Use of fuzzy statistical technique in change periods detection of nonlinear time series

Berlin Wu [a,*], Mei-Hui Chen [b]

[a] *Department of Mathematical Sciences, National Chengchi University,*
Wen-Shen District, Taipei, Taiwan, ROC
[b] *Department of Statistics, National Chengchi University, Taipei, Taiwan, ROC*

Abstract

Many papers have been presented on the study of change points detection. Nonetheless, we would like to point out that in dealing with the time series with switching regimes, we should also take the characteristics of change periods into account. Because many patterns of change structure in time series exhibit a certain kind of duration, those phenomena should not be treated as a mere sudden turning at a certain time. In this paper, we propose a procedure about change periods detection for nonlinear time series. The detecting statistical method is an application of fuzzy classification and a generalization of Inclan and Tiao's result [J. Am. Statist. Assoc. 89 (1994) 913]. Simulation results show that the performance of the proposed procedure is efficient and successful. Finally, an empirical application about change periods detecting for Taiwan monthly visitor's arrival is demonstrated.   © 1999 Elsevier Science Inc. All rights reserved.

Keywords: Change periods; Nonlinear time series; Revised centered cumulative sums of squares (RCUSUM); Fuzzy statistics

1. Introduction

Traditional methods on the model construction for a time series are based on the Bayesian experience by choosing a "good" model, which will satisfactorily explain its dynamic behavior, from a model-base. But a fundamental question that is often asked is: Do there exist switching regimes in the series?

* Corresponding author. E-mail: berlin@math.nccu.edu.tw.

Can a single model fit the dynamic all the time? If there exists significant changes for the underlying time series, it seems natural to find out those change points or change periods before modeling the whole process.

The investigation of change points detection can be found in many papers. For example, Tsay (1986, 1990) proposed some procedures for detecting outliers, level shifts and variance changes in a univariate time series. The procedures he outlined are particularly useful and relatively easy to implement. While Balke (1993) pointed out that Tsay's procedures do not always perform satisfactorily when level shifts are present. Inclan and Tiao (1994) proposed an iterative procedure to detect variance changes based on a centered version of the cumulative sums of squares presented by Brown et al. (1975).

Those testing statistics dealing with change points detection include: MPAGE (Modified PAGE) proposed by Page (1955), CUSUM (Cumulative Sum) proposed by Hinkey (1971). Hsu (1977, 1979, 1982) investigated the detection of a variance shift at an unknown point in a sequence if independent observations, focusing on the detection of points of change one at a time because of the heavy computational burden. Menzefricke (1981) presented a Bayesian analysis of a change at a unknown time point. Wosley (1986) used ML methods to test a change in mean for a sequence of independent exponential family random variables. Saatri et al. (1989) presented a study of performance comparison for six time-series change detection procedures. Barry and Hartigan (1993) demonstrated a Bayesian analysis for change point problems.

However, those detecting techniques are based on the assumption that the underlying time series conducts a significantly change point characteristic. Nevertheless, we must indicate that in dealing with the time series with switching regimes, we had better not only consider the change points detection, but also take into account the properties of change periods. Because many patterns of change structure in time series exhibit a certain kind of duration, those phenomena should not be treated as a mere sudden turning at a certain time. For instance: (i) The exchange rate may go up or down gradually after a new financial policy perform. (ii) The supply of M_1 or M_2b may change their trend at different period of time according to the national economic conditions. In fact, the semantics of the term "change point" is not very clear and well understandable (interesting readers may refer to any popular dictionary such as Webster's New Dictionary or Oxford Comtemporary Dictionary).

Hence, in order to meet the real situation, it had been better to employ the concept of "change period" instead of "change point". In this paper, we propose a procedure about change periods detection for nonlinear time series. This paper is organized as follows. In Section 2, we develop three detecting methods and present the detection procedure with fuzzy logic to identify change periods for nonlinear time series. The performance of the proposed

statistics is investigated by means of simulations in Section 3. In addition, we applied the detecting algorithm to discriminate the real case of Taiwan monthly visitor's arrival in Section 4. Section 5 gives the conclusion and suggestions.

2. Detecting statistics and procedure

2.1. Statistics for centered cumulative sums of squares

Kao and Ross (1995) proposed a modified CUSUM test, which was asymptotically significant for structural change with an unknown change point in a linear model with serially correlated disturbances. Their idea was based on the consideration of posterior recursive residuals. In other words, it was primarily assumed that there exists a model with the estimable parameters that can be found out by statistical or numerical methods. However, in a nonlinear time series analysis, there are various types of nonlinear model families. It is thus very difficult for us to decide which model will we apply in the following model fitting process. Especially, when the pattern of a nonlinear time series changed significantly after certain time.

In order to identify the switching regimes for a nonlinear time series, we propose a revised centered CUSUM test, which is a revision of an iterated centered cumulative sums of squares presented by Inclan and Tiao (1994), and will be used in an identification procedure for multiple change points/periods in a nonlinear time series process.

Consider a general nonlinear time series model

$$
y_{t-m_{j-1}} = \sum_{j=1}^{r+1} I_j [h_j (y_{t-m_{j-1}-1}, y, \ldots, y_{t-m_j+1}, \varepsilon_{t-m_{j-1}-1}, \varepsilon_{t-m_{j-1}-2}, \ldots, \varepsilon_{t-m_j+1})
$$
$$
+ \varepsilon_{t-m_j-1}], \tag{2.1}
$$

where $m_0 = 0$, ε_{t_j} is a white noise process, i.e. a sequence of independently distributed random variables with mean μ_j and variance σ_j^2, h_j is a (nonlinear) real valued function for specified j ($j = 1, 2, \ldots, r + 1$), and

$$
I_j = \begin{cases} 1 & \text{if } y_t \in \{y_{t-m_{j-1}}, y_{t-m_{j-1}-1}, \ldots, y_{t-m_j+1}\} \\ 0 & \text{elsewhere} \end{cases} \quad j = 1, 2, \ldots, r + 1. \tag{2.2}
$$

The revised CUSUM statistic defined below will be used to detect multiple thresholds $\{m_j\}$ for the considered time series.

Definition 2.1 (*RCUSUM*). Let $\{y_t, t = 1, 2, \ldots, N\}$ be a time series, $\bar{y}_k = \sum_{t=1}^{k} y_t / k$, $k = 1, 2, \ldots, N$, and $CC_k = \sum_{t=1}^{k} (y_t - \bar{y}_k)^2$ be a centered cumulative sum of squares. Then

$$CD_k = \frac{CC_k}{CC_T} - \frac{k}{T}, \quad k = 1, 2, \ldots, N, \tag{2.3}$$

is called Revised Centered Cumulative Sums of Squares (RCUSUM).

The plot of CD_k against k will oscillate around 0 for series with homogeneous mean. When there is a change in mean, the plot of CD_k will exhibit a pattern with upward or downward trend going out of some specified boundaries with high probability. That is, the slope of CD_k takes a drastic change in trend. It leads to a peak or a trough according to the changes of the structure of the series. The use of CD_k will help us to detect a priori change periods different from that of CC_k and in Section 3 we will give more detailed discussion about the behavior of CD_k.

While in some cases, especially when the underlying time series contains outliers, the performance of RCUSUM statistic might not be sensitive and efficient for change periods. Several extreme observations may just exhibit a jump. It will return to the historical behavior after the intervention is over. For nonparametric reasons, it is reasonable to transform the original series into an ordered one to avoid misspecifying the outliers or sudden changes as change points/periods. Therefore we will also use the statistics CDR_k defined at Definition 2.2 to detect the change points or periods.

Definition 2.2 (*ORCUSUM*). Let $\{y_t, \ t = 1, 2, \ldots, N\}$ be a time series. Rank $\{y_t\}$ in order to magnitude by assigning 1 to the smallest observation in $\{y_t\}$, 2 to the second smallest observation, and so on. Denote the new series as $\{z_t\}$. Let $\bar{z}_k = \sum_{t=1}^{k} z_t/k$ $(k = 1, 2, \ldots, N)$, and $CCR_k = \sum_{t=1}^{k} (z_t - \bar{z}_k)^2$ be the revised centered cumulative sums of squares of ordered time series $\{z_t\}$. Then

$$CDR_k = \frac{CCR_k}{CCR_T} - \frac{k}{T}, \quad k = 1, 2, \ldots, N, \tag{2.4}$$

is called as the (Ordered RCUSUM) (ORCUSUM) of $\{y_t\}$.

2.2. Detection procedures for change periods

The procedures about switching regimes for a time series are based on the threshold value identified by the RCUSUM or ORCUSUM statistic. We use the unsupervised method to find how many changes this time series by the plot of CD_k or CDR_k against time k. Suppose there exist r peaks or troughs in the plot of CD_k or CDR_k against k.

2.2.1. Rules for classification

Let $\Delta CD_k = CD_k - CD_{k-1}$, be the indifference between CD_k and CD_{k-1} for $k = 2, \ldots, N$. Given λ-level confidence limits for ΔCD_k, based on the threshold values of a set of classifiers C_k of ΔCD_k, we classify the time series as follows:

$$C_k = \begin{cases} 0 & \text{if } \Delta\text{CD}_k \in (-\infty, L_\lambda), \\ 1 & \text{if } \Delta\text{CD}_k \in [L_\lambda, U_\lambda], \\ 2 & \text{if } \Delta\text{CD}_k \in (U_\lambda, \infty), \end{cases} \tag{2.5}$$

where L_λ and U_λ stand for the confidence limits for a significant level λ. Usually we take L_λ and U_λ as the two standard deviation confidence limits for ΔCD_k, i.e. $L_\lambda = $ mean of $\Delta\text{CD}_k + 2\sigma_{\Delta\text{CD}_k}$ and $U_\lambda = $ mean of $\Delta\text{CD}_k + 2\sigma_{\Delta\text{CD}_k}$.

2.2.2. Identification of change periods

To identify the change periods, we need the following definitions.

Definition 2.3 (*vague points*). For a time series $\{y_t, t = 1, 2, \ldots, N\}$. Let $\{A_i, i = 1, 2, \ldots\}$ be a set with each A_i containing a run for the category C_i, i.e. A_i is an unbroken sequence of the same classifiers. Let $\#(A_i)$ be the length of A_i. Given λ-level and a prior number of change periods r, we say that the elements of A_i are vague points if $\#(A_i) < [N\lambda]$ and $\#(A_i)/\#(A_{i-1}) \leqslant 4$. Where $[x]$ stands for the greatest integer which is less than or equal to x, N is the number of observations and $r + 1$ is the total number of sub-series.

In order to classify the vague points, we use the following process and assign them into a closest set of A_i: let V be the set of vague points for time series $\{y_t, t = 1, 2, \ldots, N\}$. If $\#(A_i)/\#(A_{i-1}) < 1$ for $|i - j| = 1$, we join the A_i into A_{i-1} and rearrange this new set sequence as $\{A_i^*, i = 1, 2, \ldots\}$, i.e. A_i^* is the ith set with the same elements after clustering of vague points.

Examples 2.1. Let $A = \{00000000011000000000010111111111\ldots\}$. Then the corresponding sets $\{A_i\}$ will be $A_1 = \{000000000\}$, $A_2 = \{11\}$, $A_3 = \{0000000000\}$ and $A_4 = \{1\}\ldots$ Suppose $N = 80$ and $\lambda = 0.1$. It follows that $\#(A_2) = 2 < 5$ and $\#(A_1)/\#(A_2) = 4.5$, which is greater than 4. Hence we say that the elements in set A_2 are vague points and joint A_2 into A_1. That is, the elements in A_2 are 0 after vague point clustering $\#A_1^* = 21$.

Definition 2.4 (*change periods*). Let $\{y_t\}$ be a time series and k a pseudo change point. A significant α-level ($0 \leq \alpha < 1$) change period for specified k is defined as

$$[k - \alpha N, k + \alpha N]. \tag{2.6}$$

Consider the testing hypothesis for

H_0:   There is no change during time interval (t_i, t_j) (2.7)

versus

H_a: There exists a change period during time interval (t_i, t_j),   (2.8)

where $t_j - t_i \geqslant 0$ and $(t_i, t_j) \in \{(t_i, t_j) \mid t_i, t_j = 1, 2, \ldots, N\}$.

Consider the rearranged run sets A_i^*, A_{i+1}^* and A_{i+2}^*. Suppose the corresponding time sequence for sets A_i^*, A_{i+1}^* and A_{i+2}^* are $\{y_{t_{i-1}+1}, \ldots, y_{t_i}\}$, $\{y_{t_i+1}, \ldots y_{t_{i+1}}\}$ and $\{y_{t_{i-1}+1}, \ldots, y_{t_{i+2}}\}$, respectively. If the number of these two consecutive sets A_i^*, and A_{i+1}^*, is greater than or equal to $[\alpha N]$ for level α, then we conclude that there exists a sudden change at time point t_i or $t_i + 1$. Given α, if $\# A_{i+1}^*$ is less than $[\alpha T]$, and the number of elements of the others are greater than or equal to $[\alpha T]$, then we conclude that there exists a change period during time interval $(t_i, t_{i+1} + 1)$. Choose Med $\{t_i + 1, t_i + 2, \ldots, t_{i+1}\}$ to be the pseudo change point k in formula (2.6), and the $100(1 - \alpha)\%$ change periods will then be defined.

An α-level identification algorithm is suggested as follows:

Algorithm for Detecting Change Periods
1. Given a time series $\{y_t\}$. Calculate the RCUSUM statistics CD_k or ORCUSUM statistics CDR_k. Plot CD_k or CDR_k against k and initially identify the prior number of change periods r.
2. Classify $\{y_t\}$ according to the value of ΔCD_k
3. Rearrange the series $\{A_i, i = 1, 2, \ldots\}$ as $\{A_i^*\}$ at the λ-level.
4. From $\{A_i^*\}$ we can clearly decide the change periods at the significant α-level.

2.3. Using fuzzy entropy

Fuzzy set theory, which was first proposed by Zadeh (1965), has received much attention recently. It has fruitful achievements not only theoretically but also in applications. For instance, the classical clustering methods separated the data to c categories, while in many cases there are elements, which cannot be contained in a specific category. They belong to two or more categories simultaneously. Applying fuzzy statistical detection techniques may constitute a new trial for this problem. Wu and Hwang (1995) are two of the pioneers that proposed a detecting procedure for the α-level of fuzzy change period classification.

In this section, the proposed method for detecting change periods may be applied by using the fuzzy entropy. The term "entropy" comes from thermodynamics. Entropy can be thought of as a measure of how close a system is to equilibrium; it can also be thought of as a measure of disorder in the system. Before proceeding with the procedure, we should illustrate some concepts of fuzzy cluster centers, α-level of fuzzy point, and fuzzy entropy.

Definition 2.5 (*fuzzy cluster centers*). Let $\{y_t, t = 1, 2, \ldots, n\}$ be a time series. Giving positive integer k, if the set $C = \{C_i; \min\{y_t\} \leq C_i \leqslant \max\{y_t\},$ $i = 1, 2, \ldots, k\}$ is to minimize the sum of squared Euclidean distance, i.e.

$$\min \sum_{i=1}^{k} \sum_{\mu_{it} \in C_i} \|\mu_{it} - C_i\|^2, \tag{2.9}$$

where μ_{it} measures the degrees of y_t belonging to each cluster C_i. Then we call $C = \{C_i, i = 1, 2, \ldots, k\}$ a set of Fuzzy cluster centers for the time series $\{y_t\}$.

Definition 2.6 (*α-level of fuzzy change point*). Let $\{y_t, \ t = 1, 2, \ldots, N\}$ be a time series. Suppose we decompose $\{y_t\}$ into k clusters according to a set of fuzzy cluster centers; where μ_{it} is the degree of membership that y_t holds for cluster center C_i. For each $y_t, t = 1, 2, \ldots, N$, if

$$\max\{\mu_{it}\} - \min\{\mu_{it}\} < 1 - k\alpha, \quad \alpha \in (0, 1/k), \tag{2.10}$$

then we call y_t an α-level of fuzzy change point. Moreover, if $\max\{\mu_{it}\} - \min\{\mu_{it}\} = 0$, we call y_t an absolutely fuzzy point. If $\max\{\mu_{it}\} - \min\{\mu_{it}\} = 1$, we call y_t a crisp point.

Definition 2.7 (*fuzzy entropy*). Let μ_{it} be the membership of y_t to the cluster $C_i, i = 1, 2, \ldots, k$. The fuzzy entropy of y_t is defined as

$$\delta(y_t) = -\frac{1}{k} \sum_{i=1}^{k} \mu_{it} \ \ln(1 - \mu_{it}). \tag{2.11}$$

The proposed fuzzy detecting procedures of fuzzy change periods with α-level include (1) fuzzy-clustering, (2) deciding α-level of fuzzy point, (3) detecting α-level of fuzzy change periods. A detailed algorithm is illustrated below:

1. Input the time series $\{y_t\}$. Find C_i $(i = 1, 2, 3)$, the set of cluster centers for $\{y_t\}$. Classify $\{y_t\}$ into three categories.
2. Let μ_{it} be the degree of membership of each observation y_t to each cluster C_i. Compute the membership of μ_{it} by

$$\mu_{it} = 1 - \frac{(y_t - C_{it})^2}{\sum_{i=1}^{3}(y_t - C_{it})^2}, \quad i = 1, \ldots, 3; \ t = 1, \ldots, N. \tag{2.12}$$

3. Calculate the fuzzy entropy of y_t by using its memberships:

$$\delta(y_t) = -\frac{1}{3} \sum_{i=1}^{3} \mu_{it}(1 - \mu_{it}). \tag{2.13}$$

4. Calculate the average of cumulative fuzzy entropy for each t by:

$$\mathrm{MS}(\delta(y_t)) = \frac{1}{t} \sum_{k=1}^{t} \delta(y_t), \quad t = 1, 2, \ldots, N, \tag{2.14}$$

and find Med $(\mathrm{MS}(\delta(y_t)))$, the median of $\mathrm{MS}(\delta(y_t))$.

5. Choose a proper λ threshold level and classify the fuzzy time series as:

$$C_i^* = \begin{cases} 0 & \text{if Med } (\text{MS}\delta(y_i)) \in [0, \text{Med } (\text{MS}\delta(y_i)) - \lambda], \\ 1 & \text{if Med } (\text{MS}\delta(y_i)) \in [\text{Med } (\text{MS}\delta(y_i)) - \lambda, \\ & \qquad\qquad\qquad \text{Med } (\text{MS}\delta(y_i)) + \lambda], \\ 2 & \text{if Med } (\text{MS}\delta(y_i)) \in [\text{Med } (\text{MS}\delta(y_i)) + \lambda, 1]. \end{cases} \qquad (2.15)$$

6. Decide the α-significant level of change periods; i.e., for each categories 0, 1 and 2, if the elements of category 1 contain sample points (successive data) greater than $[\alpha N]$, then we reject the hypothesis that there is no structural change.

3. Simulation study

Threshold autoregressive (TAR) time series can be a typical nonlinear time series for switching regimes. A general TAR model for a time series $\{y_t\}$, TAR $(r; t_1, t_2, \ldots, t_{r-1})$ can be written as:

$$y_{t_j} + I_j(\phi_0^{(j)} + \sum_{i=1}^{t_j - t_{j-1} - 1} \phi_i^{(j)} y_{t_j - i}) = \varepsilon_{t_1}^{(j)}, \quad j = 1, 2, \ldots, r, \qquad (3.1)$$

where

$$I_j = \begin{cases} 1, & \text{if } y_t \in \{y_{t_j}, y_{t_j - 1}, \ldots, y_{t_{j-1} + 1}\}, \\ 0, & \text{elsewhere}, \end{cases} \quad j = 1, 2, \ldots, r, \qquad (3.2)$$

and the $\varepsilon_{t_j}^{(j)}$ are a strict white noise process with finite variance σ_ε^2. The model has many applications in various fields, such as control of birth rate, stock market index, exchange rate or GDP, etc.

3.1. Two TAR models simulation

The purpose of this study is to investigate the efficiency of our statistics proposed in Section 2. We simulate two sets of $\text{TAR}(3; t_1, t_2)$ time series process with models (3.3) and (3.4),

$$y_t = \begin{cases} 10 + 0.8y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} & \text{if } t < 51, \\ 10 + 0.5y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1} & \text{if } t = 51, 52, \ldots, 100, \\ 10 + 0.2y_{t-1} + \varepsilon_t - 0.2\varepsilon_{t-1} & \text{if } t > 100, \end{cases} \qquad (3.3)$$

$$y_t = \begin{cases} 10 + 0.8y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} & \text{if } t < 51, \\ 8 + 0.8y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} & \text{if } t = 51, 52, \ldots, 100, \\ 6 + 0.8y_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} & \text{if } t > 100. \end{cases} \qquad (3.4)$$
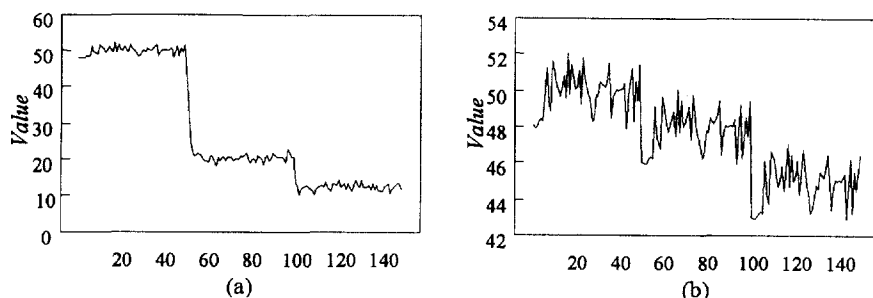
Fig. 1. Simulations for TAR models: (a) model (3.3); (b) model (3.4).

Each set contains 150 observations with nonzero mean and has trend components. The normally distributed innovation $\varepsilon_t \sim N(0,1)$ was generated by using Minitab 10.2 on a 586-PC. Fig. 1(a) and 1(b) are realizations of these two TAR models, respectively.

3.2. Change periods detection and discussion

In the performance-comparison tests of the simulation study, four change detection procedures were implemented. Sequences of the test statistics computed by these programs were output to Minitab data files for subsequent analyses. To avoid the effects of initialization transients, our detection performance analysis did not process the first observations.

Fig. 2(a) and 2(b) illustrate the series of CD_k and RCD_k for model (3.3). Fig. 2(c) and 2(d) illustrate the series of CD_k and RCD_k for model (3.4). From the plots we can find that both CD_k and RCD_k show a peak or a trough corresponding to the changing points/periods.

A comparison with four change detection statistics, CUSUM, RCUSUM, ORCUSUM and ACFE (Average of Cumulative Fuzzy Entropy) is tabulated in Table 1.

Discussion:

(i) The RCUSUM detecting statistics demonstrate the best accuracy in detecting change periods than those of CUSUM, ORCUSUM and fuzzy entropy methods for TAR model with the parameters change, model (3.4). While the ORCUSUM detecting statistics demonstrates the best accuracy in detecting change periods than those of CUSUM, RCUSUM and fuzzy entropy methods for TAR model with the scale change in model (3.4).

(ii) It seems that there is not much difference in detecting change periods between RCUSUM and ORCUSUM methods for the series simulated here. For the TAR model with huge parameter change or scale change, the ORCUSUM methods exhibit a more robust method for identifying multiple change periods.
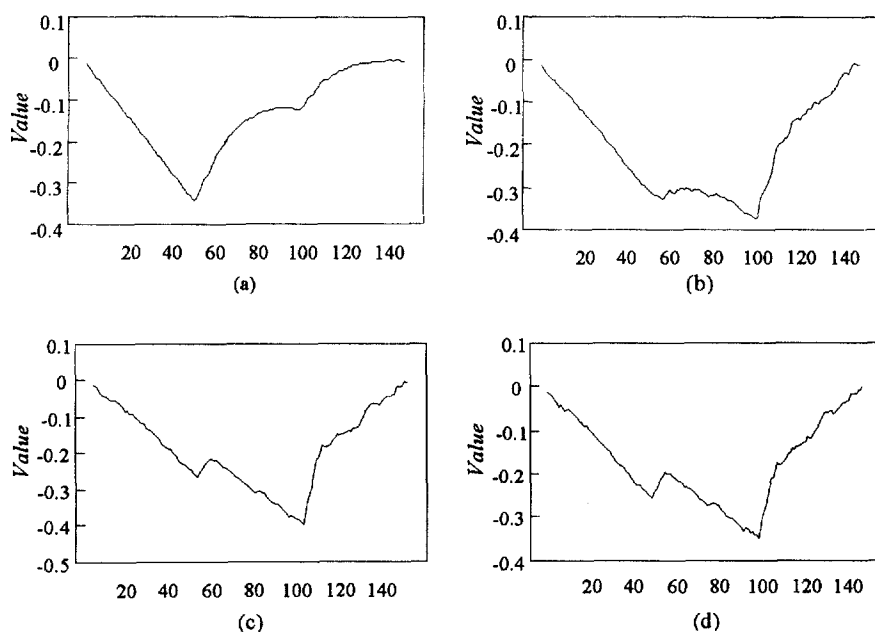
Fig. 2. Revised centered cumulative sums of squares plots: (a) CD_k, (b) RCD_k for model (3.3); (c) CD_k, (d) RCD_k for model (3.4).

Table 1
Comparison results for four detections

	TAR with parameter change: Model (3.3)	TAR with scale change: Model (3.4)
Real values	50, 100	50, 100
CUSUM	51	53
RCUSUM	48–55, 96–103	46–54, 96–104
ORCUSUM	54–63, 96–104	46–53, 98–102
ACFE	58–61, 133–135	49–53, 102–109, 119–124

(iii) It is more realistic for us to consider that a time series is changing during a period than at a time point, in view of the behaviors of nonlinear economic series. The CUSUM statistic still could detect changes for correlated series, it only could detect one change at a time.

(iv) The ACFE detecting method performs well here, we find that the detected change periods are all a little delayed comparing to the real values. The facts maybe result in the characteristic of fuzzy entropy for its slow reflection at the logarithm function.

(v) It liberates us from the modeling-based selection procedure and fewer assumptions of the sample data are made. The weakness of the change point philosophy clearly resides in the vague semantic agreements for the mutual understanding in the real world.

4. An empirical example

4.1. Identification of change-periods for Taiwan monthly visitor arrivals

As an example, we apply the detection procedures to a real data set. The series analyzed is the Taiwan monthly visitor's arrival from January 1971 to June 1993. Fig. 3 plots the trend of the Taiwan monthly visitors' arrival. Fig. 4(a) and 4(b) illustrate the standard CUSUM, developed by Inclan and Tiao (1994), and RCUSUM statistics, respectively.
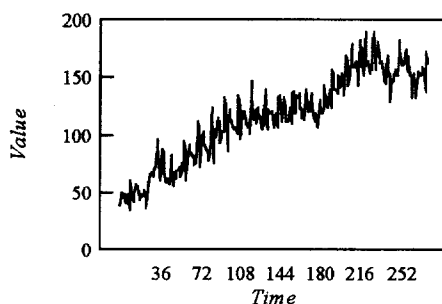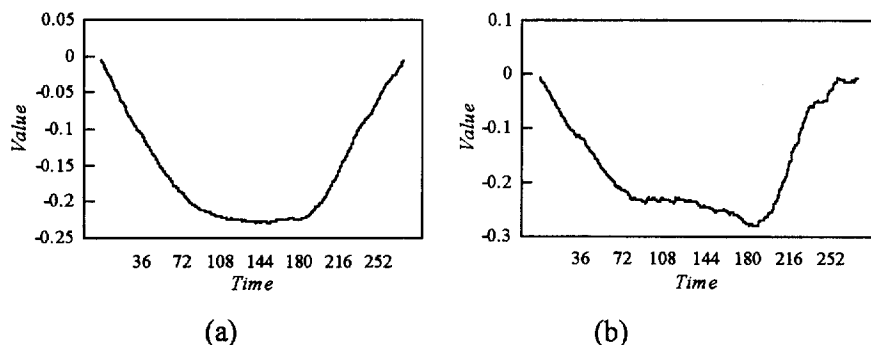

Fig. 3. Taiwan monthly visitors arrival.

(a)   (b)

Fig. 4. Cumulative sums of squares. (a) CUSUM; (b) RCUSUM.

According to Section 2.2, we calculate the RCUSUM statistic CD_k for the series of 276 observations. From Fig. 4(b), we initially identify that the prior number of change-periods is two, that is, $r = 2$. Take ∇CD_k to be the indifference between CD_k and CD_{k-1} for $k = 2, \ldots, 276$, and determine that $[-0.04, 0.04]$ is the two standard deviation confidence interval for ∇CD_k.

Classify the Taiwan monthly visitors arrival series, based on the threshold values of a set of classifiers C_k of ∇CD_k. Notice that the C_k are obtained as follows:

$$C_k = \begin{cases} 0 & \text{if } \nabla CD_k \in (-\infty, -0.04), \\ 1 & \text{if } \nabla CD_k \in [-0.04, 0.04], \\ 2 & \text{if } \nabla CD_k \in (0.04, \infty). \end{cases}$$

Let $\{A_i, i = 1, 2, \ldots\}$ be a consecutive set such that each set contains the same classifiers. Consider sets with number of elements less than or equal to three. Clustering those sets containing vague points by the procedures described in Section 2 and we identify (a) 1978/1 to 1978/4 and (b) 1986/4 to 1986/9 to be our initial change-periods. Hence 1978/2 and 1986/6 are the corresponding pseudo change points.

Finally we find out two change-periods, which are (a) October 1977 to May 1978 and (b) March 1986 to September 1986, under the $\alpha = 0.05$ level of significance for detecting change-periods for Taiwan monthly visitors' arrival.

5. Conclusions

In this paper, we introduced three detecting procedures that can effectively detect multiple change periods for a nonlinear time series. The proposed algorithm also combines with the concept of fuzzy set. We have demonstrated how to find a α-level change period to help modeling a time series model with multiple change periods.

Simulation results show that our proposed techniques of change period detection are very efficient. Our algorithm is highly recommended practically in detecting the α-level change period, which is supported from the empirical results. A major advantage of such framework is that our detecting procedures do not require any initial knowledge about the structure in the data and can take full advantage of the model-free approach.

From the above, it will be evident that the art of identification and classification of time series is still at the stage of infancy.

Certain challenging problems still remain open, such as:

1. *Problems related to change periods*

(a) The semantics about the term "stationarity" needs to be redefined carefully. It seems that the term "change points", which may stand for the

change in mean, the change in variance, the change in parameter and the change in model, needs to be classified before performing a detecting process.

(b) In the case of random walk, an appropriate detecting procedure needs to be developed for prior recognition and model identification.

(c) What knowledge basis is required to obtain specific behavior of time series under certain multivariate endogenous variables.

(d) The convergence of the algorithm for classification and the proposed statistics have not been well proved, although the algorithms and the proposed statistics are known as fuzzy criteria. This needs further investigation.

2. *Problems related to noises and interventions*

(a) In what way does the introduction of feedback affect the trend of the time series? In particular, how far does the time series react to noise and intervention?

(b) For the case of interacting noise, the complexity of multivariate filtering problems still remains to be solved.

(c) How to obtain information from chaotic trajectories about the nature trend of the time series.

Though there remain many problems to be overcome, we think that fuzzy statistical methods will be a worthwhile approach and will stimulate more empirical work in the future in time series analysis.

References

Balke, N.S., 1993. Detecting level shifts in time series. J. Business Econom. Statist. 11 (1), 81–92.

Barry, D., Hartigan, J.A., 1993. A Bayesian analysis for change point problems. J. Amer. Statist. Assoc. 88 (421), 309–319.

Brown, R., Dubin, J., Evans, J., 1975. Techniques for testing the constancy of regression relationships over time. J. Roy. Statist. Soc. Ser. B 37, 149–163.

Hinkey, D.V., 1971. Inference about the change point from cumulative sum test. Biometria 26, 279–284.

Hsu, D.A., 1977. Tests for variance shift at an unknown time point. Appl. Statist. 26, 279–284.

Hsu, D.A., 1979. Detecting shifts of parameter in gamma sequences, with applications to stock price and air traffic flow analysis. J. Amer. Statist. Assoc. 74, 31–40.

Hsu, D.A., 1982. A Bayesian robust detection of shift in the risk structure of stock market returns. J. Amer. Statist. Assoc. 77, 29–39.

Inclan, C., Tiao, G.C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. J. Amer. Statist. Assoc. 89 (427), 913–924.

Kao, C., Ross, S.L., 1995. A CUSUM test in the linear regression model with serially correlated disturbances. Econometric Rev. 14 (3), 331–346.

Menzefricke, U., 1981. Presented a Bayesian analysis of a change at a unknown time point.

Page, E.S., 1955. A test for change in a parameter occurring at an unknown point. Biometricka 42, 523–527.

Saatri, T., Flores, B., Valdes, J., 1989. Detecting points of ahange in time series. Comput. Open Res. 16, 271–293.

Tsay, R.S., 1986. Nonlinearity tests for time series. Biometrics 73, 461–466.

Tsay, R.S., 1990. Testing and modeling threshold autoregressive processes. J. Amer. Statist. Assoc. 84, 231–240.

Wosley, K.J., 1986. Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika 73, 91–104.

Wu, B., Hwang, J., 1995. On fuzzy identification of nonlinear time series, Taiwan Economic Association, Annual Conference Proceedings, pp. 169–190.

Zadeh, L.A., 1965. Fuzzy sets, Inf. Control 8, 338–353.