

Fuzzy Estimations for Detecting Abrupt Changes: Cases on Tourism Series

Nurhaida^{1,2,*}, Subanar¹, Abdurakhman¹, Agus Maman Abadi³

¹Department of Mathematics, Universitas Gadjah Mada, Yogyakarta, 55281, DIY, Indonesia

²Department of Mathematics and Statistics, Universitas Papua, Manokwari, 98314, Papua Barat, Indonesia

³Department of Mathematics, Universitas Negeri Yogyakarta, Yogyakarta, 55281, DIY, Indonesia

Received October 16, 2020; Revised December 5, 2020; Accepted December 20, 2020

Cite This Paper in the following Citation Styles

(a): [1] Nurhaida, Subanar, Abdurakhman, Agus Maman Abadi, "Fuzzy Estimations for Detecting Abrupt Changes: Cases on Tourism Series," *Mathematics and Statistics*, Vol. 8, No. 6, pp. 747-761, 2020. DOI: 10.13189/ms.2020.080616.

(b): Nurhaida, Subanar, Abdurakhman, Agus Maman Abadi, (2020). *Fuzzy Estimations for Detecting Abrupt Changes: Cases on Tourism Series*. *Mathematics and Statistics*, 8(6), 747-761. DOI: 10.13189/ms.2020.080616.

Copyright ©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract This article deals with problems of detecting abrupt changes in time series by Change Point Model (CPM) framework. We propose a fuzzification in a Fuzzy Time Series (FTS) model to eliminate a trend in a contaminated dependent series. The independent residuals are then inputted on the CPM method. In simulating an abrupt change, an ARIMA(1,1,1) and variance of the model are considered. The abrupt change is modelled as an AO (Additive Outlier) type of outliers. The minimum weight or breaksize of the abrupt change is defined based on the ARIMA variance formulated in this article. The percentage of uncorrelated residuals obtained by the FTS model and the percentage of correct detection of the proposed procedure are shown by simulation. The proposed detecting algorithm is implemented to detect abrupt changes in monthly tourism series in literature, i.e., in Taiwan and in Bali. The first series shows a slowly increasing trend with one abrupt change while the second series exhibits not only a slowly increasing trend but also a strong seasonal pattern with two abrupt changes. For comparison, we detect the changes in the empirical examples on an existing automatic detection procedure using `tso` package in R. For the first example, the results show that both detecting procedures give exactly a similar location of one change point where the package recognises it as an AO type of outliers. The abrupt change is related to the period of SARS outbreak in Taiwan. On the second example, the proposed procedure locates 4 change points which form two locations of changes, i.e., the first two change points are within 2 time points so do the last two change points. The locations are closed to times of Bali Bombing events. Meanwhile, the automatic procedure recognizes only one AO outlier on the series.

Keywords FTS, CPM, ARIMA, Abrupt Changes, Outlier.

1 Introduction

Structural breaks in a time series statistically mean that there is at least a shift in their distributional parameters, e.g., in mean and/or in variance. Some certain U.S series show structural changes as a result of changes of various international events including the two Oil Price Shocks and in International Monetary System [1]. A decline trend of tourist arrivals following 2002 Bali Bombing had a short run-effect on Bali and other famous tourist destinations in Indonesia [2]. A good news, however, the effects of terrorist attacks in 2002 and 2005 on the number of tourist arrivals in Bali is transitory [3]. Still in tourism research, an abrupt change is one out of four distinct patterns in tourism series discussed in [4].

We define abrupt changes as stated in [5], "By abrupt changes, we mean changes in characteristics that occur very fast with respect to the sampling period of the measurements, if not instantaneously". Further, the authors imply that these considered changes occur at any instant times during the observations, treating, in some sense, that characteristics of the series are constants before and after the changes. Due to temporary effect, structural breaks in a time series are shown by changes only on some parts in the time series plot.

In time series analysis, the study of impacts of changes when the time of changes are undefined is known as outlier detection analysis. Commonly, the first step in the outlier detection is to model the series in the presence of outliers [6], [7]. The resid-

uals are then formed into several models of outliers known as outlier effects. One may argue that the standard deviation of the residuals may be overestimated as it is obtained from a contaminated series. In the latter work, the contaminated standard deviation of the residuals is reestimated. The standardized versions of the outlier effects follow an approximately normal distribution. Therefore, the existence of an outlier in each model is detected by using a critical value determined from the distribution. Having been all detected, the remaining series is a free outlier model. The process of detecting all possible outliers involving iterations as the effects of recently detected outliers are to be subtracted from the recent data set and then the residuals are reupdated.

Change point formulation was originally introduced by Page in 1954 to detect changes in density of an independent and identically distributed (i.i.d) sample. The formula, known as CUSUM, is closely connected to the maximum likelihood (ML) method. Under assumption on the series distribution, parametric change detection analysis had been developed in two directions, Bayesian and MLE methods. Readers refer to [5], [8] for parametric approach in change point detection. On the other hand, change point model (CPM) which is statistically simple and numerically effective was proposed [9]. A two-sample statistics with unknown parameters is employed and a recursive calculation of the statistics is used. Although, originally formulated to detect changes in mean of the Gaussian series in control quality processes, the method has been expanded into a set of tools to allow for detecting arbitrary changes with and without distributional assumption on the sample. The CPMs are available in the **cpm** package in R (R team core 2013) and the manual in [10].

The CPM method is based on the standard two-sample statistic test where the assumption for the input series is of i.i.d observations. It is said on the manual that for autocorrelated series, however, the detectin of changes can be performed either on the residuals of estimations or of one-step ahead predictions of a fit model inasmuch giving i.i.d residuals. A similar notion is given in [5], pg. 203, "A traditional approach to failure detection consists of considering that the design of detection algorithms is basically made of two steps : 1. generation of 'residuals,' namely of artificial measurements that reflect the changes of interest; for example, these signals are ideally close to zero before the change and nonzero after; 2. design of decision rules based upon these residuals." Hence, two main steps are simultaeously performed in implementing the CPM in a time series analysis, i.e., generating the residuals and detecting the change points. The first step is to obtain the i.i.d residuals and the second step is to determine the change points. Henceforth, we refer the detecting procedure as the two steps, i.e., generating residuals and detecting change points.

The first step on the detection algorithm in CPM framework requires an i.i.d series for the input. Hence, prior to detecting we need to check for the i.i.d assumption on the input series. The process is similar to preliminary forecasting process in time series analysis in which the stationarity property is checked prior to modelling. The classical decomposition model represents any time series x_t as a realization of the pro-

cess,

$$x_t = m_t + s_t + Y_t$$

where m_t , s_t , and Y_t are the trend, the seasonal and the stationary random noise components, respectively. By careful examination of a time series plot, one can have a good idea of these components. After removing the trend and/or the seasonal components such that the series is stationary, we can proceed to construct models under the theory of stationary processes [11].

A preliminary transformation in forecasting involves a non-parametric method to eliminate the trend and or seasonal components. Some techniques are discussed in [11]. In this article, we apply a fuzzy time series (FTS) model on in-sample data to eliminate the trend. FTS of [12] was stated as an appropriate time series model to handle a trend in a series [32]. As a nonparametric method, estimating using FTS needs no distributional assumption on the sample. The uncorrelated residuals of FTS is of interest as it provides a chance to incorporate the estimation method on the first step of the detecting algorithm. The FTS model was introduced by [12],[13]. It takes into account the fuzziness in data gathering in estimation process. The model has been extended in all steps of the fuzzy procedures. A brief review on forecasting using FTS can be found in [14]. A comparison of some FTS models in forecasting is discussed in [31].

A number of research studies in change detection have employed fuzzy approach, i.e., [32], [33], [19],[20], [22], [15]. We review some of them which discuss the approach in time series. The first literature uses deviation of two fuzzy numbers of LR-type in regression trees framework while the second literature employs fuzzy clustering and defines an α -level of detection statistics. In the first literature [15], the authors use fuzzy approach in analysing Italian sovereign debt and temperature in Rome. The first data set is of rating scales while the second data set is crisp data. The method is based on the concept of contagious partitions introduced by Fisher in 1958 where the author proposed an exact grouping of n elements into G mutually exclusive and exhaustive subgroups by minimizing the sums of squares of within-group such that the subgroups having maximum homogeneity. When the data points are *a priori* ordered, such a case in time series, seeking minimum sum of within-group sums of squares is equivalent to segmenting a time interval into G subintervals which two successive subintervals have different model parameters. The concept was explored by [17] who apply global search using a dynamic programming to get G partitions. In contrast to the previous research which use non hyrarchi partition, Cappeli et. al. show that the partition can also be obtained hyrarcically by using regression trees which they then propose Atheoretical Regreesion Tree (ART) method [16]. An attempt to assess the use of ART directly on time series to detect changes shows that the method is robust for series with negative correlation and some small positive correlation [18]. The ART method applies binary splitting to divide an interval into two subintervals. The splitting criterion involves additions of sums of square residuals on both subintervals. By using deviation measures of two LR-fuzzy data in place of residuals of the crisp counterpart,

the authors proceed the splitting procedure. Hence, the method sounds fuzzy approach.

Another fuzzy detection to detect change periods was proposed by [20]. The method uses α -level change period detection proposed in [19] based on a fuzzy clustering and an extended CUSUM in [21]. The fuzzy detecting algorithm is applied to detect change periods in Taiwan monthly arrivals. Fuzzy clustering and the α -level change period detection were also used to detect change points and change periods [22]. Changes in the level and slope in between two consecutives regimes are detected by using Chow-test. The model is then applied to detect change points and change periods of six Asian Countries exchange rates.

The above fuzzy detection algorithms show that crisp detecting methods can be set up so that they work well in the fuzzy system, either by choosing a suitable fuzzification of the data and/or modifying the method such that it accommodates fuzzy procedures. In this article, a procedure to detect abrupt changes in a time series by using CPM is presented. A fuzzification is introduced in an existing FTS model to generate the i.i.d residuals.

Our motivations leading to the change detection framework can be summarized as follows. Theoretically, it allows us to use fuzzy approach in analyzing abrupt changes and it is a natural counterpart to the conventional framework as well; From practical point of view, as a statistical decision tool for detecting changes which is of great potential interest either in industry or in economy.

The outline of the article is as follows. Introduction is on section 1. The model of the abrupt change is explained in section 2. In section 3, the statistic methods employed in the proposed detecting algorithm are discussed and the proposed procedure is given. Simulation study showing the percentage of uncorrelated residuals obtained under FTS estimates and the percentage of correct detection of the detecting algorithm is presented in section 4. Implementation of the proposed detecting procedure on two empirical examples and comparison of the proposed detecting procedure to the existing automatic detection is given in section 5. Conclusions are provided in the last section.

2 The Model

The proposed detection algorithm may be applied for general nonseasonal and seasonal ARIMA models. To simplify, we use nonseasonal ARIMA without drift to show the performance of fuzzy estimation in generating the i.i.d residuals. Let Z_t be a stochastic process following an autoregressive integrated moving average of order p, d and q ; that is

$$\phi_p(B)(1 - B)^d Z_t = \theta_q(B)a_t \tag{1}$$

where B is the backshift operator such that $BZ_t = Z_{t-1}$; $\phi_p(B) = 1 - \phi B - \dots - \phi_p B^p$, $\theta_q(B) = 1 - \theta B - \dots - \theta_q B^q$ are two polynomials in B with zeros all outside the unit circles and a_t are i.i.d $N(0, \sigma_a^2)$.

One of the parsimonous ARIMA models is ARIMA(1,1,1). Parsimonous ARIMA models and the extended models are

widely employed to describe data set in economy and industry [34]. We use the ARIMA model, in the simulation to show the performance of the proposed detecting procedure. In order to do that, a number of series under the model, contaminated by a change point to some degree of break, are simulated. The mean and variance of the model are used to generate the contaminated series. We define the mean and variance of the model in the following subsection.

2.1 Mean and Variance of ARIMA(1,1,1)

For the purpose of simulation which we need to generate some series with a breakpoint where the underlying model is ARIMA(1,1,1), we need to determine the mean and variance of the model. Let, $p = 1, d = 1$, and $q = 1$ in (1). We rewrite the model as follows:

$$(1 - \phi B)(1 - B)Z_t = (1 - \theta)a_t; \tag{2}$$

which gives,

$$Z_t = (1 + \phi)Z_{t-1} - \phi Z_{t-2} + a_t - \theta a_{t-1} \tag{3}$$

Given that the mean and variance of ARIMA are time dependent [23], the mean of ARIMA(1,1,1) at time t is determined as follows:

$$E[Z_t] = (1 + \phi)E[Z_{t-1}] - \phi E[Z_{t-2}] + E[a_t] - \theta E[a_{t-1}];$$

Hence,

$$E[Z_t] = (1 + \phi)Z_{t-1} - \phi Z_{t-2} \tag{4}$$

Further, we set the time reference at n_0 and follow the way variance of IMA(1,1) is derived in [23]. Following (3), $Z_{t-1} = (1 + \phi)Z_{t-2} - \phi Z_{t-3} + a_{t-1} - \theta a_{t-2}$. After substituting it into (3) and rearranging the equation, we get,

$$Z_t = (1 + \phi + \phi^2)Z_{t-2} - \phi(1 + \phi)Z_{t-3} + a_t + (\phi + 1 - \theta)a_{t-1} - \theta(1 + \phi)a_{t-2};$$

Next, getting $Z_{t-2} = (1 + \phi)Z_{t-3} - \phi Z_{t-4} + a_{t-2} - \theta a_{t-3}$. After substituting it into the last equation, simplifying it and doing so until the- k step, we get,

$$\begin{aligned} Z_t &= (1 + \phi + \dots + \phi^{k-2} + \phi^{k-1})Z_{t-(k-1)} \\ &\quad - (\phi + \phi^2 + \dots + \phi^{k-2} + \phi^{k-1})Z_{t-k} \\ &\quad + a_t + (\phi + 1 - \theta)a_{t-1} + [(\phi + \phi^2) - \phi\theta + (1 - \theta)]a_{t-2} \\ &\quad + [(\phi + \phi^2 + \phi^3) - (\phi + \phi^2)\theta + (1 - \theta)]a_{t-3} \\ &\quad + \dots + [(\phi + \phi^2 + \phi^3 + \dots + \phi^{k-2}) - (\phi + \phi^2 \\ &\quad + \dots + \phi^{k-3})\theta + (1 - \theta)]a_{t-(k-2)} \\ &\quad - [1 + \phi + \dots + \phi^{k-2}]\theta a_{t-(k-1)}. \end{aligned}$$

Let $Z_{t-k} = Z_{n_0}$ and the subscripts and superscripts are changed, i.e., $t - k = n_0, t - (k - 1) = n_0 + 1$ and

$k - 1 = t - n_0 - 1$. Thus, we have

$$\begin{aligned} Z_t &= (1 + \phi + \dots + \phi^{t-n_0-1})Z_{n_0+1} \\ &\quad - (\phi + \phi^2 + \dots + \phi^{t-n_0-1})Z_{t-k} \\ &+ a_t + (\phi + 1 - \theta)a_{t-1} + [(\phi + \phi^2) - \phi\theta + (1 - \theta)]a_{t-2} \\ &\quad + [(\phi + \phi^2 + \phi^3) - (\phi + \phi^2)\theta + (1 - \theta)]a_{t-3} \\ &\quad + \dots + [(\phi + \phi^2 + \phi^3 + \dots + \phi^{t-n_0-2}) \\ &\quad - (\phi + \phi^2 + \dots + \phi^{t-n_0-3})\theta + (1 - \theta)]a_{n_0+2} \\ &\quad - [1 + \phi + \dots + \phi^{t-n_0-2}]\theta a_{n_0+1}. \end{aligned}$$

The above equation can be written in a shorter expression using sum notations as follows,

$$\begin{aligned} Z_t &= \left(1 + \sum_{i=1}^{t-n_0-1} \phi^i\right) Z_{n_0+1} - \left(\sum_{i=1}^{t-n_0-1} \phi^i\right) Z_{n_0} + a_t + \\ &\quad \sum_{j=1}^{t-n_0-2} \left[\sum_{i=1}^j \phi^i - \theta \sum_{i=1}^{j-1} \phi^i + (1 - \theta) \right] a_{t-j} - \\ &\quad \left(1 + \sum_{i=1}^{t-n_0-2} \phi^i\right) \theta a_{n_0+1}. \end{aligned}$$

Because the terms, Z_{n_0} , Z_{n_0+1} and a_{n_0+1} , are assumed as the predefined known values, they are statistically constant. Consequently, their variances are zero. Furthermore, as the noise are independently distributed, the variance of the series equals the total of their serial variances. Hence,

$$\begin{aligned} var(Z_t) &= var(a_t) + \\ &\quad \sum_{j=1}^{t-n_0-2} \left(\sum_{i=1}^j \phi^i - \theta \sum_{i=1}^{j-1} \phi^i + (1 - \theta) \right)^2 var(a_{t-j}); \end{aligned}$$

Therefore, the variance of ARIMA(1,1,1) at time t is as follows:

$$var(Z_t) = \left[1 + \sum_{j=1}^{t-n_0-2} \left(\sum_{i=1}^j \phi^i - \theta \sum_{i=1}^{j-1} \phi^i + (1 - \theta) \right)^2 \right] \sigma^2. \tag{5}$$

When we put the AR coefficient to be zero and choose the reference point at $n_0 + 1$, i.e., $\phi = 0, m_0 = n_0 + 1$, (5) reduces to the variance of IMA(1,1) that is equation (4.3.4) of [23], i.e. $var(Z_t) = [1 + (t - m_0 - 1)(1 - \theta)^2]\sigma^2$.

As the variance of ARIMA model depends on the time observation, we report the variance of ARIMA models at particular time. Let $t = 50$, we simulate ARIMA(1,1,1) on their stable condition, i.e., $-1 < \phi, \theta < 1$ with the standard normal error and then calculate the variance at time t . The plot shows that variance of the process is directly proportional to the AR parameters and inversely proportional to the MA parameters. The plot is not displayed here for reason of spaces. We focus our interest on the processes with small variances by choosing the following intervals for the parameters, $-1 < \phi \leq -0.5$ and $0.5 \leq \theta < 1$. We also compare, graphically, the simulated variance to the theoretical variance of (5) as shown in Fig. 1.

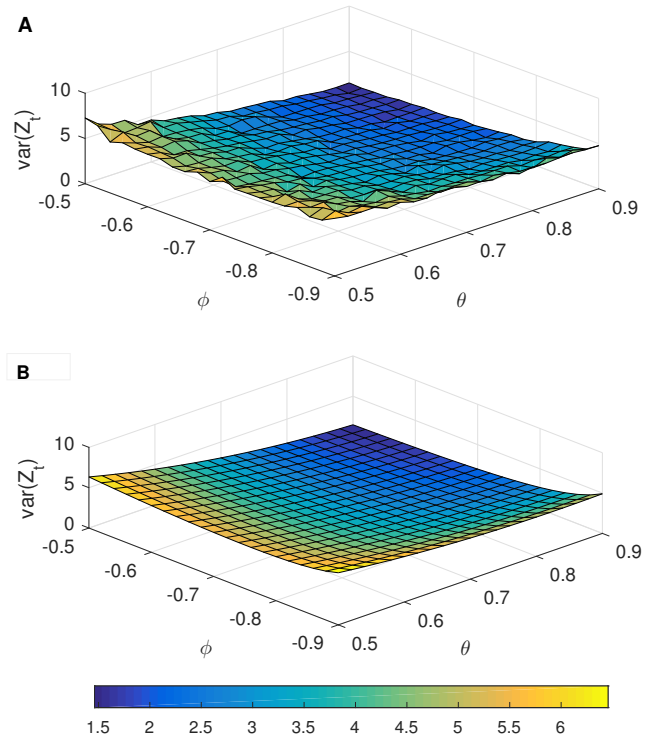


Figure 1. Variance of ARIMA(1,1,1). A. Simulation result and B. Using formula in (5).

As can be seen, overall, both subplots share a similar shape. However, the surface plot of the simulated series is rather undulating (Fig. 1A) compared to the smooth surface plot of the theoretical series (Fig. 1B). These plots qualitatively prove that our finding, the theoretical formula, in (5) is a well representative model for variance of ARIMA(1,1,1). Therefore, we can use the formula for the purpose of analysis related to the variance of ARIMA (1,1,1) at particular time t .

2.2 An ARIMA Model with A Breakpoint at Particular Time

In this subsection we define an ARIMA model which contains an abrupt change. We say that the model has an additive change as we add a degree of deviation at the mean of the observation at the targeted change point T . Given ARIMA (1,1,1) as in (2). We can write, $Z_t = \pi(B)a_t$ where $\pi(B) = \frac{1-\theta B}{1-\phi B} \frac{1}{1-B}$. Let $\sigma_{Z_t}^2$ be the process variance at time t . We define the process with a breakpoint at time $t = T$ as follows;

$$\begin{aligned} Z_t - w\sigma_{Z_t}I_t^{(T)} &= \pi(B)a_t; \\ Z_t &= w\sigma_{Z_t}I_t^{(T)} + \pi(B)a_t; \end{aligned} \tag{6}$$

where w is a constant representing the size of break or the weight and $I_t^{(T)} = 1$ for $t = T$ and zero otherwise. When, we set $w\sigma_{Z_t} = \omega$ in (6), the model is known as the AO (additive outlier) type of outliers of Fox (1972) in [7] where ω is the magnitude of the outlier effects.

One important issue in simulating series with outliers is setting the breaksize. A large breaksize produces a single outlier that may wash out the underlying characteristics of the series [23] which may result in a white noise phenomenon. On the other hand, setting a rather small breaksize may result in failure detection of the outliers as they may be masked by the underlying series. In the simulation, we set the range for w based on the variance of the observation at the change point and at the last simulated observation. This is due to the fact that the variance of ARIMA(1,1,1) increases as the time increases. Also, the outlier should lie in some distance which should be much larger than the standard deviation of the simulated series. Let $sd(Z_T)$ and $sd(Z_n)$ be standard deviation of the observation at the change point and at the last observation, respectively. We define the minimum break size for an abrupt change as follows:

$$wsd(Z_T) \geq 4sd(Z_n). \tag{7}$$

When analyzing the use of ART in detecting change points in time series, [18] concludes that the best location for simulated breakpoint is in the middle of the series under investigation. Following it, in the simulation, a change point is introduced in the middle of the simulated ARIMA.

Let $\phi = -0.5, \theta = 0.9$, and $\sigma_a^2 = 1$ be the ARIMA parameters, i.e., the model is written as follows :

$$(1 + 0.5B)(1 - B)Z_t = (1 - 0.9B)a_t. \tag{8}$$

Figure 2 shows one realization of the model and the simulated series with a break of size 4, 5, and 6. The number of observations on each series is 100 and the break is at the 50th observation.

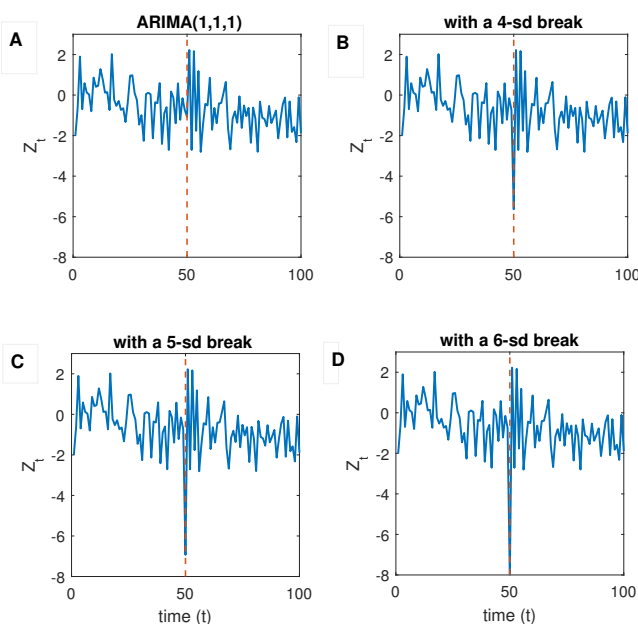


Figure 2. (A). One realization of ARIMA(1,1,1) in (7) and the model with an abrupt change at $t = 50$ in several break sizes, i.e., breaksize = 4 (B), breaksize = 5 (C), and breaksize = 6 (D).

3 Methods

In this section we discuss the methods used in constructing the algorithm of detection. We first describe the FTS models. We also describe the proposed fuzzification process in the model. As mentioned in the introduction, the model is constructed to eliminate trend on raw data in order to get i.i.d residuals. We then, review the CPM used for detecting changepoints on the residuals. We also review some tools available in the literatures related to i.i.d observations. Finally, we outline the proposed detection algorithm and give some remarks on it.

3.1 Fuzzy Time Series

We briefly review some definitions in fuzzy sets, fuzzy time series, and fuzzy logical relationship in [24], [12] and [13], respectively. We then describe step by step procedures used in the recent study.

Let $U = \{a_i; i = 1, \dots, p\}$ be a universe of discourse which is discrete and finite. A fuzzy set A of U is defined as, $A = \left\{ \frac{\mu_A(a_1)}{a_1} + \dots + \frac{\mu_A(a_p)}{a_p} \right\}$, where $\mu_A: \rightarrow [0, 1]$ is the membership function and $\mu_A(a_i)$ is the membership degree of A in a_i . The plus sign represents a union and the horizontal division line represents a separator in the fuzzy set theory. Some literatures use commas and brackets instead of plus and curly brackets in writing a fuzzy number. A fuzzy set of A , by definition, is determined by its membership degrees of $\mu_A(a_i)$.

Let $\{y_t \in R; t = 1, 2, \dots, n\}$ be a universe of discourse in which a fuzzy set $\{\mu_i(t) \in R; i = 1, 2, \dots\}$ is defined and $F(t)$ be the set of all $\{\mu_i\}$. Here, $F(t)$ is called as the fuzzy time series of $\{y_t; t = 1, 2, \dots, n\}$. In time series analysis, recent values can be described by information on previous values in the form of an analytical equation. By analogy, if a fuzzy time series at time t , $F(t)$ is explained by its previous value $F(t-1)$ then this relationship represented as $F(t-1) \rightarrow F(t)$. This expression is called as fuzzy logical relationship (FLR).

We adopt the framework in [25] in setting up the fuzzy procedures. There are 5 steps discussed in the reference. We modify the formula used in partitioning universe of discourse which is on the first step. The fuzzification is set such that heights of fuzzy numbers of the extreme values are their crisp values. It turns out that the defuzzification of these fuzzy numbers are their crisps. Hence, we are likely to produce zero residuals for these observations. We employ procedures provided in [26] for the remaining steps. Let x_1, \dots, x_n be the time series of interest.

Step 1. Partition of universal of discourse. We create subintervals $a_i = [lb_i, ub_i]$ for $i = 1, \dots, n$ based on the series. We define $D = \frac{1}{n} (\max\{x_t\} - \min\{x_t\})$ and $d = \min\{x_t\} - 0.5D$. The lower and upper bounds of the i -th subinterval are given as $lb_i = d + (i - 1)D$ and $ub_i = d + iD$, respectively.

Step 2. Define linguistic terms. Linguistic terms are given on intervals which contain at least one crisp datum. Let $p \leq n$ be the number of intervals which has at least one crisp datum. Then we are going to have p linguistic terms notated as $A_i, i = 1, \dots, p$, i.e., a set of fuzzy numbers. Fuzzy number A_i is a discrete fuzzy number which has a membership function of degree one in a_i , of degree 0.5 in both a_{i-1} and a_{i+1} and of

degree zero in other intervals. The fuzzy number is written as, $A_i = \left(\frac{0}{a_1}, \dots, \frac{0.5}{a_{i-1}}, \frac{1}{a_i}, \frac{0.5}{a_{i+1}}, \dots, \frac{0}{a_p}\right)$. Meanwhile, for $i = 1$ and $i = p$, we define fuzzy numbers A_1 and A_p as, $A_1 = \left(\frac{1}{a_1}, \frac{0.5}{a_2}, \frac{0}{a_3}, \dots, \frac{0}{a_p}\right)$ and $A_p = \left(\frac{0}{a_1}, \dots, \frac{0}{a_{p-2}}, \frac{0.5}{a_{p-1}}, \frac{1}{a_p}\right)$, respectively.

Step 3. Fuzzify crisp data. In fuzzification process, a crisp datum is assigned a fuzzy number. That is when a crisp number belong to interval a_i , the fuzzy number is the corresponding linguistic term A_i . We also collect the index of the linguistic term, i , as the weight of fuzzy number A_i .

Step 4. Set Fuzzy Logical Relationship (FLR) and Fuzzy Logical Relationship Groups (FLRGs). Let A_i and A_j be fuzzy numbers of x_{t-1} and x_t , respectively. The FLR of order one for the observations is given as $A_i \rightarrow A_j$. There are $(n - 1)$ FLRs of order one. A fuzzy number on the left hand side of an FLR arrow is also named LHS. Similarly, for the fuzzy number on the right hand side, we have RHS. Next, FLRGs are created based on the FLRs. When some FLRs have one similar LHS but of different RHSs, the FLRs are grouped into one FLRG. For example, let there be FLRs of the following forms: $A_i \rightarrow A_j, A_i \rightarrow A_k$, and $A_i \rightarrow A_l$. Then, the FLRG is of the form $A_i \rightarrow A_j, A_k, A_l$. The left hand side of an FLRG is a unique fuzzy number whereas the right hand side comprises of at least one fuzzy number. Hence, there would be less numbers of FLRGs compare with FLRs.

Step. 5. Defuzzify fuzzy numbers. Let A_i be a fuzzy number and x_t be fuzzified into A_i . Defuzzification of A_i means that we calculate \hat{x}_t , the estimate of x_t . We defuzzify all fuzzy numbers on the left hand side of FLRG one by one. Let A_i be a fuzzy number on the LHS and A_j, A_k, A_l be fuzzy numbers on the RHS of an FLRG, i.e., the FLRG is of the form: $A_i \rightarrow A_j, A_k, A_l$. Also, let the heights and weights of A_j, A_k, A_l be c_j, c_k, c_l and w_j, w_k, w_l , respectively. The formula of defuzzification of fuzzy number A_i is given as follows:

$$\hat{x}_t = \frac{1}{w_j + w_k + w_l} (c_j w_j + c_k w_k + c_l w_l).$$

The form of FLRs of an FTS model indicates the order used in the model. For FTS of order one, the recent value is determined by one previous value as in the 4^{th} step above. Further, the model gives estimates for the 2^{nd} to the last observation. Therefore, a series with n number of observations will have $(n - 1)$ residuals.

3.2 Change Point Model (CPM)

Change point detection is related to detecting change points in a sequence of data. Statistically, the changes are defined as changes in distribution. To model it, let x_1, x_2, \dots, x_n be a sequence of independent observations drawn from random variables X_1, X_2, \dots, X_n . In the absence of changes, we can model the random variables to follow a distribution, let say F_0 , i.e., $X_i \sim F_0$. Given that the sequence undergoes changes at r points of time c_1, c_2, \dots, c_r then we can model the sequence as $X_i \sim F_1$ for $i \leq c_1, X_i \sim F_2$ for $c_1 < i \leq c_2, \dots, X_i \sim F_n$ for $c_n < i \leq n$.

[9] used the conventional two-sample t -statistics test as the test to detect a change point in mean in statistical process control framework. Given n observations drawn from normally independent and identically distributed random variables, a two-sample t -test is formulated to test the hypothesis that the mean changes after some observations. The following is the test to detect one change point on a sequence of n observations which the mean changes after the- k th, $1 \leq k < n$, observation. By changes in mean, we imply that $\mu_k \neq \mu_{n-k}$ and $\sigma_k^2 = \sigma_{n-k}^2 = \sigma^2$ where μ_k and σ_k^2 are the mean and variance of the first k observations, μ_{n-k} and σ_{n-k}^2 are the mean and variance of the remaining observations and σ is the series standard deviation. The hypothesis is $\mu_k = \mu_{n-k}$ versus $\mu_k \neq \mu_{n-k}$. The statistics test is given as follows:

$$T_k = \sqrt{\frac{k(n-k)}{n}} \frac{(\bar{X}_k - \bar{X}_{n-k})}{\hat{\sigma}_k},$$

where $\bar{X}_k = \frac{1}{k} \sum_{i=1}^k x_i, \bar{X}_{n-k} = \frac{1}{n-k} \sum_{i=k+1}^n x_i, V_k = \sum_{i=1}^k (x_i - \bar{X}_k)^2 + \sum_{i=k+1}^n (x_i - \bar{X}_{n-k})^2$ and $\hat{\sigma}_k^2 = \frac{V_k}{(n-2)}$ are the MLE estimators of μ_k, μ_{n-k} and σ^2 , respectively. The last term is the sequence pooled variance. Under the null hypothesis, test statistics T_k follows Student- t distribution with $(n - 2)$ degree of freedom. Let T_k^* be the maximum of $T_k, 1 \leq k < n$. The test signals a change point at k , when $T_k^* > h_n$, a critical value defined based on the distribution.

In this article, the CPM is used in the second step in the detection algorithm that is for detecting the change point on i.i.d residuals. Unless otherwise indicated we use the CPM method as implemented in the R software of R Development Core Team (2015) contributed package **cpm** of [10].

3.3 Independent and Identically Distributed (i.i.d) Residuals

Diagnostic checking of residuals is a step in modeling time series to check for the goodness of model fitting. Residuals are differences between the model and observations. Hence, it comprises portion of information on the observations not explained by the model. The best model produces residuals that satisfy properties of i.i.d random variables, i.e., that there is no dependence among the observations. A number of tools are utilized to check for the property in a series [11]. Checking independency of a sequence qualitatively is assessed by using the properties of autocorrelation of independence series while quantitatively independent observations can be analyzed by testing null hypothesis of no autocorrelations or testing null hypothesis of random observations on the residuals. Without loss of generality, in the following, to check for independence of a sequence of observations, we review only two tools that we mainly use in the article.

We can check i.i.d series qualitatively on the autocorrelations. The sample autocorrelations of n independent observations with finite variance follow a normal distribution with zero mean and constant variance of $1/n$ [27]. In other words, at a significant level α , the $(1 - \alpha)$ confidence interval of sample autocorrelations is $[-Z_{\alpha/2} \times 1/\sqrt{n}, Z_{\alpha/2} \times 1/\sqrt{n}]$ where $Z \sim N(0, 1)$. Hence, at $\alpha = 5\%$ and $\alpha = 1\%$

, sample autocorrelations of independent residuals are within $[(-1.96)/\sqrt{n}, 1.96/\sqrt{n}]$ and $[(-2.576)/\sqrt{n}, 2.576/\sqrt{n}]$, respectively. Therefore, if there are one or more autocorrelations which lay outside the intervals, the autocorrelations of the series are too large for the series to be sampled from an i.i.d sequence.

Ljung-Box Q Test. Given Z_1, \dots, Z_n is an i.i.d series with finite variance, statistics $Q = n(n + 2) \sum_{j=1}^k \frac{\rho_j^2}{(n-k)}$ where ρ_j is the autocorrelation at lag j , can be used to test a null hypothesis of zero autocorrelation at the first k lags. Under the null hypothesis, Q follows $\chi^2(k)$ distribution. Hence, at a significant level α , we reject the null hypothesis for $Q > \chi^2_{1-\alpha}(k)$. When, we apply the test on residuals obtained by fitting a model with m parameters, we then need to adjust the degree of freedom of the Chi-square from k into $(k - m)$.

3.4 Proposed Detecting Algorithm

The proposed detecting procedure is constructed to detect abrupt changes in a time series using CPM model. The procedure consists of 4 main steps. The first step is aimed to eliminate trend by using a nonparametric FTS model. The proposed algorithm to detect change points are as follows:

- Step 1. Apply FTS of order one on the raw data using FTS model in Subsection 3.1;
- Step 2. Check the i.i.d property of the residuals using ACF and PACF analysis;
- Step 3. If the residuals show some correlated series, apply ARIMA on the residuals. Go to Step 2.
- Step 4. Apply CPM method on the i.i.d residuals.

The first step in the proposed procedure using a nonparametric method may result in correlated residuals. Therefore, diagnostic checking for i.i.d assumption has to be implemented on the residuals. In the following section, we show how sampling behavior of the autocorrelation function of the FTS residuals in terms of the percentage of uncorrelated residuals. Given that some autocorrelations in the second step are still significant, we proceed to ARIMA modelling. In the preliminary forecasting, transformed series with significant correlations at some lags either in ACF and/or PACF indicates a pattern of ARIMA models. Following the methodology in ARIMA modelling, we are able to get the i.i.d residuals. The detecting procedure using CPM method is on the last step. In order to have an efficient detection using CPM, we need background knowledge related to the changes in the raw data, i.e., one change point versus multiple change points, types of structural changes in the series and underlying assumption on the raw data if it is available.

4 Simulation Study

The proposed procedure is designed to detect abrupt changes in time series. A simulation study is carried out to analyze the performance of the proposed procedure. The analysis is conducted for single outlier of AO type which is located in the middle of the simulated series. The underlying series in the simulation is of ARIMA(1,1,1).

We use the ARIMA model as defined in (8). Let $T = 100$, and $n_0 = 0$ be the simulated change point and the reference point. A simulated series with an abrupt change at time T is constructed as follows: 1) we simulate 150 observations of the ARIMA model; 2) we set the 100th observation to be the expectation minus a weight of its standard deviation. The mean and variance of the model at time $t = T$ are determined according to (4) and (5), respectively. In other words, we define the breakpoint at $t = 100$ as follows: $Z_{100} = (1 + \phi)Z_{99} - \phi Z_{98} - wsd(Z_{100})$; 3) In order to avoid the pre sample effects on simulated data, we use the last remaining 100 observations for simulation, i.e., we use the 51th to the end observation; By doing this, we have a series of length 100 contaminated with a breakpoint of size w at the 50th observation.

4.1 Sampling Behavior of Sample ACF of Residuals

As it has been said in Section 2, we set the break size such that the deviation of the change point is around 4 times larger than the maximum standard deviation on the simulated series, that is the standard deviation of the last observation, i.e., see (7). In the case with the ARIMA model chosen, the breaksize is greater than 4.2. Further, we use break size greater than or equal to 5, but we may also display some results using break size less than 5 in order to see the pattern.

One realization of the ARIMA model with a break size of 6 and the sample ACF and PACF plots are given in Fig. 3. The nonstationarity shown by a slowly decaying pattern in the sample ACF. FTS of order one is then applied on the simulated

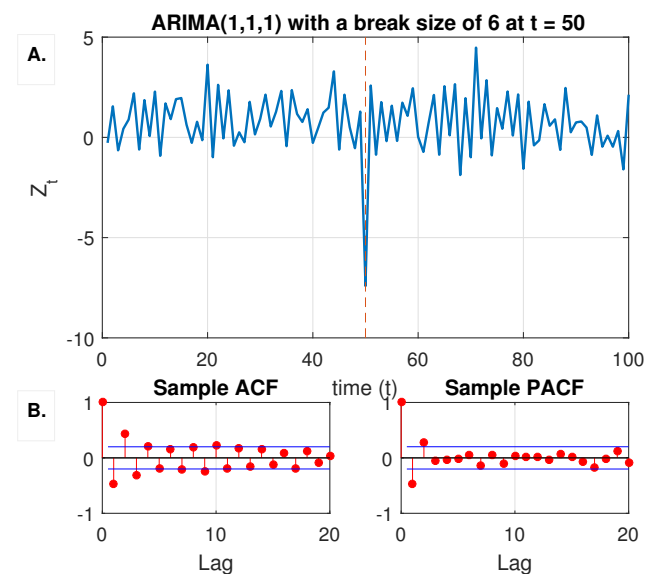


Figure 3. (A). The ARIMA(1,1,1) model with a 6 size of breakpoint at $t=50$ and (B). The sample ACF and PACF of the series.

series. The residuals is shown in Fig. 4A and the sample ACF and PACF plot is in Fig. 4B. Although, plot of the residuals show a similar pattern to that of the raw series, the sample ACF and PACF of the residuals show an i.i.d sequence.

In the simulation, the sampling behavior of the autocorrelations of the residuals is shown as the percentage of uncor-

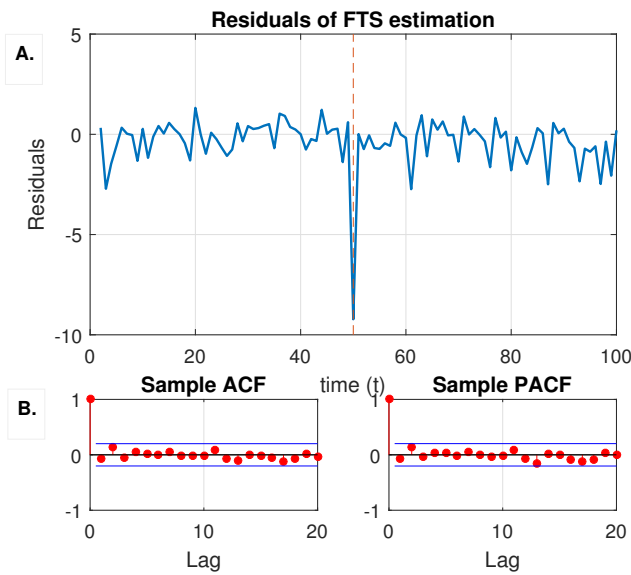


Figure 4. (A). Residuals of FTS model and B. The sample ACF and PACF of the residuals.

related residuals. We present the percentage of the uncorrelated residuals according to the number of realizations and the size of breaks. We generate a number of realizations, i.e., $N = 100, N = 500, N = 1000$ and $N = 5000$. On each realization there are 100 observations following the ARIMA model and a weighted breakpoint with size $w; w = 3, 4, 5,$ and 6 at the mid of the observation, i.e., the 50th observation. We apply FTS on the simulated series and get the residuals. Further, we count the percentage of uncorrelated residuals on 0.01 and 0.05 significant level. That is to say that to be uncorrelated on the levels, the autocorrelations of the residuals must lie on $[-0.197, 0.197]$ and $[-0.259, 0.259]$ at all lags, respectively. We report some of the results for $N = 100$ in Fig. 5. We record all the results in Table 1 and Table 2.

There are 4 subfigures in Fig 5. On each subfigures, there are 100 sample ACFs of residuals from 100 simulated series with particular breaksize. The sample ACF plots are displayed as line plots. We plot this way in order to see the trend more clear graphically. The 1% significant autocorrelation bands (shown as green horizontal-dashed line) are added on the plots to compare it to the default significant level (shown as the red-dashed line). As shown in the figures, overall, the larger the insignificant intervals are, the more the number of uncorrelated residuals is. This trend is similar to the size of breaks. In other words, the percentage of uncorrelated residuals is an increasing function of the break size.

Table 1. Percentage of the uncorrelated residuals at 0.01 significant level.

	N=100	N=500	N=1000	N=5000
$w = 3$	83	83.2	83.9	83.68
$w = 4$	78	84	82.8	84.24
$w = 5$	92	88.2	86.5	86.04
$w = 6$	84	87.4	90.1	87.96
$w = 7$	90	89.2	90.4	90.72

For higher number of realizations, on average, the percent-

Table 2. Percentage of the uncorrelated residuals at 0.05 significant level.

	N=100	N=500	N=1000	N=5000
$w = 3$	41	45.8	48.7	44.78
$w = 4$	46	47.8	46.4	47.02
$w = 5$	58	54.4	49	51.64
$w = 6$	55	54.2	55	55.2
$w = 7$	62	58	59.6	61.58

age of uncorrelated residuals on each case of w does not change. As mention previously, we define abrupt changes with size $w \geq 5$. As seen on Table 2, at the break size of $w = 5$ on the default significant level, i.e. $\alpha = 0.05$, only around 49% to 58% of the residuals are uncorrelated. These percentages increase to around 86% to 92% using lower significant level (see Table 1).

4.2 Percentage of Correct Detection

The next step is to apply CPM method on the residuals to detect the change points. As mention previously, some background information about the raw data are needed in using the detection method effectively. We use detectChangePointBatch function because we want to detect only one change point. Given that we set the changes in variance of the change point and the residuals are generated using normal standard error, we set the type of CPM model as Bartlett type and use the default parameter for significant level to determined the treshold for the test statistics, i.e. $\alpha = 0.05$. The model is used to detect changes in variance of an i.i.d normal sequence.

Table 3. Percentage of detection on FTS residuals.

	cp	cp±1	cp±2	cp±4	miss	undetec	total
$w = 5$	17.98	31.32	35.42	41.38	34.74	23.88	100
$w = 5.5$	21.41	37.3	42.14	48.77	32.00	19.23	100
$w = 6$	25.23	43.45	48.74	55.75	28.73	15.52	100
$w = 6.5$	28.1	48.72	54.66	61.76	25.52	12.72	100
$w = 7$	31.3	53.13	59.36	66.8	22.16	11.04	100
$w = 7.5$	33	56.91	63.21	70.3	19.89	9.81	100
$w = 8$	35.54	60.52	66.99	74.11	17.31	8.58	100
$w = 8.5$	37.6	63.78	70.04	77.08	14.93	7.99	100
$w = 9$	39.13	66.02	72.07	79.15	13.28	7.57	100
$w = 9.5$	40.36	68.4	75.07	81.5	11.18	7.32	100

Table 4. Percentage of detection on FTS residuals uncorrelated at 0.01 significant level.

	cp	cp±1	cp±2	cp±4	miss	undetec	total
$w = 5$	19.14	33.24	37.44	43.33	32.23	24.43	100
$w = 5.5$	22.71	39.56	44.59	51.21	29.7	19.09	100
$w = 6$	26.9	46.20	51.57	58.65	26.16	15.19	100
$w = 6.5$	29.57	51.29	57.45	64.48	23.36	12.16	100
$w = 7$	32.93	55.83	62.27	69.58	20.47	9.96	100
$w = 7.5$	34.41	59.29	65.74	72.72	18.38	8.90	100
$w = 8$	37.04	63.10	69.65	76.76	15.90	7.34	100
$w = 8.5$	38.96	66.04	72.40	79.38	13.84	6.79	100
$w = 9$	40.45	68.19	74.36	81.35	12.26	6.39	100
$w = 9.5$	41.53	70.30	77.14	83.55	10.37	6.08	100

We generate 10,000 realizations of the ARIMA model and set the break size from 5 to 9.5 with increment 0.5. For each case, we apply the FTS and collect residuals. Then, we check for the uncorrelated property on the residuals on both significant levels. Next, we apply CPM method on all the residuals.

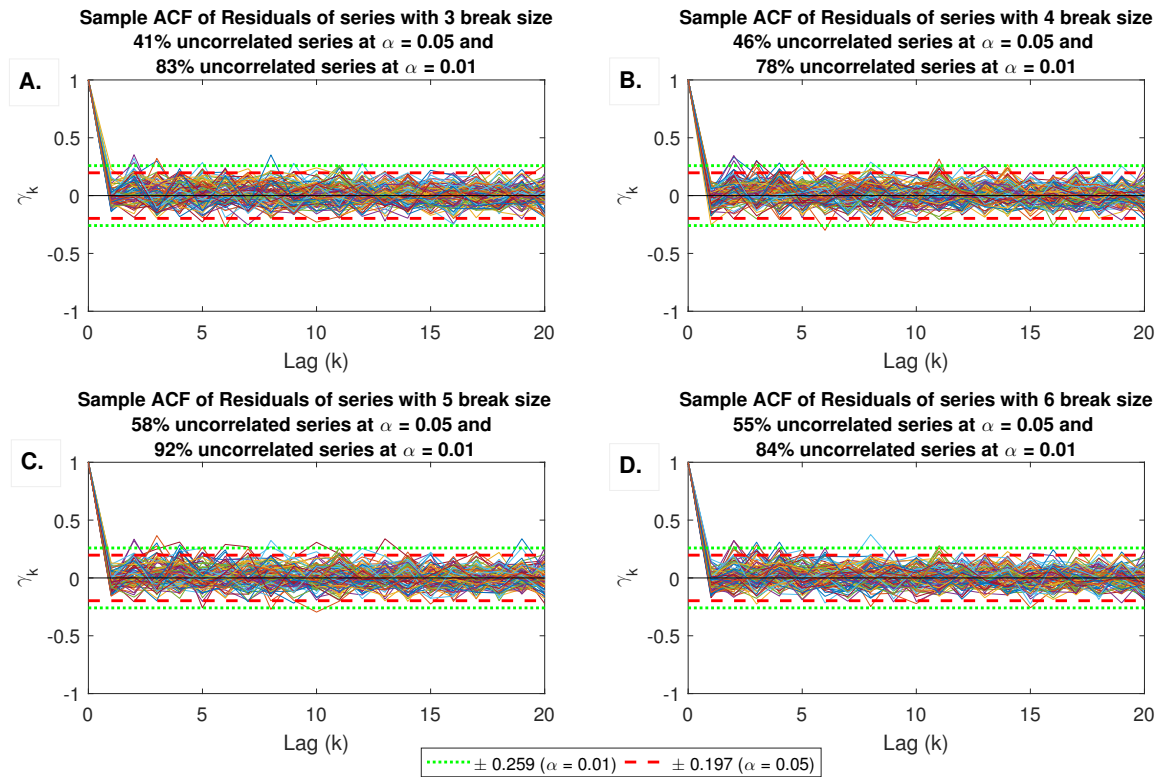


Figure 5. Percentages of uncorrelated residuals of FTS estimates on series with varying breaksize.

Table 5. Percentage of detection on FTS residuals uncorrelated at 0.05 significant level.

	cp	cp±1	cp±2	cp±4	miss	undetec	total
w = 5	21.27	36.8	41.31	47.22	28.45	24.33	100
w = 5.5	25.35	44.11	49.75	56.06	26.02	17.92	100
w = 6	29.44	50.98	56.62	63.97	22.92	13.11	100
w = 6.5	32.46	56.72	63.12	69.84	19.89	10.28	100
w = 7	36.39	61.26	68.29	75.47	16.78	7.77	100
w = 7.5	37	64.52	71.58	78.32	14.90	6.78	100
w = 8	39.78	67.72	74.68	81.62	12.85	5.52	100
w = 8.5	41.15	70.2	76.98	83.93	11.35	4.73	100
w = 9	43.26	72.79	79.21	85.93	9.82	4.25	100
w = 9.5	43.90	74.17	81.29	87.45	8.37	4.14	100

The percentage of correct detection regardless of the i.i.d assumption on the residuals is shown in Table 3. The column label 'cp' in this table summarizes the percentage of correct detection, defined as a correct identification of the change point location. Meanwhile, the column label 'cp±1', 'cp±2', and 'cp±4' record the percentage of correct period detection, defined as a correct identification of a change point located within one, two and four observations of the simulated change point, respectively. The column label 'miss' reports the percentage of misidentified change point, i.e., the number of observations in a series that are identified as change point while they are not in the 'cp±4' interval. We show the percentage of undetected change point in the column 'undetec' that is the number of series which the method fails to detect any change point. The 'total' column contains the row summation of all percentages of three columns right on the left.

A similar setting is set for the cases when we consider only uncorrelated residuals. We display the results for detecting

change points using uncorrelated residuals at 1% and 5% significant level in Table 4 and Table 5, respectively.

Overall, the percentage of correct detection is an increasing function of the break size and the significant level. Particularly, at significant level 0.05 and break size $w = 5$, percentage of correct detection of the change point is 21%, but it becomes adequate and high as we consider larger intervals around the true change point, i.e., at cp±1, cp±2 and cp±4, the percentage of correct period detection are 31%, 35% and 41%, respectively.

Figure 6 display the number of the detected residuals on their detected change points for all sizes of break. The percentage versions of these subfigures are shown previously in Table 4 and 5. As it can be seen in both figures, the detected change points are concentrated around the simulated change points. The numbers of correct detection (around the change point) in 0.01 uncorrelated residuals (bottom) seems larger than that of the 0.05 uncorrelated residuals (top). However, the number of undetected series in the 0.01 uncorrelated residuals is larger than that of the other correlated residuals. It is important to note that we have set the significant level for threshold in the **cpm** script by using the default score. All in all, a good performance of the proposed procedure is reached when the significant level for the sample autocorrelation is set at 5%.

5 Empirical Examples

In this section, we consider the abrupt changes in two tourism data mentioned in [28] and in [3],[2]. The first series exhibits an increasing trend while the other depicts both an

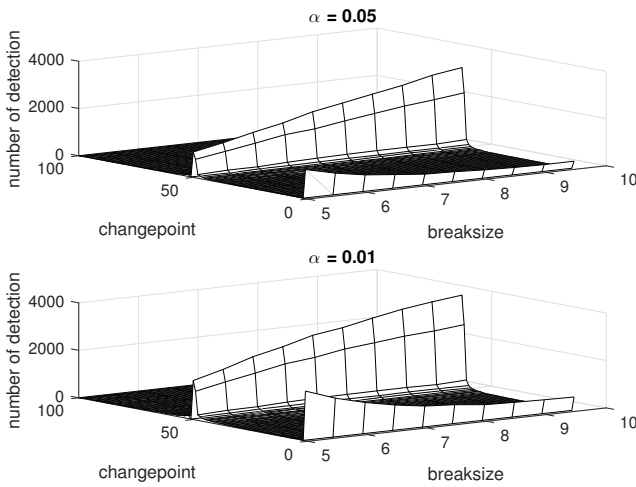


Figure 6. The number of the detected change points on 10,000 sequences of residuals which contain a single break point at $t=50$. The autocorrelations of the residuals are nonsignificant at 5% level (top) and at 1% level (bottom).

increasing trend and a strong seasonal pattern. We demonstrate the detailed steps of the proposed procedure and compare the results obtained by the proposed detecting algorithm to that of the existing detecting algorithm. i.e., automatic detecting procedure in `tsoutliers` package version 0.6-8 in R which is based on the detecting algorithm in [7].

Forecasting Taiwan tourism series using fuzzy time series is discussed in [28],[29]. The authors acknowledge that there is a possible outbreak of Severe Acute Syndrome (SARS) reported in 2002/04 on the series. Further, the authors consider the SARS period to be from 2002/11 to 2003/06. Meanwhile, it has been reported that there are two terrorist attacks in Bali (popularly known as Bali bombing) happened at a quite closed time. The first attack was in October 2002 and the second was in October 2005 [2]. The unexpected outbreak and bombings had affected the number of tourist arrivals as presented in the references. We aim to detect the first time the tourist numbers undergoing abrupt changes in their structure. In other words, we aim to detect the change points due to the intervention.

5.1 Example 1 : Detecting SARS in Taiwan Tourism

In forecasting monthly Taiwan tourist arrivals using FTS in [28], the ARIMA model for the series is also discussed. The in-sample data, i.e. on time interval 1984/01 to 2000/4, is modeled as ARIMA(1,1,1). By knowing that there is an event of SARS in Taiwan, Chow-test was used to detect the break on the out-of-sample data (2000/5 to 2005/9). Having rejecting the null hypothesis of no changes, the estimation and forecast time interval are then reset to the time before the breakpoint. For our example, we apply proposed algorithm on the out-of-sample data to detect the abrupt change on the series due to the intervention of SARS. We obtain the monthly series from data CEIC.

5.1.1 Generating Residuals : Applying FTS on Raw Data

We display the Taiwan tourist series in Fig. 7. One can see that there is a dramatic drop nearly on the middle of the overall increasing trend in the series (see Fig. 7A). We highlight the short interval where the series undergoes the break on the second subfigure, Fig. 7B. The sample ACF and PACF of the series is displayed in Fig. 7C. where we can see that the series is correlated. We would like to estimate the series such that it gives i.i.d residuals conditional to the change point.

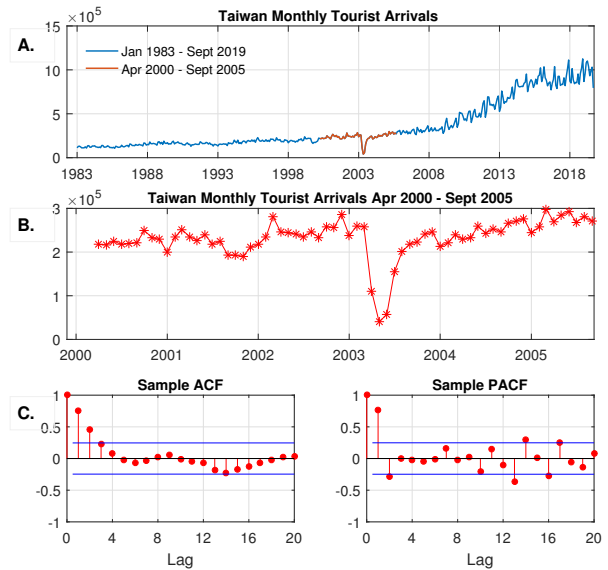


Figure 7. Taiwan monthly tourist series in longer time interval (A), the series under study (B) and the sample ACF and PACF of the series (C).

We apply FTS of order one on the Taiwan data set and display the estimates in Fig. 8A. The residuals shown on Fig. 8B still depict the similar drop pattern as does the raw data. Figure 9 shows diagnostic plots of the residuals. The ACF and PACF plots show a white noise phenomenon. Statistics test of Ljung-Box Q shows insignificant correlations for the first 20 lags of the residuals. The AICC plot shows that the minimum AICC is when the autoregressive order is 0. All plots indicate independent observations. The last two subplots show comparison of the sample distribution to the standard normal. The presence of one outlier is clearly seen on the quantile plot. However, overall spread of the residuals shows a normal distribution.

5.1.2 Detecting Change Points : CPM (Change Point Method)

As said in the introduction, `cpm` package provides a set of tools to detect changes either for parametric or nonparametric approach. We use `detectChangePointBatch` function because we want to detect change points in a fixed-sample size or we want to do *retrospective* detection problem. Based on diagnostic checking on the residuals, one can have a good idea of the type of CPM model should be used. Another option is of trial and error method. For the residuals of Taiwan series, our first attempt is to employ CPM model assuming normal distri-

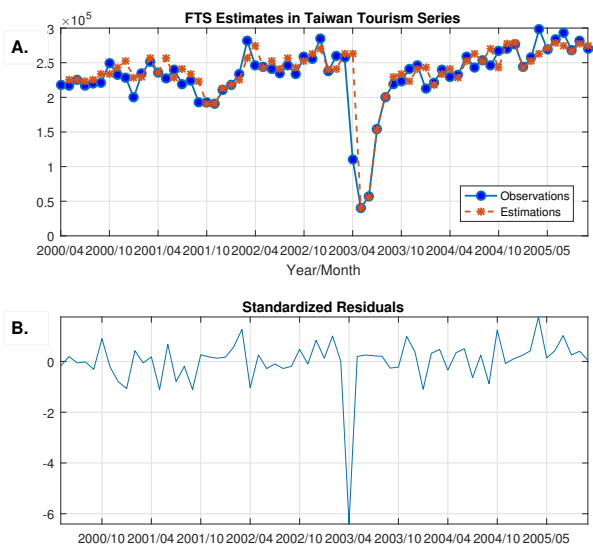


Figure 8. Estimation of Taiwan tourist arrivals using FTS (A) and the residuals (B). No trend pattern is observed on the residuals.

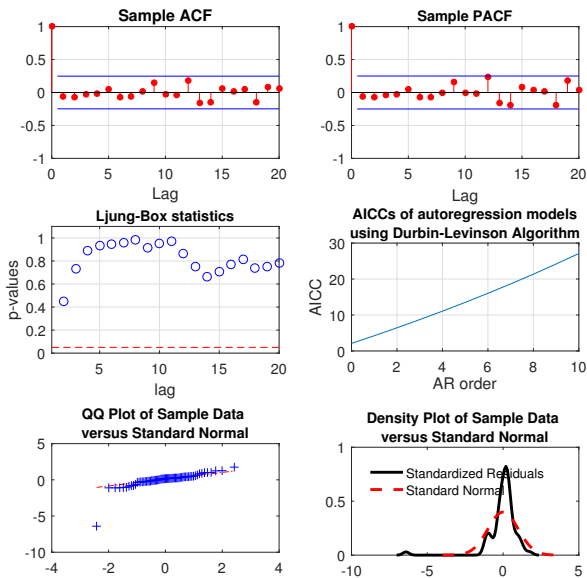


Figure 9. FTS residual diagnostic plots for Taiwan tourist series.

bution. The trials signal a change point when we use GLR type of CPM model. The R script is given on the appendix. The number of observations in Taiwan dataset is 66. Meanwhile, the FTS of order one gives estimations for the 2nd to the last observation of the dataset. Hence, the number of FTS residuals is 65. Figure 10 displays the $D_{k,n}$ statistics of the residuals. We add a vertical-dash line on the default plot to show the maximum point of the statistics which is the detected change point. The change point is located at the 35th residual. Therefore, on the raw data, location of the abrupt change is on the 36th observation, that is on 2003/03 (see Fig. 11).

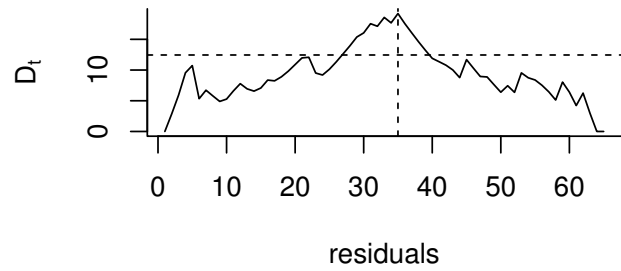


Figure 10. $D_{k,n}$ Statistics of FTS residuals. The horizontal dashed-line shows the statistics threshold while the vertical dashed-line points the maximum statistics test.

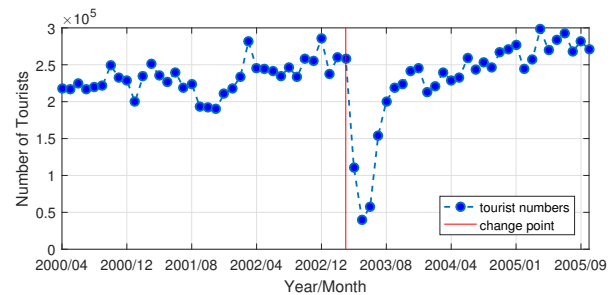


Figure 11. An abrupt change location on Taiwan tourist series.

5.2 Example 2 : Detecting Bali Bombings

Monthly numbers of the tourist arrivals in Bali from January 1982 to August 2019 are shown in Fig. 12A. Overall, the series shows an increasing trend and a seasonal pattern starting round 90s. The data set used in the example is the shorter interval shown in Fig. 12B. The series themselves and the sample ACF and PACF within the period suggest nonstationary and seasonal behavior indicated by the slow decreasing and oscillating ACFs.

5.2.1 Generating Residuals of Bali Data Set

Aiming at eliminating trend on the series which at the same time reveals the seasonal pattern, we apply FTS of order one the short interval. We plot FTS estimates superimposed by the observations and display the residuals in Fig. 13A and Fig. 13B, respectively. The sample ACF and PACF (in Fig. 13C) show a stationary pattern with significant correlations only at lag 1 and 12. One can see that the sample ACF oscillates within the significant bands. This suggests that the entertained model is a seasonal multiplicative ARIMA(1,0,0)(1,0,0)12. We estimate the parameter model using arima and estimate function in Matlab. The result below shows a significant existence of a seasonal autoregressive behavior in the series.

ARIMA(1,0,0) Model Seasonally Integrated with Seasonal AR(12) :

Conditional Probability Distribution: Gaussian

Standard Parameter	t Value	Error	Statistic
Constant	0	Fixed	Fixed
AR{1}	0.219913	0.0873665	2.51713
SAR{12}	-0.466779	0.0767058	-6.08531

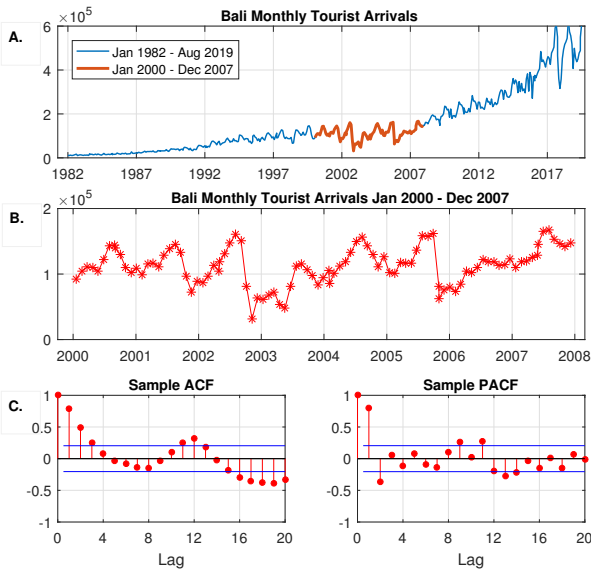


Figure 12. Bali tourist arrivals. (A) A slowly increasing trend and a strong seasonal pattern in the longer time interval. (B) The series under study contaminated by events. (C) The sample ACF and PACF of the short series. The ACFs decreases slowly indicates the existence of a trend in the series.

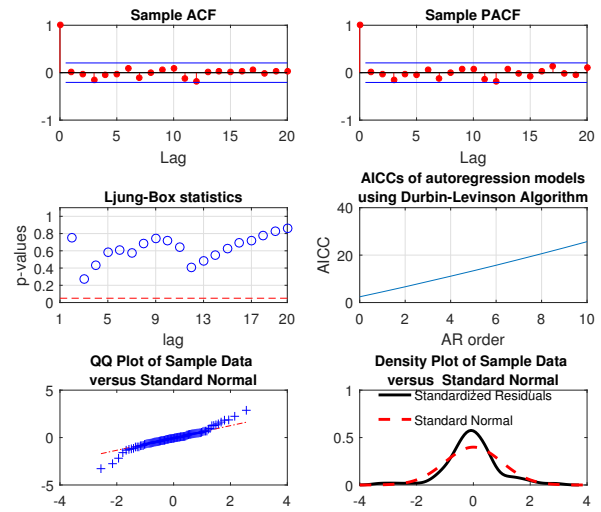


Figure 14. Diagnostic plots of Seasonal ARIMA Residuals.

5.2.2 Detecting Change Points in Bali Data Set

We input the residuals obtained on the previous subsection into the **cpm** package in R. Because we want to detect multiple changes, we treat the residuals as being a data stream which observations are received and processed sequentially. We use **processStream** function. We try to detect using all CPM types assuming normal distribution. A trial signals change points when we use GLR. The CPM model generates statistic tests to detect changes in mean and in variance of an i.i.d Gaussian stream. The R script is attached on the appendix. Plot of the residuals along with the detection times and the change points is shown in Fig. 15. The detection times are signaled at the 34, 67, 72, and 92th observation. Then, the estimated change points are found at the 32, 34, 70 and 72th observation.

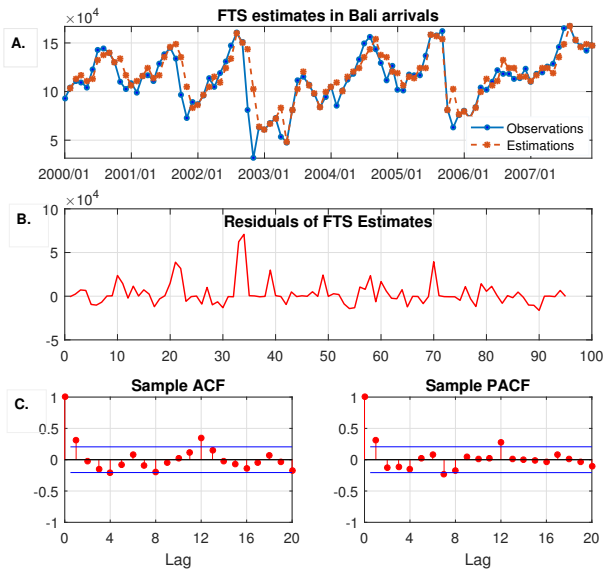


Figure 13. A. FTS estimates, B. Residuals and C. Sample ACF and PACF of the residuals.

Variance 2.26605e+08 8.98863e-11 2.52102e+18

Diagnostic plots of the seasonal ARIMA residuals in Fig. 14 show an i.i.d Gaussian series. The sample ACF and PACF show a white noise phenomenon. The fact is supported by the Ljung-Box *Q* test and the AICC plot that shows uncorrelated observations at all lags up to 20 and up to 10, respectively. The QQ plot and the comparison to the standard normal plot show that the residuals seem to be asymptotically normal.

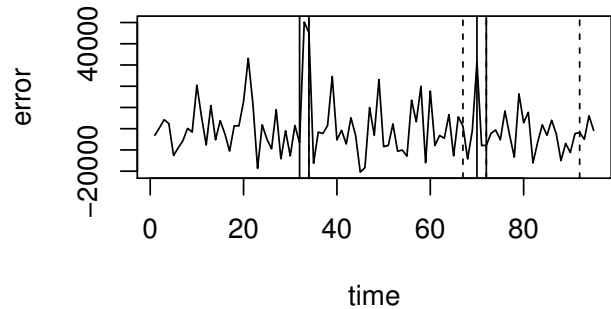


Figure 15. Detection times and change points on the SARIMA residuals. The dashed black lines show the points when the change was detected, and the solid black lines show the estimated location of change points.

We display the change points obtained by the proposed algorithm in the Bali data set. On the raw data, the multiple change points are on the 33, 35, 71 and 72th observation. We need to add 1 to the vector as the input to the **cpm** is the residuals of the second to the last observation of the raw data. Figure 16

shows plot of Bali data set superimposed by the change points. The **cpm** shows that the series undergoes changes in mean and also in variance at 4 different times where the first two times are closed apart so do the last two times. The first group is in 2002/9 and in 2002/11 while the second group is in 2005/11 and 2006/01. Given that the first attack in Bali was in 2002/10 while the second was in 2005/10, the detecting result obtained by the proposed procedure inlines with the facts.

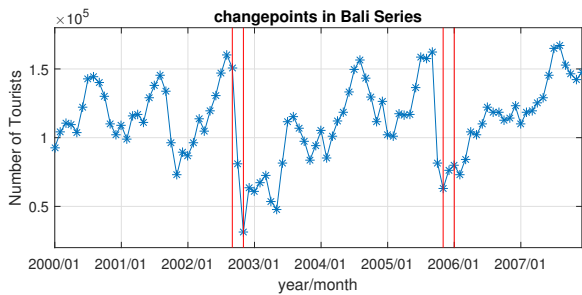


Figure 16. Locations of multiple change points detected on Bali data set detected by the proposed detecting procedure.

5.3 Comparisons to an Existing Detecting Algorithm

5.3.1 Taiwan Data Set

The following R script is used to detect the abrupt change on the Taiwan data set. We input the raw data and choose automatic detection function, **tso()**. The result gives a significant AO (additive outlier) type occurring at time 2003/03. This result is exactly equal to the detected change point obtained by the proposed procedure. This result confirms that the proposed algorithm provides an alternative tool in change detection analysis.

```
> mydata <- fread("Taiwan.csv")
> x0=mydata$shortTaiwan
> Taiwan <- ts(x0, frequency = 12, start = c(2000, 4))
> resTaiwan <- tso(y = Taiwan)
Warning message:
In locate.outliers.iloop(resid = resid, pars = pars,
  cval = cval, :
stopped when 'maxit.iloop' was reached
> resTaiwan
Series: Taiwan
Regression with ARIMA(0,1,0) errors

Coefficients:
A036      TC38
72825.00 -89668.67
s.e. 17887.31 23322.00

sigma^2 estimated as 660199103: log likelihood=-751.23
AIC=1508.45 AICc=1508.85 BIC=1514.98

Outliers:
type ind time coefhat tstat
1 AO 36 2003:03 72825 4.071
2 TC 38 2003:05 -89669 -3.845
```

5.3.2 Bali Data Set

Following [30], we detect outliers in Bali tourist arrival series using the arima framework in the **tso** package. The **tso()** function gives one AO type of outlier effects detected at time

2002/11 and one TC (temporary changes) at time 2003/04. Under the script, it seems that the function recognizes only the first event in Bali bombings.

```
> mydata <- fread("Bali.csv")
> x0=mydata$T
> library(tso)
> Bali <- ts(x0, frequency = 12, start = c(2000, 1))
> resBali <- tso(y = Bali, types = c("AO", "LS", "TC"),
+ maxit = 1, discard.method = "bottom-
+ up",
+ tsmethod = "auto.arima",
+ args.tsmethod = list(allowdrift =
+ FALSE, ic = "bic"))
> resBali
Series: Bali
Regression with ARIMA(1,0,0)(2,1,0)[12] errors

Coefficients:
ar1 sar1 sar2 A035 TC40
0.8465 -0.8802 -0.7618 -25841.77 -30431.379
s.e. 0.0647 0.0742 0.0635 7392.06 7713.324

sigma^2 estimated as 1.02e+08: log likelihood=-903.77
AIC=1819.53 AICc=1820.62 BIC=1834.12

Outliers:
type ind time coefhat tstat
1 AO 35 2002:11 -25842 -3.496
2 TC 40 2003:04 -30431 -3.945
>
```

6 Conclusions

We propose a procedure to detect abrupt changes involving fuzzy approach in change point model (CPM) framework. The proposed fuzzy procedure is used as an estimation model to capture trend in the raw data. We model the change as an AO (additive outlier) type as in time series outlier analysis. We use a parsimonious ARIMA model contaminated by a single abrupt change to simulate the sampling behavior of the auto-correlations of the Fuzzy Time Series (FTS) residuals. We also provide the detecting performance for the proposed detection procedure in the simulation. The proposed procedure gives a satisfying result in detecting the SARS outbreak in the increasing trend series of tourist arrivals in Taiwan. Meanwhile, 4 change points are detected in the seasonal Bali tourism series where the first two points are closed together, as well as the last points. The locations of the group of change points are related to the two well-known terrorist attacks in Bali, respectively. The change points of the two cases are also captured using an automatic detection package of **tso** in R. The package gives an AO type of outliers at exactly similar location to the one obtained by the proposed method for the Taiwan's series. However, the package gives only one AO type of outliers for the Bali's series.

Eliminating trend in time series analysis is a preliminary process in forecasting aimed to get the series stationary in mean. Meanwhile, eliminating trend in change detection in the framework of CPM is aimed to get the i.i.d residuals. We have shown by simulation that FTS of order one, as proposed in this article, can be used to eliminate a trend in a contaminated time series. The residuals can be used later as the input on the CPM to detect the changes. However, some notes need to mention. We define the abrupt change as the AO type of outliers. Only single break located in the middle of the simulated

series is considered in the simulation. Hence, further research on other types of outlier effects with multiple change points is a promising topic. Parameters of the ARIMA model are set such that we work on moderately small variance series. The setting are purposely chosen such that we are able to describe the series. Also more importantly, the simulated model resembles the empirical series which shows a slow increasing trend with and without a seasonal pattern. Therefore, simulating ARIMA models with high variances may offers an opportunity to observe new findings such as implementing other types of FTS models in generating i.i.d residuals. Lastly, given that the variance of ARIMA models is an increasing function of the time observation, we define the minimum breaksize of the abrupt changes as a function of the model standard deviation at the simulated change point and at the last time point. As there is no maximum size of the break, we do not discuss it on the article.

Appendix

1. R Script for Detecting Change Points on Taiwan Data Set

```
rm(list=ls())
setwd("E:/my_study/risetku/FTSmine")
library(data.table)
mydata <- fread("TaiwanTouristError1.csv")
x=mydata$residual

library(cpm)
resultsM <- detectChangePointBatch(x, "GLR", alpha =
  0.05)
#plot statistik Dk,n
plot(resultsM$Ds, type = "l", xlab = "residuals", ylab =
  expression(D[t]), bty = "n")
abline(h = resultsM$threshold, lty = 2)
abline(v = resultsM$changePoint, lty = 2)

cp = resultsM$changePoint
write.csv(cp, "changePointTaiwan.csv")
```

2. R Script for Detecting Change Points on Bali Data Set

```
library(data.table)
mydata <- fread("BaliError.csv")
x=mydata$e

library(astsa) # package for time series
y=sarima(x,1,0,0,1,0,0,12)
summary(y$fit)
SARIMA_error<-resid(y$fit)

library(cpm)
res <- processStream(SARIMA_error,"GLR", 500, 20)
plot(SARIMA_error, type = "l",)
abline(v = res$detectionTimes, lty = 2)
res$detectionTimes

abline(v = res$changePoints)
res$changePoints
```

Acknowledgements

This work was supported by Indonesian Directorate General of Resources for Science, Technology, and Higher Education (DG-RSTHE) under Grant No. 1125/DD3.2/PG/2018.

REFERENCES

- [1] J. Jouini, M. Boutahar, "Evidence on structural changes in US time series," *Economic Modelling*, vol. 22, pp. 391–442, 2005, DOI: 10.1016/j.econmod.2004.06.003
- [2] D. Pambudi, N. MaCaughey, R. Smyth, "Computable general equilibrium estimates of the impact of the Bali bombing on the Indonesian economy," *Tourism Management*, vol. 30, no. 2, pp. 232–239, 2009, DOI: 10.1016/j.tourman.2008.06.007
- [3] R. Smyth, I. Nielsen, V. Mishra, "'I've been to Bali too' (and I will be going back): are terrorist shocks to Bali's tourist arrivals permanent or transitory?," *Applied Economics*, vol. 41, no. 11, pp. 1367–1378, 2009, DOI: 10.1080/00036840601019356
- [4] G. Yu, Z. Schwartz, B. R. Humphreys, "Data patterns and the accuracy of annual tourism demand forecasts," *Tourism Analysis*, vol. 12, no. 1-2, pp. 15–26, 2007, DOI: <https://doi.org/10.3727/108354207780956744>
- [5] M. Basseville, I. V. Nikiforov, "Detection of abrupt changes: theory and application," Prentice Hall Englewood Cliffs, 1996.
- [6] R. S. Tsay, "Outliers, Level Shifts, and Variance Changes in Time Series," *Journal of Forecasting*, vol. 7, no. 1, 1988.
- [7] C. Chen, L.-M. Liu, "Joint Estimation of Model Parameters and Outlier Effects in Time Series," *Journal of the American Statistical Association*, vol. 88, no. 421, pp. 284–297, 1993, DOI: 10.1080/01621459.1993.10594321
- [8] J. Chen, A. K. Gupta, "Parametric Statistical Change Point Analysis: With Application to Genetics, Medicine and Finance," Springer Science+Business Media, 2012.
- [9] D. M. Hawkins, P. Qiu, C. W. Kang, "The changepoint model for statistical process control," *Journal of Quality Technology*, vol. 35, no. 4, pp. 355–366, 2003, DOI: 10.1080/00224065.2003.11980233
- [10] G. J. Ross, "Parametric and Nonparametric Sequential Change Detection in R: The cpm Package," *Journal of Statistical Software*, vol. 66, no. 3, pp. 1–20, 2015, URL: <http://www.jstatsoft.org/>
- [11] P. J. Brockwell, R. A. Davis, "Introduction to Time Series and Forecasting," 3rd ed, Springer, 2016.
- [12] Q. Song, B. S. Chissom, "Forecasting enrollments with fuzzy time series — Part I, Fuzzy Sets and Systems, vol. 54, pp. 1–9, 1993, DOI: 10.1016/0165-0114(93)90355-L
- [13] Q. Song, B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets and Systems*, vol. 54, pp. 269–277, 1993, DOI: 10.1016/0165-0114(93)90372-O
- [14] P. Singh, "A brief review of modeling approaches based on fuzzy time series," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 2, pp. 397–420, 2017, DOI: 10.1007/s13042-015-0332-y
- [15] C. Cappelli, P. D'Urso, F. Di Iorio, "Change point analysis of imprecise time series," *Fuzzy Sets and Systems*, vol. 225, pp. 23–38, 2013, DOI: 10.1016/j.fss.2013.03.001

- [16] C. Cappelli, R. N. Penny, W. S. Rea, M. Reale, "Detecting multiple mean breaks at unknown points in official time series," *Mathematics and Computers in Simulation*, vol. 78, pp. 351–356, 2008, DOI: 10.1016/j.matcom.2008.01.041
- [17] J. Bai, P. Perron, "Estimating and testing linear models with multiple structural changes," *Econometrica*, vol. 66, no. 1, pp. 47–78, 1998, URL: <http://digamo.free.fr/baiperron98.pdf>
- [18] W. S. Rea, "The Application of Atheoretical Regression Trees to Problems in Time Series Analysis," University of Canterbury, 2008.
- [19] B. Wu, M. H. Chen, "Use of fuzzy statistical technique in change periods detection of nonlinear time series," *Applied Mathematics and Computation*, vol. 99, no. 2–3, pp. 241–254, 1999, DOI: 10.1016/S0096-3003(98)00011-3
- [20] B. Wu, "Application of fuzzy time series analysis to change periods detection," in *FUZZ-IEEE'99. 1999 IEEE International Fuzzy Systems. Conference Proceedings (Cat. No.99CH36315)*, vol. 2, pp. 697–702, DOI: 10.1109/FUZZY.1999.793033.
- [21] C. Inclan, G. C. Tiao, "Use of cumulative sums of squares for retrospective detection of changes of variance," *Journal of the American Statistical Association*, vol. 89, no. 427, pp. 913–923, 1994.
- [22] K. Kumar, B. Wu, "Detection of change points in time series analysis with fuzzy statistics," *International Journal of Systems Science*, vol. 32, no. 9, pp. 1185–1192, 2001, DOI: 10.1080/00207720110034698.
- [23] W. W. S. Wei, "Time Series Analysis Univariate and Multivariate Methods," Pearson Education, Inc, 2006.
- [24] L. A. Zadeh. "Fuzzy Sets *," *INFORMATION AND CONTROL*, vol. 8, pp. 338–353, 1965.
- [25] S.-M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*, vol. 81, no. 3, pp. 311–319, 1996, DOI: 10.1016/0165-0114(95)00220-0
- [26] P. Singh, B. Borah, "An efficient time series forecasting model based on fuzzy time series," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 10, pp. 2443–2457, 2013, DOI: 10.1016/j.engappai.2013.07.012
- [27] P. J. Brockwell, R. A. Davis, "Time Series: Theory and Methods," Springer Science+Business Media, 1991.
- [28] K.-H. Huarng, L. Moutinho, T. H.-K. Yu, "An Advanced Approach to Forecasting Tourism Demand in Taiwan," *Journal of Travel & Tourism Marketing*, vol. 21, no. 4, pp. 15–24, 2007, DOI: 10.1300/J073v21n04
- [29] K. Huarng, T. Hui-Kuang Yu, L. Moutinho, Y. Wang, "Forecasting tourism demand by fuzzy time series models," *International Journal of Culture, Tourism and Hospitality Research*, vol. 6, no. 4, pp. 377–388, 2012, DOI: 10.1108/17506181211265095
- [30] J. López-De-Lacalle, "tsoutliers R Package for Detection of Outliers in Time Series," Online available from <https://jalobe.com>, 2019
- [31] A. K. Rana, "Comparative Study on Fuzzy Models for Crop Production Forecasting," *Mathematics and Statistics*, vol. 8, no. 4, pp. 451–457, 2020, DOI: 10.13189/ms.2020.080412
- [32] Y. H. Tseng, P. Durbin, G. H. Tzeng, "Using a fuzzy piecewise regression analysis to predict the nonlinear time-series of turbulent flows with automatic change-point detection," *Flow, Turbulence and Combustion*, vol. 67, no. 2, pp. 81–106, 2002, DOI: 10.1023/A:1014077330409
- [33] Lily Rui Liang, Carl G. Looney, "Competitive fuzzy edge detection," *Applied Soft Computing Journal*, vol. 3, No. 2, pp. 123–137, 2003, DOI: 10.1016/S1568-4946(03)00008-5
- [34] A. S. C. Ling, G. Darmesah, K. P. Chong, C. M. Ho, "Application of ARIMAX Model to Forecast Weekly Cocoa Black Pod Disease Incidence", *Mathematics and Statistics*, vol. 7, no. 4A, pp. 29–40, 2019, DOI:10.13189/ms.2019.070705