Estimation of Multiple-Regime Threshold Autoregressive Models With Structural Breaks

Chun Yip Yau, Chong Man Tang, and Thomas C. M. Lee

The threshold autoregressive ( TAR ) model is a class of nonlinear time series models that have been widely used in many areas. Due to its nonlinear nature, one major difficulty in fitting a TAR model is the estimation of the thresholds. As a first contribution, this article develops an automatic procedure to estimate the number and values of the thresholds, as well as the corresponding AR order and parameter values in each regime. These parameter estimates are defined as the minimizers of an objective function derived from the minimum description length (MDL) principle. A genetic algorithm (GA) is constructed to efficiently solve the associated minimization problem. The second contribution of this article is the extension of this framework to piecewise TAR modeling; that is, the time series is partitioned into different segments for which each segment can be adequately modeled by a TAR model, while models from adjacent segments are different. For such piecewise TAR modeling, a procedure is developed to estimate the number and locations of the breakpoints, together with all other parameters in each segment. Desirable theoretical results are derived to lend support to the proposed methodology. Simulation experiments and an application to an U.S. GNP data are used to illustrate the empirical performances of the methodology. Supplementary materials for this article are available online.

KEY WORDS: Change-point detection; Genetic algorithms; Minimum description length principle; Multiple-threshold; Nonstationary time series; Piecewise modeling

1. INTRODUCTION

The \((r + 1)\)-regime threshold autoregressive ( TAR ) model, defined by

\[
X_t = \sum_{j=1}^{r+1} \left( \Phi_j X_{t-j} + \sigma_j \varepsilon_t \right) I(\theta_{j-1} < X_{t-j} \leq \theta_j), \\
\varepsilon_t \sim \text{IID}(0, 1),
\]

where \(X_{t-j} = (1, X_{t-1}, \ldots, X_{t-p_j})\), asserts that the autoregressive structure of the process \(X_t\) is governed by the range of values of \(X_{t-j}\). The integer-valued quantity \(d\) is known as the delay parameter and the parameters \(-\infty = \theta_0 < \theta_1 < \cdots < \theta_r < \theta_{r+1} = \infty\) are called the thresholds. These thresholds partition the space of \(X_{t-j}\) into \(r + 1\) regimes, in which the \(j\)th regime follows an AR model with order \(p_j\), coefficient parameter \(\Phi_j\) and white-noise variance \(\sigma_j^2\). Since its initial proposal by Tong (1978), the TAR model has attracted enormous attention from the nonlinear time series literature. It has also been applied to study real data behaviors in many areas, especially econometrics and finance. Excellent surveys on the TAR models can be found in Tong (1990, 2011). For parameter estimation, Chan (1993) considered the least square estimation for a two-regime TAR model and established its large sample theory. Li and Ling (2012) extended the theory to a multiple-regime TAR models. In particular, it is shown that, when the number of thresholds is known, each estimated threshold is \(\sqrt{n}\)-consistent and converges weakly to the smallest minimizer of a two-sided compound Poisson process. Also, the estimated thresholds are asymptotically independent. Moreover, if the AR order is known, the AR parameter estimates 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 demands a high computational cost due to the irregular nature of the threshold parameters; for example, see Li and Ling (2012). In particular, for a \((r + 1)\)-regime TAR model, locating the global minimum of the least squares criterion requires a multi-parameter grid search over all possible values of the \(r\) threshold parameters, which is typically computational infeasible. To circumvent this difficulty, Tsay (1989) transformed (1) into a change-point model and proposes a graphical approach to help visually determine the number and values of the thresholds. Coakley, Fuertes, and Pérez (2003) used similar techniques to develop an estimation approach that is based on QR factorizations of matrices. When the number of thresholds \(r\) is unknown, Gonzalo and Pitarakis (2002) suggested a sequential estimation procedure for choosing \(r\), under the assumption that all \(\sigma_j\)'s are equal. We are not aware of any results for more general models. In particular, when the number of thresholds and the order of the AR model in each regime are unknown and the \(\sigma_j\)'s are assumed equal, no practical estimation method appears available.

In this article, motivated by the close connection between TAR model and the structural break model suggested by Tsay (1989), we propose an automatic procedure to estimate the number and values of thresholds, as well as to perform model selection in each regime. The procedure is based on an objective function derived from the minimum description length (MDL) principle, for which the best fitting TAR model is defined as its minimizer. A genetic algorithm (GA) is developed to provide

Chun Yip Yau, Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong (E-mail: cyyau@sta.cuhk.edu.hk). Supported in part by HKSAR-RGC Grants CUHK405012, 405113, and Direct Grant CUHK2069445. Chong Man Tang, Department of Statistics, Chinese University of Hong Kong, Shatin, N.T., Hong Kong (E-mail: st ➤
an efficient computational method for corresponding minimization problem. We shall establish the consistency of our MDL estimates for the number and locations of the thresholds, the delay parameter and the AR orders in each regime. Moreover, the convergence rate of the thresholds and the AR coefficients are established.

For many real life time series data, especially when they are long, the stationarity assumption is often violated. One approach to handle this issue is to perform piecewise stationary modeling via structural break detection. For example, Davis et al. (2006) considered piecewise modeling of AR processes, and their method is generalized to piecewise modeling of GARCH and stochastic volatility models in Davis, Lee, and Rodriguez-Yam (2008). Lavielle and Ludena (2000) studied structural breaks in spectral distribution in stationary time series, and Aue et al. (2009) considered break detections in multivariate time series. For TAR models, Berkes et al. (2011) and Zhu and Ling (2012) developed tests for the presence of structural breaks and demonstrated their significance in time series of soybean and stock prices. However, there seems to be no structural break estimation method available for TAR models. As a second contribution of this article, we further extend the MDL methodology to estimating structural breaks in TAR models. In particular, we provide a computationally efficient procedure for estimating all the unknown parameters in the model, including the number and locations of the structural breaks. We also provide desirable theoretical results backing up our procedure.

The rest of this article is organized as follows. In Section 2, we first derive an MDL objective function for parameter estimation in TAR models, we then develop a GA to conduct the corresponding optimization, and finally present theoretical results that lend support to our methodology. In Section 3, estimation for TAR models with structural breaks will be discussed. Simulation studies and an empirical example are given in Sections 4 and 5, respectively. Concluding remarks are provided in Section 6, while technical details and additional material are delayed to the appendix and a supplementary file.

2. ESTIMATION FOR TAR MODELS

This section considers the problem of estimating the number and values of the thresholds, and also the AR order and parameters in each regime in TAR model (1). We first apply the MDL principle to derive an objective function for which the best fitting TAR model is defined as its minimizer. Then, we construct a GA for finding this minimizer. We also provide desirable theoretical properties to support our method.

2.1 MDL for TAR Models

The MDL principle was developed by Rissanen (1989, 2007) as a general methodology for model selection. It defines the best fitting model as the one that enables the best compression of the available data. There are different versions of MDL, and we shall use the so-called “two-part MDL.” See Hansen and Yu (2001) and Lee (2001) for basic introductions.

Let M be the class of TAR models satisfying (1). Any model from this class is denoted by $F \in M$. In two-part MDL the data $\{X_t\}$ are decomposed into two components: a fitted model $\hat{F}$ and the corresponding residuals $\hat{e}$. Notice that once $\hat{F}$ and $\hat{e}$ are known, the data $\{X_t\}$ can be completely retrieved. Therefore, to store $\{X_t\}$, one could instead store $\hat{F}$ and $\hat{e}$, and the hope is that storing $\hat{F}$ and $\hat{e}$ is cheaper, memory-wise, than storing $\{X_t\}$. This leads to the following fundamental expression in two-part MDL:

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

where $CL(z)$ is a generic notation to denote the code-length for $z$, typically measured in bits. Equation (2) merely states that the code-length of the data $CL(\{X_t\})$ is the sum of the code-length of the fitted model $CL(\hat{F})$ and the code-length of the residuals given the fitted model $CL(\hat{e}|\hat{F})$. Note that $CL(\hat{e}|\hat{F})$ can be interpreted as a measure of lack of fit while $CL(\hat{F})$ can be regarded as a penalty for model complexity. The MDL principle defines that the best model is the one that minimizes $CL(\{X_t\})$.

To proceed, we first derive an expression for $CL(\hat{F})$. Since a TAR model is completely specified by $r, \theta_j$‘s, $\sigma_j^2$‘s, $p_j$‘s, and $\Phi_j$‘s, $CL(\hat{F})$ can be decomposed into

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

where $\Psi_j := (\sigma_j^2, \Phi_j)' = (\sigma_j^2, \phi_j_0, \phi_j_1, \ldots, \phi_{p_j})'$ is the parameter vector in the $j$th regime. In general, to encode an arbitrary integer $I$, approximately $log_2(I)$ bits are required. Thus $CL(r) = log_2(r)$ and $CL(p_j) = log_2(p_j)$. To calculate $CL(\Psi_j)$, the following result of Rissanen (1989) could be used: $\frac{1}{2}log_2(N)$ bits are required to encode a maximum likelihood estimate of a real-valued parameter computed from $N$ observations. Let $n_j$ be the number of observations in the $j$th regime. Then, each of the $p_j + 2$ parameters of $\Psi_j$ can be viewed as the Gaussian likelihood estimate computed from $n_j$ observations, giving $CL(\Psi_j) = (p_j + 2)log_2(n_j)/2$; see (5) below. Since each threshold $\theta_j$ divides the domain of $\{X_t\}$ into regimes, they can be represented by the order statistics of the series $\{X_{t-d}\}$. As a result, $\theta_j$ can be encoded 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, $CL(\theta_1, \theta_2, \ldots, \theta_r) = \sum_{j=1}^r log_2(n_j)/2$. Putting these together, we obtain

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

Next, we derive an expression for $CL(\hat{e}|\hat{F})$. It was shown by Rissanen (1989) that the code-length of $\hat{e}$ is well approximated by the negative log-likelihood of the fitted model $\hat{F}$. Given the thresholds $\theta_j$‘s and the order of autoregressive models $p_j$‘s, the log-likelihood of the data can be approximated by the conditional log-likelihood

$$l(\theta_1, \ldots, \theta_r, \Psi_1, \ldots, \Psi_{r+1}; x) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{r+1} \left( \log \left( 2\pi \sigma_j^2 + \frac{(X_t - \Phi_j'X_{t-1})^2}{\sigma_j^2} \right) \right)$$

$$\times I(\theta_j - 1 \leq X_{t-d} < \theta_j)$$

$$= -\frac{1}{2} \sum_{j=1}^{r+1} \left( n_j \log \left( 2\pi \sigma_j^2 + \frac{\sum_{i=1}^{n_j} (Y_{ij} - \Phi_j'Y_{ij-1})^2}{\sigma_j^2} \right) \right).$$ (5)
where \( Y^{(j)} \) are the observations in the \( j \)-th regime, sorted in an ascending of \( X_{t-d} \) and \( Y^{(j-1)} \) is a vector containing the \( p_j \) previous observations of \( Y^{(j)} \). For example, if \( X_a < X_b \) are the two smallest observations greater than \( \theta_j-1 \), then \( Y^{(j)} = \{ X_{a+d}, X_{a+d-1}, \ldots, X_{a+d-p_j} \} \). \( Y^{(j)} = \{ X_{b+d}, X_{b+d-1}, \ldots, X_{b+d-p_j} \} \). For simplicity, we take \( X_1 = 0 \) for \( t \leq 0 \). It can be checked readily that minimizing the function inside the summation in (5) gives \( \hat{\sigma}^2 = \sum_{j=1}^{n_j} (Y_j^{(j)} - \hat{\Psi}^{(j)}Y_j^{(j-1)})^2/n_j \) and \( \hat{\Psi}^{(j)} \) is the least square estimator in Li and Ling (2012). Thus, the codelength of the residuals given the fitted model is

\[
\text{CL}(\hat{\theta} | \hat{\mathbf{Y}}) = \frac{n}{2} + \frac{1}{2} \sum_{j=1}^{r+1} n_j \log(2\pi \hat{\sigma}^2_j). \tag{6}
\]

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

\[
\text{MDL}(r, \theta_1, \ldots, \theta_p, p_1, \ldots, p_{r+1}) := \text{CL}(\{X_t\}) = \log_2(n) + \sum_{j=1}^{r+1} \log_2(n_j) + \sum_{j=1}^{r+1} \log_2(p_j) + \sum_{j=1}^{r+1} \frac{p_j + 2}{2} \log_2(n_j) + \sum_{j=1}^{r+1} \frac{n_j \log(2\pi \hat{\sigma}^2_j) + \frac{n}{2}}{2}. \tag{7}
\]

The best TAR model is then chosen by minimizing MDL\((r, \theta_1, \ldots, \theta_p, p_1, \ldots, p_{r+1})\) with respect to \( r, \theta_j \)-s, and \( p_j \)-s. 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 Minimization of (7) using GAs

Given the complexity of the TAR model (1), practical minimization of the MDL criterion (7) is difficult. Given the successes of Davis, Lee, and Rodriguez-Yam (2006, 2008); Lee (2001); and Lu, Lund, and Lee (2010), we develop a GA to minimize (7) efficiently. However, we note that our GA has some critical differences from those developed in these references.

The basic idea of GAs can be described as follows. First, a set, or sometimes known as population, of candidate solutions to the optimization problem is generated. These possible solutions, known as chromosomes, are typically presented in vector forms and are free to “evolve” in the following way. Parent chromosomes are randomly chosen from the population with probability inversely proportional to the ranks of their objective criterion values. Then, offsprings are produced by mixing two parent chromosomes through a crossover operation, so that the good features from the parents may be combined. Offsprings could also be produced from a mutation operation from a single parent chromosome so that all possible solutions may be explored. With crossover and mutation, a second-generation of offsprings is produced. This process is repeated for a number of generations until the an individual chromosome is found to be minimizing the objective function. Details of our implementation of the GA for minimizing (1) is given below.

#### 2.2.1 Chromosome Representation.

When using GAs, a chromosome should carry complete information about the model under consideration. In the current context, this corresponds to the number of threshold \( r \), the threshold values \( \theta_j \)-s and the AR orders \( p_j \)-s. Notice that as long as these quantities are specified, one could uniquely calculate the maximum likelihood estimates of other model parameters, so these other model parameters do not need to be present in the chromosome. In our implementation a chromosome is a vector \( \delta = [r, p_1, (R_1, p_2), \ldots, (R_r, p_{r+1})] \) of length \( 2(r+1) \), where \( R_1 < R_2 < \cdots < R_r \) are integers from \( \{1, \ldots, n - d \} \), correspond to the ordered values of \( \{X_{t-d}\}_{t=d+1, \ldots, n} \). That is, the \( j \)-th threshold \( \theta_j \) is represented by the \( R_j \)-th smallest value of \( \{X_{t-d}\}_{t=d+1, \ldots, n} \). Note that the number of observations in the \( j \)-th regime is \( n_j = R_{j+1} - R_j + 1 \). We shall impose a minimum span constrain where \( R_{j+1} - R_j > n_A \) for some integer \( n_A \), so that we can avoid having too few observations in any regime. This constrain was also used by Davis, Lee, and Rodriguez-Yam (2006, 2008) and Lu, Lund, and Lee (2010). However, we note that the chromosome representation adopted here is fundamentally different than those in Davis, Lee, and Rodriguez-Yam (2006, 2008) and Lu, Lund, and Lee (2010): in their work the length of a chromosome is always fixed to be \( n \), while ours is of variable length. Consequently, the method for initializing the first generation, and the definitions for crossover and mutation are necessarily different in our implementation.

#### 2.2.2 Initialization of the First 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 \( R_j \)-s that violates the minimum span condition with \( n_A = 20 \). Delete one at a time randomly until the condition is satisfied. Finally, 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 flexiblity and to capture possible seasonal effects.

#### 2.2.3 Propagation of New Generations.

Once an initial population of random chromosomes are obtained, new chromosomes can then be produced by the crossover and the mutation operations. In our implementation, the probability for conducting a crossover operation is set to be \( \pi_C = 0.9 \).

During crossover, the first step is to randomly select two parent chromosomes from the current population of chromosomes. These parents were selected 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 new 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 \([3, 0, (212, 4), (349, 2), (788, 1)]\). We first 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 \([3, 0, (105, 2), (333, 0), (788, 1)]\).

The goal of the mutation operation is to prevent the GA being trapped in local optima. It is done by introducing randomness to one parent to form one offspring. In our implementation, we
first draw a parent from the current population of chromosome. Then, we generate a second random parent using the method described in Section 2.2.2. Finally, an offspring is produced by applying crossover to these two parents. To allow for a higher degree of mutation, the probabilities of selecting a threshold are, respectively, \( \pi_P = 0.3 \) for the first parent and \( 1 - \pi_P = 0.7 \) for the second random parent. 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 step, is performed. That is, the worst chromosome of the next generation is replaced with the best chromosome of the current generation.

2.2.4 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 (e.g., Alba and Troya 1999, 2002). The basic idea is, rather than running GA with one single population, the island model simultaneously runs \( N_I \) GAs in \( N_I \) different subpopulations, or islands. The distinct feature is to first set up a migration policy and then the chromosomes are allowed to migrate periodically among the islands using this policy. Here, the following migration policy is used: at the end of every \( M_I \) generations, the worst \( M_N \) chromosomes from the \( j \)th island are replaced by the best \( M_N \) chromosomes from the \( (j - 1) \)th island, \( j = 2, \ldots, N_I \). For \( j = 1 \), the worst chromosomes are replaced from the best chromosomes from the \( N_I \)th island. In our simulations, we used \( N_I = 50, M_I = 5, M_N = 2, \) and a subpopulation size of 100.

2.2.5 Declaration of Convergence. At the end of each migration, the overall best chromosome (i.e., the chromosome with the smallest MDL) is noted. This best chromosome is taken as the final solution to our optimization problem if it remains the same for 10 consecutive migrations, or if the total number of migrations exceeds 20. This convergence criterion works really well in all our numerical experiments; that is, different runs of the algorithm often converge to the same solution for the same dataset. Similar convergence criteria have been used by Davis, Lee, and Rodriguez-Yam (2006, 2008).

Finally, we point out that further computational savings can be achieved by the following. In the GA, once the thresholds are given, there is no need to check the indicator function \( I(\theta_{t-1} < X_{t-d} < \theta_t) \) 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 \( X_t \)'s 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 \( X_{t-d} \) in the order of \( \{X_{t-d}\} \), see Tsay (1989). This greatly reduces the computational burden.

2.3 Theoretical Properties

In this section, we establish the consistency of the estimates defined by the MDL criterion (7). Let \( p = \max_{j \in [1, \ldots, r+1]} P_j \) be the maximum AR order of the TAR model among all regimes and \( x_n = (x_n, x_{n-1}, \ldots, x_{n-p+1})' \). By construction, \( \{x_n\} \) is a Markov chain. Denote the \( l \)-step transition probability of \( \{x_n\} \) by \( P^l(x, A) \), where \( x \in \mathbb{R}^p \) and \( A \) is a Borel set. We impose the following assumptions.

**Assumption 1.**

(a) The Markov chain \( \{x_n\} \) admits a unique invariant measure \( \pi(\cdot) \) such that for some \( K > 0, \rho \in [0, 1) \), for positive integer \( n \) and any \( x \in \mathbb{R}^p \), \( \|P^n(x, \cdot) - \pi(\cdot)\| \leq K(1 + |x|^\rho) \), where \( \|\cdot\| \) the total variation norm and \( |\cdot| \) is the Euclidean norm.

(b) The error \( \epsilon_t \) is absolutely continuous with a uniformly continuous and positive p.d.f. Also, \( E(\epsilon_t^4) < \infty \) for some \( \delta > 0 \).

(c) The process \( \{x_t\} \) is stationary with \( E(\epsilon_t^4) < \infty \) for some \( \delta > 0 \).

(d) The autoregressive function is discontinuous in the sense that for each \( j = 1, \ldots, r + 1 \), there exist some \( z^* = (1, z_{p-1}, z_{p-2}, \ldots, z_0)' \) such that \( \Phi_j - \Phi_{j-1} \) \( z^* \neq 0 \).

(e) The number of thresholds \( r \) is unknown but has an upper bound \( R \). The distance between any two thresholds is lower bounded by some \( \epsilon_0 > 0 \).

Assumption 1(a) to 1(d) are similar to that in Chan (1993) and Li and Ling (2012). Assumption 1(e) is an additional assumption required for consistent estimation for the number of thresholds. The proof of the following theorem can be found in the appendix.

**Theorem 1.** Let \( \{x_t\}_{t=1, \ldots, n} \) follows the TAR model (1) with true parameters \( r_0, d_0, \theta, p_0 = (\theta_0(1), \ldots, \theta_0(\theta_0)) \), \( p_0 = (p_0(1), \ldots, p_0(\theta_0)) \) and \( \Psi_0 = (\Psi_0(1), \ldots, \Psi_0(\theta_0)) \). Let the corresponding estimators that minimize the MDL criterion (7) subject to Assumption 1(e) be \( \hat{r}_n, \hat{d}_n, \hat{\theta}_n, \hat{p}_n, \hat{\Psi}_n \), respectively. If Assumption 1 holds, then

\[
(\hat{r}_n, \hat{d}_n, \hat{\theta}_n, \hat{p}_n, \hat{\Psi}_n) \xrightarrow{p} (r_0, d_0, \theta, p_0, \Psi_0).
\]

Moreover, for any \( \delta > 0 \),

\[
\|\hat{\theta}_n - \theta\| = O_p(n^{-1}) \quad \text{and} \quad \|\hat{\Psi}_n - \Psi_0\| = O_p(n^{-1/2}).
\]

The above theoretical results are derived under the assumption that for all \( j = 1, \ldots, r + 1 \), the AR model parameter \( \Psi_j = (\sigma_j^2, \Phi_j) \) in the \( j \)th regime belongs to some compact space \( \Psi_j \in \mathbb{R}^{d_j+2} \). These parameter constraints can be achieved by restricting the optimization in (7) in such a way that the characteristic roots of the resulting AR polynomial are all greater than \( 1 + c \) and the noise variance lies in the interval \([c, 1/c]\), where \( c \) a small constant. In this way, explosive AR models with near unit roots are excluded. We note that our GA does not impose these constraints when optimizing (7). Nevertheless, as \( c \) can be arbitrarily small, say \( 10^{-5} \), it does not affect the results in most practical applications.

3. ESTIMATION OF STRUCTURAL BREAKS IN TAR MODELS

In this section, we extend the above methodology to the estimation of TAR models with (unknown number of) structural breaks. To be precise, we consider the fitting of the following model, which can be seen as several TAR models concatenated
together at different break points:

\[ X_i = \sum_{i=1}^{m+1} \sum_{j=1}^{l} \Phi_{ij} X_{t-i}^{(j)} + \sigma_{ij} \epsilon_t I(\theta_{i,j-1} < X_{t-d_j} \leq \theta_{ij}) \times I(\tau_{i-1} \leq t < \tau_i). \]  

(8)

Here, \( m \) is the number of structural breaks, \( 0 = \tau_0 < \tau_1 < \cdots < \tau_m < \tau_{m+1} = n + 1 \) are the locations of the structural breaks that partition the time series into \( m + 1 \) TAR segments, \( \theta_{i,j} \) are the thresholds that divides the \( i \)th segment into \( r_i + 1 \) regimes with delay parameter \( d_j \). Also, \( X_{t-i}^{(j)} = (X_t, X_{t-1}, \ldots, X_{t-d_j}) \), \( \epsilon_t \) is iid with zero mean, unit variance and independent of the past observations \( X_t \).

Finally, the parameters \( \Phi_{ij} \) and \( \sigma_{ij}^2 \) are, respectively, the AR coefficients and white-noise variance for the \( AR(p_{ij}) \) model in the \( j \)th regime of the \( i \)th stationary segment. We suggest using chromosomes of this form in the GA for TAR models. Once the current population is obtained by sampling \( m \) elements from the discrete uniform distribution on \( \{1, \ldots, n\} \), where \( n \) is the total length of the time series. Analog to the GA previously developed Section 2.2, a “minimum span” constrain has to be imposed to ensure sufficient number of observations in each segment. Here, we use the condition \( \tau_{i-1} < \tau_i < n \) and \( \tau_i = \tau_{i+1} - \delta \) with \( n = 40 \). Again, we find the set of \( \tau_i \)s that violates the minimum span condition and delete one at a time randomly until the condition is satisfied. Once the locations of breaks are specified, the TAR chromosomes for each segment are generated from the procedure discussed in Section 2.2.

In (10), we have ignored the constant \( n/2 \) which does not affect the optimization. As before, the best fitting model is chosen as the one that minimizes (10) with respect to \( m, r_i \)'s, \( \tau_i \)'s, \( \theta_{ij} \)'s, and \( p_{ij} \)'s.

3.2 Optimization Using GA

This subsection provides the implementation details of the GA tailored for minimizing (10). We first remark that the GAs developed by Davis, Lee, and Rodriguez-Yam (2006, 2008) for break point detection cannot be readily adopted here. It is because in Davis, Lee, and Rodriguez-Yam (2006, 2008), a chromosome only needs to record the structural break locations and the order of the model in each segment, as parameters in each segment can be estimated via maximum likelihood once the structural breaks are given. For model (8), however, the estimation in each segment involves the fitting of a standard TAR model (1) which requires one execution of the GA proposed in Section 2.2, hence imposes huge computational burden. To overcome this issue, one possibility is to let the chromosomes contain both the structural break and threshold regime information. The details are as follows.

3.2.1 Chromosome Representation. The chromosomes now should carry complete information about the structural breaks, threshold values, and the AR orders within each regime of each segment. We suggest using chromosomes of the form \( \eta = (m, \delta_1, (\tau_1, \delta_2), \ldots, (\tau_m, \delta_{m+1})) \), where \( \delta_i \) is the “TAR chromosome” given in Section 2.2 that specifies a TAR model (1).

3.2.2 Initialization of First Generation. First, the number of structural breaks \( m \) is generated from the Poisson distribution with mean 2. Then, the locations of structural breaks are obtained by sampling \( m \) elements from the discrete uniform distribution on \( \{1, \ldots, n\} \), where \( n \) is the total length of the time series.

3.2.3 Propagation of New Generations. Similar to Section 2.2, a new generation of chromosomes are produced by crossover and mutation with probability \( \pi_C = 0.9 \) and \( 1 - \pi_C = 0.1 \), respectively.

During crossover, two parent chromosomes are chosen from the current population in a similar fashion as before. From these two parents, we combine and sort their break locations and select each break with probability 0.5. This process is similar to combining the thresholds values in the GA for TAR models. Once the set of offspring breakpoints is obtained, we then assign each breakpoint with one TAR chromosome. In contrast to the GA in Section 2.2 where the AR order is inherited from the parent’s threshold value, we employ an extra crossover step to obtain a new TAR chromosome: at each breakpoint in the offspring, we have two TAR chromosomes from the two parents correspond...
to the TAR model when the break occurs. Then, a crossover step is conducted between these two TAR chromosomes to give a new TAR chromosome. For example, given two parents \( \Psi^a = [3, \delta^a_1, (95, \delta^a_2), (343, \delta^a_3)] \) and \( \Psi^b = [1, \delta^b_0, (208, \delta^b_1), (900, \delta^b_2)] \), the sorted breakpoints set is \((95, 208, 343, 900)\). Suppose that after selection the locations of breaks are \((95, 208, 343)\). Then, crossover operations will be conducted between the four pairs \((\delta^a_1, \delta^b_0), (\delta^a_2, \delta^b_1), (\delta^a_3, \delta^b_2)\), and \((\delta^b_0, \delta^a_1)\), giving four offspring, say \(\delta^j, j = 1, \ldots, 4\), respectively. The offspring chromosome is then given by \(\Psi^a = [3, \delta^a_1, (95, \delta^a_2), (208, \delta^a_3)]\).

In the mutation operation, we follow the idea in Section 2.2: a new random parent is generated to crossover with a parent selected from the current population. As similar to before, the probability of selecting a breakpoint is \(\pi_P = 0.3\) for the selected parent and \(1 - \pi_P = 0.7\) for the new parent. Once the breakpoints are selected, a mutation operation will be conducted for each associated TAR chromosome as discussed in Section 2.2.

Of course, at the end of each crossover and mutation operation, the procedure that ensures the minimum span condition is performed. Also, the elitist step is performed in the same fashion as before to guarantee monotonicity of the algorithm.

3.2.4 Migration and Declaration of Convergence. The same island model, migration procedure and convergence criterion described in Section 2.2 are adopted here.

We close this section with the following remark. When both structural breaks and thresholds are present, the sorting of variable \(\tau\) and \(\theta\) can be computed from the extracted observations.

3.3 Theoretical Properties

Here, we establish the consistency of our parameter estimates defined by the MDL criterion (10). Similar to Davis and Yau (2013), we require the following extra assumption to ensure consistency estimation of the number and locations of structural breaks.

Assumption 2. The location of the \(i\)th structural break satisfies \(\lim_{n \to \infty} \frac{1}{n} = \lambda_i \in (0, 1)\) for \(i = 1, \ldots, m\), where \(\lambda_1 < \cdots < \lambda_m\).

Theorem 2. Let \(\{x_t\}_{t=1,\ldots,n}\) follows model (8) with true parameter \((m_0), \lambda_0 = (\lambda_0^{(1)}, \ldots, \lambda_0^{(m_0)}), r_0 = (r_0^{(1)}, \ldots, r_0^{(m_0 + 1)}), d_0 = (d_0^{(1)}, \ldots, d_0^{(m_0 + 1)}), \theta_0 = (\theta_0^{(1)}, \ldots, \theta_0^{(m_0 + 1)}), p_0 = (p_0^{(1)}, \ldots, p_0^{(m_0 + 1)}), \) and \(\Psi_0 = (\psi_0^{(1)}, \ldots, \psi_0^{(m_0 + 1)}), \) where \(\theta_0^{(1)}\) and \(\psi_0^{(1)}\) are the parameter vectors for the \(i\)th stationary TAR segment, \(i = 1, \ldots, m_0 + 1\). Denote the estimators that minimize the MDL criterion (10) by \(\hat{m}_n, \hat{\lambda}_n = (\hat{\lambda}_n^{(1)}, \ldots, \hat{\lambda}_n^{(m_0)}), \hat{r}_n = (\hat{r}_n^{(1)}, \ldots, \hat{r}_n^{(m_0 + 1)}), \hat{d}_n = (\hat{d}_n^{(1)}, \ldots, \hat{d}_n^{(m_0 + 1)}), \) \(\hat{\theta}_n = (\hat{\theta}_n^{(1)}, \ldots, \hat{\theta}_n^{(m_0 + 1)}), \hat{p}_n = (\hat{p}_n^{(1)}, \ldots, \hat{p}_n^{(m_0 + 1)}), \) and \(\hat{\Psi}_n = (\hat{\psi}_n^{(1)}, \ldots, \hat{\psi}_n^{(m_0 + 1)}).

If Assumption 2 holds and Assumption 1 holds for each of the stationary segments, then \((\hat{m}_n, \hat{\lambda}_n, \hat{r}_n, \hat{d}_n, \hat{\theta}_n, \hat{p}_n, \hat{\Psi}_n)^P \to (m_0, \lambda_0, r_0, d_0, \theta_0, p_0, \Psi_0)\).

Moreover, \(\|\hat{\lambda}_n - \lambda_0\| = O_p(n^{-1}), \|\hat{\theta}_n - \theta_0\| = O_p(n^{-1})\)
and \(\|\hat{\Psi}_n - \Psi_0\| = O_p(n^{-1/2})\).

The proof is delayed to Section S.2 of the supplementary file.

4. SIMULATION EXPERIMENTS

In this section, we evaluate the empirical performance of the proposed methodology via three simulation experiments. Additional simulations are provided in Section S.1 of the supplementary file.

4.1 TAR Model Without Break

In this first example a series is of length 2000 is generated from the four-regime TAR model

\[ x_t = (0.2x_{t-1} - 0.48x_{t-2})I(\langle -\infty, -0.8\rangle) + (1 - 0.9x_{t-1})I(-0.8, -0.3) \]
\[ + (0.4x_{t-1} + 0.3x_{t-2} - 1.5x_{t-3} + 0.5x_{t-4})I(-0.3, 0.5) \]
\[ + (1 - 1.5x_{t-1} + 0.5x_{t-2} - 0.5x_{t-3})I(0.5, \infty) + \epsilon_t, \]

where \(\epsilon_t \sim N(0, 1)\) and \(I_{a, b} = I(a < x_t < b)\).

Figure 1 shows a realization of model (11). We applied the methodology in Section 2 to this realization and obtained three thresholds located at \(\hat{\theta}_1 = -0.8008, \hat{\theta}_2 = -0.3011\), and \(\hat{\theta}_3 = 0.5013\). Also, the proposed procedure correctly identified the AR orders \((\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4) = (2, 1, 4, 3)\) for this realization.

To study the accuracy, we applied the same procedure to 500 realizations of the process (11). Table 1 lists the percentages of the estimated number of regimes. Note that the number of thresholds are correctly identified in all the 500 realizations. Table 1 also reports the sample means and standard errors of the three estimated thresholds. All estimated thresholds have small bias and standard errors.

Table 2 summarizes the relative frequencies of the estimated AR orders in the four regimes. Of the 500 realizations, 94% of them correctly identified the model; that is, detected three thresholds with AR orders equal one in all regimes. From

Figure 1. A realization generated from the TAR model (11).
Consider the following two-regime TAR model

\[ x_t = \begin{cases} 
(-0.7x_{t-1} - 0.9x_{t-2} + 0.2x_{t-3}) \cdot I_{(\infty, -0.8)} + (0.8x_{t-1} - 0.3x_{t-2}) \cdot I_{(-0.8, -0.3)} - 1.25x_{t-1} \cdot I_{(-0.3, \infty)} + \epsilon_t, & 1 \leq t \leq 750, \\
(0.5x_{t-1} - 0.3x_{t-2}) \cdot I_{(-\infty, -1)} - 0.6x_{t-1} \cdot I_{(-1, \infty)} + \epsilon_t, & 750 < t \leq 1400, \\
(0.9x_{t-1} - 0.3x_{t-2}) \cdot I_{(-\infty, 0.5)} - 2x_{t-1} \cdot I_{(0.5, \infty)} + \epsilon_t, & 1400 < t \leq 2000.
\]

where \( \epsilon_t \sim N(0, 1) \) and \( I_{(a, b)} = I(a < x_{t-1} \leq b) \). A realization of (12) is plotted in Figure 2. Again, 500 realizations from the process (12) are simulated and estimated. Table 3 summarizes the results of the structural break estimation, including the number of breaks and the sample mean and standard deviation of the estimated break points when the number of breaks was correctly specified. The number of breakpoints were correctly identified in all of the 500 realizations.

Finally, we investigate the results of structural break estimation for the cases where the number of breakpoints were correctly identified. Table 4 summarizes the estimation results for the TAR models in each of the three segments. Despite the complexity of the model, the 500 realizations, 98% have simultaneously identified the correct numbers of structural breaks and the correct numbers of thresholds in all segments.

### 4.3 Nonlinearity and Nonstationarity

In this last example, we illustrate the behavior of the MDL model selection procedure on nonlinear and nonstationary time series. Consider the following two-regime TAR model

\[ x_t = 3I(x_{t-1} > 0) - 3I(x_{t-1} \leq 0) + \epsilon_t, \quad (13) \]

where \( \epsilon_t \sim N(0, 1) \). Figure 3(a) shows a realization generated from this model. Notice that this realization could be easily misinterpreted as generated from a nonstationary model with two breakpoints, although it was originated from the stationary pure threshold model (13). For this realization, the methodology in Section 2 detected the true two-regime TAR model. This is reasonable because a two-regime model requires coding one threshold and two AR models, which is simpler than the structural break model that requires coding two breakpoints and three AR models.

Consider another realization of model (13) displayed in Figure 3(b). It can be interpreted as a two-regime TAR model or a structural break model with one breakpoint. Recall that in the derivation of the MDL criterion, the codelength required to store a breakpoint is \( \log_2(\tau_i) \), while the codelength required to store a threshold is \( \frac{1}{2} \log_2(n_{ij}) \), where \( n_{ij} \) is the sample size of the corresponding regime. Note that \( \tau_i = n_{ij} \) for this realization. Consequently, the penalty associated with a TAR model is less than the penalty associated with the structural break model. In other words, the MDL procedure favors a stationary threshold model than a nonstationary structural break model, which seems to be an attractive feature.

To study the accuracy, we applied the same procedure to 500 realizations of the process (13) with \( n = 1000 \). In all the 500 realizations, no break point was detected. Table 5 lists the percentages of the estimated number of regimes.

Note that 24.4% of the 500 realizations were misclassified as no threshold. These correspond to those cases where no regime switch was present in the entire realization (e.g., see Figure 3(c)). The same experiment was then repeated with sample size increased to \( n = 2000 \). With a larger \( n \), the chance for regime-switching to occur within a realization increases (e.g., see Figure 3(d)), and therefore it can be seen from Table 5 that the percentage of correct model identification was increased substantially.

---

**Table 1.** Estimated numbers and locations of thresholds for the TAR process (11)

| Number of thresholds | % | Mean | s.e. | Mean | s.e. | Mean | s.e. |
|----------------------|---|------|-----|------|-----|------|-----|
| 3                    | 100.0 | -0.803 | 0.009 | -0.302 | 0.005 | 0.497 | 0.006 |
| ≥ 4                  | 0    |       |     |       |     |       |     |
| All                  | 100.0 |       |     |       |     |       |     |

Tables 1 and 2, one can see that the proposed procedure performs very well for the TAR process (11), especially in locating the thresholds.

**Table 2.** Relative frequencies of the estimated AR orders of the TAR process (11)

| Order | 0 | 1 | 2 | 3 | 4 | ≥ 5 |
|-------|---|---|---|---|---|-----|
| \( p_1 \) | 0 | 0 | 99.6 | 0.4 | 0 | 0 |
| \( p_2 \) | 0 | 94.2 | 5.8 | 0 | 0 | 0 |
| \( p_3 \) | 0 | 0 | 0 | 100.0 | 0 |
| \( p_4 \) | 0 | 0 | 0 | 98.2 | 1.8 | 0 |

**Table 3.** Estimated numbers and locations of breakpoints for the process (12)

| Number of segments | % | Mean | s.e. | Mean | s.e. |
|-------------------|---|------|-----|------|-----|
| 3                 | 100.0 | 750.51 | 8.55 | 1401.87 | 14.58 |
Table 4. Estimated number and values of thresholds for the process (12)

| Number of thresholds | First segment | Second segment | Third segment |
|----------------------|---------------|----------------|--------------|
| % \( \hat{\theta}_{11} \) | % \( \hat{\theta}_{12} \) | % \( \hat{\theta}_{21} \) | % \( \hat{\theta}_{31} \) |
| 0                    | 0             | 0              | 0            |
| 1                    | 1.4           | 99.6           | -0.931       |
|                      |               |                | (0.150)      |
| 2                    | 98.6          | -0.802         | -0.294       |
|                      |               |                | (0.027)      |
| 3                    | 0             | 0              | 0            |
| ≥ 4                  | 0             | 0              | 0            |

Note: The means of the estimated thresholds \( \hat{\theta}_{ij} \) are reported when the number of thresholds were correctly identified. Standard errors of the estimates are in parentheses.

We also studied the pure structural break model

\[
x_t = \begin{cases} 
-3 - 0.7x_{t-1} - 0.3x_{t-2} + 0.2x_{t-3} + e_t, & 1 \leq t \leq 350, \\
3 + 0.5x_{t-1} + e_t, & 350 < t \leq 800, \\
-2 - 0.3x_{t-1} + 0.2x_{t-2} + e_t, & 800 < t \leq 1000.
\end{cases}
\]  

Although the realizations of models (13) and (14) could be very similar (e.g., see Figures 3(a) and 3(e)), the first and the third segment in (14) follow different models and thus they cannot be grouped into the same regime as in (13). It can be seen from Table 6 that the MDL procedure correctly identified the model 90% of the time. In conclusion, the MDL procedure performed very well in distinguishing between structural break models and threshold models with few regime switches.

Figure 3. Realizations generated from the TAR model (13) and structural break model (14).
The MDL value corresponds to the three-regime model of Li and Ling (2012) is 295.74. In each regime of this TAR model, a high order AR model (up to order 10) is required to model the data, giving the total number of parameters in the model as 31. On the other hand, the MDL value of our estimated model is 274.13, with a total number of parameters of 21. Moreover, from Table 7, one can see that simple AR models are sufficient for the data. Thus, given the lower MDL value and few number of parameters, it appears that a TAR model with structural breaks is a better parsimonious alternative to describe the GNP data than a pure TAR model.

The estimated structural breaks correspond to year 1960, 1969, and 1983, respectively. Looking into the history, these three breaks could be associated with substantial changes in the U.S. economy. For examples, between 1960–1969, the Vietnam war broke out and there were complex cultural and political disorder across the globe. In the 1980s, there were great changes in business and economic situations as the production industries migrated to newly industrializing economies.

6. CONCLUDING REMARKS

This article considered the problems of parameter estimation and structural break detection for TAR models. In particular, a new methodology that combines the MDL principle and GAs is proposed to handle these tasks. Desirable consistency properties of the parameter estimates are established. The good finite sample performances of the methodology is illustrated with simulation experiments and an application to a U.S. GNP dataset.

APPENDIX A: PROOF OF THEOREM 1

First we define some notation. Following Chan (1993), the parameter space for the thresholds \( \theta_j, j = 1, \ldots, r \) is taken to be the compact space \( \mathcal{R} = \mathbb{R} \cup [\infty, \infty] \) equipped with the metric \( \tilde{d}(x, y) = |\arctan(x) - \arctan(y)| \). We assume that the model parameter \( \Psi_j = (\sigma_j^2, \Phi_j) \) in the \( j \)th regime belongs to some compact space \( \Psi_j \in \mathbb{R}^{p_j+2} \) for \( j = 1, \ldots, r + 1 \). Denote \( \Theta = (\theta_1, \ldots, \theta_r) \) and \( p = (p_1, \ldots, p_{r+1}) \). Also, rewrite the MDL in (7) as

\[
\text{MDL}(r, \Theta, p) = \text{Pen}(r, p) + \frac{1}{2} \sum_{j=1}^{r+1} L_n(\Psi_j^{(j)}, \theta_{j-1}, \theta_j, p_j, d),
\]

where

\[
\text{Pen}(r, p) = \log(r) + \sum_{j=1}^{r} \log(n_j) + \sum_{j=1}^{r+1} \log(p_j)
\]

\[
+ \sum_{j=1}^{r+1} p_j + 2 \sum_{j=1}^{r+1} \log(n_j)
\]

is the penalty for model complexity and

\[
L_n(\Psi_j, \theta_{j-1}, \theta_j, p_j, d) = n_j \log(2\pi \sigma_j^2) + \frac{\sum_{i=1}^{n_j} (Y_i^{(j)} - \Phi_j Y_{i-1}^{(j)})^2}{\sigma_j^2}
\]

is minus 2 times the conditional log-likelihood for the AR process in the \( j \)th regime; see (5). Here, \( \Psi_j^{(j)} \) is the minimizer of \( L_n(\Psi, \theta_{j-1}, \theta_j, p_j, d) \) w.r.t. \( \Psi \). Also, for any \( \theta_i < \theta_a \), define

\[
Q(\theta_i, \theta_a) = \text{E}(I(\theta_i < X_i \leq \theta_a)),
\]

\[
L(\Psi, \theta_i, \theta_a, p, d) = \lim_{n \to \infty} \text{E}(L_n(\Psi, \theta_i, \theta_a, p, d))
\]
where $\Psi_1$ and Leibler distance, it can be shown that expected proportion of sample size falling into the regime $(2012)$ on the growth rate series of U.S. real GNP. The vertical dash lines correspond to the breakpoint estimates, while the horizontal lines correspond to the threshold estimates. (C): the corresponding residuals of the two fittings.

![Figure 5](image)

**Table 7. Estimated parameters for the GNP growth rate using the MDL procedure**

| Segment | Model | $\hat{\tau}_1 = 56$ | $\hat{\tau}_2 = 91$ | $\hat{\tau}_3 = 150$ |
|---------|-------|---------------------|---------------------|---------------------|
| 1st     | 2nd   | 3rd                 | 4th                 |
| parameter estimates | $\hat{\phi}_{110} = 1.41$ | $\hat{\phi}_{210} = 1.83$ | $\hat{\theta}_1 = 1.93$ | $\hat{\phi}_{310} = 1.24$ |
|          | $\hat{\phi}_{111} = 0.548$ | $\hat{\phi}_{211} = 0.143$ | $\hat{\phi}_{310} = 2.60$ | $\hat{\phi}_{410} = 1.91$ |
|          | $\hat{\phi}_{112} = 0.214$ | $\hat{\sigma}_21 = 0.453$ | $\hat{\phi}_{320} = 1.91$ | $\hat{\sigma}_21 = 0.27$ |
|          | $\hat{\phi}_{113} = -0.377$ | $\hat{\delta}_1 = 1.59$ | $\hat{\phi}_{312} = 0.29$ | $\hat{\delta}_1 = 0.33$ |

Note: The AR coefficient of lag $k$ in the $i$th TAR regime of $i$th segment is denoted by $\hat{\phi}_{k,j}$. The case $k = 0$ corresponds to the constant term of the AR model.

Next we set down some propositions for proving Theorem 1. The proofs of all propositions are given in Section S.3 of the supplementary file.

**Proposition A.1.** Under Assumption 1, for $k = 0, 1, 2, j = 1, \ldots, r + 1$ and fixed $p_j$ and $d$,

$$
\sup_{\phi \in \phi_j} \sup_{\psi \in \psi_j} \left| \frac{1}{n} L_n^{(k)}(\Psi, \theta, \theta^*, p, d) - L_n^{(k)}(\Psi, \theta, \theta^*, p, d) \right| \xrightarrow{a.s.} 0,
$$

where for $k = 1, 2, L_n^{(k)}$ represents the $k$th derivative of $L_n$ with respect to $\Psi_j$, $L_n^{(0)} \equiv L_n$, and $\sup_{\phi \in \phi_j}$ denotes the supremum taken over $(\theta, \theta^*), \phi \in \phi_j$, $\psi \in \psi_j$.

$\phi_j$ and $\phi_j^*$ are converging to 0 as $n \to \infty$.

Figure 5. Estimation results from (A) the proposed MDL procedure in Section 3, and (B) the 3-regime TAR model considered in Li and Ling (2012) on the growth rate series of U.S. real GNP. The vertical dash lines correspond to the breakpoint estimates, while the horizontal lines correspond to the threshold estimates. (C): the corresponding residuals of the two fittings.
For fixed \( p, d \) and \( \theta < \theta_n \), define
\[
\Psi_a = \Psi_a(\theta, \theta_a, p, d) = \arg \min L_n(\Psi_a, \theta_a, p, d),
\]
\[
\Psi_s = \Psi_s(\theta, \theta_a, p, d) = \arg \min L(\Psi_s, \theta_a, p, d).
\]

The next proposition shows that the estimated AR coefficients using a proportion of data inside a segment still converge to the true value.

**Proposition A.2.** Suppose that Assumption 1 holds. Then, for \( j = 1, \ldots, r + 1 \) and fixed \( p, d \),
\[
\sup_{\Psi_a \in \mathcal{A}_j} \left| \Psi_a(\theta_a, p, d) - \Psi_s(\theta_a, p, d) \right| \xrightarrow{\mathbb{P}} 0.
\]

In particular, given that \( d = d_0 \), then \( \Psi_s = \Psi(0)^j \) if \( p = p(0)^j \), and \( \Psi_s = \Psi(0)^j, \theta_0 \) if \( p - p(0)^j = m > 0 \).

The next proposition shows that the number of thresholds and the AR order in each regime cannot be underestimated.

**Proposition A.3.** Under Assumption 1, we have
\[ (a) \] The number of thresholds cannot be underestimated, that is, \( r_n \geq r_0 \) a.s. Also,
\[
\max_{i=1, \ldots, r_0} \min_{j=1}^{r_n} \left| \hat{\theta}_n(0)^j - \theta_0(0)^j \right| \xrightarrow{\mathbb{P}} 0. \tag{A.2}
\]
\[ (b) \] If the \( j \)-th estimated regime is nested in the \( i \)-th true regime, then \( d_n \rightarrow d_0 \) and \( \hat{\theta}_n(0)^j \geq p(0)^j \) almost surely.

The next proposition follows from similar arguments used to derive equation (4.4) of Chan (1993).

**Proposition A.4.** For any \( j = 1, \ldots, r, \epsilon > 0 \) and \( n > 0 \), there exist some \( K > 0 \) and \( \Delta \in (0, 1) \) such that for sufficiently large \( n \),
\[
P \left( \sup_{\Delta \geq 2\Delta} \sum_{i=r_0+1}^{n} y_i I(\theta_{0(n)}^j < X_{i-d_n} \leq \theta_{0(n)}^j + \epsilon) \right) \left( nQ(\theta_{0(n)}^j, \theta_{0(n)}^j + \epsilon) + z \right) \xrightarrow{\mathbb{P}} 0. \tag{A.3}
\]
where \( y_i, \hat{y}_i = (1, 1), (\epsilon, 0) \), \( (X_{i-d_n}, 0) \), or \( (\epsilon^2, \sigma_{\epsilon^2}) \) for \( k = 1, \ldots, p \). Also, similar conclusions (with \( \epsilon^2, \sigma_{\epsilon^2} \) replaced by \( \epsilon^2, \sigma_{\epsilon^2} \)) hold if \( (\theta_{0(n)}^j, \theta_{0(n)}^j + \epsilon) \) is replaced by \( (\theta_{0(n)}^j - \epsilon, \theta_{0(n)}^j) \). It follows that for some \( c > 0 \),
\[
P \left( \sup_{\Delta \geq 2\Delta} \sum_{i=r_0+1}^{n} \left( l(\hat{\Psi}_{n(0)}^{(i+1)}) - l(\hat{\Psi}_{n(0)}^{(i)}) \right) \right) \times I(\hat{\theta}_{n(0)}^j < X_{i-d_n} \leq \hat{\theta}_{n(0)}^j) \xrightarrow{\mathbb{P}} 0 \quad \to 1, \tag{A.4}
\]
and
\[
P \left( \sup_{\Delta \geq 2\Delta} \sum_{i=r_0+1}^{n} \left( l(\hat{\Psi}_{n(0)}^{(i+1)}) - l(\hat{\Psi}_{n(0)}^{(i)}) \right) \right) \times I(\hat{\theta}_{n(0)}^j - \epsilon < X_{i-d_n} \leq \hat{\theta}_{n(0)}^j) \xrightarrow{\mathbb{P}} 0 \quad \to 1, \tag{A.5}
\]
where for \( \Psi = (\sigma^2, \Phi) \), \( l(\Psi) = \log(\sigma^2) + (X_i - \Phi X_{i-1})^2/\sigma^2 \).

Finally, the following proposition shows that each threshold can be identified by a threshold estimate around its \( n^{-1/2} \) neighborhood. Thus, each estimated regime must be nested in one of the true regime asymptotically. Also, the estimated AR coefficients converge to the true values at a \( n^{-1/2} \) rate.

**Proposition A.5.** Suppose that Assumption 1 holds. Then,
\[
\max_{i=1, \ldots, r_0} \min_{j=1}^{r_0} \left| \hat{\theta}_n(0)^j - \theta_0(0)^j \right| = O_p \left( \sqrt{n} \right) \tag{A.6}
\]
\[
\left| \hat{\Psi}_n(0)^j - \Psi(0)^j \right| = O_p \left( \sqrt{n} \right), \tag{A.7}
\]
for \( i = 1, \ldots, r_0 + 1 \) and some \( s \geq 0 \), where \( i^* \) is the index such that the \( i^* \)-th piece of estimated regime is nested inside the \( i \)-th true regime.

**Proof of Theorem 1**

From Propositions A.3 and A.5, we know that each of the true thresholds is identified by an estimated threshold in a \( n^{-1/2} \) neighborhood, and the AR orders will not be underestimated. Therefore, we only need to show that the number of thresholds or the AR orders are not overestimated. That is, it suffices to prove that for any integer \( s = 1, \ldots, R - r_0 \) and any sequence \( \theta_0 = (\theta_0(1), \ldots, \theta_0(s)) \) such that \( \theta_0(0)^j - \theta_0(0)^j \) is \( O(n^{-1}) \) for \( j = 1, \ldots, r_0 \), the quantity
\[
\min_{\theta_0(n)} \frac{2}{n} \text{MDL}(r_0 + s, \theta_0, p_0) = \frac{2}{n} \text{MDL}(r_0, \hat{\theta}_n, p_0), \tag{A.8}
\]
is positive with probability approaching 1, where MDL denotes the MDL evaluated at the true AR coefficients \( \Psi(0)^j \) instead of the maximum likelihood estimators, and \( p = (p(1), \ldots, p(r_0 + s + 1)) \) is any integer valued vector such that \( p(0)^j \geq \hat{p}(0)^j \) if \( (\theta_0(0)^j, \theta_0(0)^j) \leq (\hat{\theta}_n(0)^j, \hat{\theta}_n(0)^j) \). Let \( \hat{\theta}_n = (\hat{\theta}_n(1), \ldots, \hat{\theta}_n(s + 1)) \) be the minimizer for the first term in (A.8). Note that \( \hat{\theta}_n \in \hat{\theta}_n \) by construction. Using Taylor’s series expansions on the likelihood function, for sufficiently large \( n \), the quantity in (A.8) can be expressed as
\[
\frac{2}{n} (\text{Pen}(r_0 + s, p) - \text{Pen}(r_0, p_0)) + \frac{1}{n} \left( \sum_{i=r_0+1}^{r_0+s+1} L_n(\Psi(0)^i, \hat{\theta}_n(i-1), \hat{\theta}_n(i), p(i), d_0) \right) - \frac{1}{n} \left( \sum_{i=r_0+1}^{r_0+s+1} L_n(\Psi(0)^i, \hat{\theta}_n(i-1), \hat{\theta}_n(i), p(i), d_0) \right) \tag{A.9}
\]
\[
+ \frac{1}{n} \left( \sum_{i=r_0+1}^{r_0+s+1} \left( \left| \Psi_i(0)^i - \Psi_i(0)^i \right| \right) \right) \tag{A.10}
\]
where \( \left| \Psi(0)^i - \Psi(0)^i \right| < \left| \Psi(0)^i - \Psi(0)^i \right| \). Note from Proposition A.3 that \( 2(\text{Pen}(r_0 + s, p) - \text{Pen}(r_0, p_0)) \) is positive and of order \( O(\log n/n) \). Using Proposition A.2, it follows that the quantity in (A.9) is exactly 0 since \( (\Psi(0)^i, \hat{\theta}_n(i-1)) \) and \( (\Psi(0)^i, \hat{\theta}_n(i)) \) specify the same model if \( (\theta_0(0)^j, \theta_0(0)^j) \leq (\hat{\theta}_n(0)^j, \hat{\theta}_n(0)^j) \) and \( p(0)^j \geq \hat{p}(0)^j \). Also, from Proposition A.5, the term in (A.10) is of order \( O(n^{-1}) \). Therefore, \( 2(\text{Pen}(r_0 + s, p) - \text{Pen}(r_0, p_0))/n \), which is of order \( O(\log n/n) \), dominates the expression and the quantity in (A.8) is indeed positive with probability approaching 1. This completes the proof of Theorem 1.

**SUPPLEMENTARY MATERIALS**

S.1: Additional Simulations: TAR Model with One Structural Break
S.2: Proof of Theorem 2
S.3: Proofs of Propositions

[Received September 2013. Revised July 2014.]
REFERENCES

Alba, E., and Troya, J. M. (1999), “A Survey of Parallel Distributed Genetic Algorithms,” Complexity, 4, 31–52. [1178]
——— (2002), “Improving Flexibility and Efficiency by Adding Parallelism to Genetic Algorithms,” Statistics and Computing, 12, 91–114. [1178]

Aue, A., Hörmann, Horváth, L., and Reimherr, M. (2009), “Break Detection in the Covariance Structure of Multivariate Time Series Models,” The Annals of Statistics, 37, 4046–4087. [1176]

Berkes, I., Horváth, L., Ling, S., and Schauer, J. (2011), “Testing for Structural Change of AR Model to Threshold AR Model,” Journal of Time Series Analysis, 32, 547–565. [1176]

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. [1175,1178,1183,1185]

Coakley, J., Fuertes, A.-M., and Pérez, M.-T. (2003), “Numerical Issues in Threshold Autoregressive Modeling of Time Series,” Journal of Economic Dynamics and Control, 27, 2219–2242. [1175]

Davis, R. A., Lee, T. C. M., and Rodriguez-Yam, G. A. (2006), “Structural Break Estimation for Nonstationary Time Series Models,” Journal of the American Statistical Association, 101, 223–239. [1176,1177,1178,1179]

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 Time Series Analysis, 29, 834–867. [1176,1177,1178,1179]

Davis, R. A., and Yau, C. Y. (2013), “Consistency of Minimum Description Length Model Selection for Piecewise Stationary Time Series Models,” Electronic Journal of Statistics, 7, 381–411. [1180]

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. [1175]

Hansen, M. H., and Yu, B. (2001), “Model Selection and the Principle of Minimum Description Length,” Journal of the American Statistical Association, 96, 746–774. [1176]

Hansen, M. H., and Yu, B. (2001), “Model Selection and the Principle of Minimum Description Length,” Journal of the American Statistical Association, 96, 746–774. [1176]

Lavielle, M., and Ludena, C. (2000), “The Multiple Change-Points Problem for the Spectral Distribution,” Bernoulli, 6, 845–869. [1176]

Lee, T. C. M. (2001), “An Introduction to Coding Theory and the Two-Part Minimum Description Length Principle,” International Statistical Review, 69, 169–183. [1176,1177]

Li, D., and Ling, S. (2012), “On the Least Squares Estimation of Multiple-Regime Threshold Autoregressive Models,” Journal of Econometrics, 167, 240–253. [1175,1177,1178,1183]

Lu, Q., Lund, R., and Lee, T. C. M. (2010), “An MDL Approach to the Climate Segmentation Problem,” The Annals of Applied Statistics, 4, 299–319. [1177]

Rissanen, J. (1989), Stochastic Complexity in Statistical Inquiry, Singapore: World Scientific. [1176]

Rissanen, J. (2007), Information and Complexity in Statistical Modeling, New York: Springer-Verlag. [1176]

Tong, H. (1978), “On a Threshold Model,” in Pattern Recognition and Signal Processing. NATO ASI Series E: Applied Sc. (29), ed. C. Chen, The Netherlands: Sijthoff & Noordhoff, pp. 575–586. [1175]

——— (1990), Non-Linear Time Series: A Dynamical System Approach, Oxford: Oxford University Press. [1175]

——— (2011), “Threshold Models in Time Series Analysis—30 Years On,” Statistics and its Interface, 4, 107–118. [1175]

Tsai, R. S. (1989), “Testing and Modeling Threshold Autoregressive Processes,” Journal of the American Statistical Association, 84, 231–240. [1175,1178]

Zhu, K., and Ling, S. (2012), “Likelihood Ratio Tests for the Structural Change of an AR(p) Model to a Threshold AR(p) Model,” Journal of Time Series Analysis, 33, 223–232. [1176]