COMPARISON BETWEEN ITERATIVE AND RECURSIVE ALGORITHM APPLICATION IN CHANGE POINT ANALYSIS FOR EXTREME VALUES

Agra Yuriandana[1)]**Aji Hamin Wigena**[2)]

[1)] Department of Statistics, Faculty of Mathematics and Natural Science, Bogor Agricultural University
Jl. Raya Darmaga, Bogor, Jawa Barat 16680
email : agra.statisticians@gmail.com
[2)] Department of Statistics, Faculty of Mathematics and Natural Science, Bogor Agricultural University
Jl. Raya Darmaga, Bogor, Jawa Barat 16680
email : ajiwigena@ymail.com

ABSTRACT

Change Point Analysis (CPA) method with Generalized Pareto Distribution (GPD) approach can detect extreme points of changes. However, this method can only detect a single extreme point. To be able to detect more than one extreme point, the steps of the method need to be repeated. To accomplish that, we need to modify the CPA program using a recursive algorithm or iterative algorithms. This paper discusses CPA with GPD approach for daily Jakarta Composite Index in 2014 using recursive and iterative algorithms. The results of both algorithms are compared. CPA method with recursive and iterative algorithms generated the same extreme points of changes. The Recursive algorithm required longer execution time than the iterative algorithm.However the first algorithm is shorter and simpler than the second algorithm.

Keywords

CPA, GPD, iterative algorithm, recursive algorithm

1. Backgrounds

Stock is one of the financial market instruments well known by the public. The instrument is widely used by people to make profits. The stock price will fluctuate over time. The fluctuation caused by many factors. One of them is the sale and purchased by investors. To prevent losses, investors have to be smart to choose the company issuing stocks and choosing the right time to sell and buy stocks. Jakarta Composite Index (JCI) is an indicator of stock price movement in Jakarta Stock Exchange.The index includes common stock price movement throughout and preferred stocks are listed on the Indonesia Stock Exchange.

Changes in the JCI values can be used as an indicator of financial market in Indonesia. To see the change in these values, there is a method called the Change Point Analysis (CPA) founded by Taylor. He introduced this method based on the cumulative sum (CUSUM). The advantage of the method is able to determine some point of changes at once. Then, Prayogo conducted a research using the CPA on rainfall data based on Generalized Pareto Distribution (GPD). AccordingtoDierckx and Teugels, CPA method with GPD approach is only able to capture one point of changes.If there are more than one point, this method can only detect the most significant change [1]. To capture all points of changes from a dataset, it requires repeating all steps in the method. The repetition can be performed by a recursive algorithm and an iterative algorithm. Both algorithms have their advantages and disadvantages. In this research, both algorithms were applied in the CPA method for daily JCI 2014.

2. Literature Review

2.1 Change Point Analysis

Change point analysis is a method to find a point when a change of value occurred. In the year of 2000, Taylor introduces this CPA method. The change that CPA method analysed is the change of the average value. CPA is the combination of cumulative sum method (CUSUM) and bootstrap. CUSUM is used to detect changes and bootstrap is used to inspect hypothesis. This CPA method cannot be used to detect extreme values sothis research uses GPD approach.

Detection of extreme points of changes using a statistical test based on likelihood.If X_i ~ GPD with parameters $\theta_i = (\xi_i, \sigma_i)$ so the hypothesis is:

$H_0: \theta_1 = \theta_2 = \cdots = \theta_n$ (changes not occurred)

$H_1: \theta_1 = \theta_m \neq \theta_{m+1} = \cdots = \theta_n$ (changes occurred at m)

with the following test statistic,according to Csörgö and Horváth [2]:

$$Z_n = \sqrt{\max_{1<m\leq n}(-2\log \Lambda_m)}$$

According to Dierckx and Teugels, the value of $-2\log\Lambda_m$ is [3]:

$$-2\log\Lambda_m = 2[L_m(\hat{\theta}_m) + L_m^+(\hat{\theta}_m^+) - L_n(\hat{\theta}_n)]$$

with

$$L_m(\hat{\theta}_m) = -m\log\hat{\sigma}_m - (\frac{1}{\hat{\xi}_m}+1)\sum_{i=1}^{m}\log(1+\hat{\xi}_m\frac{E_i}{\hat{\sigma}_m})$$

$$L_m^+(\hat{\theta}_m^+) = -(n-m)\log\hat{\sigma}_m^+ - \left(\frac{1}{\hat{\xi}_m^+}+1\right)\sum_{i=m+1}^{n}\log\left(1+\hat{\xi}_m^+\frac{E_i}{\hat{\sigma}_m^+}\right)$$

$$L_n(\hat{\theta}_n) = -n\log\hat{\sigma}_n - \left(\frac{1}{\hat{\xi}_n}+1\right)\sum_{i=1}^{n}\log\left(1+\hat{\xi}_n\frac{E_i}{\hat{\sigma}_n}\right)$$

The value of $(\hat{\sigma}_m, \hat{\xi}_m)$ and $(\hat{\sigma}_m^+, \hat{\xi}_m^+)$ is obtained based on likelihood estimator from $X_1, X_2, \ldots, X_m$ and $X_{m+1}, \ldots, X_n$ while $E_i$is obtained from the equation

$$E_i = \begin{cases} x_i - u, & x_i > u \\ 0, & x_i < u \end{cases}$$

According to CsörgödanHorváth, p-value is calculated by the following equation [4]:

$$\frac{Z_n^2\exp\left(-\frac{Z_n^2}{2}\right)}{2}\left[log\frac{(1-c)(1-d)}{cd} - \frac{2}{Z_n^2}log\frac{(1-c)(1-d)}{cd} + \frac{4}{Z_n^2} + O\left(\frac{1}{Z_n^4}\right)\right]$$

withZ_n is the test statistic value, $O\left(\frac{1}{Z_n^4}\right) = \frac{1}{Z_n^4}$ and the value of $c = d = \frac{(\log k)^{3/2}}{k}$.

2.2 Generalized Pareto Distribution

Coles says that such $X_1, \ldots, X_n$ is random variables and u is a threshold value, then the distribution function of $y = X - u$ with $X > u$ is [5]:

$$F(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\tilde{\sigma}}\right), & \xi = 0 \end{cases}$$

σ is scale parameter, ξ is shape parameter.

F(y) is a cumulative function form of GPD with y = x-u and greater than 0, $\left(1+\frac{\xi y}{\tilde{\sigma}}\right)>0$. If k is the amount of the sample point that exceeds the limit u, then $y_1, \ldots, y_k$ will have a density function as below

$$f(y) = \begin{cases} \frac{1}{\tilde{\sigma}}\left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\tilde{\sigma}}\exp\left(-\frac{y}{\tilde{\sigma}}\right), & \xi = 0 \end{cases}$$

To analyse the parameter of GPD, the maximum likelihood method can be used. The probability of logarithmic function from$y_1, \ldots, y_k$ is:

$$l(\tilde{\sigma}, \xi) = \begin{cases} -k\log\tilde{\sigma} - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi y_i}{\tilde{\sigma}}\right), & \xi \neq 0 \\ -k\log\tilde{\sigma} - \tilde{\sigma}^{-1}\sum_{i=1}^{k}y_i, & \xi = 0 \end{cases}$$

$F(y)$iscumulative function of GPD with y = x - u and y > 0, $\left(1+\frac{\xi y}{\tilde{\sigma}}\right)>0$. Let k is the number of observations which their values exceed the threshold u, then $y_1, \ldots, y_k$ will have the GPD density function as follows:

$$f(y) = \begin{cases} \frac{1}{\tilde{\sigma}}\left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\tilde{\sigma}}\exp\left(-\frac{y}{\tilde{\sigma}}\right), & \xi = 0 \end{cases}$$

Parameter estimation of the GPD can be used a maximum likelihood method. Log likelihood function of $y_1, \ldots, y_k$ is:

$$l(\tilde{\sigma}, \xi) = \begin{cases} -k\log\tilde{\sigma} - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi y_i}{\tilde{\sigma}}\right), & \xi \neq 0 \\ -k\log\tilde{\sigma} - \tilde{\sigma}^{-1}\sum_{i=1}^{k}y_i, & \xi = 0 \end{cases}$$

3. Results

In the previous research,Prayogodeveloped the program in R to perform CPA method with GPD approach. However, this function can only detect one extreme point of changes as developed by Dierckx and Teugels. That program needs to be modified such that the program can detect all extreme points of changes in a dataset. The program developed by Prayogo will be the base of the proposed program to be developed in this paper.

In general, the logic of the proposed programis as follows. First, create a vector **m** that will be filled by the index points of changes of extreme values. Then specify the lower limit (*a*) and the upper limit (*b*) then the interval of both limits will be analysed by CPA.The value of *a* starts from 1 and the value of*b*isequal to the number of data (*n*). Then,these steps are repeated with the following requirements:

1) if changes in extreme values at the **m** points are found, then a = a and b = the smallest **m** value that is greater than a

2) if changes in extreme value at the a point is found, then a = a + 1 and b = the smallest **m** value that is greater than a
3) if changes in extreme values is not found, then a= b values of the previous process and b = the smallest **m** value that is greater than a
4) if changes in extreme values is not found, while the smallest **m** value greater than b in the previous process does not exist; then a = b values of the previous process and b = n.

The above procedure is implemented usingthe following iterative function as shown below:

```
DTI<-function(x,u){
n<-length(x)
a<-1
b<-n
M<-rep(0,n)
while(a<b){
 M<-sort(M)
 m<-DT(x[a:b],u)
if (!is.null(m)){
M[1]<- m+a-1
if (any(M==a)) a=a+1
  b1<-M>a
if(any(b1)==F){
   M<-M[M>0]
   }
  b2<-M[b1]
  b<-min(b2)
 } else {
  a<-b+1
if(a>n){
   M<-M[M>0]
   }
  b1<-M[1]>a
if (any(b1)){
   b2<-M[b1]
   b<-min(b2)
   }
else
   b<-n
  }
 }
 M
}
```

Beside the iterative function, it can also be implemented using therecursive function as follows:

```
DTR <- function (x,u,a,b) {
 m<-DT(x[a:b],u)
if (b>a&& !is.null(m)) {
 p<- c(a-1+m,DTR(x,u,a,a+m-1),
     DTR(x,u,a+m,b))
  } else {p<- NULL}
 P
}
```

There are few modifications related to the variables x, u, a, and b of DT functiondevelopedby Prayogousedin the DTR recursive function. The variable x is data that will be used, the variable u is the threshold value, the variable a is the lower limit, andhe variable b is the upper limit value.The iterative function and recursive function above will produce extreme points of changes butthe codes in the iterative function are longer than that in the recursive function.

The results of CPA for daily JCI in 2014 with the threshold of 5.172,288 are in Table 1 and the execution time of both functions is in Table 2.

Table 1The Results of CPA for Daily JCI in 2014

Points of Change (days)	Date	Value	Previous Value	Afterward Value	Zn
162	04/09/2014	5205.32	5224.13	5217.33	11.74397
156	27/08/2014	5165.25	5146.55	5184.48	13.29458
174	22/09/2014	5219.80	5227.58	5188.11	6.578243
170	16/09/2014	5130.50	5144.90	5188.18	3.730053
227	04/12/2014	5177.16	5166.04	5187.99	4.646202
177	25/09/2014	5201.38	5174.01	5132.56	6.545606

Table 2 The CPA for Daily JCI in 2014's Execution Time

Method	User	System	Elapsed
Recursive	226.68	1.65	229.11
Iterative	223.89	1.22	226.04

Table 1 shows that there are six points (156, 162, 170, 174, 177, and 227) of change are detected. The same results were obtained using either the recursive algorithm or iterative algorithm. However, their differences are only the execution time.Theelapsed time of iterative algorithm was faster than that of recursive algorithm.

The values of the detected points of change are not necessarily the highest values compared to the values before and after. For example on 27 August 2014, the detected value is 5165.25 while the previous value (5146.55) and the afterward value (5184.48), and the value (5130.50) on 16 September 2014 is lower than the value before (5144.90) and after (5188.18). This is due to the nature of the CPA method that will detect changes in the maximum extreme value and did not seek the maximum extreme value. This detection is derived from the adjustment of data of the daily JCI in 2014 with the Generalized Pareto Distribution. So it can be said that the detected value is the point when a change which could be higher or lower. The results can also be seen in Figure 3.

Figure 3 shows the time series plot of daily JCI in 2014 with the CPA results. The points of changes are detected not only in the local maximum extreme values as described previously. There are six points that are detected as the points of maximum extreme changes, but five points are adjacent one to another and the other is at the right end. This is possible because the data from points 143 to 183 have the change of each value which is large enough and many values have largevaluesthat exceeds the threshold, so that the maximum extreme value changes are mostly detected in that interval.

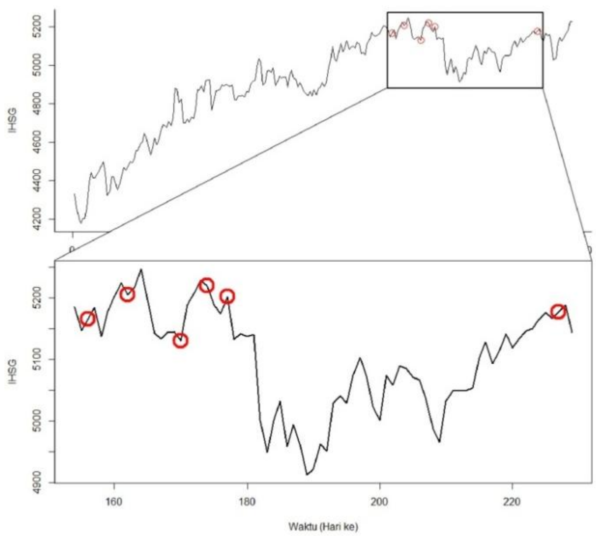

Figure 3 Time Series Plot of Daily JCI in 2014 with The CPA Result

The detection of the extreme point of change is based on the calculated test statistic (Zn). Every time CPA is implemented, the value of Zn at each point of data is computed. A Zn value can be interpreted as the value of test statistic that a change occurs at the point. The greater the value of Zn means the bigger the change. For example, in the first repetition the highest Zn value (11.74397) is at the point of 162 shown in Figure 3. The highest Zn value could be the extreme point of change. However, the highest value of Zn in the first repetition is not necessarily to be the highest value of Zn compared to the highest value of Zn in the next repetitions.

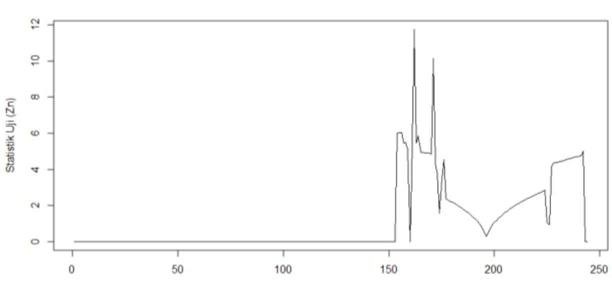

Figure 4 Test Statistic (Zn) Plot

4. Conclusion

The Change Point Analysis method with the Generalized Pareto Distribution approach is able to detect changes of the maximum extreme values. For daily data of JCI in 2014 using 5172,288 as the threshold, the change in maximum extreme values are obtained on August 27, September 4, September 16, September 22, September 25, and December 4, 2014. The use of the recursive and iterative algorithms to detect of change in the maximum extreme values gave the same results. However, the recursive algorithm needs a longer execution time compared to the iterative algorithm. This difference is due to the recursive algorithm takes more time and memory to call the function and takes time to make the return value. The advantage of the recursive algorithm is more elegant and easier to read than the iterative algorithm. Of course, to develop a program with recursive algorithm requires more thought to have a concise program without reducing the content and capabilities of the program.

REFFERENCE

[1] Dierckx G., Teugels JL, 2010, Change Point Analysis of Extreme Values, Environmetrics, 21(7-8): 661–686.
[2] Csörgö M., Horváth L, 1997, Limit Theorems in Change-Point Analysis, Chichester (EN): J Wiley.
[3] Dierckx G., Teugels JL, 2010, Change Point Analysis of Extreme Values, Environmetrics. 21(7-8): 661–686.
[4] Csörgö M., Horváth L, 1997, Limit Theorems in Change-Point Analysis, Chichester (EN): J Wiley.
[5] Coles S, 2001, An Introduction to Statistical Modelling of Extreme Values, London (EN): Springer.

Agra Yuriandana,S.StatdegreefromDepartment of Statistics at Bogor Agricultural University in 2015.
AjiHamimWigena,Irdegree fromBogor Agricultural University in 1976, Master degree fromAsian Institute of Technology,Thailand in 1981, andDoctoral degreefromBogor Agricultural University in 2006;and currently as lecturer at Department of Statistics, Bogor Agricultural University.