neq d \leqslant p} |\phi_{1j} - \phi_{2j}|.$$

Hence, it seems plausible to choose

$$w_n = w_n(X_1, \ldots, X_n) = h(\hat{\vartheta}_n)$$

as the test statistic, where the estimation of ϑ is based on model (1). The following theorem shows that this idea indeed leads to a test with asymptotically correct level. Moreover, as it should be, under the alternative hypothesis the power tends to 1.

Theorem 5.1. *Base the estimation of* ϑ *on estimating model* (1). *If the model is discontinuous, assume Regularity Conditions* 3.1. *If the model is continuous, assume Regularity Conditions* 3.2. *Denote by* $\mathscr{P}_0$ *the class of continuous SETAR models and by* $\mathscr{P}_1$ *the class of discontinuous SETAR models. Let* $w_n = h(\hat{\vartheta}_n)$ *and* $\tau_n = n^{1/2}$. *Further, assume that* $b \to \infty$ *and* $b/n \to 0$ *as* $n \to \infty$.

(i) *If the underlying SETAR model is continuous, then the subsampling test based on* (5) *has asymptotic size equal to* α.
(ii) *If the underlying SETAR model is discontinuous, then the subsampling test based on* (5) *has asymptotic power equal to* 1.

6. Choice of the block size

The application of the subsampling method requires a choice of the block size b; the problem is very similar to the choice of the bandwidth in applying smoothing or kernel methods. Unfortunately, the asymptotic requirements $b \to \infty$ and $b/n \to \infty$ as $n \to \infty$ give little guidance when faced with a finite sample. Instead, we propose to exploit the semi-parametric nature of SETAR models to estimate a 'good' block size in practice. The approach will be detailed for the use of subsampling for confidence interval construction. An analogous approach can be used when hypothesis tests are to be constructed; see Remark 6.1.

To illustrate the idea, assume the goal is to construct a $1-\alpha$ confidence interval for the univariate parameter of interest θ (the threshold parameter r or one of the regression parameters ϕ_{ij}). In finite samples, a subsampling interval will typically not exhibit coverage probability exactly equal to $1-\alpha$; moreover, the actual coverage probability generally depends on the block size b. Indeed, one can think of the actual coverage level $1-\lambda$ of a subsampling confidence interval as a function of the block size b, conditional on the underlying probability mechanism P—that is, the fully specified SETAR model in our application—and the nominal confidence level $1-\alpha$. The idea is now to adjust the 'input' b in order to obtain the actual coverage level close to the nominal one. Hence, one can consider the block size calibration function $g: b \to 1-\lambda$. If $g(\cdot)$ were known, one could construct an 'optimal' confidence interval by finding $\tilde{b}$ that minimizes $|g(b)-(1-\alpha)|$ and use $\tilde{b}$ as the block size; note that $|g(b)-(1-\alpha)|=0$ may not always have a solution.

Of course, the function $g(\cdot)$ depends on the underlying probability mechanism P and is therefore unknown. We now propose a semi-parametric bootstrap method to estimate it. The idea is that in principle we could simulate $g(\cdot)$ if P were known by generating data of size n according to P and computing subsampling confidence intervals for θ for a number of different block sizes b. This process is then repeated many times and for a given b one estimates $g(b)$ as the fraction of the corresponding intervals that contain the true parameter. The method we propose is identical except that P is replaced by an estimate $\hat{P}_n$.

For our application, P is the completely specified SETAR model. It depends on ϑ, σ_1, σ_2, and the marginal distribution of ε_t. The natural estimator of ϑ is $\hat{\vartheta}_n$, either based on estimating model (2) in case the model is known to be continuous or based on estimating model (1) otherwise. In principle, the remaining components could be estimated explicitly as well. Instead, we opt for an 'implicit estimation' by bootstrapping the residuals from the two distinct regimes. To this end, define, for $t = p+1, \ldots, n$,

$$\hat{u}_{t,n} = \begin{cases} X_t - \hat{\phi}_{10,n} - \hat{\phi}_{11,n} X_{t-1} - \cdots - \hat{\phi}_{1p,n} X_{t-p} & \text{if } X_{t-d} \leqslant \hat{r}_n, \\ X_t - \hat{\phi}_{20,n} - \hat{\phi}_{21,n} X_{t-1} - \cdots - \hat{\phi}_{2p,n} X_{t-p} & \text{if } X_{t-d} > \hat{r}_n, \end{cases}$$

$$\hat{U}_1 = \{\hat{u}_{t,n} : X_{t-d} \leqslant \hat{r}_n\}, \quad n_{\hat{U}_1} = |\hat{U}_1|$$

and

$$\hat{U}_2 = \{\hat{u}_{t,n} : X_{t-d} > \hat{r}_n\}, \quad n_{\hat{U}_2} = |\hat{U}_2|,$$

where, necessarily, $n_{\hat{U}_1} + n_{\hat{U}_2} = n - p$. Now, the estimated SETAR model, denoted by $\hat{P}_n$, gives rise to a sequence $X_1^*, \ldots, X_n^*$ in the following manner.

Algorithm 6.1 (*sampling from estimated SETAR model*).

1. Generate sequences $u_{i1}^*, \ldots, u_{in}^*$ by sampling with replacement from $\hat{U}_i$, for $i = 1, 2$.
2. $X_t^* = X_t$ for $t = 1, \ldots, p$.
3. $X_t^* = \begin{cases} \hat{\phi}_{10,n} + \hat{\phi}_{11,n} X_{t-1}^* + \cdots + \hat{\phi}_{1p,n} X_{t-p}^* + u_{1t}^* & \text{if } X_{t-d}^* \leqslant \hat{r}_n \\ \hat{\phi}_{20,n} + \hat{\phi}_{21,n} X_{t-1}^* + \cdots + \hat{\phi}_{2p,n} X_{t-p}^* + u_{2t}^* & \text{if } X_{t-d}^* > \hat{r}_n \end{cases}$ for $t = p+1, \ldots, n$.

Having specified how to generate data from estimated SETAR model, we next detail the algorithm to determine the block size b.

Algorithm 6.2 (*choice of the block size*).

1. Fix a selection of reasonable block sizes b between limits b_{low} and b_{up}.
2. Generate K pseudo sequences $X_{k1}^*, \ldots, X_{kn}^*$, $k = 1, \ldots, K$, according to Algorithm 6.1. For each sequence, $k = 1, \ldots, K$, and for each b, compute a subsampling confidence interval $\text{CI}_{k,b}$ for θ.
3. Compute $\hat{g}(b) = \#\{\hat{\theta}_n \in \text{CI}_{k,b}\}/K$.
4. Find the value $\tilde{b}$ that minimizes $|\hat{g}(b) - (1-\alpha)|$.

Remark 6.1. If subsampling is used to construct hypothesis tests rather than confidence intervals, then an analogous algorithm can be used by focusing on the size of the test rather than the confidence level of the interval. Of course, in doing so it is important that the estimated SETAR model $\hat{P}_n$ satisfy the null hypothesis. For example, for the continuity test of Section 5, one needs to base the estimation of P on estimating model (2).

Remark 6.2. Strictly speaking, the Theorems of Section 3 require an a priori determined sequence of block sizes b as $n \to \infty$. In practice, however, the choice of b will typically be data-dependent, such as given by Algorithm 6.2. As discussed in PRW (1999, Section 3.6), this does not affect the asymptotic validity of subsampling inference with strong mixing data as long as $b_{\text{low}} \to \infty$ and $b_{\text{up}}/n^{1/2} \to 0$ as $n \to \infty$. This result also implies the consistency of the subsampling inference for r when the true model is continuous but a discontinuous model is estimated in practice. While $\hat{P}_n$ will be discontinuous with probability one, the data-dependent choice of block size will result in confidence intervals with asymptotically correct coverage probability as long as the afore-mentioned conditions on b_{low} and b_{up} are satisfied.

7. Simulation evidence

The goal of this section is to examine the small sample performance of our methods via a simulation study. To reduce the computational burden, we consider the simplest case $d = p = 1$. The following two SETAR models are included in the study:

$$X_t = \begin{cases} 0.52 + 0.6X_{t-1} + \varepsilon_t & \text{if } X_{t-d} \leqslant 0.8, \\ 1.48 - 0.6X_{t-1} + 2\varepsilon_t & \text{if } X_{t-d} > 0.8 \end{cases} \tag{7}$$

and

$$X_t = \begin{cases} 0.7 - 0.5X_{t-1} + \varepsilon_t & \text{if } X_{t-d} \leqslant 0, \\ -1.8 + 0.7X_{t-1} + \varepsilon_t & \text{if } X_{t-d} > 0, \end{cases} \tag{8}$$

where the ε_t are i.i.d. N(0, 1). Model (7) is the continuous model used in Chan and Tsay (1998) with $\sigma_1 = 1$ and $\sigma_2 = 2$. The discontinuous model (8) is taken from Tong (1990, Section 5.5.3) with $\sigma_1 = \sigma_2 = 1$. Since it would be cumbersome to simulate X_1 directly from the stationary distribution of X_t, we start the simulations at $X_{-99} = 0$ and then discard the first 100 observations to avoid startup effects. Fig. 1 shows 500 data points from the two models, where X_t is plotted against X_{t-1} and the true autoregressive functions are overlaid.

7.1. Confidence intervals for SETAR model parameters

Performance of confidence intervals is judged by estimated coverage probabilities of nominal 90% and 95% two-sided symmetric subsampling intervals. The parameters of interest are r and ϕ_{11}. When intervals for r are constructed, we use both the true and the estimated rate of convergence. The former corresponds to knowing the (dis)continuity of the model while the latter corresponds to the general case. The three sample sizes considered are $n = 100$, 200, and 500.

Some words about the rate estimation are in order. We started out with the estimator $\hat{\beta}_{I,J}$ defined in (6), using $I = J = 4$. The quantiles t_j were evenly distributed

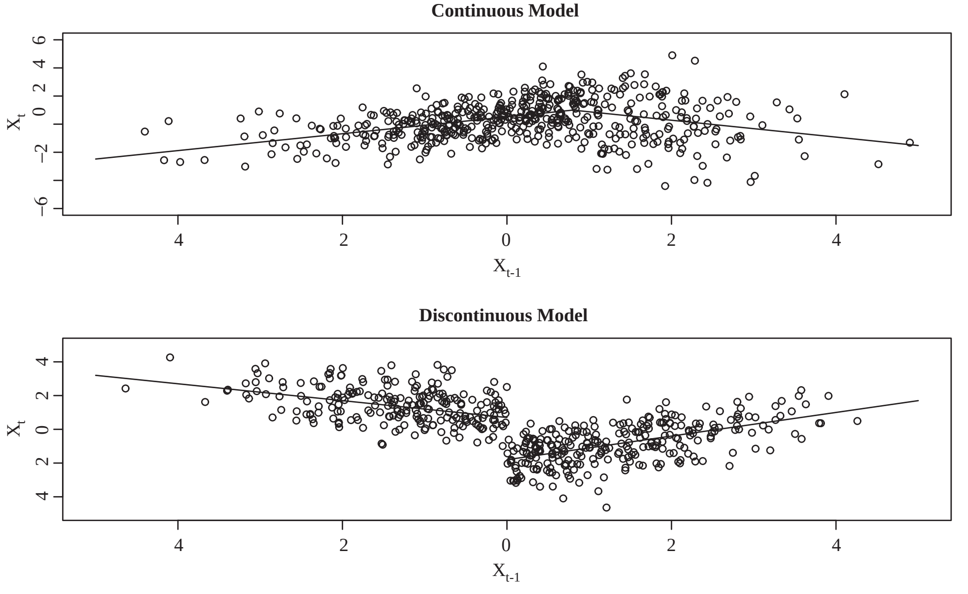

Fig. 1. 500 data points were generated from models (7) and (8), respectively. The plots show X_t against X_{t-1}, with the true autoregressive functions overlaid.

between 0.7 and 0.99. The block sizes b_i were chosen according to the rule

$$b_i = \lfloor n^{\gamma_i} \rfloor \quad \text{with} \quad \gamma_i = \kappa * [1 + \log((i+1)/(I+1))/\log 100], \; i = 1, \ldots, I, \tag{9}$$

where $0 < \kappa < 1$ is a model parameter. In small to moderate samples, this produced 'overdispersed' estimates. This means that in the continuous model, $\hat{\beta}_{I,J}$ tended to be less than 0.5 and in the discontinuous model, $\hat{\beta}_{I,J}$ tended to be bigger than 1. We therefore switch to the truncated estimator

$$\hat{\beta}^{\text{Trunc}}_{I,J} = \begin{cases} 0.5 & \text{if } \hat{\beta}_{I,J} \leqslant 0.75, \\ 1 & \text{if } \hat{\beta}_{I,J} > 0.75. \end{cases} \tag{10}$$

For an application, the model parameter κ in (9) has to be chosen. Table 1 reports how often the correct rate was identified in the two models for the parameters $\kappa = 0.7, 0.8$, and 0.9 and the sample sizes considered in our simulation study. It is seen that the method is not very reliable for $n = 100$ but that starting at $n = 200$, the choice $\kappa = 0.8$ yields a quite good estimator. In the simulations that follow, we employ the choice $\kappa = 0.8$ throughout.

Remark 7.1. Based on the simulation experience of Table 1, we make the following recommendations concerning the use of the rate estimator $\hat{\beta}^{\text{Trunc}}_{I,J}$ in empirical applications. Choose $I = J = 4$ and the quantiles t_j evenly distributed between 0.7 and 0.99. Choose the block sizes b_i as in (9) and $\kappa = 0.8$. Do not attempt rate estimation for sample sizes $n < 200$.

Table 1
Proportions of times the estimator $\hat{\beta}_{I,J}^{\text{Trunc}}$ selected the correct rate as a function of the model parameter κ and the sample size n

n	$\kappa = 0.7$	$\kappa = 0.8$	$\kappa = 0.9$
Continuous model			
100	0.82	0.82	0.73
200	0.95	0.95	0.86
500	1.00	0.99	0.98
Discontinuous model			
100	0.25	0.63	0.89
200	0.55	0.92	0.97
500	0.96	1.00	0.99

The continuous model is (7) and the discontinuous model is (8). The estimation is based on model (1) always. The results are based on 1000 replications.

Remark 7.2. The truncated estimator $\hat{\beta}_{I,J}^{\text{Trunc}}$ can be used as the basis of an alternative method to construct a confidence interval for the threshold parameter r in the general case. When $\hat{\beta}_{I,J}^{\text{Trunc}} = 0.5$, construct the interval assuming a continuous model; when $\hat{\beta}_{I,J}^{\text{Trunc}} = 1$, construct the interval assuming a discontinuous model. It is straightforward to see that this method yields consistent inference. For example, consider the case of the true model being continuous. Since $\hat{\beta}_{I,J}$ is a consistent estimator of $\beta = 0.5$, the event $\{\hat{\beta}_{I,J}^{\text{Trunc}} = 0.5\}$ has probability tending to one. Hence, with probability tending to one, the confidence interval will be based on a continuous model, which is the true state of nature. The same reasoning applies to case of the true model being discontinuous.[3]

For all scenarios, we include three fixed block sizes in addition to the 'optimal' block size chosen according to Algorithm 6.2. Since this algorithm is computationally rather expensive, we had to limit the input block sizes to the corresponding three fixed block sizes; note that in a concrete application a finer grid should be chosen. Also, for the parameter K of the algorithm, $K = 200$ is employed; in a concrete application, we suggest to employ $K = 1000$. All estimated coverage probabilities are based on 1000 repetitions. The results are presented in Tables 2 and 3.

[3]Note that this alternative method is different from a pretest method. A pretest method would start with a test for the null hypothesis of a continuous model. If the null is not rejected, the interval is constructed assuming a continuous model; if the null is rejected, the interval is constructed assuming a discontinuous model. The pretest method is troubled, though. Consider the case of the true model being continuous. Even when the sample size tends to infinity, the test will reject the null hypothesis with a positive probability, the nominal level of the test. Hence, the probability that the inference will be based on a continuous model remains strictly less than one.

Table 2
Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals for the threshold parameter r based on 1000 replications

Continuous model, $n = 100$						
Target	$b = 15$	$b = 25$	$b = 35$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.77	0.82	0.80	0.81	0.51	0.67
0.95	0.83	0.86	0.84	0.86	0.61	0.72
Continuous model, $n = 200$						
Target	$b = 30$	$b = 45$	$b = 60$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.88	0.89	0.87	0.88	0.55	0.86
0.95	0.82	0.93	0.91	0.92	0.67	0.91
Continuous model, $n = 500$						
Target	$b = 70$	$b = 100$	$b = 130$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.93	0.91	0.88	0.91	0.70	0.91
0.95	0.97	0.95	0.92	0.95	0.79	0.95
Discontinuous model, $n = 100$						
Target	$b = 10$	$b = 20$	$b = 30$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.89	0.94	0.87	0.91	0.92	0.61
0.95	0.93	0.97	0.92	0.95	0.96	0.69
Discontinuous model, $n = 200$						
Target	$b = 10$	$b = 25$	$b = 40$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.87	0.95	0.86	0.89	0.90	0.81
0.95	0.91	0.98	0.92	0.95	0.95	0.85
Discontinuous model, $n = 500$						
Target	$b = 15$	$b = 35$	$b = 55$	$\tilde{b}$	$\tilde{b}$ and $\hat{\tau}_n$	$\tilde{b}$ and $\hat{\beta}^{\text{Trunc}}$
0.90	0.91	0.87	0.79	0.90	0.90	0.90
0.95	0.94	0.95	0.87	0.94	0.94	0.94

The continuous DGP is (7) and the discontinuous DGP is (8). Columns 2–4 list the results for fixed block sizes using the true rate τ_n; column 5 lists the results for the adaptive choice of block size in conjunction with the true rate τ_n; in columns 2–5 the (dis)continuity of the model is assumed known. Column 6 lists the results when the estimation is based on the discontinuous model (1) always but the estimated rate of convergence is used. Column 7 lists the results when either the continuous model (1) or the discontinuous model (2) is used, depending on the outcome of the truncated rate estimator (10); see Remark 7.2. Both columns 6 and 7 use the adaptive choice of block size.

First, we discuss the confidence intervals for the threshold parameter r in the continuous model. The results for fixed block sizes (columns 2 to 4) and the first results for the data-dependent choice of block size (column 5) are based on estimating model (2). This approach corresponds to knowing the continuity of the model and hence the numbers should be compared to the simulations of Chan and Tsay (1998). It is seen that the intervals undercover for $n = 100$ and $n = 200$. Still, our results for $n = 100$ are comparable to those of Chan and Tsay (1998) for $n = 200$; and our results for $n = 200$ are comparable to those of Chan and Tsay (1998) for $n = 1000$. When $n = 500$ the intervals work satisfactorily while those of Chan and Tsay (1998) still undercover when $n = 2000$. Hence, in this context, subsampling offers improved finite-sample performance compared to the asymptotic method

Table 3
Estimated coverage probabilities of nominal 90% and 95% subsampling confidence intervals for the regression parameter ϕ_{11} based on 1000 replications

Continuous model, $n = 100$				
Target	$b = 15$	$b = 25$	$b = 35$	$\tilde{b}$
0.90	0.95	0.90	0.85	0.87
0.95	0.98	0.95	0.90	0.92
Continuous model, $n = 200$				
Target	$b = 20$	$b = 35$	$b = 50$	$\tilde{b}$
0.90	0.95	0.91	0.87	0.88
0.95	0.98	0.96	0.92	0.93
Continuous model, $n = 500$				
Target	$b = 50$	$b = 80$	$b = 110$	$\tilde{b}$
0.90	0.92	0.89	0.87	0.89
0.95	0.97	0.95	0.92	0.95
Discontinuous model, $n = 100$				
Target	$b = 20$	$b = 25$	$b = 30$	$\tilde{b}$
0.90	0.94	0.91	0.87	0.91
0.95	0.96	0.94	0.92	0.95
Discontinuous model, $n = 200$				
Target	$b = 30$	$b = 40$	$b = 50$	$\tilde{b}$
0.90	0.92	0.88	0.84	0.89
0.95	0.96	0.93	0.89	0.94
Discontinuous model, $n = 500$				
Target	$b = 40$	$b = 60$	$b = 80$	$\tilde{b}$
0.90	0.94	0.90	0.86	0.90
0.95	0.97	0.94	0.92	0.95

The continuous DGP is (7) and the discontinuous DGP is (8). Columns 2–4 list the results for fixed block sizes and column 5 lists the results for the adaptive choice of block size. The estimation is based on the discontinuous model (1) always.

based on normality. The results for the data-dependent choice of block size in column 6 are based on estimating model (1) in conjunction with estimating the rate of convergence; this approach corresponds to the general case. The results are certainly disappointing, though they get less disappointing as the sample size n increases.[4] The results for the data-dependent choice of block size in column 7 are based on the technique described in Remark 7.2. The coverage probabilities are much improved and for $n = 500$ are identical to column 5. This is no surprise, since according to Table 1 the truncated rate estimator will indicate the continuous model with probability 0.99 in this case.

Second, we discuss the confidence intervals for the threshold parameter r in the discontinuous model. When the discontinuity of the model is known, the coverage probabilities are satisfactory starting at $n = 100$ already. See column 5 for the results

[4]Part of the disappointment seems to be due to the unequal innovation standard deviations in model (7). When the innovation in both regimes have equal standard deviation one the coverage probabilities are much higher, see Gonzalo and Wolf (2001).

for the data-dependent choice of block size which compare very favorably to the simulations of Hansen (2000). The results for the data-dependent choice of block size in column 6 are based on estimating model (1) in conjuction with estimating the rate of convergence; this approach corresponds to the general case. The empirical coverage probabilities differ very little from those of column 5, so little is lost in not knowing the true state of nature in this approach. The results for the data-dependent choice of block size in column 7 are based on the technique described in Remark 7.2. The coverage probabilities are worse than those in column 6 for $n = 100, 200$ but for $n = 500$ they are identical to column 5. This is no surprise, since according to Table 1 the truncated rate estimator will indicate the continuous model with probability 1 in this case.

Based on these simulation results, we would recommend the technique described in Remark 7.2 for the general case. Starting at $n = 500$, basically nothing is lost in not knowing the (dis)continuity of the model, as the truncated rate estimator (10) will indicate the true state of nature with probability very close to one. For smaller values of n the coverage probabilities do suffer. On the other hand, for smaller values of n the confidence intervals for r are not very trustworthy even in the case of a continuous model with the continuity known; see the simulations in Chan and Tsay (1998) and column 5 of Table 2. If we require $n \geqslant 500$ to make inference for r in the continuous case[5], there is no further detriment if we require $n \geqslant 500$ in the general case.

Table 4 sheds some light on the distribution of the data-dependent choice of block size in selected scenarios.

Third, we discuss the intervals for the regression parameter ϕ_{11}. In both models, the results are always based on estimating model (1). The data-dependent choice of block size works quite satisfactorily, though the intervals undercover somewhat in the continuous model.[6] The results for the discontinuous model compare favorably to the ad hoc method of Hansen (2000) who employs a Bonferroni-type method (see Section 4).

7.2. Test for continuity

A similar simulation setup is used to judge the performance of the subsampling test for the null hypothesis of a continuous SETAR model; see Remark 6.1 for the data-dependent choice of the block size. The results are presented in Table 5. Note that the test over-rejects for small sample sizes but as the sample size increases, the actual level tends to the nominal level and the power tends to one in accordance with the theory. Moreover, the data-dependent choice of the block size performs well.

[5]While this might seem a strong requirement, recall that the inference based on asymptotic normality of Chan and Tsay (1998) would require $n > 2000$.

[6]While not reported, estimated coverage probabilities in the continuous model improve if model (2) is estimated. When the continuity of the model is known, the intervals works satisfactorily for $n = 100$ already.

Table 4
Proportions of times the input block sizes b were chosen by Algorithm 6.2 in the determination of the data-dependent block size $\tilde{b}$ in various scenarios

CI for r, continuous model, $n = 100$			
Target	$b = 15$	$b = 25$	$b = 35$
0.90	0.16	0.62	0.22
0.95	0.14	0.70	0.16
CI for r, discontinuous model, $n = 100$			
Target	$b = 10$	$b = 20$	$b = 30$
0.90	0.39	0.38	0.23
0.95	0.28	0.58	0.14
CI for ϕ_{11}, continuous model, $n = 100$			
Target	$b = 15$	$b = 25$	$b = 35$
0.90	0.09	0.21	0.70
0.95	0.13	0.38	0.49
CI for ϕ_{11}, discontinuous model, $n = 100$			
Target	$b = 20$	$b = 25$	$b = 30$
0.90	0.24	0.54	0.22
0.95	0.39	0.50	0.11

The continuous model is (7) and the discontinuous model is (8). The results for the confidence interval for r correspond to column 5 of Table 2 and the results for the confidence interval for ϕ_{11} correspond to column 5 of Table 3. The results are based on 1000 replications.

8. Empirical application

Chan and Tsay (1998) fitted the following continuous SETAR(2) model to the first differences of the quarterly U.S. unemployment rates from 1948 to 1993 ($T = 184$):

$$\hat{X}_t = 0.0888 + 0.7870X_{t-1} + \begin{cases} 0.1060\,(X_{t-2} - r) & \text{if } X_{t-2} \leqslant 0.134, \\ -0.5582\,(X_{t-2} - r) & \text{if } X_{t-2} > 0.134, \end{cases} \tag{11}$$

where the sample sizes for the two regimes are 130 and 52, respectively. As a comparison, they also employed a discontinuous SETAR(2) model to the same data. The following is the estimated model:

$$\hat{X}_t = \begin{cases} 0.0207 + 0.6011X_{t-1} + 0.0801X_{t-2} & \text{if } X_{t-2} \leqslant 0.034, \\ 0.2280 + 0.8815X_{t-1} - 0.6903X_{t-2} & \text{if } X_{t-2} > 0.034, \end{cases} \tag{12}$$

where the sample sizes for the two regimes are 115 and 67, respectively. Comparing with the continuous model in (11), Chan and Tsay (1998) observed that the two models are similar but were not able to formally test the null hypothesis of a continuous model.

We now apply the test of Section 5. Table 6 presents the estimated rejection probabilities of the test under the null for various block sizes and nominal levels. (The smallest block size included is $b = 30$, since for values smaller than that the estimation of a SETAR(2) model becomes problematic.) The numbers in the table indicate that the test tends to over-reject. Given the relatively small sample size of $n = 184$, this is consistent with the simulation study in the previous section. For

Table 5
Estimated rejection probabilities of nominal 10% and 5% subsampling hypothesis tests for the null hypothesis of a continuous SETAR model based on 1000 replications

Continuous model, $n = 100$				
Target	$b = 20$	$b = 30$	$b = 40$	$\tilde{b}$
0.10	0.20	0.18	0.21	0.18
0.05	0.12	0.12	0.15	0.12
Continuous model, $n = 200$				
Target	$b = 20$	$b = 40$	$b = 60$	$\tilde{b}$
0.10	0.14	0.13	0.13	0.12
0.05	0.07	0.07	0.08	0.07
Continuous model, $n = 500$				
Target	$b = 90$	$b = 120$	$b = 150$	$\tilde{b}$
0.10	0.08	0.09	0.11	0.09
0.05	0.04	0.05	0.06	0.05
Discontinuous model, $n = 100$				
	$b = 20$	$b = 30$	$b = 40$	$\tilde{b}$
	0.89	0.90	0.91	0.91
	0.96	0.96	0.94	0.96
Discontinuous model, $n = 200$				
	$b = 20$	$b = 40$	$b = 60$	$\tilde{b}$
	0.98	0.99	0.99	0.99
	1.00	1.00	1.00	1.00
Discontinuous model, $n = 500$				
	$b = 30$	$b = 60$	$b = 90$	$\tilde{b}$
	1.00	1.00	1.00	1.00
	1.00	1.00	1.00	1.00

The continuous DGP is (7) and the discontinuous DGP is (8). Columns 2–4 list the results for fixed block sizes and column 5 lists the results for the adaptive choice of block size.

Table 6
Estimated rejection probabilities of the subsampling test for continuity for the unemployment data of Section 8 as a function of the nominal level α and the block size b

α	$b = 30$	$b = 40$	$b = 50$
0.025	0.05	0.1	0.12
0.05	0.09	0.14	0.15
0.1	0.14	0.18	0.20

The probabilities are estimated using the calibration method of Section 6.

example, according to the estimation, a test with nominal level $\alpha = 0.025$ and block size $b = 30$ has an actual level of about 0.05. And a test with nominal level $\alpha = 0.05$ and block size $b = 30$ has an actual level of about 0.09. Table 7 presents the subsampling P-values for the null hypothesis of a continuous model and various block sizes. All the P-values are well above 0.1. This fact together with test being

Table 7
Subsampling P-values for the null hypothesis of a continuous SETAR(2) model for the unemployment data of Section 8 as a function of the block size b

$b = 30$	$b = 40$	$b = 50$
0.37	0.25	0.16

The P-values are computed as described in Remark 2.1.

somewhat anticonservative implies that we cannot reject the null hypothesis of a continuous SETAR(2) model.

9. Discussion

We have proposed the subsampling methodology as a unified inference method in SETAR models. First, it solves several problems that had not been solved before: consistent confidence intervals for the threshold parameter r when the model is discontinuous; and consistent confidence intervals for r and for regression parameters ϕ_{ij} when the (dis)continuity of the model is unknown. Second, it improves the finite sample performance of some previous approaches: confidence intervals for r when the model is continuous (Chan and Tsay, 1998), and confidence intervals for ϕ_{ij} when the model is discontinuous (Chan, 1993; Hansen, 2000). Third, it considers and solves a problem that had been neglected so far: a hypothesis test for the continuity of a SETAR model.

Acknowledgements

We are grateful for helpful comments of two anonymous referees that have led to an improved presentation of the paper. Moreover, we thank Professor Kung-Sik Chan for providing us with the unemployment data analyzed in Section 8.

Appendix A. Proofs of technical results

Proof of Theorem 3.1. The weak convergence of $n^{1/2}(\hat{r}_n - r)$ to a normal distribution follows from Theorem 2.2 of Chan and Tsay (1998). One of the regularity conditions of the theorem is that the underlying sequence $\{X_t\}$ is β-mixing, which in return implies that $\{X_t\}$ is strong mixing (Doukhan, 1994). The result now follows from Corollary 2.1. $\square$

Proof of Theorem 3.2. The weak convergence of $n(\hat{r}_n - r)$ to a continuous limiting distribution follows from Theorem 2 of Chan (1993). Next, consider $\boldsymbol{Z}_t = (X_t, X_{t-1}, \ldots, X_{t-p+1})'$. Then $\{\boldsymbol{Z}_t\}$ is a Markov Chain. The regularity conditions of

Theorem 2 of Chan (1993) imply that the chain is geometrically ergodic (Chan, 1993), which in return implies that $\{\boldsymbol{Z}_t\}$ is β-mixing (Chan and Tsay, 1998), which in return implies that $\{X_t\}$ is strong mixing (Doukhan, 1994). The result now follows from Corollary 2.1. □

Proof of Theorem 3.3. It suffices to show that both in the discontinuous and in the continuous case the assumptions of Corollary 2.2 are satisfied. In the continuous case, this follows from Theorem A.1 at the end of this Appendix and the fact that $\{X_t\}$ is strong mixing, as discussed in the proof of Theorem 3.2. In the discontinuous case, this follows from Theorem 2 of Chan (1993) and the fact that $\{X_t\}$ is strong mixing. □

Proof of Theorem 3.4. It suffices to show that both in the discontinuous and in the continuous case the assumptions of Corollary 2.1 are satisfied. In the continuous case, this follows from Theorem A.1 and the fact that $\{X_t\}$ is strong mixing, as discussed in the proof of Theorem 3.2. In the discontinuous case, this follows from Theorem 2 of Chan (1993) and the fact that $\{X_t\}$ is strong mixing. □

Proof of Theorem 5.1. The almost sure convergence of w_n to $w(P) = h(\vartheta)$ both under the null and alternative hypothesis follows immediately from Theorem 1 of Chan (1993). Obviously, under the null hypothesis $h(\vartheta)$ is equal to zero and under the alternative it is positive. The convergence in distribution of $n^{1/2}w_n$ under the null hypothesis to a normal distribution with mean zero follows from Theorem A.1 and the Delta Method. Finally, as discussed in the proof of Theorem 3.2, the sequence $\{X_t\}$ is strong mixing. The result now follows from Corollary 2.3. □

Theorem A.1. *Base the estimation of* ϑ *on estimating model* (1). *Assume Regularity Conditions* 3.2, *which include that the true model is continuous.*

Then $n^{1/2}((\hat{\Phi}'_{1,n}, \hat{\Phi}'_{2,n}, \hat{r}_n)' - (\Phi'_1, \Phi'_2, r)')$ *converges weakly to a normal distribution with mean zero.*

Proof. By the reasoning of Chan (1993, p. 527) and of Chan and Tsay (1998, p. 416), we may assume without loss of generality that d is known for the asymptotic analysis.[7] We proceed by mimicking/extending the proof of Theorem 2.2 of Chan and Tsay (1998), abbreviated by CT henceforth. To this end, write the general model (1) in the equivalent form of model (2) plus an extra intercept term for the second regime.

$$X_t = \phi_0 + \phi_{00}1\{X_{t-d} > r\} + \sum_{j=1, j\neq d}^{p} \phi_j X_{t-j} + \begin{cases} \phi_{d-}(X_{t-d} - r) + \sigma_1\varepsilon_t & \text{if } X_{t-d} \leqslant r, \\ \phi_{d+}(X_{t-d} - r) + \sigma_2\varepsilon_t & \text{if } X_{t-d} > r. \end{cases} \tag{13}$$

To match the notation of the proof of CT, introduce the parameter

$$\theta = (\phi_0, \ldots, \phi_{d-1}, \phi_{d-}, \phi_{d+}, \ldots, \phi_p, r, \phi_{00})'$$

[7]This follows from the discreteness of d and the consistency of $\hat{d}_n$. Hence, $\hat{d}_n$ will be equal to d almost surely for all sufficiently large n.

and denote the true parameter by θ^*. We assume that the true model is continuous, that is, $\phi^*_{00} = 0$. It is obviously sufficient for our purposes to demonstrate the asymptotic normality of $n^{1/2}(\hat{\theta}_n - \theta^*)$. Next, introduce the error term

$$\begin{aligned} e_t(\theta) &= X_t - E(X_t|\mathscr{F}_{t-1};\theta) \\ &= X_t - \phi_0 - \phi_{00}1\{X_{t-d} > r\} \\ &\quad - \sum_{j=1, j\neq d}^{p} \phi_j X_{t-j} - \begin{cases} \phi_{d-}(X_{t-d} - r) & \text{if } X_{t-d} \leqslant r, \\ \phi_{d+}(X_{t-d} - r) & \text{if } X_{t-d} > r \end{cases} \end{aligned}$$

and let $e_t = e_t(\theta^*)$. Finally, $H_t(\theta)$ is the vector of partial derivatives of $e_t(\theta)$ with respect to the elements of θ and $H_t = H_t(\theta^*)$.

Now consider the original proof of CT (given in their Appendix). We shall indicate all quantities that appear in CT by the subscript $_{\mathrm{CT}}$. Since they consider continuous models only and do not have the extra parameter ϕ_{00}, their terms are 'smaller'; for example, $\theta = (\theta'_{\mathrm{CT}}, \phi_{00})'$ and $H_t(\theta) = (H'_{t,\mathrm{CT}}(\theta_{\mathrm{CT}}), -1\{X_{t-d} > r\})'$.

As do CT, we can decompose

$$e_t(\theta) = e_t + H'_t(\theta - \theta^*) + |\theta - \theta^*| R_t(\theta),$$

where our remainder term $R_t(\theta)$ is related to the one in CT in the following fashion:

$$R_t(\theta) = \frac{|\theta_{\mathrm{CT}} - \theta^*_{\mathrm{CT}}| R_{t,\mathrm{CT}}(\theta_{\mathrm{CT}}) + \phi_{00}1\{r^* < X_{t-d} \leqslant r\}}{|\theta - \theta^*|}.$$

Next, the decomposition of $e_t^2(\theta)$ and the definition of $W_t(\theta)$ are exactly as in CT.

To show asymptotic normality now, we need to check conditions (i)–(iii) of CT. The verifications of (ii) and (iii) are analogous to those in CT and hold no matter what the value of ϕ^*_{00}. On the other hand, the verification of (i) requires that $\phi^*_{00} = 0$, that is, that the true model be continuous. To see why, note that in a continuous model ϕ_{00} tends to zero as $|\theta - \theta^*|$ tends to zero and so the verification of (i) in CT goes through. On the other hand, if the model is discontinuous, ϕ_{00} is bounded away from zero as $|\theta - \theta^*|$ tends to zero and the verification of (i) in CT no longer holds; for example, $R_t(\theta)$ no longer is a bounded function over a bounded neighborhood of θ^*. $\square$

Remark A.1. More specifically, it follows from the extension of the proof of CT that the limiting covariance matrix of $n^{1/2}(\hat{\theta}_n - \theta^*)$ is given by $U^{-1}VU^{-1}$ where $U = E(H_t H'_t)$ and $V = E(e_t^2 H_t H'_t)$, which is a $(p+3) \times (p+3)$ matrix. Since the last element of H_t is non-deterministic, the upper $(p+2) \times (p+2)$ block of this matrix is 'larger' than $U^{-1}_{\mathrm{CT}} V_{\mathrm{CT}} U^{-1}_{\mathrm{CT}}$, the limiting $(p+2) \times (p+2)$ covariance matrix of CT. (This is easiest to see when $\sigma_1 = \sigma_2$ and the limiting covariance matrices simplify to $\sigma^2 U^{-1}$ and $\sigma^2 U^{-1}_{\mathrm{CT}}$, respectively, but is also true in general.) The implication is that when the true model is continuous but the general, discontinuous model is estimated, then one pays a price in terms of the efficiency of the estimator. This finding is not surprising and in agreement with the conjecture of CT in their Section 5:

'In practice, it may not be known that the autoregressive function is continuous. Instead of fitting model (2), one may fit the more general model (1) to the data. It is then interesting to investigate the asymptotics of the conditional LS estimators of a general [SE]TAR model when the true autoregressive function is continuous everywhere. Preliminary study suggests that the asymptotics depend on whether or not the estimation scheme assumes the a priori information that the autoregressive function is continuous.'

References

Chan, K.S., 1993. Consistency and limiting distribution of the least square estimator of a threshold autoregressive model. Annals of Statistics 21, 520–533.

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

Doukhan, P., 1994. Mixing: Properties and Examples, Lecture Notes in Statistics, vol. 85. Springer, New York.

Gonzalo, J., Wolf, M., 2001. Subsampling inference in threshold autoregressive models. Working Paper 573. Department of Economics and Business, Universitat Pompeu Fabra. Available at http://www.econ.upf.es/eng/research/research2.php.

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

Politis, D.N., Romano, J.P., 1994. Large sample confidence regions based on subsamples under minimal assumptions. Annals of Statistics 22, 2031–2050.

Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.

Potter, S.M., 1995. A nonlinear approach to US GNP. Journal of Applied Econometrics 2, 109–125.

Tiao, G.C., Tsay, R.S., 1994. Some advances in non-linear and adaptive modeling in time series. Journal of Forecasting 13, 109–131.

Tong, H., 1990. Non-Linear Time Series: A Dynamical System Approach. Oxford University Press, New York.