

## **ABSTRAK**

### **PENDETEKSIAN TITIK UBAH MENGGUNAKAN RUNTUN WAKTU FUZZY**

Oleh

NURHAIDA

15/390541/SPA/00551

Fuzzy time series (FTS) telah banyak digunakan dalam peramalan runtun waktu, terutama pada runtun waktu dengan trend. Mengingat bahwa tidak ada asumsi yang diperlukan dalam menerapkan model tersebut, maka menjadi menarik untuk menerapkan model tersebut dalam pencarian titik ubah pada data runtun waktu yang terkontaminasi. Namun, FTS yang konvensional perlu dimodifikasi sedemikian rupa sehingga prosedurnya menjadi kurang subjektif dan karenanya dapat digunakan untuk kumpulan data apa pun. Sementara itu, Pendeteksian titik ubah menjadi pendeteksian pencilan dalam analisis runtun waktu atau pendeteksian perubahan dalam statistik proses kontrol. Keduanya berkaitan dengan pencarian lokasi perubahan karena intervensi dalam suatu urutan pengamatan. Distribusi residual masih merupakan masalah dalam analisis pencilan runtun waktu. Di sisi lain, meskipun change point model (CPM), sebuah model untuk penentuan titik ubah yang didasarkan pada uji-dua sampel dalam statistik proses kontrol, menawarkan banyak metode tentatif (dalam mendeteksi perubahan) sehubungan dengan distribusi, CPM memerlukan pengamatan-pengamatan yang saling bebas sebagai input. Jika data mentahnya berkorelasi, CPM dapat diterapkan pada residual atau residual prediksi satu langkah asalkan residualnya saling bebas. Dengan kata lain, diperlukan suatu metode untuk mengestimasi data mentah yang terdapat titik ubah sedemikian rupa sehingga residualnya saling bebas. Dalam disertasi ini akan digunakan model ARIMA. Model tersebut menghasilkan deret nonstasioner homogen yang memiliki variansi error konstan. Pada beberapa nilai parameter, pola tren terlihat jelas pada plot runtun waktunya. Karena adanya titik ubah, data mentah mungkin mengalami perubahan parameter. Oleh karena itu, sebagai metode nonparametrik, mengestimasi data mentah yang terkontaminasi menggunakan FTS tampaknya sebagai suatu solusi. Seperti yang diklaim bahwa FTS memberikan hasil terbaiknya pada pemodelan data dengan tren, orang mungkin berpendapat bahwa FTS dapat digunakan untuk memodelkan runtun waktu dengan tren yang terkontaminasi oleh titik ubah sedemikian sehingga residualnya memenuhi syarat sebagai input dalam CPM. Pernyataan berikut adalah tujuan disertasi ini: merancang metode pendeteksian titik ubah menggunakan FTS; menerapkan metode pendeteksian tersebut pada data riil. Disertasi ini menggunakan FTS yang dimodifikasi dalam mendeteksi titik ubah pada kerangka CPM. Metode pendeteksian terdiri dari dua bagian utama: (1) Mengha-



silkan residual yang i.i.d (*independently identically distributed*) menggunakan FTS, (2) Mendeteksi titik perubahan menggunakan CPM. Untuk menunjukkan performa metode pendeteksian yang diusulkan, maka dirancang sebuah studi simulasi. Model ARIMA (1,1,1) pada kondisi stabilnya disajikan sebagai runtun waktu yang mendasari data mentah. Model titik ubah yang mirip dengan model tipe pencilan Additive Outlier (AO) didefinisikan. Standar deviasi process ARIMA(1,1,1) pada waktu T digunakan dalam model titik ubah tersebut. Variansi ARIMA (1,1,1) diberikan dalam teorema dan dibuktikan. Beberapa akibat terkait variansi ARIMA(1,1,1) diberikan dan dibuktikan. *Break size* yang menyatakan dinamika intervensi diberikan dalam teorema. Performa pendeteksian diberikan dalam bentuk persentase pendeteksian titik ubah yang tepat dan persentase pendeteksian interval yang tepat. Dalam simulasi, sebuah titik ubah pada waktu T terletak di tengah runtun waktu digunakan. Hasil penelitian menunjukkan bahwa penerapan FTS yang dimodifikasi pada model ARIMA (1,1,1) yang terkontaminasi oleh titik ubah dapat menghasilkan residual yang i.i.d. Persentase residual yang i.i.d ini lebih tinggi ketika model ARIMA (1,1,1) memiliki pola tren, khususnya, yaitu ketika parameter AR,  $\phi$  positif dan parameter MA,  $\theta$ , negatif. Selanjutnya, metode pendeteksian yang diusulkan memberikan persentase pendeteksian yang lebih tinggi ketika parameter model  $\phi$  positif dan  $\theta$  negatif. Khususnya, jika runtun waktunya adalah ARIMA (1,1,1) dengan  $\phi = 0.9$  dan  $\theta = (-1, 0]$ , power dari prosedur untuk mendeteksi titik ubah pada waktu T, dengan *break size* sama dengan 3,5 kali standar deviasi proses pada waktu T, berkisar antara 49% hingga 60%. Sebuah peningkatan yang substansial dalam persentase pendeteksian terjadi jika pendeteksian diamati pada level interval perubahan. Persentase meningkat menjadi 81%–87%, 90%–95%, dan 94%–99%, untuk interval  $T \pm 1$ ,  $T \pm 2$  dan  $T \pm 4$  di sekitar titik ubah, berturut-turut. Juga, jelas bahwa persentase pendeteksian menaik seiring meningkatnya *break size*. Penerapan metode pendeteksian yang diusulkan pada data riil dilakukan pada dua data runtun waktu pariwisata, yaitu di Taiwan dan di Bali. Perbandingan dengan metode pendeteksian lainnya dilakukan dengan mendeteksi titik ubah data riil tersebut menggunakan paket **tsoutlier**, sebuah paket pendeteksian titik ubah otomatis yang tersedia pada program R. Hasilnya menunjukkan bahwa metode yang diusulkan berkinerja lebih baik daripada metode pendeteksian yang telah ada.

**Kata Kunci:** FTS, CPM, ARIMA, titik ubah, outlier.

## ABSTRACT

### CHANGE POINT DETECTION USING FUZZY TIME SERIES

By

NURHAIDA

15/390541/SPA/00551

Fuzzy time series (FTS) has been largely deployed in forecasting time series, particularly for series with trend. Given that no assumption is needed in applying the model, it is interested to employ the model to search for changes in a contaminated series. However, the conventional FTS needs to be modified such that the procedure is less subjective and hence it can be used for any data set. Meanwhile, detecting changes becomes detecting outliers in time series or detecting change points in statistical process control. Both deals with searching for the location of changes due to intervention in a sequence of observations. It has been said that distribution on the residuals of a time series model is an issue in time series outlier analysis. On the other hand, even though change point model (CPM), a change point formulation which is based on the two sample t-test in statistical process control, offers lots of tentative methods in detecting change points with respect to distributions, CPM requires independent observations for the input. In case that the raw data is correlated, CPM is applied on the residuals or the one-step ahead prediction residuals inasmuch the residuals are independent. In other word, there is a need of a method to estimate raw data such that the residuals are independent conditional on the change points. In this dissertation, ARIMA models are considered. The models generate homogeneous nonstationary series which have constant error variance. In some values of the parameters, patterns of trend are clearly seen on the time series plots. Due to the existence of change points, raw data may experience changes in the parameter(s). Hence, as a nonparametric method, estimating the contaminated raw data by FTS seems as a solution. As claimed that FTS is at its best in modelling data with trend, one may argue that FTS can be used to model a series with trend contaminated by change points which the residuals are eligible for the input in CPM to detect for the points. The following statements state the targets for this dissertation: to design a change detection method using FTS; to apply the detection method on the real data set. This dissertation employs a modified FTS in detecting change points in the framework of CPM. The detecting method consists of two main tasks: (1) Generating i.i.d residuals involving FTS, (2) Detecting change points using CPM. To show the performance of the proposed detecting method, a simulation study is designed. The ARIMA(1,1,1) models on their stable condition are served as the underlying series. A model of the change point which is closed to the Additive Outlier (AO) type of outliers is defined. The standard deviation of the process at time  $T$  is con-



sidered in the model. The variance of  $ARIMA(1,1,1)$  is given in a theorem and the proof is provided. A corollary related to the variance of the  $ARIMA$  is given and proved. The break size that represents the dynamic of the intervention is given in a theorem. The performance of the detecting method is given as the percentage of correct change point detection and the percentage of correct change interval detection. In the simulation, a single change point at time  $T$  in the middle of the series is set. The result shows that the modified FTS is able to generate i.i.d residuals on  $ARIMA(1,1,1)$  models which are contaminated by a change point. The percentage of generating i.i.d residuals is higher when the  $ARIMA(1,1,1)$  model shows a visible trend pattern. In particular, the model parameters are in positive values of the AR parameters,  $\phi$  and the negative values of the MA parameters,  $\theta$ . The proposed detecting method applied on  $ARIMA(1,1,1)$  model gives a higher percentage of correct detection when the model parameter  $\phi$  is positive and  $\theta$  is negative. Particularly, when the underlying series is an  $ARIMA(1,1,1)$  with  $\phi = 0.9$  and  $-1 < \theta \leq 0$ , the power of the procedure for detecting a change point at time  $T$ , which the break size is equal to 3.5 standard deviation of the process at time  $T$ , ranges between 49% to 60%. A substantial increase in the percentage of detection is observed when detection on the change interval is considered. The percentage increases to 81% – 87%, 90% – 95%, and 94% – 99%, for correct change interval at the  $T \pm 1$ ,  $T \pm 2$  and  $T \pm 4$  around the change point, respectively. Also, it is clear that the percentage is an increasing function of the break size. An application of the proposed detecting method on real data is done on two tourism series, i.e., in Taiwan and in Bali. A comparison to other detection method is done by detecting the change points on the real series in tso package, an automatic detection package in R. The result shows that the proposed method performs better than the established detection method.

**Keywords:** FTS, CPM,  $ARIMA$ , change point, outlier.