RESEARCH REPORT

Serial No. 510

December 2013

NESTED SUB-SAMPLE SEARCH ALGORITHM FOR ESTIMATION OF THRESHOLD MODELS

by

Dong Li and Howell Tong

THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

Pokfulam Road, Hong Kong. E-Mail: saas@hku.hk
Tel.: (852) 2859 2466/7 Fax: (852) 2858 9041

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 method typically requires $O(n)$ operations for a sample of size n, which 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 $O(\log n)$. We justify the new algorithm theoretically as well as demonstrate its speed and reliablity via Monte Carlo simulation studies with finite samples.

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

1 Introduction

Threshold models have attracted much attention and been widely used to model nonlinear phenomena in diverse areas, such 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 the response space 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). See also Tong and Lim (1980), Tong (1990) and the references therein. Recently Hansen (2011) has provided a fairly comprehensive review of TAR models by reference

to 75 influential papers published in the econometrics and economics literature. 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 TAR models since Tong (1978). For the former, see, e.g., Bacon and Watts (1971), Goldfeld and Quandt (1972), Maddala (1977), Quandt (1984) 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 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), Yu (2012) among others.

Despite the theoretical progress in threshold models, computational issues are somewhat lacking behind, which hinder 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, conventional numerical approach for threshold modelling incurs inevitably high cost. For example, about np^3 least squares operations are needed when fitting a threshold model with p covariates to data with sample size n. For example, if n is 1000 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.

In the time series literature, Tong (1983, Appendix A10) proposed and later Tsay (1989) re-discovered the SGS approach based on the rearranged technique, which essentially turns the 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 the SGS algorithm of threshold regression models, a program in R by Hansen (2000) is available 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 in practice. 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, the SGS algorithm remains to-date the most commonly adopted technique in practice due to its simplicity and reliability, although it is time-consuming.

In this paper, we propose a novel algorithm, namely 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 $O(n)$ down to $O(\log n)$ least-squares operations, n being the 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. Its validity is also supported theoretically. 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 following threshold stochastic regression model:

$$y_t = \beta_1' \mathrm{x}_t I(z_t \leq r) + \beta_2' \mathrm{x}_t I(z_t > r) + \varepsilon_t, \tag{2.1}$$

where $\mathrm{x}_t = (1, x_{t1}, ..., x_{tp})'$, $I(\cdot)$ is the indicator function, z_t is the threshold variable that controls regime switching according to the value of the threshold r, and β_1 and β_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 $\{(\mathrm{x}_{j+1}, z_{j+1}, \varepsilon_j) : j \leq t\}$. Let $\boldsymbol{\theta} = (\beta_1', \beta_2', r)'$ denote the parameter, whose true value $\boldsymbol{\theta}_0 = (\beta_{10}', \beta_{20}', r_0)'$. Throughout the paper, r is assumed to lie in the bounded subset $[\underline{r}, \bar{r}]$ and $\beta_{10} \neq \beta_{20}$.

For simplicity, we first introduce some notations. Let $\mathbf{y} = (y_1, ..., y_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_n)'$, $\mathbf{z} = (z_1, ..., z_n)'$ and $\mathbf{I}(a < \mathbf{z} \leq b) = (a_{ij})_{n \times (p+1)}$ with $a_{ij} = I(a < z_i \leq b)$. Denote $\mathbf{X} = (\mathrm{x}_1, ..., \mathrm{x}_n)'$, $\mathbf{X}_1(r) \equiv \mathbf{X} * \mathbf{I}(\mathbf{z} \leq r)$ (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, model (2.1) can be reformulated in matrix form

$$\mathbf{y} = \mathbf{X}_1(r)\beta_1 + \mathbf{X}_2(r)\beta_2 + \boldsymbol{\varepsilon}. \tag{2.2}$$

Given the sample $\mathbf{y}, \mathbf{X}$ and $\mathbf{z}$, our aim is to estimate $\boldsymbol{\theta}$. For each fixed r, model (2.2) is linear in β_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

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

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

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

Before discussing asymptotic properties of $J_n(r)$, we first introduce two assumptions.

Assumption 2.1 (i) *The minimum eigenvalues of* $\frac{1}{n}\underline{\mathbf{X}}_\eta'\underline{\mathbf{X}}_\eta$ *and* $\frac{1}{n}\bar{\mathbf{X}}_\eta'\bar{\mathbf{X}}_\eta$ *are bounded away from zero in probability as* $n \to \infty$ *for any* $\eta > 0$, *where* $\underline{\mathbf{X}}_\eta = \mathbf{X} * \mathbf{I}(r_0 - \eta < \mathbf{z} \leq r_0)$ *and* $\bar{\mathbf{X}}_\eta = \mathbf{X} * \mathbf{I}(r_0 < \mathbf{z} \leq r_0 + \eta)$.
(ii) *The threshold variable* z_t *has a positive density on* $[\underline{r},\, \bar{r}]$.

Assumption 2.2 *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) $\sup_{r \in \mathbb{R}} \left(\left|\frac{1}{n}\mathbf{X}_1(r)'\varepsilon\right| + \left|\frac{1}{n}\mathbf{X}_2(r)'\varepsilon\right|\right) \xrightarrow{p} 0$,

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 2.1 requires that there are enough observations in the neighbourhood of the threshold r_0 so that it is identifiable. Assumption 2.2 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, when model (2.1) is a self-exciting TAR model, if $\{y_t\}$ is strictly stationary and ergodic with $Ey_t^2 < \infty$ and the density of y_t is continuous and positive on $\mathbb{R}$, then Assumptions 2.1 and 2.2 hold with $\mathbf{G}_x = E\{\mathbf{y}_{t-1}\mathbf{y}_{t-1}' I(y_{t-d} \leq x)\}$ and $\mathbf{y}_{t-1} = (1, y_{t-1}, ..., y_{t-p})'$; see Chan (1990, 1993).

By Lemma 2.1 with $m = 1$ in Gonzalo and Pitarakis (2002), we have

Theorem 2.1 *If Assumptions 2.1-2.2 hold, then*

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

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

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

with $\boldsymbol{\rho} = \boldsymbol{\beta}_{10} - \boldsymbol{\beta}_{20}$. □

Furthermore, the above $J(r)$ is unimodal.

Theorem 2.2 *$J(r)$ is strictly monotonically increasing in $[\underline{r},\ r_0]$ and strictly monotonically deceasing in $[r_0,\ \bar{r}]$.*

Proof. Using the expression of $J(r)$ and the symmetry and positive definiteness of $\mathbf{G}_x$, we have, after some calculations,

$$J(r) - J(s) = \boldsymbol{\rho}'(\mathbf{G} - \mathbf{G}_{r_0})\{(\mathbf{G} - \mathbf{G}_r)^{-1} - (\mathbf{G} - \mathbf{G}_s)^{-1}\}(\mathbf{G} - \mathbf{G}_{r_0})\boldsymbol{\rho} > 0$$

for $\underline{r} \leq s < r \leq r_0$, and

$$J(r) - J(s) = \boldsymbol{\rho}'\mathbf{G}_{r_0}\{\mathbf{G}_r^{-1} - \mathbf{G}_s^{-1}\}\mathbf{G}_{r_0}\boldsymbol{\rho} < 0$$

for $r_0 \leq s < r \leq \bar{r}$. Thus, the result holds. □

A common definition is as follows: a function $f(x)$ is said to be unimodal if there exists a real number m so that it is monotonically increasing for $x \leq m$ and monotonically decreasing for $x \geq m$.

From Theorems 2.1 and 2.2, we conclude that $J_n(r)/n$ is monotonically increasing in $[\underline{r},\ r_0]$ and monotonically deceasing in $[r_0,\ \bar{r}]$ with probability tending to one as the sample size goes to infinity. To illustrate Theorem 2.2, we next give four examples.

Example 2.1 *Consider the threshold regression model:*

$$y_t = \begin{cases} \beta_{10} + \beta_{11}x_t + \varepsilon_t, & \text{if } z_t \le r_0, \\ \beta_{20} + \beta_{21}x_t + \varepsilon_t, & \text{if } z_t > r_0, \end{cases} \tag{2.6}$$

where $\{(x_t, z_t, \varepsilon_t)'\}$ *is i.i.d.;* x_t, z_t *and* ε_t *are mutually independent;* ε_t *has zero mean and finite variance* σ_ε^2*;* x_t *has zero mean and finite variance* σ_x^2*; the density of* z_t *is continuous and positive in the neighbourhood of* r_0*, and* $\beta_{10} \neq \beta_{20}$ *or* $\beta_{11} \neq \beta_{21}$.

It is not hard to see that Assumptions 2.1 and 2.2 hold. By calculation, we have $\mathbf{G}_r = \mathrm{diag}(1, \sigma_x^2)F_z(r)$, where $F_z(\cdot)$ is the cumulative distribution function of z_t, and

$$J(r) = \left\{(\beta_{10} - \beta_{20})^2 + \sigma_x^2(\beta_{11} - \beta_{21})^2\right\} J^*(r),$$

where

$$J^*(r) = F_z(r)\{1 - F_z(r)\}\left(\frac{F_z(r \wedge r_0)}{F_z(r)} + \frac{F_z(r \wedge r_0) - F_z(r_0)}{1 - F_z(r)}\right)^2.$$

Figure 1 gives the curves of $J^*(r)$ when z_t is Cauchy distribution with $r_0 = 1$ and when $z_t \sim U[0,1]$ with $r_0 = 0.4$, respectively. From Figure 1, we can see that $J^*(r)$ is unimodal, and so is $J(r)$ since $J(r)$ is proportional to $J^*(r)$.

Example 2.2 *Consider the threshold regression model:*

$$y_t = \begin{cases} \alpha_1 x_{t1} + \alpha_2 x_{t2} + \varepsilon_t, & \text{if } x_{t1} \le r_0, \\ \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t, & \text{if } x_{t1} > r_0, \end{cases} \tag{2.7}$$

where $(\alpha_1 - \beta_1)r_0 \neq 0$, $\{(x_{t1}, x_{t2}, \varepsilon_t)'\}$ *is i.i.d.,* $\varepsilon_t \sim N(0,1)$ *and independent of* $(x_{t1}, x_{t2})'$ *which is bivariate normal with mean zero and covariance matrix*

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

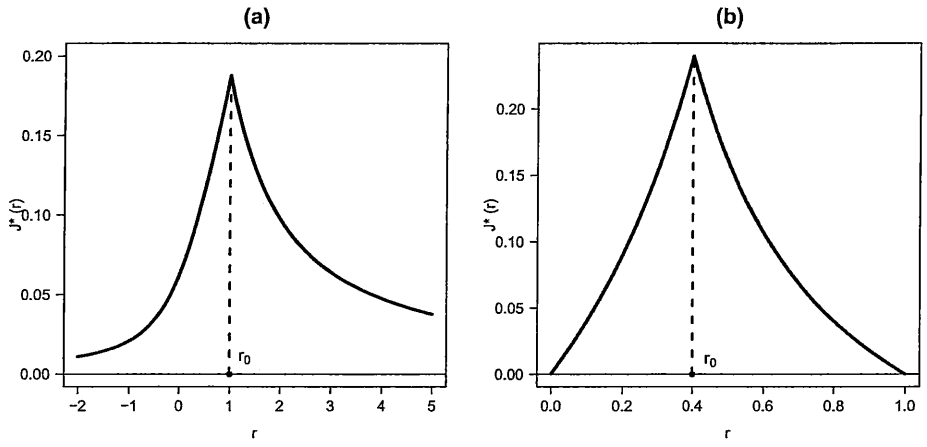

Figure 1: The curves of $J^*(r)$ in models (2.6) when z_t is Cauchy distribution with $r_0 = 1$(left) and $U[0,1]$ with $r_0 = 0.4$ (right), respectively.

Using the fact $x_{12}|x_{11} \sim N(\rho\frac{\sigma_2}{\sigma_1}x_{11},\ (1-\rho^2)\sigma_2^2)$, we have

$$\mathbf{G}_x = \begin{pmatrix} \sigma_1^2 h(x/\sigma_1) & \rho\sigma_1\sigma_2 h(x/\sigma_1) \\ \\ \rho\sigma_1\sigma_2 h(x/\sigma_1) & (1-\rho^2)\sigma_2^2\Phi(x/\sigma_1) + \rho^2\sigma_2^2 h(x/\sigma_1) \end{pmatrix},$$

where $h(x) = \Phi(x) - (2\pi)^{-1/2}x\exp(-x^2/2)$ and $\Phi(x)$ is the cumulative distribution function of standard normal. Figure 2 shows the curves of $J(r)$ for the following two cases:

(i). $\Sigma = \begin{pmatrix} 4 & 7 \\ 7 & 25 \end{pmatrix}$ and $r_0 = 1$; (ii). $\Sigma = \begin{pmatrix} 1 & -0.3 \\ -0.3 & 0.25 \end{pmatrix}$ and $r_0 = 0$

with $\boldsymbol{\rho} = (\alpha_1 - \beta_1, \alpha_2 - \beta_2)' = (1, 0.5)'$. From Figure 2, we can see that $J(r)$ is unimodal.

Example 2.3 *Consider the piecewise constant TAR model:*

$$y_t = \alpha_0 I(y_{t-1} \le r_0) + \beta_0 I(y_{t-1} > r_0) + \varepsilon_t, \tag{2.8}$$

where $\varepsilon_t \sim_{i.i.d.} N(0,1)$.

From Li, Ling and Tong (2012), we know that model (2.8) is always strictly stationary. Let $F(x) = P(y_t \le x)$. Then it follows that

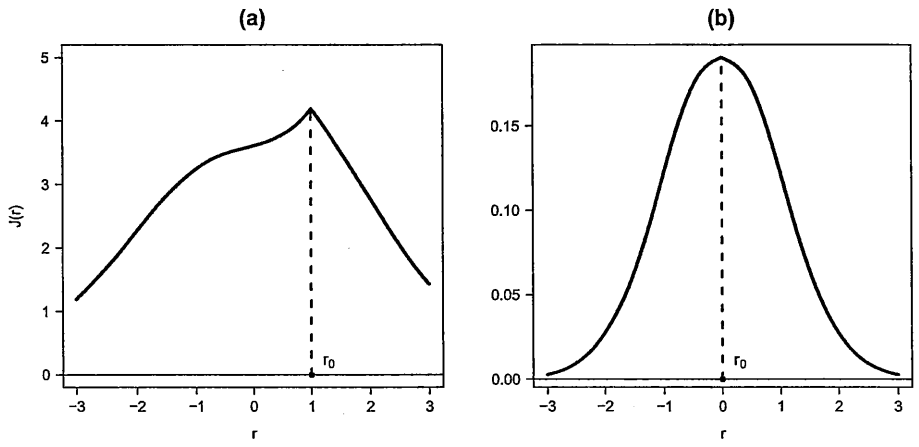

Figure 2: The curves of $J(r)$ in models (2.7) for the case (i) (left) and the case (ii) (right), respectively.

$$F(x) = \frac{\Phi(r_0 - \beta_0)\Phi(x - \alpha_0)}{1 - \Phi(r_0 - \alpha_0) + \Phi(r_0 - \beta_0)} + \frac{\{1 - \Phi(r_0 - \alpha_0)\}\Phi(x - \beta_0)}{1 - \Phi(r_0 - \alpha_0) + \Phi(r_0 - \beta_0)},$$

where $\Phi(x) = P(\varepsilon_t \leq x)$. A simple calculation yields that

$$J(r) = (\beta_0 - \alpha_0)^2 \left\{ \frac{F(r \wedge r_0)}{F(r)} + \frac{F(r \wedge r_0) - F(r_0)}{1 - F(r)} \right\}^2 F(r)\{1 - F(r)\}.$$

Figure 3 (a) shows the curve of $J(r)$ in model (2.8) with $(\alpha_0, \beta_0, r_0) = (-1,\ 2,\ 0.7)$ for $r \in [-4,\ 5]$. Clearly, $J(r)$ is unimodal.

Example 2.4 *Consider the TAR model:*

$$y_t = \begin{cases} 1 - 0.3y_{t-1} + 0.5y_{t-2} + \varepsilon_t, & \text{if } y_{t-2} \leq 1, \\ -1 + 0.6y_{t-1} - 0.3y_{t-3} + \varepsilon_t, & \text{if } y_{t-2} > 1, \end{cases} \tag{2.9}$$

where $\varepsilon_t \sim_{i.i.d.} N(0, 1)$.

For this TAR model, $J(r)$ has no closed form, but we can simulate it by Theorem 2.1. Figure 3 (b) shows the simulated curve of $J(r)$ by 50 replications each with sample size 2000. Clearly, $J(r)$ is unimodal.

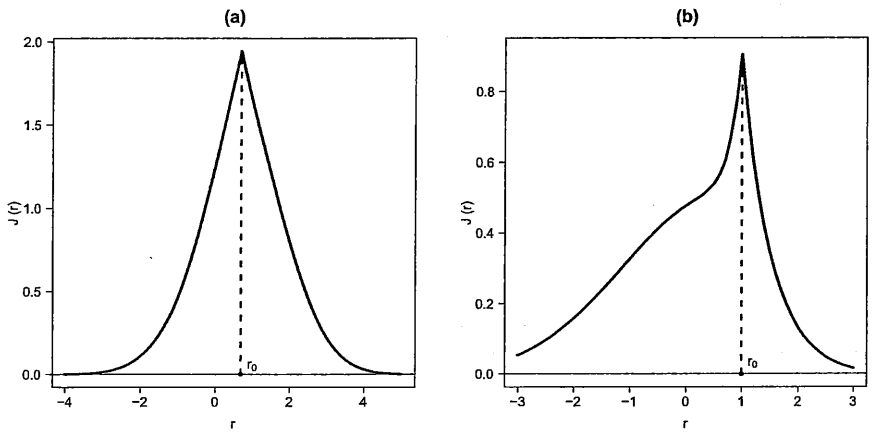

Figure 3: The curves of $J(r)$ in models (2.8) (left) and (2.9) (right).

3 Nested sub-sample search algorithm

Suppose the sample $\mathbf{y}$, $\mathbf{X}$ and $\mathbf{z}$ is available. Clearly, $J_n(r)$ is a step function, namely $J_n(r) = J_n(z_{(i)})$ for $r \in [z_{(i)},\ z_{(i+1)})$, where $z_{(1)} \leq ... \leq z_{(n)}$ is the order statistics of $\{z_1, ..., z_n\}$. We adopt the approach of Tong and Lim (1980) by considering the empirical percentiles as candidates for the threshold values. The SGS algorithm maximizes $J_n(r)$ defined in (2.4) over the *feasible set* $\{z_{(1)}, ..., z_{(n)}\}$ 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 $\{z_{(1)}, ..., z_{(n)}\}$. We first maximize $J_n(r)$ over the subset $\{z_{(k)}, z_{(2k)}, ..., z_{(qk)}\}$, where $k = [n/(q+1)]$ and $[a]$ is the largest integral part of a. Then we get the maximizer $z_{(j_0 k)}$ for some $j_0 \in \{1, ..., q\}$ and a new feasible set $(z_{((j_0-1)k)},\ z_{((j_0+1)k)}) \cap \{z_{(1)}, ..., z_{(n)}\}$.

Repeat the procedure above by updating the feasible set. After m steps, we get a feasible set that contains $[2^m n/(q+1)^m]$ candidates out of $\{z_{(1)}, ..., z_{(n)}\}$ and over which we maximize $J_n(r)$ by enumeration. Thus, the total required number of least squares operations is about $mq + [2^m n/(q+1)^m]$. If, in the last step on maximizing $J_n(r)$, the required number of least squares operations is not beyond δ, which is a pre-assigned positive integer, that is, $[2^m n/(q+1)^m] \leq \delta$, then the number m of iterations satisfies

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

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

$$\frac{q}{\log(\frac{q+1}{2})} \log(\frac{n}{\delta+1}) + \delta. \tag{3.10}$$

Since the minimizer of $q/\log((q+1)/2)$ in (3.10) over the set of positive integers is 3, we take $q = 3$. As for the choice of δ, we can set $\delta = 50$ empirically when sample size $n \geq 200$. If the sample size is less than 200, we can set $\delta = 85$ or use the SGS algorithm to get the estimate $\widehat{r}$ since the computational cost is not high in this case. Figure 4 gives the total required number of least squares operations

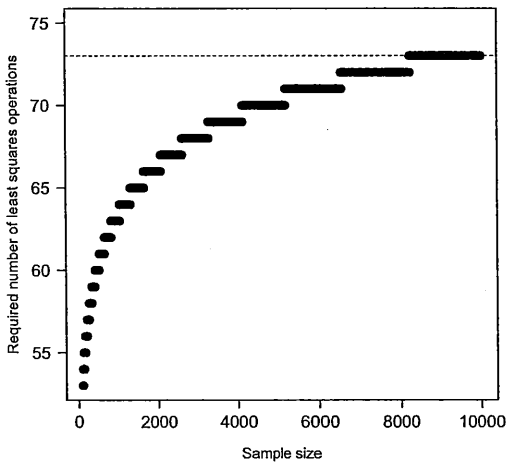

Figure 4: The total required number of least squares operations in (3.10).

in (3.10) 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 $\mathfrak{F} = \{z_{(1)}, ..., z_{(n)}\}$
  while $(\#\mathfrak{F} > \delta)\{$
    calculate q=quantile($\mathfrak{F}$, c(0.25, 0.5, 0.75)) and $J_n(q[i])$ for $i = 1, 2, 3$;
      if $(J_n(q[1]) \geq \max\{J_n(q[2]), J_n(q[3])\})$
        $\mathfrak{F} \leftarrow \mathfrak{F}[\mathfrak{F} \leq q[2]]$
      else if $(J_n(q[2]) \geq \max\{J_n(q[1]), J_n(q[3])\})$
        $\mathfrak{F} \leftarrow \mathfrak{F}[q[1] \leq \mathfrak{F} \leq q[3]]$
      else
        $\mathfrak{F} \leftarrow \mathfrak{F}[\mathfrak{F} \geq q[2]]$
      }
Maximizing $J_n(r)$ over $\mathfrak{F}$ and then getting $\widehat{r}$. □

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

Finally, we should mention that the Fibonacci algorithm, which is optimal for optimizing *deterministic* unimodal functions in numerical analysis, is not optimal for optimizing $J_n(r)$ due to its randomness.

4 Simulation studies

To assess the performance of our algorithm in finite samples, we conduct simulation studies, using sample size $n = 200$, 400, 800, 1600 and 3200 for model (2.9) and model (2.7) with $(\alpha_1, \alpha_2) = (0.5, 1.2)$, $(\beta_1, \beta_2) = (-0.5, 0.7)$, $r_0 = 1$, $(x_{t1}, x_{t2})' \sim N(\mathbf{0}, \Sigma)$ with

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

In all simulations, the innovation $\varepsilon_t \sim_{i.i.d.} N(0,1)$. The program is written in R. For threshold regression models, like model (2.6) and (2.7), the program in R by Hansen (2000) is available on the website:

http://www.ssc.wisc.edu/~bhansen/progs/ecnmt_00.html.

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 time in optimizing $J_n(r)$ for model (2.7) by the SGS and our algorithm with 100 replications, as well as that for model (2.9) by `tar`. 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.

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

Model	n	200	400	800	1600	3200
Model (2.7)	SGS	2.62	6.26	20.62	61.90	211.62
	NeSS	0.83	1.11	2.28	3.00	5.52
Model (2.9)	`tar`	6.46	12.31	24.23	48.88	100.93
	NeSS	1.34	2.05	3.48	6.52	15.13

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 ratio of the total numbers of times that the two algorithms produce the same maximizer to one thousand in 1000 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 (2.7), the matching rates are both still 1. However, for model (2.9), the matching rate is 0.992 for NeSS, which is very close to 1. Of course, when $n < 200$, we can use either the SGS algorithm or the `tar` directly since the computational cost is not high.

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 1000 replications.

Model	200	400	800	1600	3200
Model (2.7)	1.000	1.000	1.000	1.000	1.000
Model (2.9)	0.999	1.000	1.000	1.000	1.000

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. Its validity has been supported theoretically and its reliability demonstrated numerically.

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) and Yu (2012). In this case, we can use the NeSS algorithm instead to get the LSE of the parameter as the pre-estimate or to initiate the optimization of the object function to get the MLE or LADE of the parameter. The NeSS algorithm can also be applied to T-CHARM of Chan *et al.* (2014) and 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 the multiple thresholds, one at a time, by the NeSS algorithm. However, it is known that the limiting distributions 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.

Acknowledgment

Tong's research was partially supported by the Distinguished Visiting Professorship at the University of Hong Kong.

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. *Econometric 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** (to appear).

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/.

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. Econometrics* **110**, 319–352.

Gonzalo, J. and Wolf, M. (2005). Subsampling inference in threshold autoregressive models. *J. Econometrics* **127**, 201–224.

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

Hansen, B. E. (1997). Inference in TAR models. *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. Interface* **4**, 123–127.

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

Li, D., Ling, S. and Tong, H. (2012). On moving-average models with feedback. *Bernoulli* **18**, 735–745.

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. Ass.* **53**, 873–880.

Quandt, R. E. (1984). Computational problems and methods, *Handbook of Econometrics* Vol. 1, ed. Z. Griliches and M.D. Intriligator, North-Holland, Amsterdam, 701–764.

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. Econometrics* **141**, 704–735.

Tong, H. (1978). On a Threshold Model. In: *Pattern Recognition and Signal Processing*, ed. C.H. Chen, Sijthoff and Noordhoff, Amsterdam, 575–586.

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. R. Stat. Soc.* 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. Econometrics* **167**, 274–294.

Mathematical Sciences Center, Tsinghua University, Beijing 100084, China.

E-mail: dongli@math.tsinghua.edu.cn

Department of Statistics, London School of Economics and Political Science.

E-mail: howell.tong@gmail.com

Current Departmental Research Reports

Serial No.	Date	Research Report Title	Author(s)
498	Oct-12	Buffered threshold autoregressive time series models	Guodong Li, Bo Guan, Wai Keung Li and Philip L.H. Yu
499	Nov-12	On a rescaled fractionally integrated GARCH model	Muyi Li, Wai Keung Li and Guodong Li
500	Nov-12	On the sphericity test with large-dimensional observations	Qinwen Wang and Jianfeng Yao
501	Nov-12	On generalized expectation based estimation of a population spectral distribution from high-dimensional data	Weiming Li and Jianfeng Yao
502	Nov-12	A note on multivariate CUSUM charts	Hualong Yang and Jianfeng Yao
503	Apr-13	A geometric process credibility model	Yeh Lam
504	Jun-13	On mixture double autoregressive time series models	Zhao Liu and Guodong Li
505	Jun-13	A simple formula for mixing estimators with different convergence rates	Stephen M.S. Lee and Mehdi Soleymani
506	Nov-13	On singular value distribution of large-dimensional autocovariance matrices	Zeng Li, Guangming Pan and Jianfeng Yao
507	Nov-13	Threshold models in time series analysis-some reflections	Howell Tong
508	Nov-13	An essay on statistics and dynamical phenomena to high school children	Howell Tong
509	Dec-13	Exact distribution of spurious correlations, Minimax Sample Size and direct selection of predictors	Kai Wang Ng
510	Dec-13	Nested sub-sample search algorithm for estimation of threshold models	Dong Li and Howell Tong

❖❖❖ *The complete listing can be found at* ❖❖❖

http://www.saasweb.hku.hk/research/staff_research_report.php

Requests for off prints may be sent to saas@hku.hk by e-mail