



World Scientific News

An International Scientific Journal

WSN 125 (2019) 217-229

EISSN 2392-2192

An Application of Bootstrapping for CUSUM Test in Mean Change-Point Model and Forecasting

Rumana Rois*, Afsana Tasnim Prima, Munia Afroza Shanta

Department of Statistics, Jahangirnagar University, Savar, Dhaka - 1342, Bangladesh

*E-mail address: rois@juniv.edu

ABSTRACT

Application of change-point analysis increases as more data sets are collected in a wide variety of fields. Detection of change-point is useful in modeling and prediction of time series, especially, it has a significant impact on forecasting. The critical value of a test is required to conduct that test in detecting change-point. The calculation of the critical values is based on the distributional asymptotics of the test statistics under the null hypothesis. Cumulative sum (CUSUM) test is a popular change-point test in location model. The convergence of the limit distribution of the CUSUM test statistic is rather slow. Antoch and Hušková (2001) suggested that the critical value of the permutation test, a test based on the bootstrap principle, performs better than the asymptotic critical value of CUSUM test in location model. Inspired by them, we consider a change in the mean with i.i.d. errors to evaluate the performance of the bootstrap and the asymptotic critical values of CUSUM test to the simulated and real data. We used the monthly average rainfall in Cumilla, a district in Bangladesh, from 1948 to 2013 as a real data. We also motivated to develop a forecasting model taking into account of the detected change-point. The result demonstrates that the performance of the bootstrap critical value of CUSUM test is better than the asymptotic one for both the simulated and real data. Moreover, the accuracy of the monthly average rainfall forecasting in Cumilla is improved by considering the valid change-point in modeling.

Keywords: change-point, CUSUM test, AMOC model, asymptotic critical value, bootstrap critical value, extreme value distribution

1. INTRODUCTION

Change-point analysis is usable to detect if there is any change and to locate the change-point in a model. Application of change-point analysis increases as more data sets are collected in a wide variety of fields, for instance, time series (Franke, *et al.*, 2012), economics (Bai and Perron, 2003), finance (Andreou and Ghysels, 2002), climatology (Reeves, *et al.*, 2017), engineering (Stoumbos, *et al.*, 2000), image analysis (Aston and Kirch, 2012) and epidemiology (Sonesson and Bock, 2003).

The problem of abrupt changes in general arises in quite a variety of models within time-ordered observations, for instance, changes in mean with known or unknown starting value and variance, changes in variance, changes in mean and variance, changes in location and/or scale, changes in correlation coefficient, changes in regression coefficient. In particular, we are interested in the well-known cumulative sum (CUSUM) test, a nonparametric test, to apply for detecting change in the At-Most-One-Change (AMOC) mean model.

Change-point detection is closely related to the two well-known problems of statistical inference, i.e., test of hypothesis (change-point testing) and estimation (change-point estimation). In detecting a change-point, the critical value(s) of the test statistic is required to conduct the test.

The calculation of the critical values is based on the distributional asymptotics of the test statistics under the null hypothesis. Cumulative sum (CUSUM) test is a popular change-point test in location model with errors different distributions. We consider a change in the mean with independent and identically distributed (i.i.d.) errors, more specifically, the AMOC model with i.i.d. errors. The asymptotic distribution of the CUSUM test statistic for the single mean change-point model is either follow extreme value type distribution or Brownian Bridge type distribution. The convergence of the limit distribution of the CUSUM test statistic is rather slow. Antoch and Hušková (2001) suggested that the critical value of the permutation test, a test based on the bootstrap principle, performs better than the asymptotic critical value of the CUSUM test in location model. Inspired by them, we are motivated to calculate the bootstrap critical value of the CUSUM test to the simulated and real data.

Bangladesh is an agricultural country and the prosperity of agriculture depends on rainfall and humidity. Consequently, rainfall forecasting is an important topic in Bangladesh economy. Various types of forecasting models are available for rainfall prediction in Bangladesh, but no one taking the change-point into account. However, the existence of change-point in the data decreases the accuracy of time series forecasting. Hence, we consider the monthly average rainfall in Cumilla (formally addressed as Comilla), a district in Bangladesh, during the period Jan 1948 – Dec 2013 as our real data. Our work focuses on (i) the change-point detection using the bootstrap critical value rather than the asymptotic one - hence evaluating the performance of these two types of critical values, (ii) estimation of that change-point, and (iii) evaluation of a forecasting model for the average rainfall in Cumilla by considering the detected change-point – which will be improved and simplified forecasting process compared to its existing counterparts.

The organization of this paper is as follows: Section 2 reviews different CUSUM tests with their asymptotic critical values. The bootstrap critical value is illustrated with an algorithm in Section 3. Result and analysis is explored in Section 4. Finally, the paper concludes with some discussions in Section 5.

2. CUSUM TEST FOR SINGLE CHANGE-POINT IN MEAN

The classical change-point model with single change in mean, known as At-Most-One-Change (AMOC) mean model, is defined by

$$X_i = \begin{cases} \mu + e_i & 1 \leq i \leq m \\ \mu + \delta + e_i & m < i \leq n, \end{cases} \quad (2.1)$$

where: μ , δ and $m \leq n$ are unknown, n is the total number of observations and known, and m is called the change-point. In nonparametric settings, we assume that errors $\{e_i : i = 1, \dots, n\}$ are i.i.d. but non-observable with

$$E(e_i) = 0, \quad 0 < E(e_i^2) = \sigma^2 < \infty, \quad E(e_i)^\nu < \infty \text{ for some } \nu > 2. \quad (2.2)$$

Most of the statistics were originally developed for independent normal errors, but it can be shown that the statistics work for all non-degenerate sequences of i.i.d. errors as long as the ν^{th} moment ($\nu > 2$) exists, for details confer Csörgö and Horváth (1997). Hereafter, we need to define a hypothesis to detect the valid change-point m and then conduct the test of hypothesis with the suitable test statistic.

2. 1. Test Hypothesis

In the detection or testing of change-point in mean, we decide whether the sequence of observations is homogeneous or not with the following test hypothesis

$$\begin{aligned} H_0 : \mu_1 = \dots = \mu_n = \mu, \quad \text{against} \\ H_1 : \mu_1 = \dots = \mu_m = \mu; \quad \mu_{m+1} = \dots = \mu_n = \mu + \delta, \end{aligned} \quad (2.3)$$

where $m = 1, \dots, n-1$ is an unknown index of change-point, the initial value μ and δ (either $\delta > 0$ or $\delta \neq 0$) are also unknown. The decision whether the sequence of observations is homogeneous or not is based on the test statistics. In the following subsection, we make a brief review on different types of CUSUM test statistics used to detect a change in the AMOC mean model in context of offline nonparametric inferences.

2. 2. Test Statistics

The most common statistics usually applied for the location model in nonparametric inferences are the pseudo maximum-likelihood method and the pseudo-Bayesian method, which are the max-type statistics and the sum-type statistics, respectively. To derive pseudo maximum-likelihood statistics, a common practice is to consider independent and identically distributed (i.i.d.) standard normal errors first and then prove that the statistic derived under assumption (2.2) still gives valid results for different distribution. In particular, we are interested in the well known cumulative sum (CUSUM) statistics to apply for testing change in the AMOC mean model (2.1).

The CUSUM test statistic was initially proposed by Page (1954, 1955 and 1957) in the context of quality control. Hereafter, various authors used that statistic in different contexts with different types, see e.g., Hušková (197), Berkes *et al.*, (2011), Horváth and Rice (2014), and Berkes *et al.*, (2009), etc. We focus on the classical CUSUM, weighted CUSUM and CUSUM R-type tests to detect the change-point in the AMOC mean model (2.1) with the hypothesis (2.3).

The classical CUSUM test statistics is defined as

$$T_n^{(1)} = \max_{1 \leq m \leq n} \left\{ \frac{1}{\hat{\sigma}_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^m (X_i - \bar{X}_n) \right| \right\}, \tag{2.4}$$

where $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ and $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. The classical CUSUM test statistics follows a Brownian bridge distribution. The weighted CUSUM test statistic also based on the max-type principle with an unknown k and express as

$$T_n^{(2)} = \max_{1 \leq k < n} \left\{ \frac{1}{\hat{\sigma}_n} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k (X_i - \bar{X}_n) \right| \right\}. \tag{2.5}$$

CUSUM R-type statistic is a robust method which is based on the simple linear rank statistics. The CUSUM R-type statistic is defined by

$$T_n^{(3)} = \max_{1 \leq k < n-1} \left\{ \frac{1}{\sigma_{n,R}} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k a_n(R_i) - \bar{a}_n \right| \right\}, \tag{2.6}$$

where (R_1, \dots, R_n) is the vector of ranks corresponding to the observation X_1, \dots, X_n the scores $a_n(1), \dots, a_n(n)$ are typically defined either by Wilcoxon scores or van der Waerden scores with

$\bar{a}_n = \frac{1}{n} \sum_{i=1}^n a_n(i)$. Using Wilcoxon scores $a_n(i) = \frac{i}{(n+1)}$, $i = 1, \dots, n$, and

$$\sigma_{n,R}^2 = \frac{1}{(n-1)} \sum_{i=1}^n (a_n(i) - \bar{a}_n)^2.$$

2. 3. Asymptotic Critical Values

The approximate critical values can be calculated for large n and $t \in R^1$ from the asymptotic behavior of the probabilities under H_0 in (2.3) using Darling and Erdős (1959)

$$P(T_n^{(2)} A(\log n) \leq t + D(\log n)) \approx \exp(-2e^{-t}), \tag{2.7}$$

where: $A(x) = \sqrt{2 \log x}$, $D(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi$, and $T_n^{(2)}$ is defined in (2.5). It is assumed that $\max_{1 \leq k < n} |\hat{\sigma}_n^2 - \sigma_n^2| = o_p((\log \log n)^{-n})$, for details confer Theorems 1.3.1 and 1.4.2 of Csörgö and Horváth (1997). Hence, the asymptotic critical value at α level is

$$c_\alpha = \frac{1}{A(\log n)} \left(-\log \left(-\frac{\log(1-\alpha)}{2} \right) + D(\log n) \right). \tag{2.8}$$

Under H_0 , the CUSUM R-type statistic follows an extreme-value type distribution for large n and can be defined as

$$P(T_n^{(3)} A(\log n) \leq t + D(\log n)) \approx \exp(-2e^{-t}). \tag{2.9}$$

2. 4. Change-Point Estimation

Using a convention (see for example, Dumbgen, 1991), the estimator for change-point in mean is

$$\hat{m} = \arg \max \left\{ \sqrt{\frac{n}{k(n-k)}} \left| \sum_{i=1}^k (X_i - \bar{X}_n) \right|; k \in \{1, \dots, n-1\} \right\}. \tag{2.10}$$

and the estimator for change-point in mean using the CUSUM R-type statistic (2.6) is

$$\hat{m} = \arg \max \left\{ \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k (a_n(R_i) - \bar{a}_n) \right|; k \in \{1, \dots, n-1\} \right\}. \tag{2.11}$$

3. BOOTSTRAPPED CRITICAL VALUES

Finding good approximations to the critical values is one of the leading problems in hypothesis testing. In change-point literature, the distributions of test statistics are very complex even in those cases where the assumption of normal errors is fulfilled, so that they can be determined explicitly only for small sample sizes. Consequently, the asymptotic critical values are only good approximations for large sample sizes, otherwise they fail. Therefore, resampling becomes one of the reasonable possibilities to approximate the critical values. In the context of change-point, this approach was first suggested by Antoch and Hušková in 2001 and later pursued by others, see for example, Hušková and Picek, (2005), Hušková (2004), Kirch and Steinebach (2006) and Kirch (2006). Permutation and bootstrap tests have been developed in such approximations.

Antoch and Hušková (2001) developed the permutation distribution of $T_n^{(2)}$, which is described as the conditional distribution (given X_1, X_2, \dots, X_n) of

$$T_{nR}^{(2)} = \max_{1 \leq k < n} \left\{ \frac{1}{\hat{\sigma}_n} \left| \sqrt{\frac{n}{k(n-k)}} \sum_{i=1}^k (X_{R_i} - \bar{X}_n) \right| \right\}, \quad (2.12)$$

where: $R = (R_1, \dots, R_n)$ is a random permutation of $(1, \dots, n)$. They proved Theorem 1 and showed that the conditional and unconditional limit distribution of $T_{nR}^{(2)}$ is the same, both under the null hypothesis and fixed alternative. Moreover, under H_0 the distributions of $T_n^{(2)}$ and $T_{nR}^{(2)}$ coincide.

Theorem 1: Let the observations X_1, X_2, \dots, X_n follow the model (2.1), the assumptions (2.1) be satisfied and let $|\delta| \leq D_0$ with some $D_0 > 0$. If $n \rightarrow \infty$, then for all $t \in R^1$ we have

$$P(T_{nR}^{(2)} A(\log n) \leq t + D(\log n) | X_1, X_2, \dots, X_n) \approx \exp(-2e^{-t}), \quad (2.13)$$

where $A(x) = \sqrt{2 \log x}$, and $D(x) = 2 \log x + \frac{1}{2} \log \log x - \frac{1}{2} \log \pi$.

This permutation distribution, $F_p(x; T_n^{(2)}) = \frac{1}{n!} \#\{r \in \mathfrak{R}_n; T_n^{(2)}(r) \leq x\}$, where \mathfrak{R}_n is the set of all permutations of $(1, \dots, n)$ and $\#\{A\}$ denotes the cardinality of set A . The critical region with the level α of the permutation test based on $T_n^{(2)}$ has the form

$$T_n^{(2)} \geq x_{1-\alpha, n}, \quad (2.14)$$

where $x_{1-\alpha, n}$ is the $100(1-\alpha)\%$ quantile of the permutation distribution $F_p(\cdot; T_n^{(2)})$, for details confer Antoch and Hušková (2001). Using the permutation principle, a bootstrap algorithm can be developed to approximate the critical value of CUSUM test.

Algorithm 1: Algorithm for approximating bootstrap critical value of $T_n^{(2)}$

1: Draw a bootstrap sample $X_j^* = (x_1^*, \dots, x_n^*)$ from x_1, \dots, x_n with replacement.

2: Calculate the test statistic $T_{ni}^{(2)*}$.

3: Repeat step 1. and step 2. for $i = 1, \dots, B$, where $B = 10,000$ (or 1000 times).

4: Calculate the empirical critical value, α -quantile = c , such that

$$\frac{1}{B} \sum_{i=1}^B 1_{\{T_{ni}^{(2)*} \leq c\}} \geq (1 - \alpha).$$

5: Reject the null hypothesis if $T_n^{(2)} > c$.

The bootstrap procedure simply resamples from the empirical distribution defined by one sample. If the original sample size is n , then the sampling is done at random with replacement. Resamples are all of size n . Bootstrap without replacement is coinciding with permutation principles. There are many types of bootstrap, e.g., percent bootstrap, naive bootstrap, block bootstrap, residual bootstrap, etc. A good discussion of such resampling procedures and about their accuracy can be found in Efron (1987) and Singh (1981), respectively.

Using the permutation principle of CUSUM test developed by Antoch and Hušková (2001), we are describing an algorithm for finding bootstrap critical values in the weighted CUSUM test (2.5). This algorithm can be carried out for finding the bootstrap critical value using CUSUM R-type test (2.6) as well.

4. RESULT AND ANALYSIS

4. 1. Simulated Data Analysis

Antoch and Hušková (2001) suggested the permutation critical value, a kind of bootstrap test, performs better than the asymptotic critical value in CUSUM test. We are also interested to explore the behavior of the bootstrap and asymptotic critical values for the known change-point in mean using CUSUM test (2.5). So we use generated data from two parametric distributions i.e., Normal and Gamma. We have generated 3000 random observations from Normal and Gamma distributions considering the change in mean at the middle, at the beginning and at the end point of the data.

Table 1. Estimated Asymptotic Critical Value and Bootstrap Critical Value of CUSUM Test (2.5) at Different Level of Significance for the AMOC model (2.1) with Gamma Distributed Errors and Sample Size $n = 3000$.

n	m	δ	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
			CV _{Asym}	CV _{Boot}	CV _{Asym}	CV _{Boot}	CV _{Asym}	CV _{Boot}
3000	250	5	4.5338	3.8870	3.7347	3.4539	3.3818	3.1338
3000	1500	50	4.5338	3.7408	3.7347	3.3222	3.3818	3.0837
3000	2650	75	4.5338	4.1773	3.7347	3.6526	3.3818	3.2864

To evaluate the performance of the asymptotic critical values and the bootstrap critical value of the weighted CUSUM test, we prepared the simulation experiment in which we generated data from the model (2.1) with $n = 3000$, $m = 250$ (change-point at beginning), $m = 1500$ (change-point at middle) and $m = 2650$ (change-point at ending). We generated independent errors from Normal and Gamma distribution with the fixed variance $\sigma^2 = 16$. As a result, we have used different means for our generated data. Tables 1-2 illustrate the better performance of the bootstrap critical value than that of the asymptotic critical value in all cases.

The generated errors have different μ_1 (mean of the 1st segment) and μ_2 (mean of the 2nd segment), where μ_1 and μ_2 is differing with the value δ .

Table 2. Estimated Asymptotic Critical Value and Bootstrap Critical Value of CUSUM Test (2.5) at Different Level of Significance for the AMOC model (2.1) with Normally Distributed Errors and Sample Size $n = 3000$.

n	m	δ	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.10$	
			CV _{Asym}	CV _{Boot}	CV _{Asym}	CV _{Boot}	CV _{Asym}	CV _{Boot}
3000	250	190	4.5338	4.5067	3.7347	3.7044	3.3818	3.2907
3000	1500	200	4.5338	3.8343	3.7347	3.2930	3.3818	3.0623
3000	2650	0.03	4.5338	4.0012	3.7347	3.4816	3.3818	3.2513

4. 2. Real Data Analysis

Surface area monthly average rainfall data in Cumilla district of Bangladesh during the period 1948-2013 have been used to detect a change-point in the model (2.1). The observational data for this work has been collected from Bangladesh Meteorological Department (BMD).

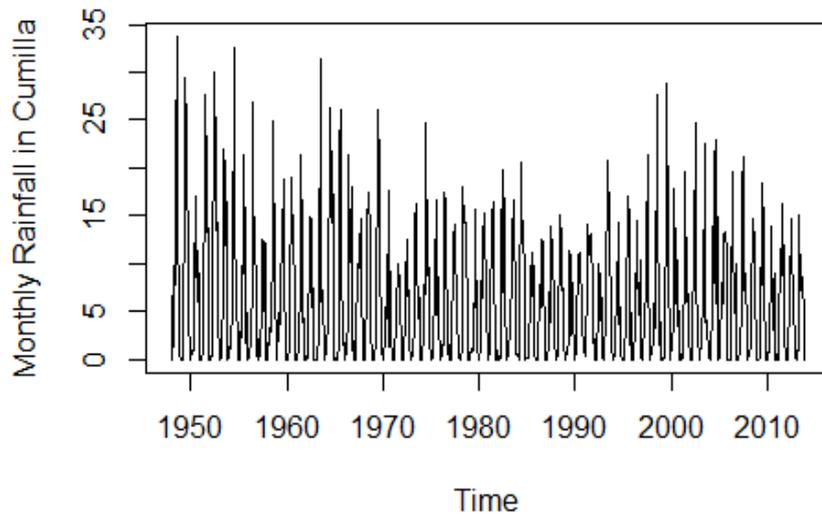


Figure 1. Monthly Amount of Average Rainfall in Cumilla District from 1948-2013.

Precipitation in Cumilla averaged 6.11 mm from 1948 until 2013, reaching an all-time high of 33.71 mm in July of 1948 and a record low of 0.03 mm in December of 1957, 1992, 1996, 2005 and in January in 1961 and 1993. It is observed that there was usually no rain in

December and January, after that, the lowest precipitation months are November and February during the period 1948-2013 in Cumilla. There is no consistent trend (upward or downward) in Figure 1 over the entire time span. However, seasonal variation may occur in the data. Figure 1 explores that the monthly average rainfall in Cumilla have a different mean with constant variance. Hence, we are interested to detect the change in mean using CUSUM test (2.5), which will be better to predict the average rainfall in Cumilla. To perform the CUSUM test (2.5), we need to calculate the asymptotic critical value (2.8) and bootstrap critical value using Algorithm 1 for the month average rainfall data.

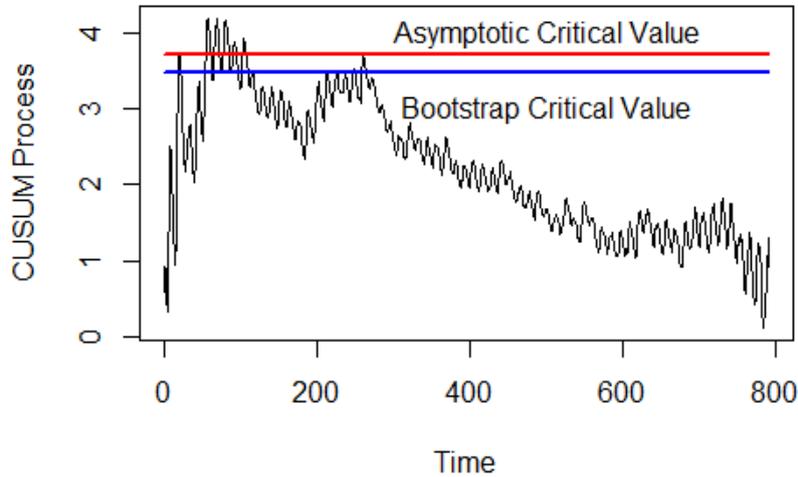


Figure 2. CUSUM Process with Asymptotic Critical Value and Bootstrap Critical Value at 5% Level of Significance.

Figure 2 shows that the bootstrap critical value (the blue line) is more effective than the asymptotic critical value (the red line). We realize the bootstrap critical value is 3.525707 and the asymptotic critical value is 3.699311.

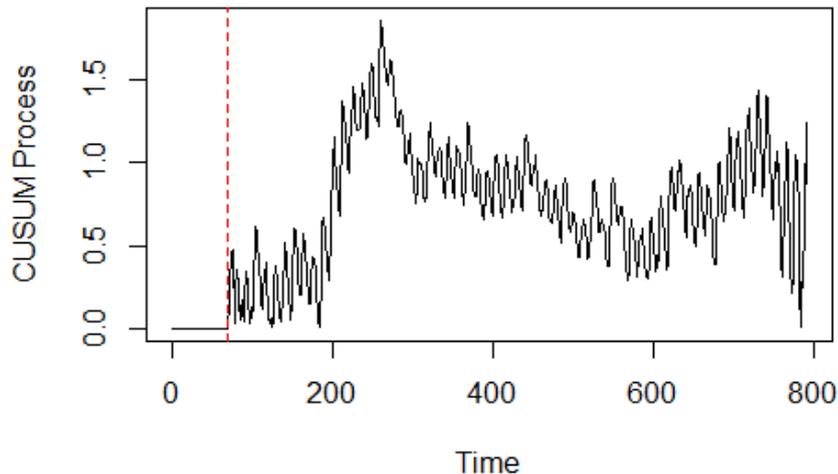


Figure 3. CUSUM process (2.4) for detecting single change-point in the model (2.1).

Figure 3 represents the change-point in mean by the red dotted vertical line at the data point 69 (September 1953). Hence, we observed that there are two different average rainfalls in Cumilla for the two different data segments, ‘January 1948 to September 1953’ and ‘November 1953 to December 2013’. Because of that change-point in mean, there are two estimated means i.e., $\mu_1 = 9.253$ mm and $\mu_2 = 5.811$ mm, which are marked in the Figure 4 with the two different red horizontal lines.

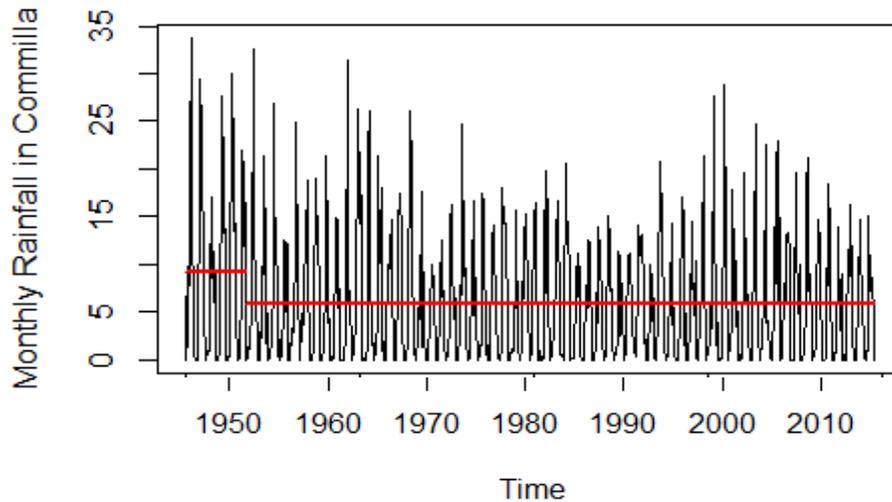


Figure 4. CUSUM process (2.5) for detecting single change-point and their estimated means in the model (2.1).

As the data contains a valid change-point in mean September 1953, so we are motivated to evaluate different time series models for different segments of data to forecast average rainfall in Cumilla. Hence, we used **auto.arima** function in the **forecast package** of **R** software for each segments of data to find the best fitted model for forecasting. To evaluate the impact of change-point in time series forecasting, we first used the entire data (Jan 1948 – Dec 2013) and then used the segmented data (Jan 1948 – Sep 1953 and Oct 1953 – Dec 2013, due to valid change-point at Sep 1953) for modeling purpose.

Implementing the **auto.arima** function we found, the best fitted model for the entire data of the monthly average rainfall in Cumilla is the AutoRegressive Integrated Moving Average of order $p = 5$, $I = 1$, and $q = 2$, i.e., ARIMA(5, 1, 2), whereas, the best fitted model for Jan 1948 – Sep 1953 is the AutoRegressive of order $p = 1$, i.e., AR(1), and for the data segment Oct 1953 – Dec 2013 is the AutoRegressive of order $p = 1$, i.e., AR(1).

The estimated ARIMA(5, 1, 2) model for Jan 1948 – Dec 2013 with the coefficient’s Standard error (SE) in the bracket of the 2nd line of the equation (4.1). The best fitted time series model for Jan 1948 – Sep 1953 is AR(1) model, the estimated model with the coefficient’s SE in the bracket is illustrated in (4.2). Finally, equation (4.2) reveals the estimated AR(1) model which is the best fitted time series model for Oct 1953 – Dec 2013.

$$\begin{aligned} & \left(\begin{matrix} 1 - 0.9093B + 0.1447B^2 + 0.0773B^3 + 0.1316B^4 + 0.1793B^5 \\ (0.0494) \quad (0.0504) \quad (0.0480) \quad (0.0476) \quad (0.0406) \end{matrix} \right) (1 - B)X_t \\ & = \left(\begin{matrix} 1 + 1.5778B + 0.6038B^2 \\ (0.0379) \quad (0.0390) \end{matrix} \right) Z_t, \end{aligned} \tag{4.1}$$

with $Z_t \sim \text{WN}(0, 19.92)$, and where B is the backward shift operator and $B^i X_t = X_{t-i}$.

$$X_t - 0.6803X_{t-1} = Z_t, \quad \text{with } Z_t \sim \text{WN}(0, 45.3). \tag{4.2}$$

(0.0305)

$$X_t - 0.5698X_{t-1} = Z_t, \quad \text{with } Z_t \sim \text{WN}(0, 26.4). \tag{4.3}$$

(0.0305)

Table 3. Estimated Time Series Models for Different Segments of Data to Forecast the monthly average rainfall in Cumilla with AIC, BIC, and AICc information criteria.

Data Segment	Model	AIC	BIC	AICc
JAN 1948 - DEC 2013	ARIMA(5, 1, 2)	4625.3	4662.69	4625.48
JAN 1948 - SEP 1953	AR(1)	463.53	470.23	463.9
OCT 1953 - DEC 2013	AR(1)	4422.69	4436.44	4422.73

Table 3 illustrates that the Akaike information criterion (AIC), Bayesian information criterion (BIC) and Akaike information criterion corrected (AICc) for the three estimated model (4.1), (4.2), and (4.3). To forecast the monthly average rainfall in Cumilla we may use either the model ARIMA(5, 1, 2) or the model AR(1). However, the model AR(1) for Oct 1953 – Dec 2013 is giving the minimum information criterion values of AIC, BIC and AICc compare to that of ARIMA(5,1,2). As a result, we can suggest to use a simple model AR(1) to forecast monthly average rainfall in Cumilla, which is reduced the calculation complexity and make efficient forecast values. Hence, considering the impact of change-point in forecasting we can find a much simpler efficient forecasting model.

5. CONCLUSIONS

Crop production and rainfall have a significant role in Bangladesh economy. Although a wide variety of rainfall forecast methods are available in Bangladesh, considering the impact of change-point leads researchers to find a simple and efficient time series model in forecasting.

CUSUM test requires the critical value to detect the change-point in mean. As there is no exact critical value, so we calculated the bootstrap critical value, and hence compare that with the asymptotic one. Simulated results illustrated that the bootstrap critical value performed better than the asymptotic critical value in all cases. For the monthly average rainfall in Cumilla, we have found the bootstrap critical value is 3.5257 and the asymptotic critical value is 3.6993. So the bootstrap critical value is better than its counterpart.

Therefore, the bootstrap critical value of the CUSUM test is performed better than that of asymptotic critical value in the real and simulated data. Instead of using monthly average rainfall 6.11 mm for the entire data in Cumilla, we can split the average rainfall into more homogeneous groups, i.e., 9.25 mm for Jan 1948 – Sep 2013 and 5.81 mm for Oct 1953 – Dec 2013. Finally, we can say that consideration of the impact of the change-point in forecasting results in a significant and efficient forecasting model. From the result of our real data, we confirmed that the monthly average rainfall in Cumilla district can be forecasted with a simple model AR(1) instead of using an advanced model ARIMA(5, 1, 2). Such a simple and effective forecasting model will be helpful for the farmers and the decision makers to manage the crop production in a positive and constructive way.

References

- [1] Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17, 579-600
- [2] Antoch, J., and Hušková, M. (2001). Permutation tests in change point analysis. *Statistics & probability letters*, 53(1), 37-46
- [3] Antoch, J., Hušková, M., and Veraverbeke, N. (1995). Change-point problem and bootstrap. *Journaltitle of Nonparametric Statistics*, 5(2), 123-144
- [4] Aston, J. A., and Kirch, C. (2012). Evaluating stationarity via change-point alternatives with applications to fMRI data. *The Annals of Applied Statistics*, 6(4), 1906-1948
- [5] Bai, J., and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1), 1-22
- [6] Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P., (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5), 927-946
- [7] Berkes, I., Horváth, L., and Schauer, J. (2011). Asymptotics of trimmed CUSUM statistics. *Bernoulli*, 17(4), 1344-1367
- [8] Csörgö, M., and Horváth, L., (1997). Limit theorems in change-point analysis. John Wiley & Sons Inc., New York.
- [9] Darling, D.A., and Erdős, P. (1956). A limit theorem for the maximum of normalized sums of independent random variables. *Duke Math. J* 23(1), 143-155
- [10] Dumbgen, L. (1991). The asymptotic behavior of some nonparametric change-point estimators. *The Annals of Statistics*, 19(3), 1471-1495

- [11] Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397) 171-185
- [12] Franke, J., Kirch, C., Kamgaing, J.T. (2012). Changepoints in times series of counts. *Journal of Time Series Analysis*, 33(5), 757-770
- [13] Horváth, L., and Rice, G. (2014). Extensions of some classical methods in change point analysis. *Test*, 23(2), 219-255
- [14] Hušková, M., (1997). Limit theorems for rank statistics. *Statistics & probability letters*, 32(1), 45-55
- [15] Hušková, M. (2004). Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun*, 44, 273-291
- [16] Hušková, M., and Picek, J. (2005). Bootstrap in detection of changes in linear regression. *Sankhya: The Indian Journal of Statistics*, 67(2), 200-226
- [17] Kirch, C., and Steinebach, J. (2006). Permutation principles for the change analysis of stochastic processes under strong invariance. *Journal of computational and applied mathematics*, 186(1), 64-88
- [18] Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2), 100-115
- [19] Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42(3/4), 523-527
- [20] Page, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, 44(1/2), 248-252
- [21] Reeves, J., Chen, J., Wang, X.L., Lund, R., and Lu, Q.Q. (2007) A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, 46(6), 900-915
- [22] Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 1187-1195
- [23] Sonesson, C., and Bock, D. (2003). A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 166(1), 5-21
- [24] Stoumbos, Z.G., Reynolds Jr, M.R., Ryan, T.P., and Woodall, W.H. (2000). The state of statistical process control as we proceed into the 21st century. *Journal of the American Statistical Association*, 95(451), 992-998