Statistica Sinica **26** (2016), 1543-1554
doi:http://dx.doi.org/10.5705/ss.2013.394t

NESTED SUB-SAMPLE SEARCH ALGORITHM FOR ESTIMATION OF THRESHOLD MODELS

Dong Li and Howell Tong

Tsinghua University and London School of Economics & Political Science

Abstract: Threshold models have been popular for modelling nonlinear phenomena in diverse areas, in part due to their simple fitting and often clear model interpretation. A commonly used approach to fit a threshold model is the (conditional) least squares method, for which the standard grid search typically requires $O(n)$ operations for a sample of size n; this is substantial for large n, especially in the context of panel time series. This paper proposes a novel method, the *nested sub-sample search algorithm*, which reduces the number of least squares operations drastically to $O(\log n)$ for large sample size. We demonstrate its speed and reliability via Monte Carlo simulation studies with finite samples. Possible extension to maximum likelihood estimation is indicated.

Key words and phrases: Least squares estimation, maximum likelihood estimation, nested sub-sample search algorithm, standard grid search algorithm, threshold model.

1. Introduction

Threshold models have attracted much attention and been widely used to model nonlinear phenomena in such diverse areas as ecology, economics, finance, and others. Their success is partly due to their simple fitting and often clear interpretation. Threshold models are typically characterized by piecewise linearization via partitioning a complex system into regimes by some threshold (or covariate) variable, thereby providing a relatively easy-to-handle approximation of a complex system. When the model within each regime is a linear regression, we have the well-known two-phase regression of Quandt (1958). On the other hand, when the model within each regime is a linear autoregression, we have the well-known threshold autoregressive (TAR) model of Tong (1978), including the self-exciting threshold autoregressive model and its smooth cousin, the smooth threshold (or transition) autoregressive model, as special cases. See also Tong and Lim (1980), Chan and Tong (1986), Tong (1990), and the references therein. Recently Hansen (2011) has provided a fairly comprehensive review of the impacts of TAR models on econometrics and economics by reference to 75 influential papers published in the literature. Chen, So, and Liu (2011) has

provided a similar review of the impacts on finance. More recently, Chan, Yau, and Zhang (2015) has adopted the LASSO method to estimate TAR models with multiple thresholds, with promising results. A concise overview of the history and prospects of threshold models is given by Tong (2011).

As far as theoretical results are concerned, much progress has been in two-phase regression since Quandt (1958), and in TAR models since Tong (1978). For the former, see, e.g., Bacon and Watts (1971), Goldfeld and Quandt (1972), Maddala (1977), Quandt (1983), and others. For the latter, see, e.g., Chan (1993), who first showed that the least squares estimator (LSE) of the threshold parameter is super-consistent and obtained its limiting distribution theoretically; Hansen (1997, 2000), who presented an alternative approximation to the limiting distribution of the estimated threshold when the threshold effect diminishes as the sample size increases; Gonzalo and Pitarakis (2002), who developed a sequential estimation approach that makes the estimation of multiple threshold models computationally feasible and formally discussed the large sample properties; Li and Ling (2012), who established the asymptotic theory of LSE in multiple threshold models and proposed a resampling method for implementing the limiting distribution of the estimated threshold directly when the threshold effect is fixed. Other significant results related to threshold models include Tsay (1989, 1998), Hansen (1996), Caner and Hansen (2001), Gonzalo and Wolf (2005), Seo and Linton (2007), and Yu (2012), among others.

Despite the theoretical progress in threshold models, computational issues are somewhat lacking behind, which hinders wider practical applications. A key issue is computational cost.

A commonly used approach to fit a threshold model is the (conditional) least squares method. When the threshold is known, the threshold model is piecewise linear in the remaining parameters and thus linear estimation techniques can be applied. However, when the threshold is unknown, the ordinary least squares method for linear regression cannot be applied immediately since the threshold parameter lies in an indicator function. This issue has been commonly tackled by using the single grid search (SGS) algorithm over a feasible threshold space; see Tong and Lim (1980), Chan (1993), Hansen (1997, 2000), Gonzalo and Pitarakis (2002), Li and Ling (2012), Yu (2012), and others. The SGS algorithm requires least squares operations of order $O(n)$ for single threshold models, where n is the sample size. If n is small, the SGS algorithm can be effectively used to search for the estimate of the threshold over a set of threshold candidates by enumeration. However, when n is large, this algorithm can be rather time-consuming. The situation is worse when we wish to fit threshold models to a panel of observations. Gonzalo and Wolf (2005) considered subsampling inference of threshold models and massive computations are needed in the choice of the block

size. Similarly, massive computations are also needed in bootstrap estimation of single threshold models in Seijo and Sen (2011). In practice, a conventional numerical approach for threshold modelling incurs inevitably high cost with large samples. For example, about np^3 least squares operations are needed when fitting a threshold model with p covariates to data with sample size n. To illustrate, if n is 1,000 and p is 10, then we need about one million least squares operations. Thus, it is crucially important to find ways to reduce the computational cost when fitting a threshold model for large n.

In the time series literature, Tong (1983, Appendix A10) proposed and later Tsay (1989) re-discovered the SGS approach based on the rearranged technique, that essentially turns threshold estimation into a change-point problem of the associated order statistics obtained from the observations. See also Ertel and Fowlkes (1976). This method is now available by calling the function `tar` in the package `TSA` in R; see Chan and Ripley (2012). For threshold regression models, the SGS algorithm is available by calling the program in R developed by Hansen (2000) on the website: `http://www.ssc.wisc.edu/~bhansen/progs/ecnmt_00.html`. Wu and Chang (2002) proposed a genetic algorithm for TAR models. However, this algorithm has many limitations, as recognised by the above authors, so it is not widely used. Coakley, Fuertes, and Pérez (2003) presented an algorithm based on the QR decomposition of matrices for a particular class of TAR models (called the band-type TAR model). For general threshold models, it is fair to say that the SGS algorithm remains to-date the most commonly adopted technique in practice due to its simplicity and reliability, although it is time-consuming for large n.

In this paper, we propose the *nested sub-sample search algorithm*, or the NeSS algorithm for brevity, to produce a much faster search that is reliable in the context of threshold estimation. Compared with existing algorithms, the NeSS algorithm reduces the computational cost drastically, from $O(n)$ to $O(\log n)$ least squares operations for large sample size. The idea is simple. We shrink the nested feasible set step by step and finally maximize $J_n(r)$ in (2.4) over a small feasible set by enumeration so that it is expected to save computational costs. The performance of our method is evaluated via Monte Carlo simulation studies in finite samples.

The remainder of the paper is organized as follows. Section 2 addresses the model and estimation issues. Section 3 presents our new algorithm. Section 4 evaluates the performance of our algorithm via Monte Carlo simulation studies and Section 5 concludes the paper.

2. Model and Least Squares Estimation

Consider the threshold stochastic regression model

$$y_t = \boldsymbol{\beta}_1' \mathbf{x}_t I(z_t \le r) + \boldsymbol{\beta}_2' \mathbf{x}_t I(z_t > r) + \varepsilon_t, \tag{2.1}$$

where $\mathbf{x}_t = (1, x_{t1}, \ldots, x_{tp})'$, $I(\cdot)$ is the indicator function, z_t is the threshold variable that controls regime switching according to the value of the threshold r, and $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are the coefficients. The innovation ε_t is a real-value martingale difference with respect to an increasing sequence of σ-fields $\mathcal{F}_t$ generated by $\{(\mathbf{x}_{j+1}, z_{j+1}, \varepsilon_j) : j \le t\}$. Let $\boldsymbol{\theta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', r)'$ denote the parameter, with true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_{10}', \boldsymbol{\beta}_{20}', r_0)'$. Throughout, r is assumed to lie in the bounded subset $[\underline{r},\ \bar{r}]$ and $\boldsymbol{\beta}_{10} \neq \boldsymbol{\beta}_{20}$.

We introduce some notation. Let $\mathbf{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\mathbf{z} = (z_1, \ldots, z_n)'$ and $\mathbf{I}(a < \mathbf{z} \le b) = (a_{ij})_{n \times (p+1)}$ with $a_{ij} = I(a < z_i \le b)$. Write $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)'$, $\mathbf{X}_1(r) \equiv \mathbf{X} * \mathbf{I}(\mathbf{z} \le r)$ (i.e., $a = -\infty$) and $\mathbf{X}_2(r) \equiv \mathbf{X} * \mathbf{I}(\mathbf{z} > r)$ (i.e., $b = \infty$), where '$*$' denotes the Hadamard product operator of matrices. Then, (2.1) can be reformulated in matrix form as

$$\mathbf{y} = \mathbf{X}_1(r)\boldsymbol{\beta}_1 + \mathbf{X}_2(r)\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}. \tag{2.2}$$

Given the sample $(\mathbf{y}, \mathbf{X}, \mathbf{z})$, our aim is to estimate $\boldsymbol{\theta}$. For each fixed r, (2.2) is linear in $\boldsymbol{\beta}_i$'s and the application of the ordinary least squares principle yields the sum of squared errors function

$$S_n(r) = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}_1(r)(\mathbf{X}_1(r)'\mathbf{X}_1(r))^{-1}\mathbf{X}_1(r)'\mathbf{y} - \mathbf{y}'\mathbf{X}_2(r)(\mathbf{X}_2(r)'\mathbf{X}_2(r))^{-1}\mathbf{X}_2(r)'\mathbf{y},$$

from which r can be estimated as

$$\widehat{r} = \arg \min_{r \in [\underline{r},\ \bar{r}]} S_n(r). \tag{2.3}$$

For convenience, we consider an alternative objective function

$$J_n(r) = S_n - S_n(r), \tag{2.4}$$

where $S_n = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Note that $\mathbf{X} = \mathbf{X}_1(r) + \mathbf{X}_2(r)$ and $\mathbf{X}_i(r)'\mathbf{X}_j(r) \equiv \mathbf{0}$ for $i \neq j \in \{1, 2\}$. After simple calculations, it follows that

$$J_n(r) = (\widehat{\boldsymbol{\beta}}_2(r) - \widehat{\boldsymbol{\beta}}_1(r))'\mathbf{X}_2(r)'\mathbf{X}_2(r)(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}_1(r)'\mathbf{X}_1(r)(\widehat{\boldsymbol{\beta}}_2(r) - \widehat{\boldsymbol{\beta}}_1(r)),$$

where $\widehat{\boldsymbol{\beta}}_j(r) = (\mathbf{X}_j(r)'\mathbf{X}_j(r))^{-1}\mathbf{X}_j(r)'\mathbf{y}$ for $j = 1, 2$. Now, (2.3) is equivalent to

$$\widehat{r} = \arg \max_{r \in [\underline{r},\ \bar{r}]} J_n(r). \tag{2.5}$$

To quickly obtain $\widehat{r}$, we explore the shape of the limit of $J_n(r)$, which provides some useful information in seeking $\widehat{r}$. If Assumptions A.1−A.2 in the Appendix hold, by Lemma 2.1 with $m = 1$ in Gonzalo and Pitarakis (2002), it follows that

$$\sup_{r\in[\underline{r},\ \bar{r}]} \left| \frac{J_n(r)}{n} - J(r) \right| \xrightarrow{p} 0, \tag{2.6}$$

where $J(r)$ is a non-stochastic continuous function over $[\underline{r},\ \bar{r}]$ defined by

$$\begin{aligned} J(r) =&(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20})'\{\mathbf{G}_{r\wedge r_0}\mathbf{G}_r^{-1} + (\mathbf{G}_{r\wedge r_0} - \mathbf{G}_{r_0})(\mathbf{G} - \mathbf{G}_r)^{-1}\}(\mathbf{G} - \mathbf{G}_r)\mathbf{G}^{-1}\mathbf{G}_r \\ &\times \{\mathbf{G}_r^{-1}\mathbf{G}_{r\wedge r_0} + (\mathbf{G} - \mathbf{G}_r)^{-1}(\mathbf{G}_{r\wedge r_0} - \mathbf{G}_{r_0})\}(\boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20}), \end{aligned}$$

with $\mathbf{G}$ and $\mathbf{G}_x$ defined in the Appendix. Furthermore, $J(r)$ is *unimodal*, strictly monotonically increasing in $[\underline{r},\ r_0]$ and strictly monotonically deceasing in $[r_0,\ \bar{r}]$. Thus, we conclude that $J_n(r)/n$ is unimodal over $[\underline{r},\ \bar{r}]$ with probability tending to one as $n \to \infty$, which implies that $J_n(r)/n$ may be unimodal over the set $\{k\triangle : k \in \mathbb{Z}\} \cap [\underline{r},\ \bar{r}]$ for some suitable $\triangle > 0$. This provides useful information for seeking the maximizer of $J_n(r)$.

It is well known in numerical analysis that the Fibonacci algorithm is optimal for optimizing deterministic unimodal functions. It is natural to expect that there exists similar algorithms for optimizing $J_n(r)$, at least for large sample size.

3. Nested Sub-sample Search Algorithm

Suppose the sample $(\mathbf{y},\ \mathbf{X},\ \mathbf{z})$ is available. We have $J_n(r) = J_n(z_{(i)})$ for $r \in [z_{(i)},\ z_{(i+1)})$, where $z_{(1)} \le \cdots \le z_{(n)}$ is the order statistics of $\{z_1, \ldots, z_n\}$. We adopt the approach of Tong and Lim (1980) by considering the empirical percentiles as candidates for the threshold values. The SGS algorithm maximizes $J_n(r)$ at (2.4) over the *feasible set* $\{z_{(1)}, \ldots, z_{(n)}\}$ by enumeration. To get the global maximizer of $J_n(r)$, the required number of least squares operations is n.

Now, we propose a new algorithm and call it the *nested sub-sample search* (NeSS) algorithm since the feasible set shrinks by a half after each iteration. The idea is simple. We shrink the nested feasible set step-by-step and finally maximize $J_n(r)$ over a small feasible set by enumeration so that it is expected to save computational costs. Specifically, suppose the initial feasible set is $\mathcal{D}_0 \equiv \{z_{(1)}, \ldots, z_{(n)}\}$. We first maximize $J_n(r)$ over $\{z_{(k_0)}, z_{(2k_0)}, \ldots, z_{(qk_0)}\}$, where $k_0 = [\#\mathcal{D}_0/(q+1)]$ and $[a]$ is the largest integral part of a. Then we get the maximizer $z_{(j_0k_0)}$ for some $j_0 \in \{1, \ldots, q\}$ and a new feasible set $\mathcal{D}_1 \equiv (z_{((j_0-1)k_0)},\ z_{((j_0+1)k_0)}) \cap \mathcal{D}_0$. Repeat the procedure above by updating the feasible set. After m steps, we get a feasible set $\mathcal{D}_m = (z_{((j_{m-1}-1)k_{m-1})},\ z_{((j_{m-1}+1)k_{m-1})}) \cap \mathcal{D}_0$ that contains $[2^m n/(q+1)^m]$ candidates out of $\{z_{(1)}, \ldots, z_{(n)}\}$, where $k_{m-1} = [\#\mathcal{D}_{m-1}/(q+1)]$ and all $j_i \in \{1, \ldots, q\}$. Figure 1 illustrates the procedure.

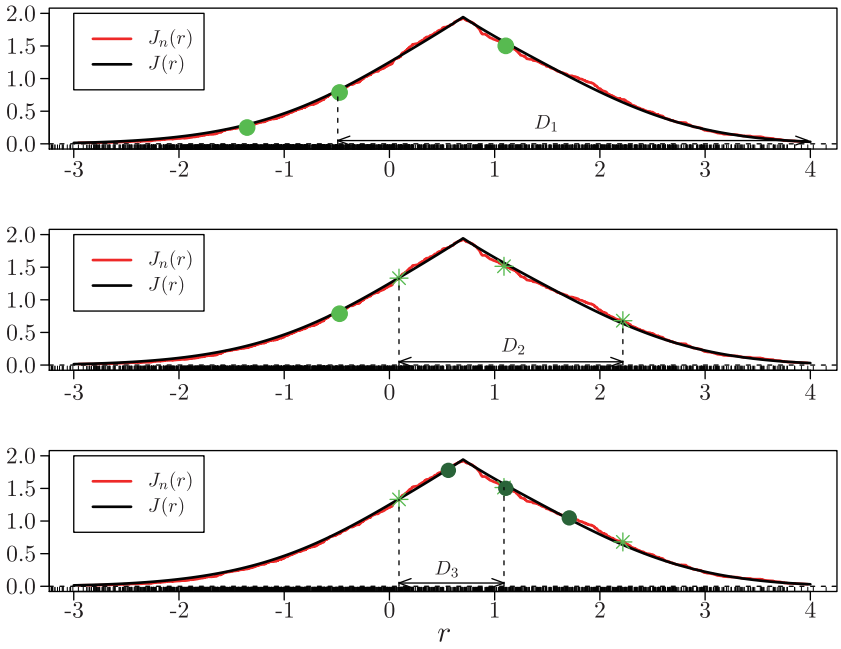

Figure 1. An illustration of the NeSS algorithm.

Unlike deterministic unimodal functions, due to its randomness, $J_n(r)$ may not be unimodal over the set $\{z_{(k_m)}, z_{(2k_m)}, \ldots, z_{(qk_m)}\} \subset \mathcal{D}_m$ if $k_m = \#\mathcal{D}_m$ is small, although $J(r)$ is unimodal. Thus, we must use the enumeration method to maximize $J_n(r)$ over $\mathcal{D}_m$ for some m. If, in the last step of maximizing $J_n(r)$, the required number of least squares operations is not larger than δ, which is a preassigned positive integer, $[2^m n/(q+1)^m] \leq \delta$, then the number m of iterations satisfies

$$m \geq \frac{\log(n/(\delta+1))}{\log((q+1)/2)}.$$

Thus, the total required number of least squares operations is about

$$mq + \left[\frac{2^m n}{(q+1)^m}\right] = \frac{q}{\log((q+1)/2)} \log\left(\frac{n}{\delta+1}\right) + \delta. \tag{3.1}$$

Since the minimizer of $q/\log((q+1)/2)$ in (3.1) over the set of positive integers is 3, we take $q = 3$. As for the choice of δ, if it is set small, then we may not get the global maximizer of $J_n(r)$ since the randomness of $J_n(r)$ can obscure the unimodality such that $\mathcal{D}_{m+1}$ does not cover the global maximizer. Empirically, we can set $\delta = 50$ when the sample size $n \geq 200$. If the sample size is less than 200, we can directly use the SGS algorithm to get the estimate $\widehat{r}$ since the computational cost is not high in this case. Figure 2 gives the total required

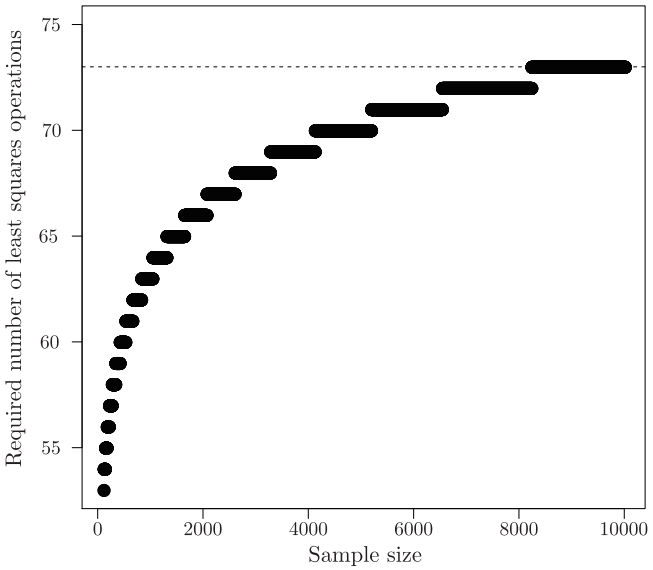

Figure 2. The total required number of least squares operations in (3.1).

number of least squares operations in (3.1) after taking the ceiling function when $\delta = 50$ and the sample size varies from 100 to 10,000.

Summarizing the above discussion, we have the following algorithm.

NeSS algorithm

Given the initial feasible set $\mathcal{D} = \{z_{(1)}, \ldots, z_{(n)}\}$
  while $(\#\mathcal{D} > \delta)\{$
    calculate q=quantile($\mathcal{D}$, c(0.25, 0.5, 0.75)) and $J_n(q[i])$ for $i = 1, 2, 3$;
      if $(J_n(q[1]) \geq \max\{J_n(q[2]), J_n(q[3])\})$
        $\mathcal{D} \leftarrow \mathcal{D}[\mathcal{D} \leq q[2]]$
      else if $(J_n(q[2]) \geq \max\{J_n(q[1]), J_n(q[3])\})$
        $\mathcal{D} \leftarrow \mathcal{D}[q[1] \leq \mathcal{D} \leq q[3]]$
      else
        $\mathcal{D} \leftarrow \mathcal{D}[\mathcal{D} \geq q[2]]$
      }
Maximizing $J_n(r)$ over $\mathcal{D}$ and then getting $\widehat{r}$.

Generally, $\#\mathcal{D}$ obtained in the last iteration is less than δ so that it is possible that there are not enough data for us to get the genuine global estimate $\widehat{r}$. For example, suppose we set $\delta = 50$ and $\mathcal{D} = \{z_{(51)}, \ldots, z_{(120)}\}$ in the penultimate step. Clearly, $\#\mathcal{D} = 70 > \delta$. Then we further shrink the feasible set according to the above algorithm and get the final feasible set $\mathcal{D}_o = \{z_{(68)}, \ldots, z_{(103)}\}$ say, with $\#\mathcal{D}_o = 36$. For this case, we had better extend $\mathcal{D}_o$ both forward and backward equally so that $\#\mathcal{D}_o = \delta$. For example, $\mathcal{D}_o$ can be extended to $\mathcal{D}_* = \{z_{(61)}, \ldots, z_{(110)}\}$.

Table 1. Total elapsed time (in seconds) for 100 replications.

Model	n	200	400	800	1,600	3,200
Model (4.1)	`tar`	6.46	12.31	24.23	48.88	100.93
	NeSS	1.34	2.05	3.48	6.52	15.13
Model (4.2)	SGS	2.62	6.26	20.62	61.90	211.62
	NeSS	0.83	1.11	2.28	3.00	5.52

Finally, we mention that the Fibonacci algorithm is not optimal for optimizing $J_n(r)$ due to the randomness of the latter. For the Fibonacci algorithm, $q = 2$ is enough, two golden points for each iteration, while $q = 3$ is required in our NeSS algorithm.

4. Simulation Studies

To assess the performance of our algorithm in finite samples, we conducted simulation studies, using sample size $n = 200$, 400, 800, 1,600 and 3,200 for the model

$$y_t = \begin{cases} 1 - 0.3y_{t-1} + 0.5y_{t-2} + \varepsilon_t, \text{if } \ y_{t-2} \le 1, \\ -1 + 0.6y_{t-1} - 0.3y_{t-3} + \varepsilon_t, \text{if } \ y_{t-2} > 1, \end{cases} \tag{4.1}$$

and the model

$$y_t = \begin{cases} 0.5x_{t1} + 1.2x_{t2} + \varepsilon_t, \text{if } \ x_{t1} \le 1, \\ -0.5x_{t1} + 0.7x_{t2} + \varepsilon_t, \text{if } \ x_{t1} > 1, \end{cases} \tag{4.2}$$

where $(x_{t1}, x_{t2})' \sim_{i.i.d.} N(\mathbf{0}, \Sigma)$ with

$$\Sigma = \begin{pmatrix} 4 & 7 \\ 7 & 25 \end{pmatrix}$$

and independent of $\{\varepsilon_t\}$. In all simulations, the innovation $\varepsilon_t \sim_{i.i.d.} N(0, 1)$. The program[1] is written in R. For threshold regression models, we used the program in R by Hansen (2000) mentioned before. For TAR models, the SGS algorithm is available by calling the function `tar` in the package `TSA` in R; see Chan and Ripley (2012).

Table 1 reports the total elapsed times in optimizing $J_n(r)$ for model (4.1) by `tar` and our algorithm with 100 replications, as well as those for model (4.2) by the SGS. Here, we search for the estimate of r_0 within the 90% inner sample range and set $\delta = 50$. From Table 1, we can see that the NeSS algorithm saves substantial time when the sample size is large.

To examine whether the NeSS algorithm and the SGS algorithm can produce an identical global maximizer of $J_n(r)$ or not, we define the *matching rate* as the

[1]The program is run on a personal computer with a 3.30GHz Intel® Core(TM)i3-3220 CPU, 4GB RAM and 64-bit Operating system.

Table 2. Matching rate in 1,000 replications.

Model	200	400	800	1,600	3,200
Model (4.1)	0.999	1.000	1.000	1.000	1.000
Model (4.2)	1.000	1.000	1.000	1.000	1.000

ratio of the total numbers of times that the two algorithms produce the same maximizer to one thousand in 1,000 replications. The matching rates are reported in Table 2. From Table 2, we can see that the NeSS algorithm and the SGS have identical matching rates when the sample size $n \geq 200$. We also did simulations for $n = 100$. For model (4.2), the matching rates are both still 1. However, for model (4.1), the matching rate is 0.992 for NeSS. Of course, when $n < 200$, we can use either the SGS algorithm or the `tar` directly since the computational cost is not high. For this case, we do not recommend NeSS because it does not have any advantage in reliability.

Finally, when the threshold is not identified or the threshold effect is very small, the matching rate drops considerably. For example, it is only 0.451 for $n = 200$ for the AR(2) model

$$y_t = 1 + 0.3y_{t-1} - 0.5y_{t-2} + \varepsilon_t, \qquad \{\varepsilon_t\} \overset{i.i.d.}{\sim} N(0, 1).$$

This is not surprising because the limit of $J_n(r)/n$ is zero in probability, which cannot provide any useful information for our algorithm. In addition, when the proportion of observations in one regime to the whole is less than 5%, the estimator of threshold obtained by either the SGS algorithm or the NeSS algorithm may not be reliable since the usual choice of 90% inner sample range may have ruled out the genuine maximizer of $J_n(r)$.

5. Concluding Remarks

This paper has developed a new algorithm that can search for an estimate of the threshold parameter within the framework of threshold stochastic regression models, at a substantially faster rate than all existing algorithms that we are aware of and with demonstrable reliability.

In the literature, the maximum likelihood estimation (MLE) and the least absolute deviations estimation (LADE) are also considered for threshold models. Usually they are obtained by the SGS algorithm; see, e.g., Caner (2002), Samia and Chan (2011), and Yu (2012). As a referee has pointed out, our NeSS algorithm can be extended to the nonlinear optimization problem associated with maximum likelihood estimation; our numerical experimentations have lent support to his/her observation. Moreover, the NeSS algorithm can be applied to T-CHARM of Chan et al. (2014) and the multivariate threshold models studied by Tsay (1998).

For multi-threshold stochastic regression models (e.g., Ertel and Fowlkes (1976), Liu, Wu, and Zidek (1997), Gonzalo and Pitarakis (2002), Li and Ling (2012)), we can use the NeSS algorithm to obtain a sequential estimate of multiple thresholds, one at a time, by the NeSS algorithm. However, it is known that the limiting distribution of such a sequential estimate is different from that of a joint estimate. If our particular interest is in getting a joint estimate of all thresholds, then how to reduce the computational burden remains a challenge.

Acknowledgements

The authors would like to thank the Editor and an anonymous referee for their very helpful comments. Li's research was partially supported by the Start-up Fund of Tsinghua University (No.553310013) and the National Natural Science Foundation of China (NSFC) under Grant (No.11401337). Tong's research was partially supported by the Distinguished Visiting Professorship at the University of Hong Kong and a Senior Fellowship at the Institute of Advanced Studies, the University of Bologna, Italy.

Appendix

In this section, we give two assumptions for (2.6) holding.

Assumption A.1.

(i) *The minimum eigenvalues of* $(1/n)\underline{\mathbf{X}}_\eta'\underline{\mathbf{X}}_\eta$ *and* $(1/n)\bar{\mathbf{X}}_\eta'\bar{\mathbf{X}}_\eta$ *are bounded away from zero in probability as* $n \to \infty$ *for any* $\eta > 0$, *where* $\underline{\mathbf{X}}_\eta = \mathbf{X} * \mathbf{I}(r_0 - \eta < \mathbf{z} \le r_0)$ *and* $\bar{\mathbf{X}}_\eta = \mathbf{X} * \mathbf{I}(r_0 < \mathbf{z} \le r_0 + \eta)$.

(ii) *The threshold variable* z_t *has a positive density on* $[\underline{r},\ \bar{r}]$.

Assumption A.2. As $n \to \infty$,

(i) $\sup_{r\in\mathbb{R}} \left|\frac{1}{n}\mathbf{X}_1(r)'\mathbf{X}_1(r_0) - \mathbf{G}_{r\wedge r_0}\right| \xrightarrow{p} 0$,

(ii) $\sup_{r\in\mathbb{R}} \left|\frac{1}{n}\mathbf{X}_2(r)'\mathbf{X}_2(r_0) - \{(\mathbf{G} - \mathbf{G}_r) - (\mathbf{G}_{r_0} - \mathbf{G}_{r\wedge r_0})\}\right| \xrightarrow{p} 0$,

(iii) $\sup_{r\in\mathbb{R}} \left(\left|\frac{1}{n}\mathbf{X}_1(r)'\boldsymbol{\varepsilon}\right| + \left|\frac{1}{n}\mathbf{X}_2(r)'\boldsymbol{\varepsilon}\right|\right) \xrightarrow{p} 0$,

where $\mathbf{G}_x$ is a symmetric and positive-definite matrix, which is absolutely continuous and strictly increasing in x, with $\mathbf{G}_{-\infty} \equiv \mathbf{0}$ and $\mathbf{G}_\infty \equiv \mathbf{G}$, and $r \wedge r_0 = \min\{r, r_0\}$.

Assumption A.1 requires that there are enough observations in the neighbourhood of the threshold r_0 so that it is identifiable. Assumption A.2 is a type of condition related to the uniform law of large numbers, which holds if

$\{(\mathbf{x}_t, z_t, \varepsilon_t)\}$ is strictly stationary and ergodic with finite second moment and z_t has a continuous distribution. See, e.g., Lemma 1 in Hansen (1996). In particular, if (2.1) is specialized to a self-exciting TAR model, with $\{y_t\}$ strictly stationary and ergodic with a continuous and positive density on $\mathbb{R}$ and $Ey_t^2 < \infty$, then Assumptions A.1 and A.2 hold with $\mathbf{G}_x = E\{\mathbf{y}_{t-1}\mathbf{y}_{t-1}' I(y_{t-d} \le x)\}$ and $\mathbf{y}_{t-1} = (1, y_{t-1}, \ldots, y_{t-p})'$; see Chan (1990, 1993).

References

Bacon, B. W. and Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika* **58**, 525-534.

Caner, M. (2002). A note on least absolute deviation estimation of a threshold model. *Econom. Theory* **18**, 800-814.

Caner, M. and Hansen, B. E. (2001). Threshold autoregression with a unit root. *Econometrica* **69**, 1555-1596.

Chan, K. S. (1990). Testing for threshold autoregression. *Ann. Statist.* **18**, 1886-1894.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.* **21**, 520-533.

Chan, K. S., Li, D., Ling, S. and Tong, H. (2014). On conditionally heteroscedastic AR models with thresholds. *Statist. Sinica* **24**, 625-652.

Chan, K. S. and Ripley, B. (2012). TSA: Time Series Analysis. R package version 1.01, URL `http://cran.r-project.org/web/packages/TSA/`.

Chan, K. S. and Tong, H. (1986). On estimating thresholds in autoregressive models. *J. Time Series Anal.* **7**, 179-190.

Chan, N. H., Yau, C. Y. and Zhang, R. M. (2015). LASSO estimation of threshold autoregressive models. *J. Econometrics* **189**, 285-296.

Chen, C. W. S., So, M. K. P. and Liu, F. C. (2011). A review of threshold time series models in finance. *Stat. and Its Interface* **4**, 167-182.

Coakley, J., Fuertes, A.-M. and Pérez, M.-T. (2003). Numerical issues in threshold autoregressive modeling of time series. *J. Econom. Dynam. Control* **27**, 2219-2242.

Ertel, J. E. and Fowlkes, E. B. (1976). Some algorithms for linear spline and piecewise multiple linear regression. *J. Amer. Statist. Assoc.* **71**, 640-648.

Goldfeld, S. M. and Quandt, R. E. (1972). *Nonlinear Methods in Econometrics*. North-Holland, Amsterdam.

Gonzalo, J. and Pitarakis, J. Y. (2002). Estimation and model selection based inference in single and multiple threshold models. *J. Econom.* **110**, 319-352.

Gonzalo, J. and Wolf, M. (2005). Subsampling inference in threshold autoregressive models. *J. Econom.* **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. *Stud. Nonlinear Dyn. Econom.* **2**, 1-14.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica* **68**, 575-603.

Hansen, B. E. (2011). Threshold autoregression in economics. *Stat. and Its Interface* **4**, 123-127.

Li, D. and Ling, S. (2012). On the least squares estimation of multiple-regime threshold autoregressive models. *J. Econom.* **167**, 240-253.

Liu, J., Wu, S. and Zidek, J. V. (1997). On segmented multivariate regression. *Statist. Sinica* **7**, 497-525.

Maddala, G. S. (1977). *Econometrics*. McGraw-Hill, New York.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Stat. Assoc.* **53**, 873-880.

Quandt, R. E. (1983). Computational problems and methods. *Handbook of Econometrics* **1** (Edited by Z. Griliches and M. D. Intriligator), 699-764. North-Holland, Amsterdam.

Samia, N. I. and Chan, K. S.(2011). Maximum likelihood estimation of a generalized threshold stochastic regression model. *Biometrika* **98**, 433-448.

Seijo, E. and Sen, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39**, 1580-1607.

Seo, M. H. and Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *J. Econom.* **141**, 704-735.

Tong, H. (1978). On a threshold model. In: *Pattern Recognition and Signal Processing* (Edied by C. H. Chen), 575-586. Sijthoff and Noordhoff, Amsterdam.

Tong, H. (1983). *Threshold Models in Nonlinear Time Series Analysis.* Lecture Notes in Statistics, **21**. Springer-Verlag, New York.

Tong, H. (1990). *Nonlinear Time Series. A Dynamical System Approach*. Oxford University Press, New York.

Tong, H. (2011). Threshold models in time series analysis — 30 years on. *Stat. Interface* **4**, 107-118.

Tong, H. and Lim, K. S. (1980). Threshold autoregression, limit cycles and cyclical data (with discussion). *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.* **84**, 231-240.

Tsay, R. S. (1998). Testing and modeling multivariate threshold models. *J. Amer. Statist. Assoc.* **93**, 1188-1202.

Wu, B. and Chang, C.-L. (2002). Using genetic algorithms to parameters (d, r) estimation for threshold autoregressive models. *Comput. Statist. Data Anal.* **38**, 315-330.

Yu, P. (2012). Likelihood estimation and inference in threshold regression. *J. Econom.* **167**, 274-294.

Center for Statistical Science, Tsinghua University, Beijing 100084, China.

E-mail: malidong@tsinghua.edu.cn

Department of Statistics, London School of Economics and Political Science, London, WC2A 2AE, UK.

E-mail: howell.tong@gmail.com

(Received December 2013; accepted January 2015)