On estimating multiple-regime threshold autoregressive models

Chun Yip Yau*
Chinese University of Hong Kong

Chong Man Tang†
Chinese University of Hong Kong

September 22, 2013

Abstract

Threshold autoregressive (TAR) model is a wildly used class of nonlinear time series models in many fields. However, when the number of thresholds is large, the estimation of the thresholds is often computationally infeasible. In this work we employ the Minimum Description Length (MDL) Principle to develop a criterion function to estimate the number of thresholds and the corresponding AR order and parameter values in each regime. A genetic algorithm is developed to efficiently solve this optimization problem. This can be interpreted as the spatial version of AutoPARM of Davis, Lee and Rodriguez-Yam (2006).

keywords: Genetic Algorithm, Minimum Description Length (MDL) Principle, Multiple-threshold

1 Introduction

The $r+1$-regime threshold autoregressive (TAR) model, given by

$$X_t = \sum_{j=1}^{r+1} \left(\Phi_j \mathbf{X}_t^{(j)} + \sigma_j e_t \right) I(\theta_{j-1} < X_{t-d} \leq \theta_j), \quad e_t \sim \text{IID}(0,1), \tag{1}$$

where $\mathbf{X}_t^{(j)} = (1, X_{t-1}, \ldots, X_{t-p_j})$, describes the data by different autoregressive structure depending on the range of values of X_{t-d}. The integer d is known as the *delay* parameter

*Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Email: cyyau@sta.cuhk.edu.hk.

†Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Email: s1010097819@sta.cuhk.edu.h

and the parameters $-\infty = \theta_0 < \theta_1 < \cdots < \theta_r < \theta_{r+1} = \infty$ are called the thresholds. Proposed by Tong (1978), the TAR model has received great attention in the nonlinear time series literature and has been widely used in many areas including finance, econometrics and engineering. Surveys on TAR models can be found in Tong (1990, 2012).

For estimating Threshold models, Chan (1993) studied the Least Square Estimation (LSE) of a two-regime TAR model established its asymptotic properties. The theory is extended by Li and Ling (2011) to multiple-regime TAR models. In particular, when the number of threshold is known, the estimated thresholds are n-consistent and asymptotically independent. Also, each threshold estimate converges weakly to the smallest minimizer of a one-dimensional two-sided compound Poisson process. Moreover, the AR parameters in each regime are $\sqrt{n}$-consistent and asymptotically normal.

Despite the well developed theoretical background of the estimation theory, the estimation procedure of TAR models is still computational demanding due to the irregular nature of the threshold parameters. (see, e.g., Li and Ling (2011)). In particular, for a $r+1$-regime TAR model, the global minimum of the least squares criterion requires a multi-parameter grid-based search over all possible values of r threshold parameters, which is of order C_r^n. To tackle this problem, Tsay (1989) connects (1) to a change-point model and proposed a graphical approach to determine the number and the locations of the thresholds. Coakley, Fuertes and Perez (2003) used similar techniques to provide an efficient estimation approach using QR factorizations of matrices. When the number of threshold r is unknown, Gonzalo and Pitarakis (2002) suggest a binary segmentation procedure for choosing r when σ_js are equal. We are not aware of any results for more general models.

In this paper, motivated by the connection between change-point model and TAR model suggested by Tsay (1989), we propose an automatic procedure to estimate the number and location of thresholds and order selection of autoregressive model in each regime. The procedure is based on an objective function derived from the minimum description length (MDL) principle (Davis et al. (2006)).

The paper is organized as follows. In Section 2 we develop the objective function for TAR model based on MDL and present a genetic algorithm for the optimization. Some simulation studies are given in Section 3.

2 Estimation for TAR models

2.1 MDL for TAR models

In this section, we derive an objective function for estimating TAR models based on MDL and propose a genetic algorithm for the optimization. The MDL principle was developed by Rissanen (1989) as a general method for model selection by choosing the model which enables the best compression of data. See Hansen and Yu (2001) and Lee (2001) for basic introductions.

Let $\mathcal{M}$ be the class of TAR model satisfying (1). Any model from this class is denoted by $\mathcal{F} \in \mathcal{M}$. The MDL principle asserts that the information storage of the observations $\{X_t\}$ requires a codelength given by

$$\mathbf{CL}(\{X_t\}) = \mathbf{CL}(\hat{\mathcal{F}}) + \mathbf{CL}(\hat{\mathbf{e}}|\hat{\mathcal{F}})\,, \tag{2}$$

where $\hat{\mathcal{F}}$ is the fitted model and $\hat{\mathbf{e}}$ is the residuals under $\hat{\mathcal{F}}$. That is, the codelength of the data is the sum of the codelength of the fitted model and codelength of the residuals given the fitted model. Note that $\mathbf{CL}(\hat{\mathbf{e}}|\hat{\mathcal{F}})$ can be interpreted as a measure of lack of fit and $\mathbf{CL}(\hat{\mathcal{F}})$ can be regarded as a penalty for model complexity. The MDL principle suggests that the best model is the one that minimizes $\mathbf{CL}(\{X_t\})$.

To proceed, we derive an expression for $\mathbf{CL}(\hat{\mathcal{F}})$. Since a TAR model (1) is completely specified by r, θ_js, σ_j^2s, p_js and Φ_js, $\mathbf{CL}(\hat{\mathcal{F}})$ can be decomposed into

$$\begin{aligned}\mathbf{CL}(\hat{\mathcal{F}}) = \mathbf{CL}(r) + \mathbf{CL}(\theta_1, \theta_2, \cdots, \theta_r) + \mathbf{CL}(p_1, \cdots, p_{r+1}) \\ + \mathbf{CL}(\Psi_1) + \mathbf{CL}(\Psi_2) + \ldots + \mathbf{CL}(\Psi_{r+1})\,,\end{aligned} \tag{3}$$

where $\Psi_j := (\Phi_j, \sigma_j^2) = (\phi_{j0}, \phi_{j1}, \ldots, \phi_{jp_j}, \sigma_j^2)$ is the parameter vector in the j-th regime. In general, to encode an arbitrage integer I, approximately $log_2(I)$ bits are needed. Thus $\mathbf{CL}(r) = \log_2(r)$ and $\mathbf{CL}(p_j) = \log_2(p_j)$. To calculate $\mathbf{CL}(\Psi_j)$, we use the result of Rissanen (1989): A maximum likelihood estimate of a real parameter computed from N observations can be effectively encoded with $\frac{1}{2}\log_2(N)$ bits. Let n_j be the number of observations in the j-th regime. Then each of the $p_j + 2$ parameters of Ψ_j can be viewed as the Gaussian likelihood estimate computed from n_j observations, giving $\mathbf{CL}(\Psi_j) = (p_j + 2)\log_2(n_j)/2$ (see (5) below). Since each threshold θ_j divides the domain of $\{X_{t-d}\}$ into regimes, they can be represented by the order statistics of the series $\{X_{t-d}\}$. As a result, θ_j can be coded by the

maximum of the n_j observations in the j-th regime. As the maximum value of a set of real number is the maximum likelihood estimate of the upper bound of a uniform distribution, $\mathbf{CL}(\theta_1, \theta_2, \cdots, \theta_r) = \sum_{j=1}^{r} \log_2(n_j)/2$. Putting together, we obtain

$$\mathbf{CL}(\hat{\mathcal{F}}) = \log_2(r) + \frac{1}{2}\sum_{j=1}^{r} \log_2(n_j) + \sum_{j=1}^{r+1} \log_2(p_j) + \sum_{j=1}^{r+1} \frac{p_j + 2}{2} \log_2(n_j). \tag{4}$$

Next, we derive an expression for $\mathbf{CL}(\hat{\mathbf{e}}|\hat{\mathcal{F}})$. From Shannon's classical results in information theory, Rissanen (1989) demonstrated that the codelength of $\hat{\mathbf{e}}$ is given by the negative log-likelihood of the fitted model $\hat{\mathcal{F}}$. Given the thresholds θ_js and the order of autoregressive models p_js, the log-likelihood of the data can be approximated by the conditional log-likelihood

$$\begin{aligned} & l(\Psi_1, \cdots, \Psi_{r+1}; \mathbf{x}) \\ = & -\frac{1}{2}\sum_{t=1}^{n}\sum_{j=1}^{r+1}\left(\log(2\pi\sigma_j^2) + \frac{(X_t - \Phi_j \mathbf{X}_t)^2}{\sigma_j^2}\right) I(\theta_{j-1} \le X_{t-d} < \theta_j) \\ = & -\frac{1}{2}\sum_{j=1}^{r+1}\left(n_j log(2\pi\sigma_j^2) + \frac{\sum_{t=1}^{n_j}(Y_t^{(j)} - \Phi_j \mathbf{Y}_t^{(j)})^2}{\sigma_j^2}\right), \end{aligned} \tag{5}$$

where $\{Y_t^{(j)}\}$ are the observations in the j-th regime, sorted in an ascending of X_{t-d} and $\mathbf{Y}_t^{(j)}$ is a vector containing the p_j previous observations of $Y_t^{(j)}$. For example, if $X_a < X_b$ are the two smallest observations greater than θ_{j-1}, then $Y_1^{(j)} = X_{a+d}$, $\mathbf{Y}_1^{(j)} = (1, X_{a+d-1}, \ldots, X_{a+d-p_j})$, $Y_2^{(j)} = X_{b+d}$, $\mathbf{Y}_2^{(j)} = (1, X_{b+d-1}, \ldots, X_{b+d-p_j})$. For simplicity, we take $X_t = 0$ for $t \le 0$. It can be checked readily that minimizing the function inside the summation in (5) gives $n_j(\log(2\pi\hat{\sigma}_j^2) + 1)$, where $\hat{\sigma}_j^2 = \sum_{t=1}^{n_j}(Y_t^{(j)} - \hat{\Phi}_j \mathbf{Y}_t^{(j)})^2/n_j$ and $\hat{\Phi}_j$ is the least square estimator in Li and Ling (2011). Thus the codelenght of the residuals given the model is

$$\mathbf{CL}(\hat{\mathbf{e}}|\hat{\mathcal{F}}) = \frac{n}{2} + \frac{1}{2}\sum_{j=1}^{r+1} n_j \log(2\pi\hat{\sigma}_j^2)\,. \tag{6}$$

Combining (2), (4) and (6), we obtain

$$\begin{aligned} & MDL(r, \theta_1, \cdots, \theta_r, p_1, \cdots, p_{r+1}) := \mathbf{CL}(\{X_t\}) \\ = & \log_2(r) + \sum_{j=1}^{r}\frac{\log_2(n_j)}{2} + \sum_{j=1}^{r+1}\log_2(p_j) + \sum_{j=1}^{r+1}\frac{p_j+2}{2}\log_2(n_j) \\ & + \frac{1}{2}\sum_{j=1}^{r+1} n_j \log(2\pi\hat{\sigma}_j^2) + \frac{n}{2}\,. \end{aligned} \tag{7}$$

The best model is then chosen by minimizing $MDL(r, \theta_1, \cdots, \theta_r, p_1, \cdots, p_{r+1})$ with respect to r, θ_js and p_js. For simplicity we have assumed that d is known. In general the estimation procedure can be repeated for various d to select the best model.

2.2 Optimization Using Genetic Algorithms (GA)

In this section, following Davis, Lee and Rodriguez-Yam (2006, 2008) and Lu, Lund and Lee (2010), we propose a Genetic Algorithm (GA) to conduct the optimization in (7) efficiently. The basic idea of genetic algorithm (GA) can be described as follows. An initial set (*population*) of possible solutions to the optimization problem is generated. The possible solutions, known as *chromosomes*, are presented in vector forms and are free to "evolve" in the following way: Parent chromosomes are randomly chosen from the initial population with probability inversely proportional to the objective criterion to be minimized. Then offspring are produced by mixing two parent chromosomes through a *crossover* operation, so that the good features from the parents may be combined. Offspring may also be produced from a *mutation* operation of a single parent chromosome so that all possible solutions may be explored. With crossover and mutation, a second-generation of offspring is produced. This process is repeated for a number of generations until the an individual chromosome is found to be optimizing the objective function.

The details implementation of the GA is described as follows.

Chromosome Representation.

In GA, a chromosome should carry complete information about the model, i.e., the number of threshold r, the threshold values θ_j and the AR orders p_j. Once these quantities are specified, maximum likelihood estimates of other model parameters can be uniquely determined. Here a chromosome is a vector $\delta = (r, p_1, (R_1, p_2), \cdots, (R_r, p_{r+1}))$ of length $2(r+1)$, where $R_1 < R_2 < \ldots < R_r$ are integers from $\{1, \ldots, n-d\}$ corresponding to the ordered value of $\{X_{t-d}\}_{t=d+1,\ldots,n}$. That is, the j-th threshold θ_j is represented by the R_j smallest value of $\{X_{t-d}\}_{t=d+1,\ldots,n}$, $j = 1, \ldots, r$. Note that the number of observations in the i-th regime is $n_i = R_{i+1} - R_i + 1$. By imposing a *minimum span* constrain where $R_{i+1} - R_i > n_A$ for some integer n_A, we can avoid having too few observations in one regime.

Initial Population Generation.

First the number of threshold r is generated from the Poisson distribution with mean equals 2. Then R_1 to R_r are sampled from $1, \ldots, n-d$ with uniform probability. Next, we find the set of R_is that violates the minimum span condition with $n_A = 20$. Delete one at a time randomly until the condition is satisfied. Next, for $j = 1, \ldots, r+1$, the AR order p_j for the j-th regime is generated uniformly from the integers $\{0, 1, \ldots, P_0\}$. We use $P_0 = 12$ to allow sufficient flexibility and to capture possible seasonal effects.

New generations.

Once a set of initial random chromosomes is generated, new chromosomes are generated by either one of the crossover or the mutation operations. In our implementation the probability for conducting a crossover operation is set to be $\pi_C = 0.9$.

For the crossover operation, two parent chromosomes are chosen from the current population of chromosomes with probabilities inversely proportional to their ranks sorted by their MDL values. Thus, chromosomes with smaller MDL will have a higher chance of being selected. From these two parents, a now offspring is produced as follows. First, the offspring's AR order of the first segment is chosen from one of the parents with equal probabilities. Then we combine and sort the parents' threshold values and select each threshold and its associated AR order with probability 0.5. For example, consider two parents $(2, 1, (105, 2), (333, 0))$ and $(3, 0, (212, 4), (349, 2), (788, 1))$. First we select the order of the first segment randomly form the set $(1, 0)$, say giving 0. Then, the set of thresholds from the parents are combined and sorted to be $(105, 212, 333, 349, 788)$. Next, each of the threshold value is selected with probability 0.5, say yielding $(105, 333, 788)$. Then the offspring chromosome is constructed as $(3, 0, (105, 2), (333, 0), (788, 1))$.

For mutation, one offspring is reproduced from one parent. In particular, a new parent is generated to crossover with the selected parent. Contrary to the ordinary crossover operation, the probabilities of selecting a threshold is π_P for the selected parent and $1 - \pi_P$ for the new parent. To allow a higher degree of mutation, we take $\pi_P = 0.3$. After an offspring is produced, for each regime, with probability $\pi_N = 0.3$ we randomly draw an AR order to replace the existing one. Of course, at the end of each crossover and mutation operation, the procedure that ensures the minimum span condition will be performed.

To guarantee the monotonicity of the algorithm, an additional step, the *elitist*, is per-

formed. That is, the worst chromosome of the next generation is replaced with the best chromosome of the current generation.

Migration.

To gain computational efficiency from parallel computing, we implement the *island model* that can speed up the convergence rate and to reduce the chance of converging to suboptimal solutions (Forrest (1991); Martin, Lienig, and Cohoon (2000); Alba and Troya (1999, 2002)). Rather than running GA in one giant population, the island model simultaneously runs N_I GAs in N_I different subpopulations (Islands). The key feature is that some chromosomes are migrated periodically among the islands according to some migration policy. In this article we adopt the following migration policy: After every M_i generations, the worst M_N chromosomes from the j-th island are replaced by the best M_N chromosomes from the $(j-1)$st island, $j = 2, \ldots, N_I$. For $j = 1$, the best M_N chromosomes are migrated from the N_Ith island. In our simulations we used $N_I = 50, M_i = 5, M_N = 2$, and a subpopulation size of 100.

Declaration of Convergence.

At the end of each migration, the overall best chromosome (i.e., the chromosome with the smallest MDL) is noted. If the best chromosome does not change for 10 consecutive migrations, or if the total number of migrations exceeds 20, then this chromosome is taken as the solution to the optimization problem.

Remark 1. In the GA, once the thresholds are given, there is no need to check the indicator function $I(\theta_{j-1} < X_{t-d} < \theta_j)$ for each t to categorize the observations into regimes. Recall that the thresholds are specified by $(R_1, \ldots, R_r)$, the order of the X_{t-d}s. Thus, the observations in the j-th regimes are those X_ts with X_{t-d} ranking between $R_{j-1}+1$ and R_j. Therefore, to categorize the observations into regimes, we only need to sort $\{X_{t-d}\}$ and arrange the corresponding values of X_t and $\mathbf{X}_t$ in the order of $\{X_{t-d}\}$, see Tsay (1989). This greatly reduces the computational burden.

3 Simulation experiments

In this section we demonstrate the MDL procedure of TAR models with a simulation experiments. In particular, a series is of length 2000 is generated from the four-regime TAR model

$$\begin{aligned} x_t = &-0.7x_{t-1}I(x_{t-1} \le -0.8) + 0.8x_{t-1}I(-0.8 < x_{t-1} \le -0.3) \\ &- 1.25x_{t-1}I(-0.3 < x_{t-1} \le 0.5) - 2x_{t-1}I(0.5 < x_{t-1}) + e_t, \end{aligned} \tag{8}$$

where $e_t \overset{iid}{\sim} N(0,1)$. Figure 5.1 shows a realization of model (8). We applied the MDL procedure to that realization and obtained three threshold breaks located at $\hat{\theta}_1 = -0.798$, $\hat{\theta}_2 = -0.299$ and $\hat{\theta}_3 = 0.502$. Also, the procedure correctly identified the AR orders ($\hat{p}_1 = \hat{p}_2 = \hat{p}_3 = \hat{p}_4 = 1$) for this realization.

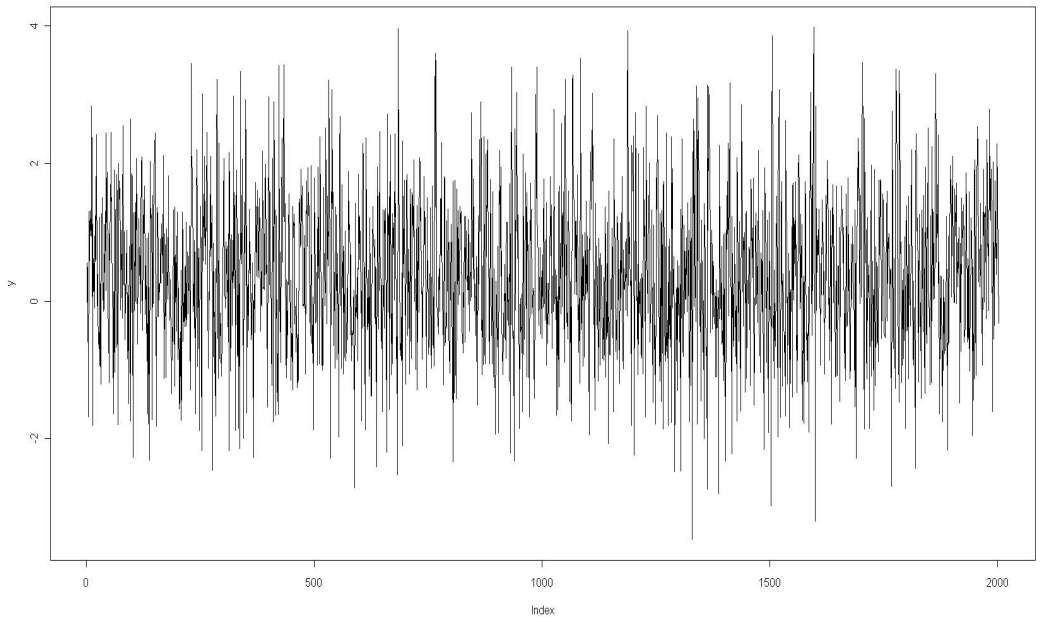

Figure 1: *A realization from the TAR Model* (8).

To study the accuracy, we applied the MDL procedure to 200 realizations of the process (8). Table 1 lists the percentages of the estimated number of regimes. Note that the number of thresholds are correctly identified in all the 200 realizations. Table 1 also reports the sample

mean and standard deviation of the three estimated thresholds. All estimated thresholds have small bias and standard deviations.

Table 2 summarizes the relative frequencies of the estimated AR orders in the four regimes. Of the 200 realizations, 96% correctly identify the model, i.e., detecting three thresholds and specifying that the AR orders in all regimes are 1. From Tables 1 and 2, we can see that the proposed method performs well for the TAR process (8), especially in locating the thresholds.

Table 1: *Estimated number and locations of thresholds for the TAR process* (8)

Number of Thresholds	%	Threshold Values Mean	SE	Mean	SE	Mean	SE
3	100.0	-0.804	0.011	-0.3	0.026	0.499	0.015
4	0						
≥ 5	0						
All	100.0						

Table 2: *Relative frequencies of the estimated AR order of the TAR process* (8)

Order	0	1	2	3	≥ 4
p_1	0	97.0	2.0	1.0	0
p_2	1.0	98.0	1.0	0	0
p_3	0	98.0	2.0	0	0
p_4	0	99.0	1.0	0	0

Acknowledgements

The authors would like to thank Prof. Thomas C.M. Lee for introducing the problem. The research of Chun Yip Yau is supported in part by HKSAR-RGC Grants CUHK405012, 405113 and Direct Grant CUHK2060445.

References

[1] Alba, E., and Troya. J. M. (1999). *A Survey of Parallel-Distributed Genetic Algorithms.* Complexity, **4**, 31-52.

[2] Alba, E., and Troya. J. M. (2002). *Improving Flexibility and Efficiency by Adding Parallelism to Genetic Algorithms.* Statistics and Computing, **12**, 91-114.

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

[4] Coakley, J., Fuertes, A.M., and Perez, M.T. (2003) *Numerical issues in threshold autoregressive modeling in time series.* Journal of Economic Dynamics and Control, **27**, 2219–2242.

[5] Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2006). *Structural break estimation for non-stationary time series models.* Journal of the American Statistical Association, **101**, 223–239.

[6] Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2008). *Break detection for a class of nonlinear time series models.* Journal of the Time Series Analysis, **29**, 834–867.

[7] Forrest, S. (1991). *Emergent Computation.* Cambridge, MA:MIT Press.

[8] Gonzalo, J. and Pitarakis, J.Y. (2002). *Estimation and model selection based inference in single and multiple threshold models.* Journal of Econometrics, **110**, 319–352.

[9] Hansen, M.H. and Yu, B. (2001). *Model selection and the principle of minimum description length.* Journal of American Statistical Assocation, **96**, 746–774.

[10] Lee, T.C.M. (2001). *An introduction to coding theory and the two-part minimum description length principle.* International Statistical Review, **69**, 169–183.

[11] Li, D. and Ling, S. (2011). *On the least squares estimation of multiple-regime threshold autoregressive models.* Journal of Econometrics, **167**, 240–253.

[12] Lu, Q, Lund, R.B. and Lee (2010). *An MDL approach to the climate segmentation problem.* Annals of Applied Statistics, **4**, 299–319.

[13] Martin, W.N., Lienig, J., and Cohoon, J. P. (2000), *Island (Migration) Models: Evolutionary Algorithm Based on Punctuated Equilibria.* Evolutionary Computation, Vol. 2, Advanced Algorithms and Operators, ed. D. B. Fogel, Philadelphia: Bristol, pp. 101-124.

[14] Rissanen, J. (1989). *Stochastic Complexity in Statistical Inquiry.* Singapore: World Scientific.

[15] Tong, H (1978). *On a threshold model*. In: Chen, C, (ed.) Pattern recognition and signal processing. NATO ASI Series E: Applied Sc. (29). Sijthoff & Noordhoff, Netherlands, pp. 575–586.

[16] Tong, H. (1990). *N*onlinear Time Series: A dynamic system approach. Oxford University Press, New York.

[17] Tong, H. (2012). *T*hreshold models in time series analysis 30 years on. Statistics and Its Interface, **4**, 107–118.

[18] Tsay, R. S. (1989). *T*esting and Modeling threshold Autoregressive processes. Journal of American Statistical Association, **84**, 231–240.