Change Point Analysis of Correlation Structure: Evidence from Selected Real Data and Monte Carlo Simulation

by

Sharia Iqbal

PIDE2016FPMHILETS12

Supervised by: Dr. Amena Urooj

Co-Supervised by: Ms. Uzma Zia

Department of Economics and Econometrics

Pakistan Institute of Development Economics

Islamabad

2020

Pakistan Institute of Development Economics

CERTIFICATE

This is to certify that this thesis entitled: **"Change Point Analysis of Correlation Structure: Evidence from Selected Real Data and Monte Carlo Simulation"** submitted by Ms. Sharia Iqbal is accepted in its present form by the Department of Economics & Econometrics, Pakistan Institute of Development Economics (PIDE), Islamabad as satisfying the requirements for partial fulfillment of the degree in **Master of Philosophy in Econometrics**.

Supervisor:

Dr. Amena Urooj
Assistant Professor
PIDE, Islamabad

Co-supervisor:

Ms. Uzma Zia
Senior Research Economist
PIDE, Islamabad

External Examiner:

Dr. Saqlain Raza
Department of Mathematics
COMSATS, Islamabad

Head,
Department of Economics & Econometrics:

Dr. Karim Khan

Dedicated to *My parents*

&

My all Teachers

ACKNOWLEDGEMENT

"Trust in Allah with all your heart and lean not on your own understanding; in all your ways acknowledged him, and he will make your path straight."

All thanks and praises to ALLAH Almighty, the merciful, the compassionate, who provided me the opportunity and strength to complete the research work within the stipulated time.

"ONE GOOD MENTOR COULD BE MORE INFORMATIVE THAN A COLLEGE EDUCATION AND MORE VALUABLE THAN A DECADE INCOME."

I would like to express my heartiest cum warmest gratitude to my mentors, Dr. Amena Urooj and Madam Uzma Zia for obliging me highly with their scholarly guidance and patient co-operation in answering my questions. It is due to their gracious support from beginning to the end that I have completed this research study successfully.

I am extremely grateful to Ma`am Nadia Hassan, Dr. Saud Ahmad Khan, Dr. Atiq-ur-Rehman, Dr. Hafsa Hina, Prof. Azhar Hussain Shah, Prof. Manzoor Hussain, Prof. Asad Javed, Prof. Waseem Akram and Prof. Ali Akbar for providing their efficient help and singular guidance in my studies. I believe that I could not come to this level without their unfailing help.

Very special thanks to my Father whose love, support, care, patience and valuable guidance always blessed me with unforgettable memories and contributed a lot in my career building and I am at loss for words to express my thanks to my loving parents and my sister whose love, affection, prayers, care and unflinching support always remained unyielding force of inspiration during my studies.

I am also most indebted to Syed Zafar Abbas Naqvi (PhD Econometrics), Rizwan Ahmad (PhD Econometrics), Bushra Pervez and Fiza Shaheen for their frank responses to my many thirsty questions. Finally, I am pleased to express my thanks to Pakistan Institute of Development Economics (PIDE) that endowed me with noble platform and opportunity of learning.

Sharia Iqbal

LIST OF ABBREVIATION

ML	Maximum Likelihood
CUSUM	Cumulative Sum
SBS	Sparsified Binary Segmentation
KCP	Kernel change point Permutation test
E-Divisive	Energy Divisive
PELT	Pruned Exact Linear Time
TPR	Two Phase Regression
SNH	Standard Normal Homogeneity
ECP	Execution of the Change Point
Acgh	Comparative genomic hybridization array
ADJI	Average of Dow Jones Industrial
CCB	Chinese Commercial Bank
S&P	Stock and Price
GHRG	Generalized hierarchical random graph
HRG	Hierarchical random graph
MRM	Markov Random Model
EEG	Electroencephalogram
ECG	Electrocardiogram
SBP	State Bank of Pakistan
BCFT	Billion Cubic Feet
IFS	International Financial Statistics

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

ABSTRACT

The motivation of change point detection is to take into account sudden changes in distribution of time series, which remain important in social sciences, economics, medicine, environmental sciences, etc. In addition, due to distributional changes some statistical properties of time series changes such as variance, mean and correlation. The present study has been devised to explore the abrupt distributional changes in time series in multivariate frame work using non parametric method i.e. E-Divisive method, to detect multivariate change point. Our main focus is to explore the abrubt distributional changes in macro data of paskistan in multivariate framework. To check the performance of E-Divisive method for different location and number of change points in case of multivariate and individual analysis; we have applied Monto Carlo simulations. It has been found that the performance of E-Divisive method has high power in experiments locating change points at same and varying locations in multivariate scenario. In this study, attempt has been made to explore the change point in multivariate time series i.e. Macro-economic variables from energy sector and banking sector of Pakistan economy are selected. Results are drawn by assessing correlations among macro variables, unit root stationarity tests and change point detection method i.e. E-Divisive. The study spans over a period from 1990-2016. The study uses variables like production of Electricity, production of Natural gas, production Crude Oil, Cash in Pakistan, Balances with State Bank of Pakistan, Borrowing from State Bank of Pakistan, Real interest rate, Exchange rate and M2 (money supply). Change point have been identified in all of the series under study in univariate as well as multivariate framework.

Keywords: Change point detection, E-Divisive, Energy sector, Banking Sector

CHAPTER 1

INTRODUCTION

1.1 Background of the Study

One of the critical yet complex problem of time series analysis is the existence of change points and their impacts. The motivation of change point detection is to take into account sudden changes in distribution of time series, which remain important in social sciences, economics, medicine, environmental sciences, etc. In addition, due to distributional changes some statistical properties of time series changes such as variance, mean and correlation. Focus remains on finding of change point in the structure of mean and variance. However, since majority of macroeconomic variables work in association with other related variables so any sudden change gets the association among variables affected as well. Detecting correlation change using change point detection analysis focuses on methods that finds the location and extent of changes in correlation structure of a time series.

Analysis of change point is the process of detecting the distributional changes according to ordered observations. For a given observed time series, the instance where distributional changes happen are referred as change points. Change points have implications on both applied and theoretical statistics. Generally speaking, change point analysis can be performed through the use of nonparametric or parametric models. Parametric change point study assumes that observations are drawn from a class of distribution functions, who's members can be uniquely identified by parameter value. In this setting, analysis focuses on detecting changes in the parameter of interest. While parametric approaches lead to many useful theoretical results about their performance, they cannot always be used in real world settings, since adherence to the underlying model is not guaranteed. In such situations using a nonparametric[1] change point algorithm would be more appropriate.

Despite the fact that change point analysis is essential in numerous fields, majority of methodologies that have been established to date often undertake a single or multiple but known number of change points. This process estimates hierarchically or concurrently all change points. If there are k change points, concurrent methods usually optimize a unique objective function. Hawkins (2001) found likelihood function can be maximized by locating the all change points. Lavielle and Ere (2006) observed the similar results by minimizing the loss function. In Sequential methods mostly change points are estimated just once at a time as observed by Guralnik and Srivastava (1999), yet some are able to find or calculate two or more than two change points at given time span as identified by Olshan and Venkatraman (2004). Such methods are generally known as bisection approaches. Change point detection methods have been pervasively studied for the purpose to identify the structural changes in time series data. In making predictions, the change point explains unexpected variation that represents transition. The change-point problem has been considered by several researchers Srivastava (1999) who identified certain problems in introducing the CUSUM (cumulative sums)and measured a normal model test for a mean level change. Truong *et al*.,(2018) introduced approaches in multivariate time series to find multiple change points. They include reviewing methods like regression, maximum likelihood (ML) estimation, kernel techniques etc. In this broad research area, applications are diverse and abundant; many other approaches and working constraints (on difficulty and precision) prevail. A

[1] A non-parametric approach is the one which assumes no restrictions on the distribution function of the underlying data generating process. It assumes that the parameters do not follow any distribution.

recognized framework for the detection of change point is presented to make sense to the significance of research work.

Until recently, research on change point detection remain focused upon univariate time series, resilient methods to find out the changes in mean, variance and autocorrelation. Univariate change point refers to the change in the univariate series as a result of the sudden or smooth changes in the process of building of series (Jandhayala and Fotopoulos, 1999).

With the advancement of technology, recent research has focused upon multivariate time series. Change point detection in multivariate data identify more changes than in single variable. It is argued that besides the mean, the characteristics of correlation structure also has been changed when change in mean point occurs. Change point detection is useful but a difficult task in multivariate time series and has gained a lot of attention from researchers (Terien *et al*., 2009). There are many methods which have been presented for the detection of multivariate change point in both mean and correlation structure. When properties of time series show some changes, the change point detection focuses on detecting unexpected changes in data. The distributional change usually occurs in multivariate time series to compute the mean, variance and correlation structure. The importance of multivariate change point detection is that it is more useful for the development, planning and modeling of multivariate control charts. There are many non- parametric approaches to detect multivariate change point like E-Agglomerative, CUSUM test, sparcified binary segmentation (SBS), supervised & unsupervised methods, sub categories of unsupervised (likelihood ratio method, subspace model method, probabilistic method, kernel based method, graph based method, clustering method). These are used to find changes in mean, variance and correlation structure. These methods also include DeCon (Bulteel *et al*., 2014), E-divisive (Matteson and James 2014),Multirank (Fong *et al*.,2011), KCP (Arlot *et al*.,2012) and KCP permutation test(Cabrieto et al.,2018). The methods are based on different statistical approaches i.e. robust methods for DeCon, rank information for Multirank, the kernel approach for KCP, Euclidean distances for E-divisive and KCP permutation test is focus on correlation changes. This diversity makes it difficult for the applied researcher to appraise the methods. Since, these methods are based on different statistical approaches, it is still unknown which of these four methods should be preferred in particular circumstances. Sharkey & Killick (2014) have presented different approaches to detect change points in a nonparametric setup and then they are modified to account multiple change points. According to Killick *et al*., (2014) nonparametric method performs better having reliable results. This study aims to check the performance of one of these non-parametric change point detection method i.e. E-Divisive by using series of Monte Carlo experiments covering various number and locations of change points. In case of Pakistan, we are unable to find any study detecting change points in macroeconomic time series in multivariate framework. For macro-economic data of Pakistan, we will apply the method to observe any possible change in correlation structure.

1.2 Research Objectives

Specific objectives of the study are**:**

- To explore the abrupt distributional changes in macro time series data of Pakistan in multivariate frame work using non parametric method.
- To explore the performance of E-Divisive technique to detect change point in multivariate frame work and correlation structure.

1.3 Significance of the Study

Change point detection in time series has gained significant importance as it gives insight about data behavior. Among several suggested methods, the non-parametric methods are more favorable due to their wide spread applicability. Not only change point detection for mean is used, but currently change point in correlation is also gaining importance. The major task of our study is to examine the performance of the method i.e. E-Divisive method for various macro-economic time series of Pakistan. To the best of our knowledge the existing literature, lacks any study exploring change point in correlation structure of multivariate time series of Pakistan. Thus, this study will contribute by exploring change point in the structure of correlation in number of time series of Pakistan and one E-Divisive method has been used for change point detection and to explore change point in correlation structure. The selection of this method for the purpose resides in evidences from literature. E-Divisive method is claimed to perform well as compared to other methods in detecting number of change points and location of change points (James and Matteson, 2013).

1.4 Research Gap

Up till several years, research on change point detection focused almost entirely at univariate time series consisting the process to find changes in mean and variance. In recent years, researchers have focused on change point detection in multivariate framework, where not only change in mean and variances are in focus but also changes in correlation has been explored. It has been argued that not only mean and variances get affected due to sudden abrupt changes in distribution of variable but their mutual behavior also get affected and it is indicated inform of change in their correlation structure. Hence, the research from recent years like Cabrieto *et al*.,(2016 and 2018) have focused on locating change points using correlation structure. The aim, in this study is to identify the change points using association structure. Conventionally, Chow test is being used to detect mean and variance break in univariate, bivariate and multivariate data. However, Chow test or any of its modification are not suggested for change point in correlation structure, for this purpose other tests are also suggested and used in recent literature. Present study is aimed to fill the research gap in the literature as up to our knowledge, no study in Pakistan has focused on multivariate change point detection. This study provides an insight on time series data from Pakistan and attempts to fulfill this gap by identifying the change points in correlation structure of multivariate time series from Pakistan by using Non-parametric approach.

1.5 Organization of the Study

In chapter 1 we have discussed the introduction, the significance of the study, objective of the study, and research gap. Chapter 2 is based on literature review. The data and methodology is discussed in chapter 3. Chapter 4 is about change point detection based on simulation. Chapter 5 is the empirical application of selected method for time series of Pakistan. Chapter 6 is conclusion.

CHAPTER 2

LITERATURE REVIEW

This section includes the literature about the detection of change point in correlation structurein multivariate time series. A vast variety of literature is available which carefully observes the subject matter.

There exist multiple different techniques to solve change point problems, like BSP (binary segmentation procedures) or penalized likelihood methods. Multivariate change point detection methods are classified under two approaches through which change point can be detected. Parametric approach assumes a model for multivariate time series named as parametric model such as a VARMA model. Second, the non-parametric method which is our main focus. Galeano and Wied (2014) have suggested a nonparametric statistic known as CUSUM to test the existence of constant correlations among two random variables, Galeano and Wied (2014) have presented a system based on the constancy of correlation test to find both the timing and number of thinkable change points.

The literature shows that there exist several methods for the detection of change point correlation under non-parametric and parametric settings. Several methods have been reviewed in this regard including DeCon, Multirank, E-Divisive, KCP, KCP permutation test, PELT, Homogeneity testing, Wilcoxon rank statistic, CUSUM, SBS (sparsified binary segmentation), unsupervised method (likelihood ratio method, subspace model method, probabilistic method, kernel based method, graph based method, clustering method),supervised methods, log likelihood method, ED-PELT, Copula approach, Bayesian method etc. E-Divisive is a top down approach, as all observations start in one cluster and splits are performed recursively as one moves down the hierarchy. It tests the statistical significance of each hierarchically estimated change point. Divisive estimation sequentially identifies change points via a bisection algorithm. The method of “E-Divisive” chronologically test for testing the significance of every change point estimation given the earlierchange in the estimation of locations, although its output is sensitive and to be contingent on the estimated change points numbers. This method estimates both the location and change point numbers. It hierarchically test to check the significance of everydetected change points. The methods are also available for the estimation of change point’s locations, without a special information of the number of multiple change points. This method has hierarchically detected the number and location of change pointswhich other methods can not

2.1 Breaks in Multivariate

Bhattacharya (1994) has examined change-point analysis and tested the hypothesis of point and interval estimation of, ‘no change’, variation in non-parametric models, variation in regressions and examined the distributional changes in serially observed data. This leads to consider some stochastic processes, whose minimum and maximum values make the basis for likelihood methods and its integral regarding prior distributions make the ground of Bayesian methods. This study found that neither of the two procedures are better than the other when the changeoccurred in the very beginning or at the very end in the data.

Work done by Bai (1997) examined the change point in multiple regressions through least squares estimation. This research paper resolved the problem of parameter-change with an unidentified breaks in multiple regressions. The key concern was to find the change point andthe statistical model of the change-point estimator. The uncertainty was caused by the Fed's change in working events through the time period of Oct 1979 to Oct 1982 . The error method can be heteroskedastic and dependent. For instabilities or nonstationary regressors, the asymptotic distribution showed to be skewed. For overall skewed distribution , the cumulative distribution function and systematic density function was derived. The method was used to evaluate the reply of market interest rates to discount rate fluctuations.

Reeves *et al.*,(2007) have noted that in time series variation points are times of gaps which can be encouraged from changes in observation positions, equipment, measurement methods, environmental changes, and so on. In climate data series different researchers have proposed different approaches that have been planned to distinguish undocumented variation points. These include two-phase regression (TPR) procedures, standard normal homogeneity (SNH) test, inhomogeneity tests, Wilcoxon's nonparametric test and information criteria procedure. In these methods researchers tried to provide guidelines for best work procedure in different situations.

Work done by Adams and Mackay (2007) focuses on the abrupt variations in the generative parameters of different variables and also studied the online (sequential analysis) change point detection. Three different real world data set are used in this study. Watergate affair, OPEC oil embargo, and daily returns of the Dow Jones Industrial Average are used from 3rd July 1972 to 30th June 1975. This is the weekly data about the happening of disasters in coal mine that causes the death of more than ten persons in between 1851 and 1962. The study used the probability distribution by using simple message passing algorithm (of length) of the accurate time to the last point change. It focused on the predictive filtering by generating and accurate distribution in the sequence in which data is already observed or known. The study is based on three data sets of real world with the diverse requirements in this algorithm. The probability distribution of currenttime length from the last point of change by applying a simple passing algorithm will be computed in this study.

Kawahara & Sugiyama (2009) have noted that change-point detection is needed where the characteristics of time-series data can variate. Wide range of difficulties about real-world are discussed in the field of data mining and statistics. The researchers havepresented anon-parametric approach for the detection ofchange in probability distributions of arranged data. An important idea was estimation of probability densities ratio and not the probability densities themselves. This formulation has enabled to ignore the estimation of non-parametric density, thatidentified to be a hard issue, being an online estimation method, it provideda change-point discovery process based on estimationof direct density-ratio. The helpfulness of the anticipated method is established through tests using real and non-natural datasets.

Fong *et al*.,(2011) have explored that changes in the multivariate data is an important recent concern. In this study data are generated under Gaussian Distribution. The study focused on multivariate data and check two types of problems which arehomogeneity testing and change point detection based on famous Wilcoxon rank statistic. The two sample homogeneity test statistic can be used for censored or ordinal data. It is also used for testing the homogeneity of more than two samples. Reviewing the change point detection it showed the use of dynamic programming is a good choice because the approach is better for detecting a large number of change points. This method is mostly recommendable where the correlation between the coordinates of the data are moderate. To improve the method it would be desirable to provide significance of level for change point detection when it is used to detect more than a single potential change points.

Galeano and Wied (2014) has anticipated a binary segmentation process in correlation structure of arbitrary variables for change points . As much as, researchers identify the method to explain this type of problem. The method was based on a “CUSUM” test statistic given by Wied *et al*., (2012). It showed that proposed technique reliably detects location of change points and accurate number of change points. Similarly, the fixed sample assets of the process have been examined through the analysis of some simulation studies and the claim of process on the data set of real-world. The observed results in the real-world data set recommend the process to detect the fluctuations under these conditions. Due to financial crisis the relationship among financial earnings changed.

Another study done by Zhu *et al*., (2013)is based on copula approach to detect a change point with two outstanding benefits. In this study, quarterly data has beenused for financial series of sixteen Chinese Commercial Banks (CCB) . Study has used the two methods to find the change point. One method was to deal with general and important instable panel data . Whilethe second method has been used to find the multiple change points. In CCB case, they used these approaches to detect the subprime crisis. After using these approaches, the study shows that the contagion starts in second quarter of 2007 and ends in first quarter of 2009, which is relevant according to recent studies. The technique was used to find the subprime crisis time period in China bank. The result displayed that the subprime crisis start 2007 Q2 and end 2009 Q1 which was considerable according to some applicable research.

Another study by Roy *et al*.,(2017) tackled a problem of change-point estimate in the framework of high-dimensional Markov random models. These change-points denoted the main characteristic of dynamically developing network circle. The change-pointattained to make the most of outline penalized pseudo-likelihood purpose under the sparsity statement. Even in setting, study also derived a logarithmic factor for the estimate,in which the number of probable boundaries in network are more than size of sample. The act of the anticipated estimator was evaluated on sets of artificial data and used to discover voting forms in US Senate through the time period in1979 to 2012.

2.2 Change in Correlation Structure

Another study by Arlot *et al*.,(2012) has tackled the problem of change point within general data set and suggested a penalty to select a change point in thekernel algorithm. This study proposed a penalty generalizing one of the kernel change point problem and showed it satisfies a non- asymptotic oracle inequality .That penalty had generalized one of the Lebarbier (2005) to Kernel Change Point (KCP) problem and proved that it satisfied non asymptotic oracle inequality by presenting new results in the Hilbert space. Experiments showed that it can detect changes in the distribution whenever the variance and mean are constant and also illustrated the accuracy of the model on real and synthetic data[2].

In contrast James &Matteson (2013), pointed several methods in which the execution of the change point (ECP) is placed from parametric ways to those which are independent of distribution. The study used the two real datasets (micro-array aCGH data, weekly log returns of the companies that are composed by the Average of Dow Jones Industrial), the time period is used toApr 1990 to Jan 2012. The ecp is designed to perform various analysis of the chain point while keeping in view few assumptions as possible. Univariate data and many other methods are valid to perform, and for both the observations i.e. multivariate & univariate R package is appropriate. Moreover for the estimation of hierarchy the algorithmused is either divisive or agglomerative. Bisection algorithm change points are identified properly by the divisive estimation. While determining the optimal segmentation agglomerative algorithm estimates the change point. Within the data, both approaches are able to distinguish any kind of change that is related to distribution. This gives an edge to the other obtainable change points which are only capable of detecting changes within the marginal distribution.

[2] **Synthetic data** is "any production **data** applicable to a given situation that are not obtained by direct measurement".

For the implementation of analyzing the nonparametric change point in case of multivariate data, the “ecp package” is most suitable. This package gives us two primary approaches to perform analysis, each of them are capable to configure out the number of change points without operator input. Apart from the data itself the only user-provided parameter is the selection of α (The moment index used for determining the distance between and within segments). If α which is designated to fall in the (0.2) interval, then any form of distributional change within the series which are to be observed are detected by the technique provided by the respective package, providing the absolute αth moments exists. Any kind of changes in distribution are detected by these two approaches within data. This is beneficial over many change points procedures that are able to find changesin the marginal distributions.

Liu *et al*., (2013) explored the change-point detection to discover abrupt property changes lying behind time-series. Researchers paid attention on calculating the score at change-point which presents the acceptability of change points and also observed original statistical change point detection algorithm based on non-parametric deviation estimation among time-series samples from two retrospective segments. This technique used the relative Pearson deviation as a deviation measure, and the technique of direct density-ratio estimation was exactly and professionally estimated. To determine the helpfulness of the projected way , through experiments on artificial and real-world datasets which include speech, human-activity sensing, and Twitter messages. Experimental results on artificial and several real-world datasets which, include Twitter messages, speech, human-activity sensing from Feb 2010 to Oct 2010. The future work will spread over these methods to detect change-point and estimate their applied helpfulness.

A binary segmentation algorithm on the bases of CUSUM was proposed by Cho and Fryzlewicz (2015) termed as Sparsified Binary Segmentation (SBS), to find the changes in multivariate time series about second-order-structure. The study used the data about daily closing prices of stock and prices (s&p) 500 stock market index with the time of the current financial crisis. The step of SBS lessened the effect of unrelated contribution that was mostly useful in high dimensions. The segmentation of multivariate time series estimated the average on the basis of CUSUM and compared the theoretical properties was well performed from one factor of multivariate time series. The sparsifying step minimize the impact of inappropriate contribution, which is mostly helpful in high dimension, SBS automatically identifies common change point which remove the need for post-processing.

A study done by Mattesson and James, (2014) had given a method for the analysis of multiple change point that was able to estimate the change point detection of location and numbers simultaneously. The study used the financial data for Cisco Systems Inc of 262 months. This study compared the performance of three methods of change point detection namely, PELT algorithms, E-Divisive and Multirank in univariate analysis and found that E-Divisive is a most favorable method for the detection of change point in variance and mean. The analysis is again carried in multivariate setup, performance of three methods had compared, namely, Kernel Change Point (KCP), E-Divisive and Multirank and found the non-reliability of Multirank method. Cluster analysis methods are used to check the performance and also for comparison of locational estimates even when number of estimation are quite different.

Another work done by Bulteel *et al.*, (2014) is about estimated multivariate time series data in the presence of response concordance and applied robust technique. DeCon method was mainly used as a tool that detects the time series data separately for change in co-variation and mean level. DeCon method used robust statistical techniques for this purpose. It was found in robust statistics exercise that DeCon was well founded, could make a hypothetical distribution and easily detect change in multiple points within time series data.

In another study done by Sharkey & Killick (2014), the change points have been extensively examined in time series data to identify structural changes, typically when the data are of known parametric then for sequential testing two-sample hypothesis testing procedures were introduced, in which test statistics were calculated on the basis of ranks of quantities are modified for change point detection. This framework is then stretched to consider data streams and multiple change points. By comparing the impact of test statistics in a range of scenarios the performance and characteristics of testing procedure can be identified, it is found that nonparametric tests are suitable alternatives. In addition, it is found that tests that used for arbitrary distributional changes are comparable and designed to detect changes in location and scale. Overall, the performance of nonparametric hypothesis testing procedure is well, when performing change point analysis on data of no known distributional form it represents logical course of action, a common scenario that applies to a wide variety of real-world processes.

Another study has done by Peel and Clauset (2015) has denoted these quence of networks, to each provide a snapshot of the contacts over a brief time period. An imperative task to analyzed a developing networks was change-point detection, in which both recognize the times at which the large-scale form of contact changes basically and measure how huge and what kind of change happened. This method associated a general hierarchical random chart model with a Bayesian theory test to quantitatively control and exactly detect how a change point occurred. Using the two high-resolution developing social networks, this method recognized the order of change points that make parallel with known outside "shocks" to these systems. Under a probabilistic method to change-point discovery, researchers closed parametric distribution over networks and introduced the generalized hierarchical random graph (GHRG) model. Model has some structures that make it gorgeous for change-point result and generalized the common hierarchical random graph (HRG) model and applied both the simple and GHRG methods to weekly Polaroids from May 1999 to June 2002.

Non parametric change point detection provided the best segmentation for a data set. It explained which methods minimized the penalized cost function in term of minus sign within each segmentation in case of non-parametric log likelihood for data proposed by Haynes *et al*., (2017). It was observed that, minimized cost function using dynamic programing provided a sound view. They used heart-rate data, and change in heart rate during physical activity. The study focused on change point detection in univariate time series. This study, used the different dynamic programming approach known as, Pruned Exact Linear Time (PELT) (kellick *et al*.,(2012).It called new algorithm ED-PELT. This method is based on assumption that, data has identical and independent distribution (IID) within segment (i.e In probability theory and statistics, a collection of random variables is independent and identical distribution as the others and all are mutually independent). By using the ED-PELT method, the study was able to segment the data into meaningful segments using an appropriately chosen penalty value that correspond to different phases of the run and can be related to different regimes in heart rate activity.

Another study by Ali ppi *et al*.,(2015) showed that size of change divergence between pre- and post-distribution could be naturally measured by symmetric Kullback -Leibler[3] and the change point detection had given magnitude worsens when increased in data dimension. The study refers this problem to detectability loss, due to linear relationship between the data dimension and variance of log-likelihood. The theoretical and empirical analysis reveals that the famous technique of monitoring the Log Likelihood (LL) in a multivariate data stream suffers to the loss of detectability , when there are increase in dimensions of data.

In a most recent work Aminikanghahi and Cook (2017) proposed to detect the change point in time series which showed the variation had been unexpected. In this study , the detection of change point has remained useful in prediction and modeling of time series. They used data of climate change on human activity physiological variables like, heart rate, electroencephalogram (EEG), and electrocardiogram (ECG) in order to perform automated, real-time monitoring They emphasized that , the estimation of change point detection to describe the nature and degree of the known change. It includes unsupervised[4] and supervised[5] techniques ,which are selected on the algorithm of the desired outcome. Significance of change point detection has shown importance of unsupervised method. Recently, these methods (supervised method and unsupervised method) had compared to detect the change scores and to determine the starting value whether change has occurred or not occurred. Finally, change point detection for upcoming challenges covered in non-stationary time series. Calculating dissimilarity

[3] Method to measure the difference between two probability distributions.

[4] Unsupervised learning is a type of machine learning algorithm used to draw inferences from datasets consisting of input data without labeled responses. The most common unsupervised learning method is cluster analysis, which is used for exploratory data analysis to find hidden patterns or grouping in data.

[5] if weare working out our machine learning exercise for each input with given objectives. so, this is known as supervised.

measure for each feature whenever a change occurs represent one possible solutionfor finding the change source and the total dissimilarity measure can then be used to conduct a change estimation.

In recent years a study done by Cabrieto *et al.*, (2017) highlighted hypothetical data set of 50 time series points which were used to measure mean change point and correlation structure of three independent autoregressive series are generated through Data Generating Process (DGP). The performances of four newly proposed change point detection methods namely, DeCon, Multirank, E-divisive and Kernel Change Point (KCP) was assessed in nonparametric multivariate time series. The study found that KCP was the best method for change point detection in mean and correlation, in both cases. On the other hand DeCon method also inspected change points. Generally, as this study concludes, DeCon method has found performing less accurate than KCP. To detect correlation changes this method is found sensitive, particularly as it contains various noise variables in multivariate time series.

Biau *et al.*,(2016) explored that the concept of change-point detection is well-studied and there are many aspects of the problem. Its general literature vary from parametric approaches by using log-likelihood systems to nonparametric ones which rely on the type of Wilcoxon statistics, Sequential ranks and U-statistics. They also suggest the monograph for a detailed treatment of these techniques. This, leads to a simple and beneficial data-taken statistical test for the detection of change-point. Their study then apply this test to real-world and simulated data.

Arlot *et al*., (2012) introduced kernel change point (KCP) in their study and tried to see location of change-points with unknown number. KCP provide the result of change points with optimal rate. For example, detecting of change points occurs in comparative genomic hybridization (CGH) arrays is important for cancer diagnosis in early stages.

All analysis are done assuming that or kernel change point is bonded. As the result of change points through kernel, there are some consequences like KCP detect all kinds of changes in distribution. It also detects change in complex structured data. This method KCP can be used for the detection for any type of distributional changes (not only changes in variance or mean) and can also estimate in case of difficult structured data.

Some more work is done by Cabrieto *et al*., (2018) as they argue when the goal is to perceive sudden correlation changes, then change point detection is done to check when those changes exactly happened. This study proposed a permutation test for Kernel Change Point (KCP) to detect the correlation. The non-parametric variations are attractive tool in the case when there is no priori information present on the data distribution that researcher don't have. Apart from general purpose he found the other changes like variance and mean succeeding in correlation changes. Several methods are introduced which are particularly assigned to find correlation changes. Those methods are Frobenius , CUSUM & Maximum norm approach. By detecting the presumably correlation change point and identifying that weather there exist significant difference in correlation before and after, these later (are Frobenius, CUSUM & Maximum norm approach) approaches detect the correlation changes. The similar methods test the examined phase for further change point when there is found significance in change point. While if there is no significance present in change point then the following method dismisses and leads towards the conclusion which defines that the constant correlation is present in the whole time series. To test the single but significant change point that is sensible at first look, may face the power loss in case of multiple points of change exist in the time series.

Concluding this literature review, many methods of change point detection remained in focus of researchers. Mainly four techniques are in focus (KCP permutation test, DeCon, Multirank and E-Divisive method). KCP permutation test is found as the best method in literature and is used for kernel approach. It is seen that every methods is used to serve different purposes. The literature has shown use of different techniques adopted in various studies. Based on the literature review, our goal is focused on change point detection in multivariate time series using non-parametric approach. We have applied the non-parametric method and estimate the change point analysis and correlation structure in multivariate time series. This study investigations are mainly based on E-Divisive method which is available in non-parametric settings and used for correlation change point detection.

2.3 Overview of Macro Economy and importance of macroeconomic variables

Macroeconomic policy is concerned with the overall working of economy and the influence of macroeconomic variables through three sets of instruments: monetary, exchange rate and fiscal policy. Capturing the economic activity as a whole, the macroeconomic variables provide the aggregate averages of the economy which are responsible to regulate the overall development, growth and efficiency of the economy.

Macroeconomic variables help to achieve the goal of economic growth and analyze the forces which determine economic growth of a country. The targets of Macroeconomic policies are to bring sustainability in development process which is usually a main concern of economists, researchers and policy makers. To observe sustainable development of the economy there is a need to see the economic problems like: energy/oil crises, unemployment, poverty and inflation etc. Solutions to these problems are possible at the macro level if right policies are employed at the right time. Consequently, detailed knowledge of correlations among macro variables and detection of any disturbance in their functioning is needed in order to formulate the best economic policies and make them consistent with international economic policies. To carry on the analysis in this study, on the basis of importance, two important (independent) sectors of Pakistan's economy are chosen: Banking sector and Energy sector.

2.3.1 Importance of Banking Sector

Banking sector, being a key feature of financial system, is chosen for its importance observed since history. Commercial banks play an imperative role in managing financial dealings of the economy and banks distribute funds from savers to borrowers in an effectual way. Special financial services are being provided by them, which reduce the cost of attaining facts about savings and borrowing prospects. The financial services by banks support the economy and enable it to work proficiently. Banking plays an important role in financial life of a business and the importance of banks can be seen from the fact that they are considered as support of modern economy. Wealth is not created by banks, but their key activities focus on wealth issues, savings, investments and exchange. This makes the banks an effective partner in growth and economic development of a country.

In banking sector, under the head of money and credit scheduled bank's consolidated position (Based on weekly position Liabilities & Assets (All Banks)) is taken in million rupees. Monthly data of 1990-2016 has been taken for the three variables**:** Cash in Pakistan, Balances with SBP and borrowings from SBP.

Interest rate in Pakistan economy

Interest rate, in any country is used as a policy tool to bring economic stability. There are two views about interest rate: first Keynesian view, that higher interest rate lowers investment and hence growth. The second Mackinnon-Shaw hypothesis that postulates that increase in interest rate improves the efficiency of investment and accelerate economic growth. Interest rate as policy variables, is directly related to economic growth. The movement in interest rate influences the financial market operations. A rising trend in interest rate negatively impacts overall investment decisions hence distorting the overall investment structure. The central bank of Pakistan (SBP) exerts significant influence on interest rates through monetary policy measures as is the practice in most developing countries.

Money Supply(M2)

The process of money creation takes place when SBP either buys domestic assets, government securities, and foreign exchange from the banks or it directly lends to the government/ financial institutions. Money supply is usually measured as a sum of currency in circulation and deposits of general public in different financial institutions. In case of Pakistan, broad money M2 is extensively used. There exist a two way phenomena of measuring money supply, one is from liability side and the other is in terms of assets. From liability side, SBP measure M2 as a sum of currency in circulation which includes total deposits of non-government sector, residents' foreign currency deposits; and other deposits. And from asset side, M2 is a sum of net domestic assets and net foreign assets of the banking system (i.e. SBP and scheduled banks).

Exchange rate

The relationship between exchange rate and economic growth has been an important subject in economics. Exchange rate means how many units of one nation's currency can be purchased with one unit of domestic currency. Exchange rate is considered as the conversion factor that determines the rate of change of currencies.

Real exchange rate volatility shows the short term fluctuation of the real exchange rate. Different patterns of exchange rate behavior into various categories is known as exchange rate regime. A regime in which exchange rate remains fixed is called fix exchange rate regime and in which exchange rate fluctuates is known as floating exchange rate regime. The middle of fix and floating exchange rate is called managed float regime. Pakistan adopted floating exchange rate system. By currency devaluation, the foreign goods become expensive; therefore, people switch from the consumption of foreign goods to the domestic goods. Similarly, the local goods will become cheaper for the foreigners and the export will rise. Pakistan nominal exchange rate is increasing day by day this clearly indicates that more Pakistan's currency is required to buy one dollar. Hence Pakistani rupee experiences deprecation against dollar. Moreover, Pakistan's currency faces devaluation and depreciation over different periods of time depending upon official and unofficial increase in the exchange rate. Economic growth results in the currency appreciation and improves the living standard while failure of the economic growth leads to the depreciation of the currency. Official exchange rate of PKR declined from PKR 10/$ in 1980 to PKR 90/$ in 2011.

Recently, Pakistan's currency depreciated around 34% which brought economic instability in the market.

Cash in Pakistan

Banks position is usually explained in terms of liabilities and assets. Cash in Pakistan is a component of assets.

Balances with SBP

Banks position is usually explained in terms of liabilities and assets. Balances with state bank of Pakistan (SBP)is a component of assets.

Borrowing from SBP

Banks position is usually explained in terms of liabilities and assets. Borrowing from state bank of Pakistan (SBP)is a component of liabilities.

2.3.2 Importance of energy sector

The consumption and production of energy resources is enormously vital to the international economy. Energy consumption and production both are good indicators of a country's overall economic development. Economic activities are mostly dependent on energy resources, whether it's in provision of transport, manufacturing goods, working on machines/computers, industrial production and household consumption, all entails the energy resources. The energy industry is the includes all the industries which are involved in the production and sale of energy like: extraction of fuel, manufacturing, refining and distribution. The most significant measure in Pakistan`s energy balance is the total consumption of 92.33 billion kW of electric energy per year. Per capita this is an average of 469 kWh.In Pakistan oil and gas are two key components of energy mix contributing almost 65 percent (oil 15% and gas 50%) share to the 64.7 million tons of oil equaling of energy supplies during 2012 while share of coal and nuclear is almost 7 percent and 2 percent, respectively.

In energy sector production of selected manufactured goods (Minerals and Electricity Generation, variables are taken as: Electricity (Mn Kwh), Natural Gas (Billion CFT), Crude oil ('000' Barrels).

Electricity

Since independence the matter of balancing the electricity supply against its demand had remained an unresolved matter for many years. These shortages provide great incentive for electricity generator, producer which had become a necessity instead of a luxury at both domestic and commercial level. Other problem is lack of efficiency in dealing with growing demand of energy.

Natural Gas

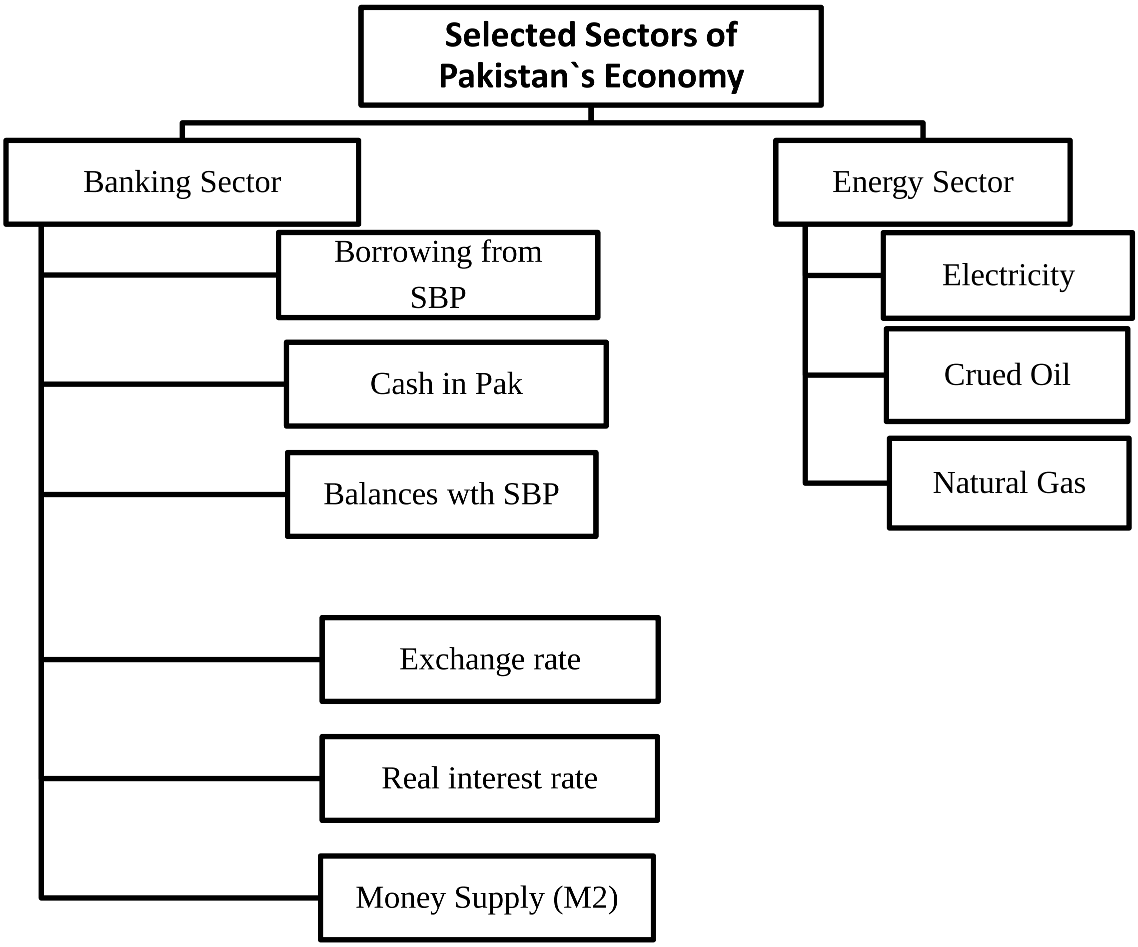
The different fields of gas are: Dhodak (1976), in Punjab, Dakni (1983), in the south west of Islamabad. Dakni started production in (1989), Qadirpur gas fields (1996). Pakistan is one of the largest consumers of gas in the region while Pakistan's proven coal reserves are the world's sixth largest. Thus the government intends to increase the share of coal in the overall energy mix. There is a claim that coal quality is inferior and having low BTU (British Thermal Unit), however, in this era of modernization, technology and boilers are available that can burn any kind of coal, still there is need of true economic cost of measurement. However, there is no doubt that the strategic location of country has the attraction and significance as an attractive market and transit route for energy, therefore merits the investor friendly policies. Pakistan is facing 1.8 billion cubic feet gas shortfall because of high demand in winter season.

Crude oil

There are different oil fields of Pakistan. The first oil field in Pakistan was determined in the province of Baluchistan near a Sui gas fields in 1952. Oil field located in the Pothwar region of Punjab were found in 1964. During Ayyub khan`s regime Pakistan petroleum and Pakistan oil fields explored and drilled the first well. Oil fields have an approximate capacity to produce 60 million barrels of oil.

In figure 2.1 have selected two sector from macroeconomy of Pakistan, one is the banking sector, we use six different series i.e. Borrowing from SBP, Cash in Pakistan, Balances with SBP, Exchange rate, Real interest rate and M2 (money supply). In energy sector, we use three series i.e. production of Electricity, production of Natural gas and Crude oil.

Figure 2.1: Structure of Selected Series



To carry on this research, we have chosen three cases as following:

Case 1: Borrowing from SBP, Cash in Pakistan and Balances with SBP.

Case 2: Electricity, Crude oil and Natural gas.

Case 3: Real interest rate, Exchange rate and Money Supply (M2).

CHAPTER 3

METHODOLOGY AND MODEL

3.1 Methodology

There is no single superlative technique for all change-point formulations and applications. But with the help of general characteristics or with specific problem detection compared between these selected macroeconomic variables. To search the possible change points in time series data of Pakistan, we have applied the E-Divisive method on time series data. We have used these series empirically and it is estimated to find out the significance change point for Pakistan`s data.

3.2 Data

To carry on this methodology, data from two important sectors, Banking sector and Energy sector of Pakistan economy have been selected. These sectors have been selected due to their importance in growth of economy. From banking sector six series Cash in Pakistan, Balances with state bank of Pakistan, Borrowings from state bank of Pakistan, M2 (money supply), Exchange rate and Real interest rate from IFS are taken. From energy sector three time series Electricity, Natural Gas and Crude Oil are taken. This study uses secondary data from statistical bulletin of State Bank using monthly data in time period 1990 to 2016. To detect the change point analysis in correlation structure using non-parametric method i.e. E-Divisive. We have applied these series and tried to find the significance change points.

3.3 Method of E-divisive

The E-divisive method recursively screens a time series & uses a permutation test to define change points, but it is computationally rigorous. The E-divisive with median algorithms exist as a way of generating a computationally controllable procedure for defining whether or not a new lump of time series data is significantly different from the previous through the use of advanced distance statistics robust to irregularities such as those present in twitter's cloud data. The method of "E-Divisive" chronologically test for testing the significance of every change point estimation given the earlier change in the estimation of locations, although its output is sensitive and to be contingent on the estimated change points numbers. This method estimates both the location and change point numbers. It hierarchically test to check the significance of every detected change points. The methods are also available for the estimation of change point's locations, without a special information of the number of multiple change points. The method adopt that observation are sovereign with limitedα^{th} absolute moments for $\alpha \in (0,2)$.

E-divisive detects change points by quantifying how different the characteristic functions of the distributions of subsequent segments of the time series are (Matteson & James, 2014). The E-divisive with median algorithms exist as a way of creating a computationally tractable procedure for determining whether or not a new chunk of time series data is considerably different from the previous through the use of advanced distance statistics robust to anomalies such as those present in twitter's cloud data. E-Divisive method sequentially test the statistical significance of each change point estimate given the previously estimated change locations, even though its running time is output sensitive and depends on the number of estimated change points. This method estimates both the number of change points and their locations. It hierarchically tests the statistical significance of each hierarchically estimated change point.

E-divisive performs the following segmentation steps.

- Segment the time series into two phases for which the characteristic functions maximally differ.

$$\hat{Q}(\tau) = \frac{\tau(n-\tau)}{n}\left[\frac{2}{\tau(n-\tau)}\sum_{i=1}^{\tau}\sum_{j=\tau+1}^{n}\left\|X_i - X_j\right\| - \binom{\tau}{2}^{-1}\sum_{i=1}^{\tau-1}\sum_{k=i+1}^{\tau}\|X_i - X_k\| - \binom{n-\tau}{2}^{-1}\sum_{j=\tau+1}^{n-1}\sum_{k=j+1}^{n}\left\|X_j - X_k\right\|\right] \quad (3.1)$$

Where $\|.\|$ denotes the Euclidean distance,(Matteson and James (2014) have indicated the option of increasing the Euclidean distances to a power α, with 0<α<2. They have used the default α- value of 1, as similar results were claimed to be obtained when other α -values were used). The optimal estimate of the change point location can be derived by inspecting which τ-value maximizes $\hat{Q}$.

- Determine if the change point is significant through a permutation test.

After estimating the change point location, its significance is tested by means of a permutation test on the maximal Q^ -value.

3.4 Change Point Estimation

This section reveals information about estimating the location of the change point and number of change point using E-Divisive estimating method.

3.4.1 Estimating the Locations of Change Points

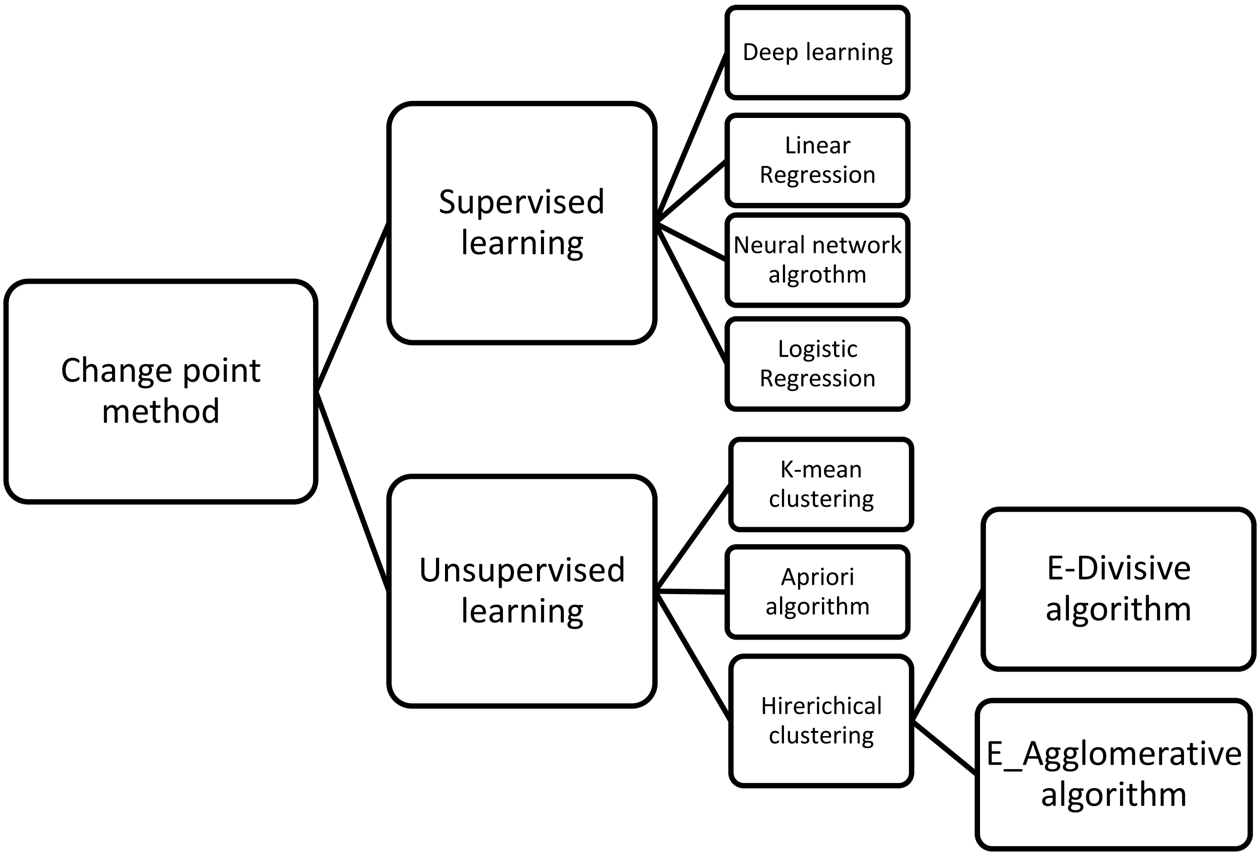
In this paragraph, we will define a change point based on the location measure change point analysis can be performed by parametric or non-parametric distribution. In present study we do not have an idea about the distribution of parameters of our banking sector and energy sectors. Therefore our multiple change point design is going to be performed in non-parametric settings. Our estimation is based upon algorithm e-divisive method. This method estimation sequentially identifies change point via a bisection algorithm.

3.4.2 Hierarchical Estimating Methods

In Cluster analysis we wish to partition the observations into homogeneous subsets. Subsets may not be contiguous in time without some constraints. Matteson and James (2013) invent a new technique for multiple change points detection based on hierarchical clustering for time-ordered data. They mainly focus on divisive clustering, i.e., clustering from top to bottom. However, in this thesis we will equally talk about agglomerative clustering, which means clustering from bottom to top, and divisive clustering. We will also keep the name for the clustering methods consistent with their, E-Divisive respectively. For E-Divisive, multiple change points are estimated by iteratively applying a procedure for locating a single change point. The statistical significance of an estimated change point is determined through a permutation test.

To explain the change point method, a detailed structure has been presented in form of figure (fig 3.1). The figure 3.1 explains the background structure of E-Divisive algorithm which is mainly used in our study.

Figure 3.1: Structure of E-Divisive Method



Most statistical learning problems fall into one of two categories: supervised or unsupervised. We have discussed so far unsupervised all fall into the supervised learning domain. For each observation of the predictor measurement(s) xi, i = 1,...,n there is an associated response measurement yi. The aim of accurately predicting the response for future observations better understanding the relationship between the response and the predictors.

Unsupervised learning describes the somewhat more challenging situation in which for every observation i = 1,...,n, . Vector of measurements xi but no associated response yi. It is not possible to fit a linear regression model, since there is no response variable to predict. The situation is referred to as unsupervised because lack a response variable that can supervise analysis.

Chapter 4

SIMULATION ANALYSIS

We arrange the simulation study to estimate the performance of the non-parametric test i.e.E-Divisive in finding the occurrence of at least one correlation variation. These calculations take the situations with no, single and multiple change points to examine the test's perform in different settings. In our simulation study, case of multiple change points andsingle change point are considered, where there are same and different starting points.

4.1 Monte Carlo Simulation Design

In this chapter mainly discusses comprises of the Monte Carlo simulation details of the E-Divisive method which is used to be detect the change points. Y objective is to explore the performance of E-Divisive method to detect the change in multivariate frame work and correlation structure on the bases of Monte Carlo simulation. The Monte Carlo comprises the following components. First is the data generating process. Second to detect the change points in mean and variance at different and same location. Third to detect the change points in covariance at same and different location.

4.1.1 Data Generating Process

We generate three series on mean and variance to detect change point. There is no specific way through which we select variances and covariances randomly. That's why in this study variances and covariances are selected non stochastically. Study allocate the random values of S.D (standard deviation) and mean, the values are:

Mean=[0, -5, 5]

S.D=[3, 7 ,1]

In order to check the change point, we picked up the pair from mean and S.D one by one.

$$Y_1 = Mean_1 + S.D_1$$

$$Y_2 = Mean_2 + S.D_2$$

$$Y_3 = Mean_3 + S.D_3$$

For first simulation study, we assume the detected change points at same location, (with correlated variables). Our simulation is based on 5000 time iterations. In second simulation experiment iteration, we consider the cases for change points at different location and repeated the process 5000 times.

4.1.2 Detect the change point in mean & variance at same and different location

In this section we select value of mean and variance one by one by which we detect the change points at same and different location.

4.1.2.1 Effect the number of change points at same location

In this section, we find the change point initiated at same location in all series. Hence we have chosen different sizes of change point by taking different k values and check how many are detected the change points and same location of change points. We have generated three series (V) with multivariate normal distribution, which is shown above that is 1,2,3.We have introduced change points (K) from 0 to 5 points to use in 5000 time simulation iteration

Change points Numbers K: 0,1, 2,3, 4,5

Variables V: 1,2,3

To calculate the running correlations, we use a minimum size of clusters. Simulations are run for 5000 iterations and the E-Divisive method performed is assessed by observing its power, which is calculated as the number of data sets declaring at least one significant change in correlation through the test. We chose the level of significance for all iterations at $\alpha = 0.05$

Table:4.1 Empirical power calculation of E-Divisive method for 5000 iterations (same location)

K (change points)	Power for multivariate case	Power for univariate cases		
		Series 1	Series 2	Series 3
1	0.97	0.88	0.89	0.86
2	0.88	0.65	0.61	0.72
3	0.89	0.48	0.42	0.52
4	0.87	0.46	0.40	0.52
5	0.89	0.38	0.46	0.42

We can see from the table 4.1 that the change points detected by the E-Divisive method are not satisfying in case of univariate analysis. We generate three series to examine the performance of the nonparametric method. From the data generated process we have generated three different stationary series with different number of change points (k), taking each one by one. Then refer these series to E-Divisive method to find the number of change point. The results shown in table 4.1 reveals that when we put k=0 i.e. no change points in generated series the E-Divisive method correctly identifies with probability of 0.96. K=0 is not null hypothesis, because this study introduce change points from 0 to 5 (where K= no of change point). When we have 1 change point in the data generating process of all series, the E-Divisive method identifies the change point with probability of 0.88. 0.89 and 0.86 which is high value. In case of multivariate analysis the e-divisive method is performing well in case of k=0 i.e. no change point as well as for k=1 i.e. one change point in all series under consideration. The power of test at k=0 is calculated as 0.96 and k=1 it is 0.97. As we increase the number of change points the performance of E-Divisive method for change point detection decrease in univariate case as well as in multivariate. Comparatively, multivariate case is as compared to univariate case because the empirical power in multivariate case remain high i.e. about 0.88 even when we have large number of change points. While in case of univariate analysis the empirical power falls drastically as k increase. The multivariate case is outperforming in each case we increase the number of change points, e-divisive method gives the best performance in case of multivariate analysis but the power of e-divisive is fails in case of univariate.

Table:4.2 Detected change points by E-Divisive (Same location)

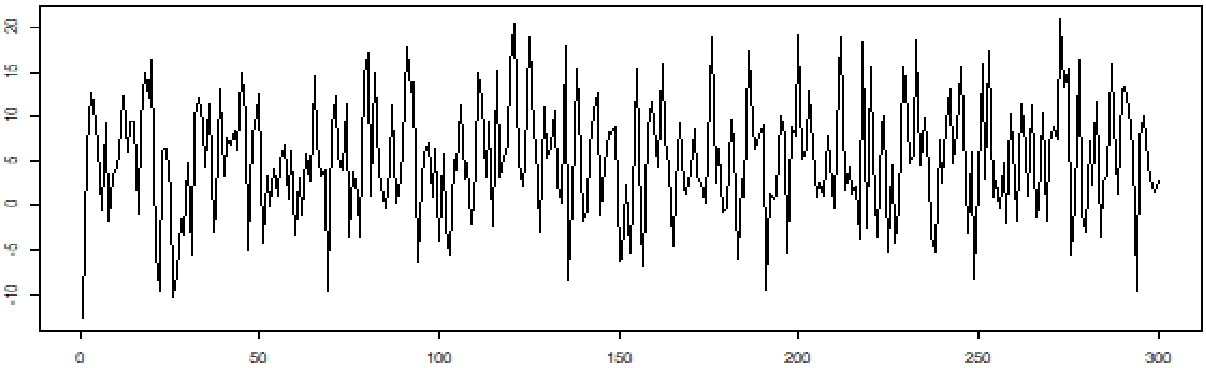
K	Actual location of change points detected				
K=1	150				
K=2	101	201			
K=3	72	151	227		
K=4	61	121	181	241	
K=5	51	101	151	201	251

Table 4.2 gives the simulation results for a number of change points detected by e-divisive method. Keeping the significance level to 5%, other change points are also detected as we change the k values to k=0,1,2,3,4,5.

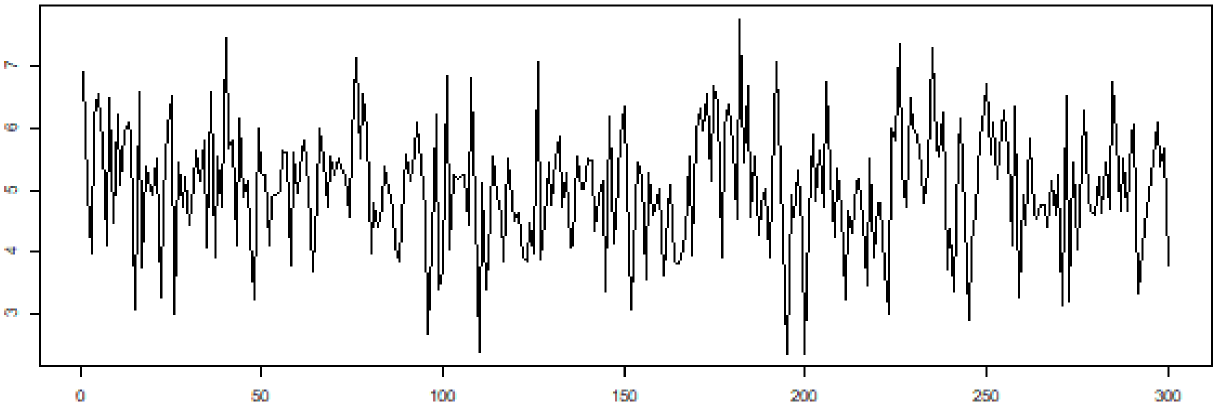
We have used the R package procedure mvt norm to generate the observations. We have generated 300 normal observations. Thus among the 300 observations there are many change points, 150, 101, 201, 72, 151 with p-value 0.00332 and change point 227 with p-value 0.02990 respectively. The power of E-Divisive test is 97% and location of change point is 150 when we detected point k=1. Similarly we select value of k=2 the E-Divisive method has detected the change points at 101 and 201 with power 88%. we have detected the change points on 72, 151 and 227 with power of e-divisive test 89% when k=3. The table 4.2 represents the detected location of change points and we keep the value of k=4 then detection point is same for all series that point is 61, 121, 181 and 241 with power 87%. When we select k=5 this method detects the five change point location, that is 51, 101, 151, 201 and 251 with power of E-Divisive method is 89% correctly detect the change points at same location (see power of calculation in table 4.1). Every time we have taken different k values and it gives the correct results for the hypothetically generated change points in different series. It has detected the different point correctly in all simulations in case of multivariate.

Moreover graphically, the analysis is also presented. The analysis is arranged in order, location and in number of change points. Assuming time series change point definition as a change in mean or standard deviation value of time series data, the simulation randomly selects a new pair of these values. There is a 0.01 probability of selecting two identical pairs consecutively, this is fairly low so we can be confident that each change point gets a new value of mean and standard deviation. The information confined in the graphs is given below.

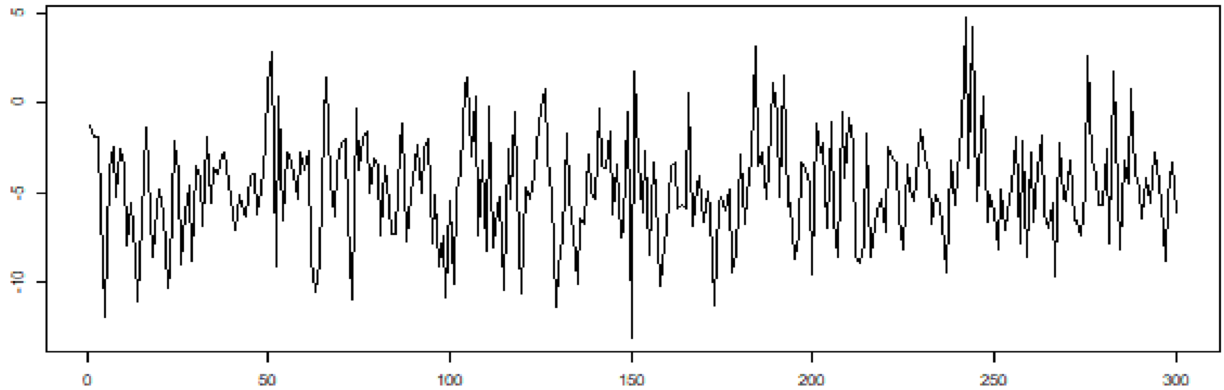
Figure:4.1 Change point detection in same location when k=0



No change point in generated (Y_1)



No change point in generated (Y_2)



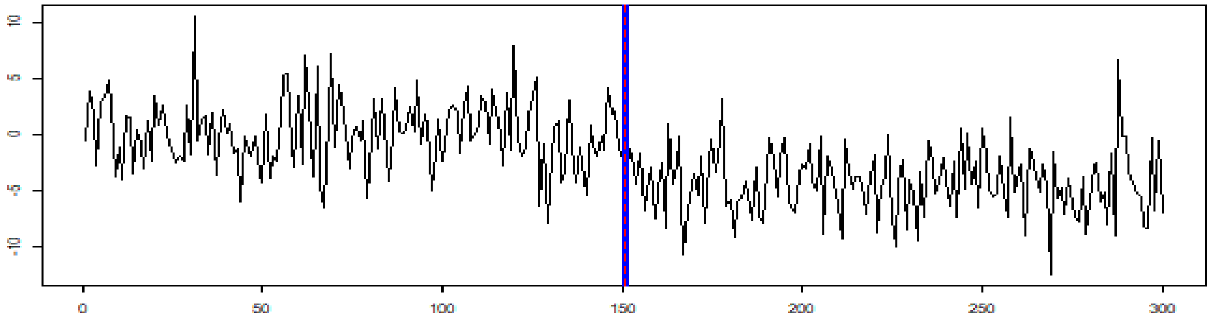
No change point in generated (Y_3)

Detect the change point at same location when we keep the k value is 0

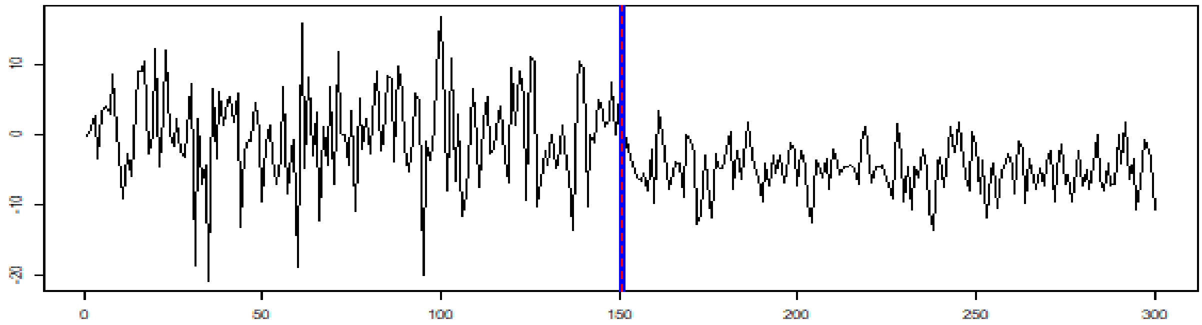
Figure 4.1. to 4.6 show the generated three series through multivariate normal distribution and detection of change points at different value of k, the multivariate normal distribution is offen used to describe at least approximately, any set of

(possibility) correlated real valued random variables each of which clusters around a mean value. In figure 4.1 it is observed that there is no change point detected because we chose the k value 0 in 5000 simulations. The graph shows that the e-divisive method has not provided any evidence of change point detection for each series.

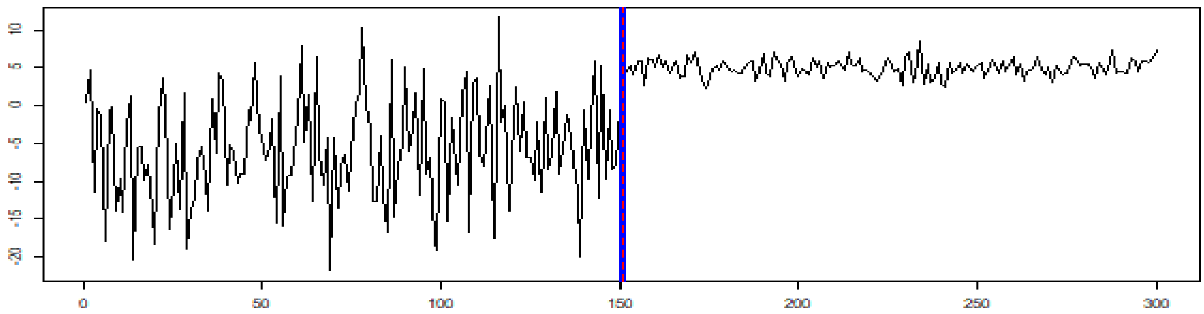
Figure:4.2 Detect the same location of change point (k=1)



One change point in generated (Y_1)



One change point in generated (Y_2)

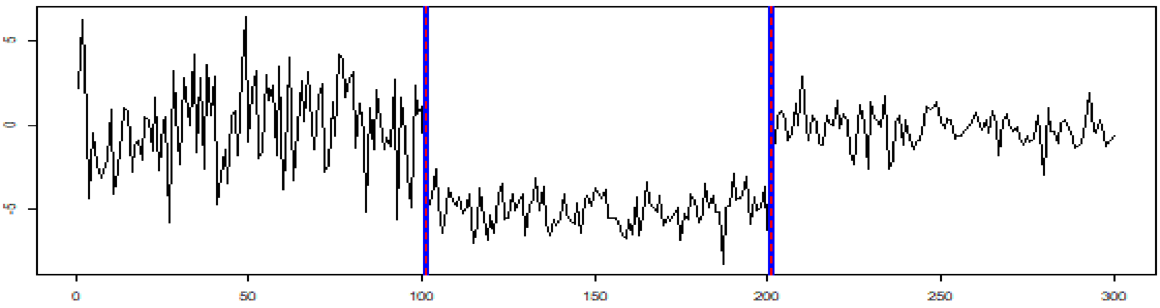


One change point in generated (Y_3)

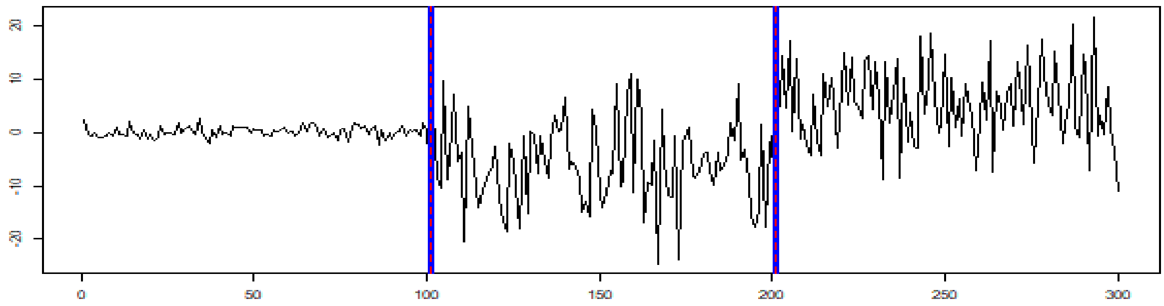
Detected only one change point at same location for each series.

The above figure 4.2 represents one change points for each series. In data generating process, e-divisive method detects one change point when k value is 1 and the simulation is done 5000 times. E-divisive method detected change point per series at same location. Therefore the most of the graphs represent the change point of all three series at same location (150) when E-Divisive method detect the single change point.

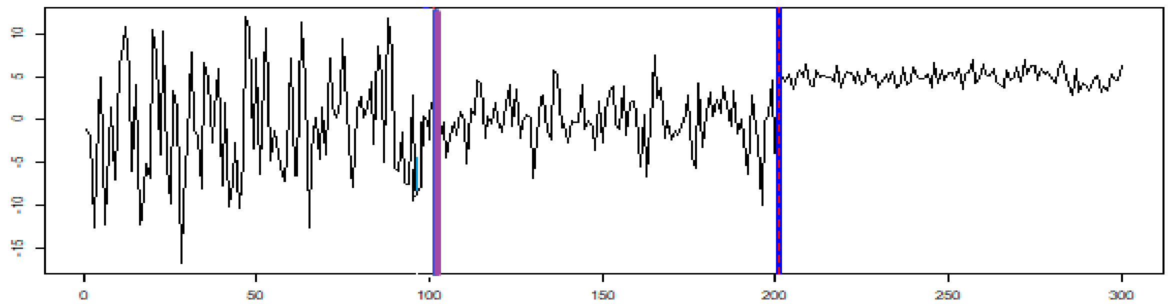
Figure:4.3 Detect the same location of change point(k=2)



Two change points in generated (Y_1)



Two change points in generated (Y_2)

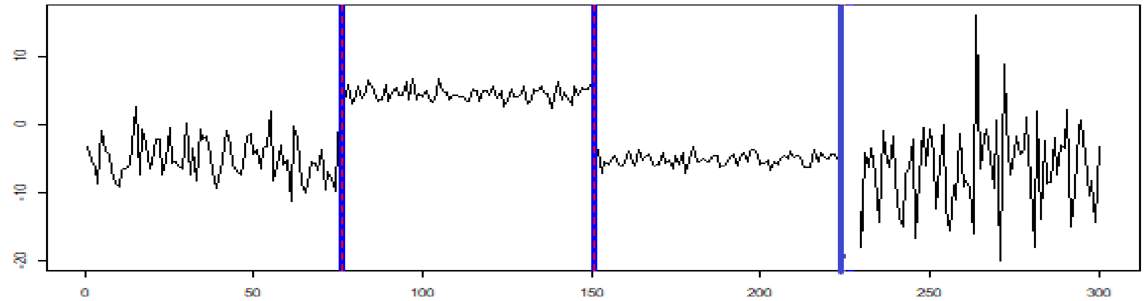


Two change points in generated (Y_3)

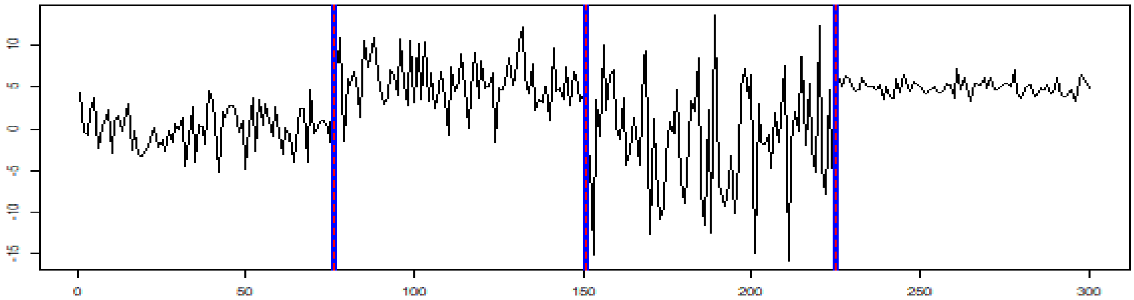
Two change point detection is at same location for each series when selected the value of k is 2

In figure, 4.3 only two change points are detected at same location because we introducek value is 2, so e-divisive method has detected the two change points at 101 and 201.

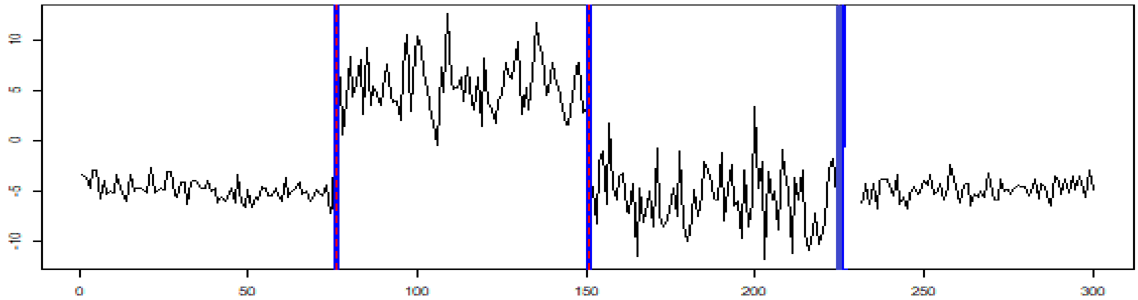
Figure:4.4 Detect the change points when (K=3)



Three change points in generated (Y_1)



Three change points in generated (Y_2)

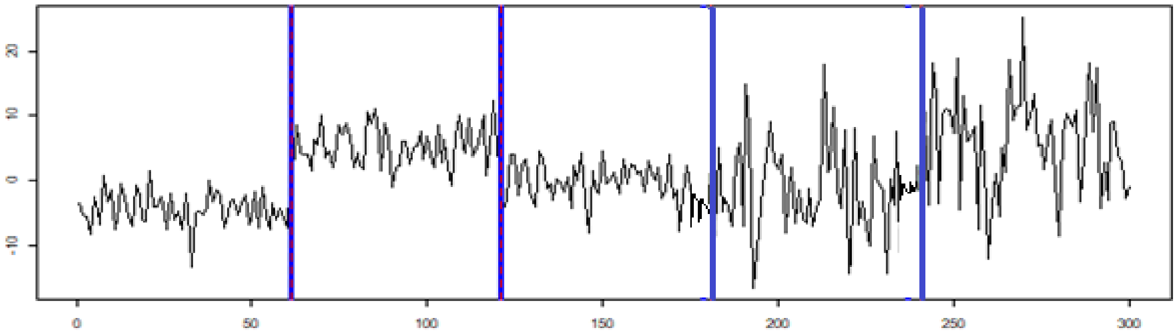


Three change points in generated (Y_3)

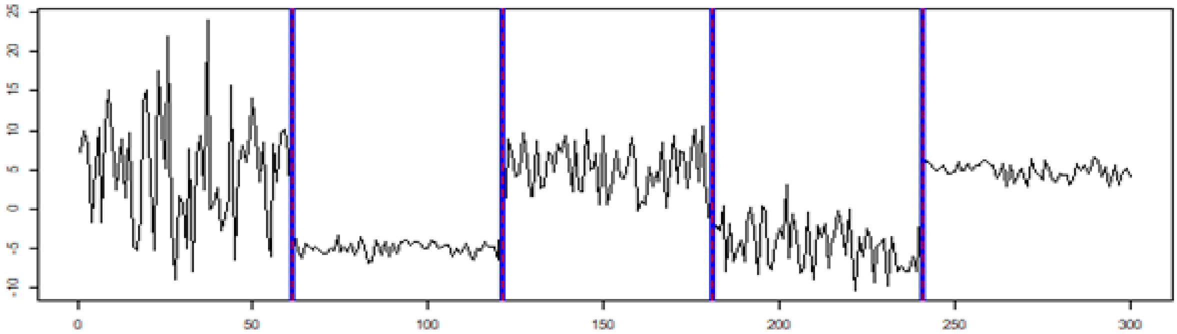
E-Divisive method is detect the three change points for per series, when k=3

Figure 4.4 is showing the three change points at same place, the location of change point is 72, 151, and 227.

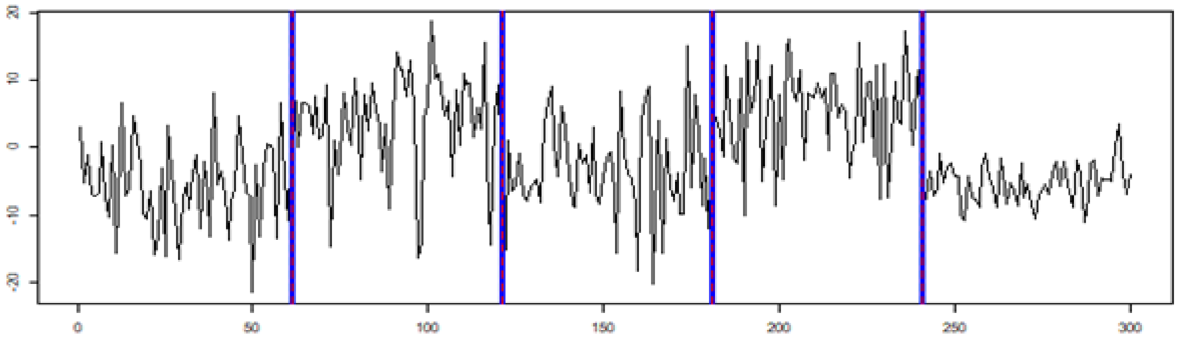
Figure:4.5 Detect the location of change point when (K=4)



Four change points in generated (Y_1)



Four change points in generated (Y_2)

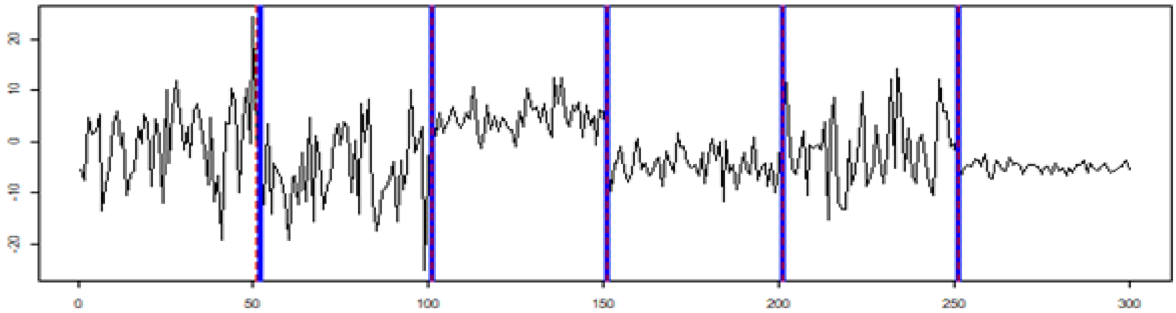


Four change points in generated (Y_3)

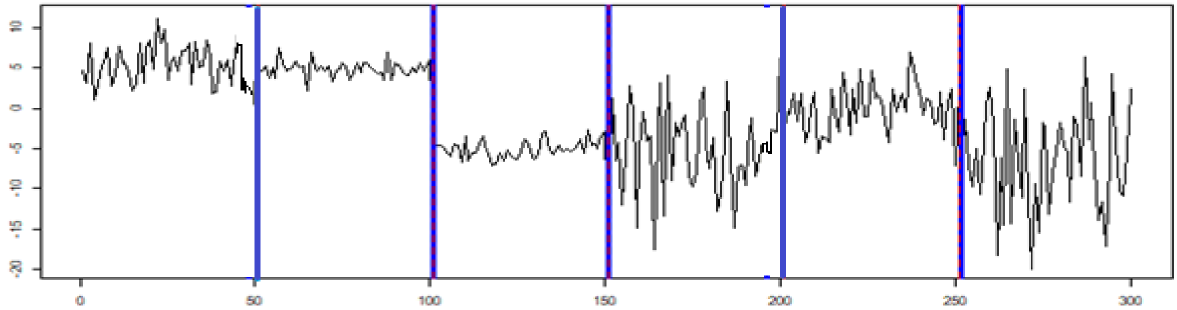
Four detected change points at same place for per series

Fig 4.5 is represents the four change points in each series, as we introduce the value of k is 4 then the e-divisive method has detected the four change points. In this scenario, this approach has detected the change points at correct location in all repeated simulation 5000 times. This method has given different results for every simulation, and detected only 4change points but these change points at same locations in each series.

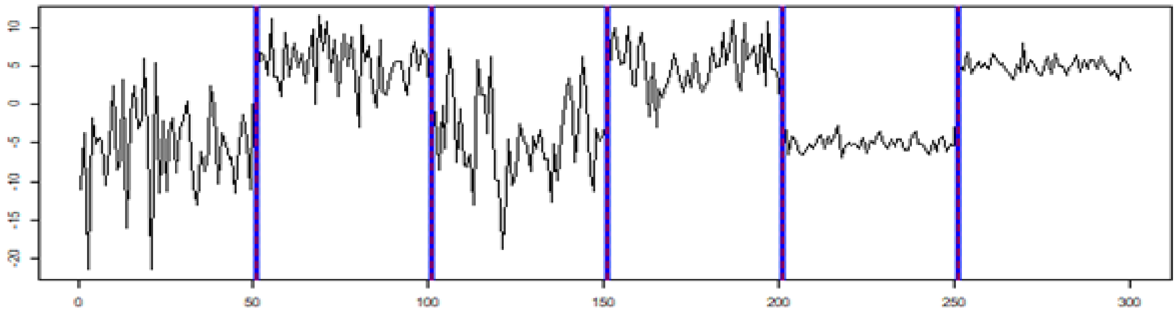
Figure:4.6 Change point detection at same location when (K=5)



Five change points in generated (Y_1)



Five change points in generated (Y_2)



Five change points in generated (Y_3)

In this graph, this approach is detected five change points at same location for each series.

Figure 4.6 five change points in each series are detected. In each series, the change points are 51, 101, 151, 201, and 251 at same location. Here DGP series is stationary.

4.1.2.2 Effect the number of change points at different location

In this section we have now moved to check change points at different locations. We have chosen the different k values and checked on different location.

Table:4.3 Empirical power calculation of E-Divisive method for 5000 iterations at different location

K	Power for multivariate case	Power for univariate cases		
		Series 1	Series 2	Series 3
1	0.92	0.86	0.87	0.80
2	0.93	0.72	0.72	0.70
3	0.9	0.46	0.52	0.50
4	0.86	0.48	0.43	0.41
5	0.84	0.31	0.40	0.42

In table 4.3, correctly detected the change points which we introduce the different k value in multivariate and for collectively series.

Results in table 4.3 are representing that whatever change points are introduced, e-divisive method has detected in a new point in every series. As in table 4.3, by keeping value of k=0, we have concluded that results are different for multivariate and individual analysis. That means, we have taken the value of k=0, how many times e-divisive method has detected the change points in case of multivariate analysis. The table 4.3 shows that this method has detected the zero break points for 92 times correctly and in univariate analysis. Moreover this method has detected the value of zero for different time correctly. Like, in first series, this method has detected zero for exactly 92 times and in the second series has detected for 97 times correctly. Same as by keeping value of k=1, this method has detected for 92 time in multivariate analysis and in univariate, this method has detected the value of 1 less as compared to multivariate. Similarly, in series 1, it has detected k=0 for 86 times, in series 2 this point has detected for 87 time and in series 3, correctly detected for 80 times. Same method has been used when we choose different k values as: k=2,3,4,5. Every time this method has detected the points less as compared to already detected values. Results are showing that performance of E-Divisive is better in multivariate as compared to univariate, that is because by calculating the power of these points, it has proved that power of detecting change points of multivariate is better as compared to individual analysis. Results enabled us to identify that for greater number of break points in series, the power of change point detection through e-divisive decrease gradually and the power of detected change points decreases speedily. Hence, we conclude by analysis that E-Divisive method is much more better in case of multivariate.

4.2.1 Detection of change points by E-divisive method

In this section, detection of change points has been done on different location of generated three series. The statistical results of these three series are given in table 4.4.

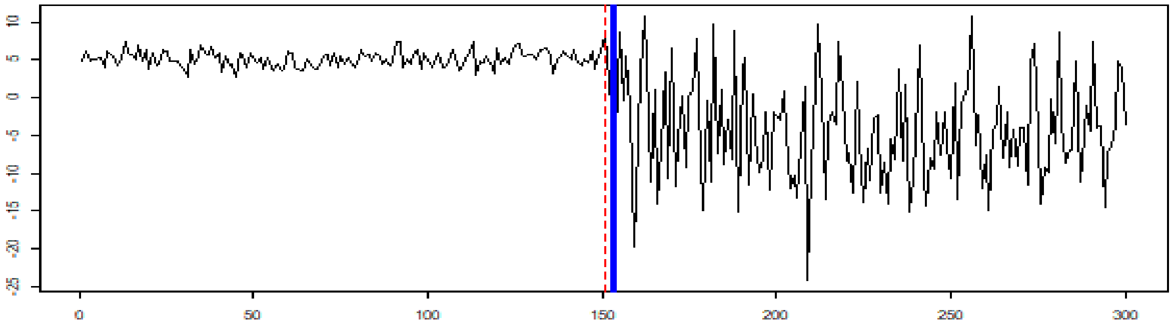
Table:4.4 Detected the change points at different location

K	Actual location of change points detected											
	Series 1					Series 2			Series 3			
K=1	153					149			151			
K=2	200					101	201		97	219		
K=3	76	214				151	231		80	151	226	
K=4	61	181	242			62	121	241	64	119	181	243
K=5	48	106	151	201	252	51	146		69	99	150	198

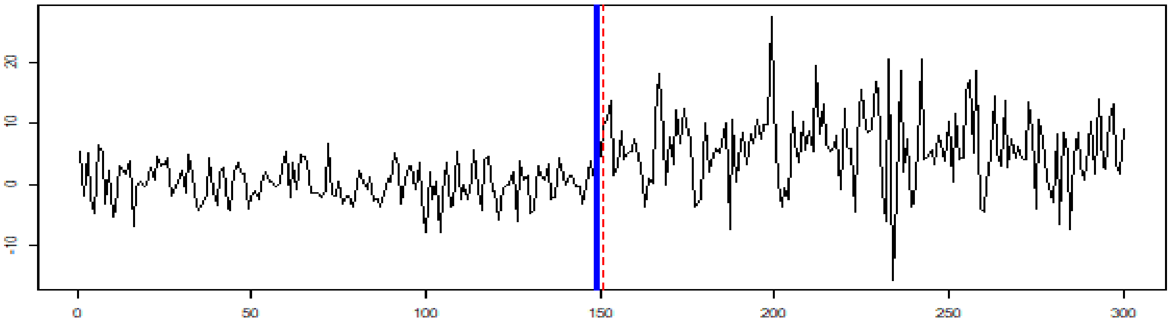
Detected the change points at different location

We have generated three data series of 300 observations with normal distribution. E-Divisive method has detected different change points at different location in series 1, 2 and 3. The table 4.4 represents the detected location of change points. If we keep the value of k=1 then detection point is different for all series that point is 153, 149, 151 with power of detected change points 92%. These points shows the location of change point at different location through e-divisive. When we chose the k=2 , E-Divisive method detects total five different change points for all series which is 200, 101, 201, 97 and 219 with power of change point detection is 93%. When we select k=3this method detects the different change points at different location, that is 76, 214 in series 1, 151 and 231 in series 2, 80, 151, 226 with power of change points is 90% in series 3. These change point gives us different detection points. Every time we have taken different k values and it gives the correct results for the hypothetically generated change points in different series. It has detected the different point correctly in all simulations at different locations in case of multivariate. Same as by setting the value of k=4 and 5 e-divisive method detects the different change points with calculate the power detection is 86% and 84% at different location in each series (see power of calculation in table 4.3).

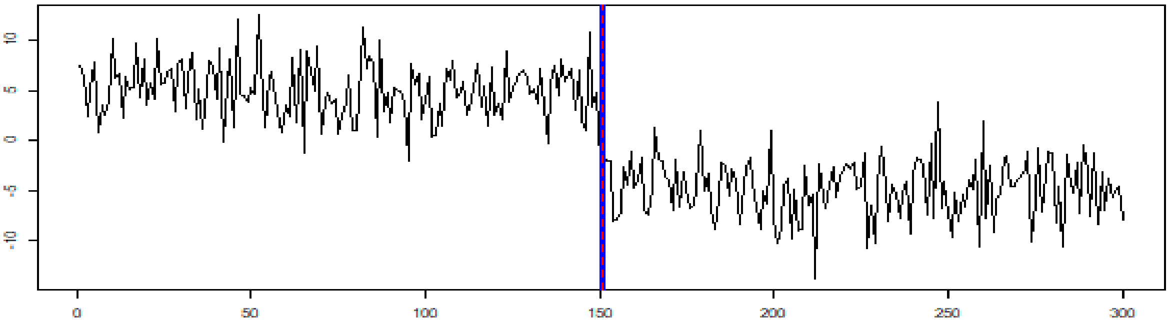
Figure:4.7 Detect the change point at different location when k=1



One change point in generated (Y_1)



One change point in generated (Y_2)

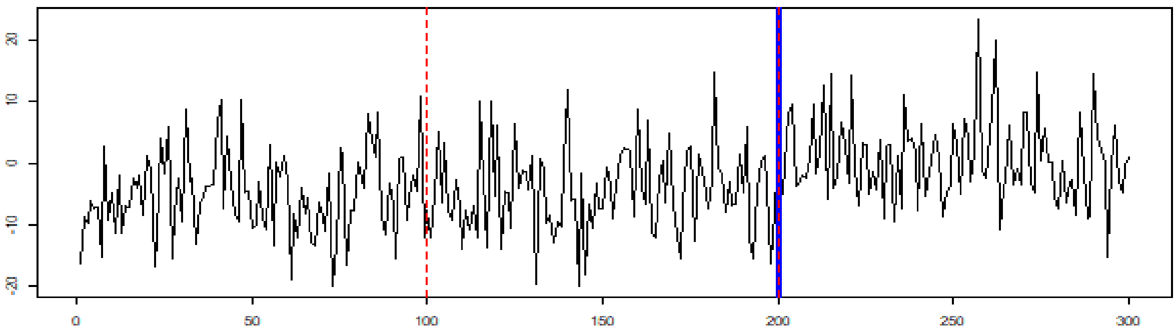


One change point in generated (Y_3)

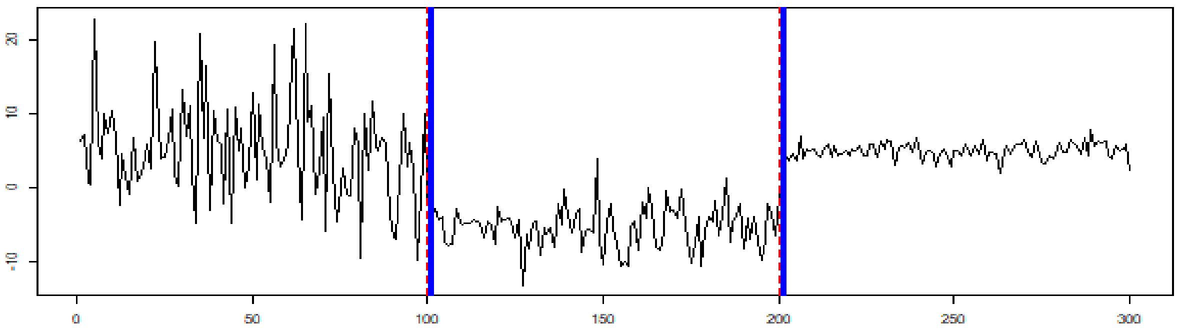
Figure:4.7 Simulation change point detection from k=1

Figure 4.7 detects the one change point for every series but at different locations. Doted lines shows the original change points whenever the solid lines represents the estimated change points. In series one the change point is at t=153,one change point is at t= 149 in series two andthere is one change point in series three at t=151.

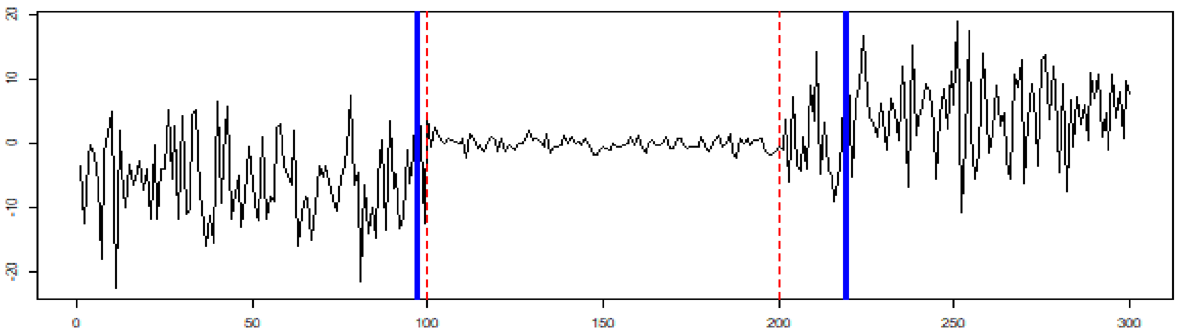
Figure:4.8 Detect the change point at different location when K=2



Two change points in generated (Y_1)



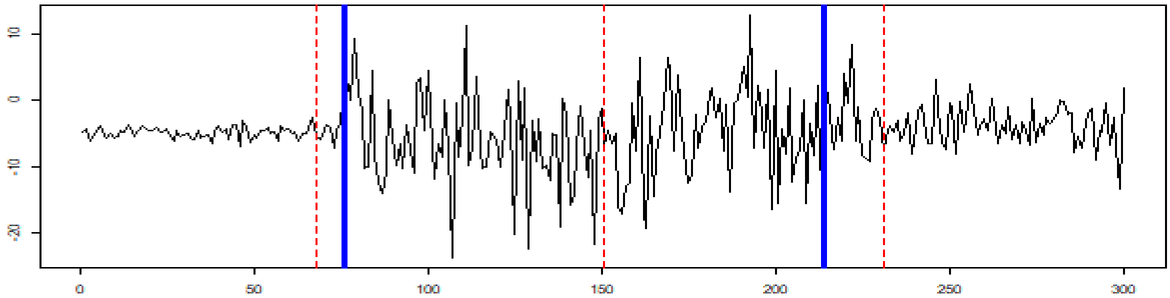
Two change points in generated (Y_2)



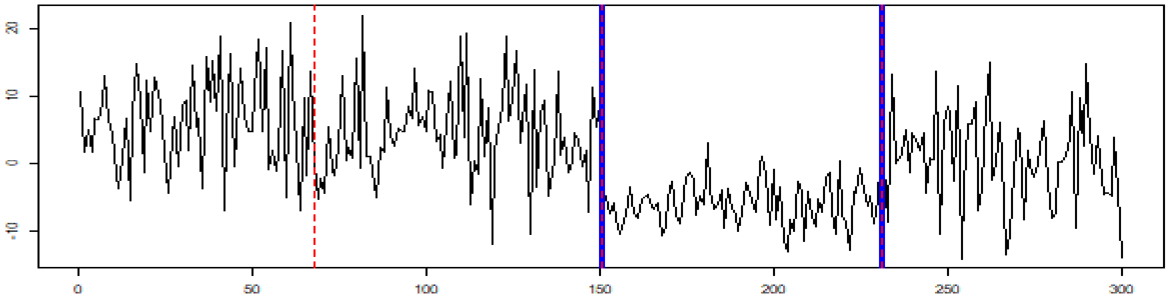
Two change points in generated (Y_3)

Figure: 4.8 Two change points has detected through e-divisive simulation at different locations in each series. E-divisive method has detected one change point in series one and detected two change point in series 2, 3 but at different locations.

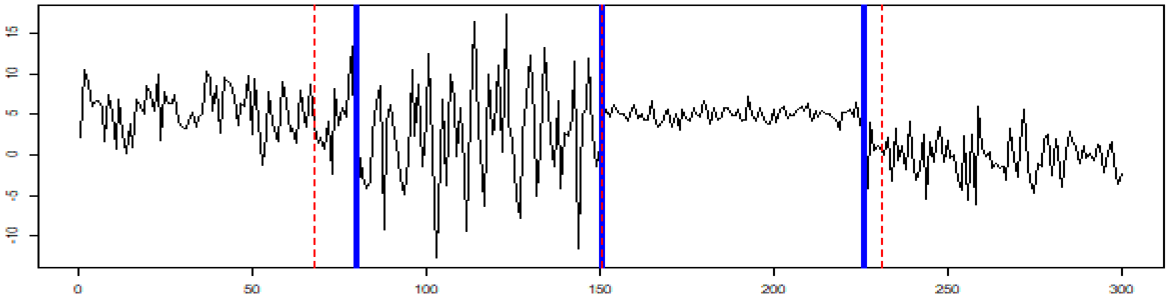
Figure:4.9 Detect the different location of change point when K=3



Three change points in generated (Y_1)



Three change points in generated (Y_2)

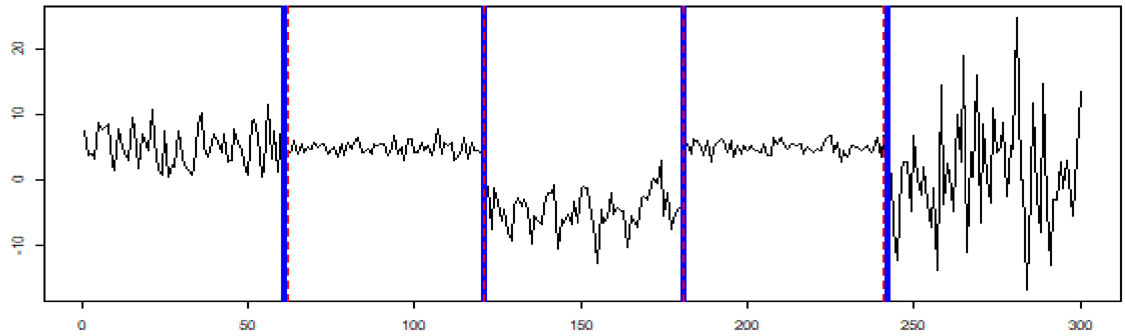


Three change points in generated (Y_3)

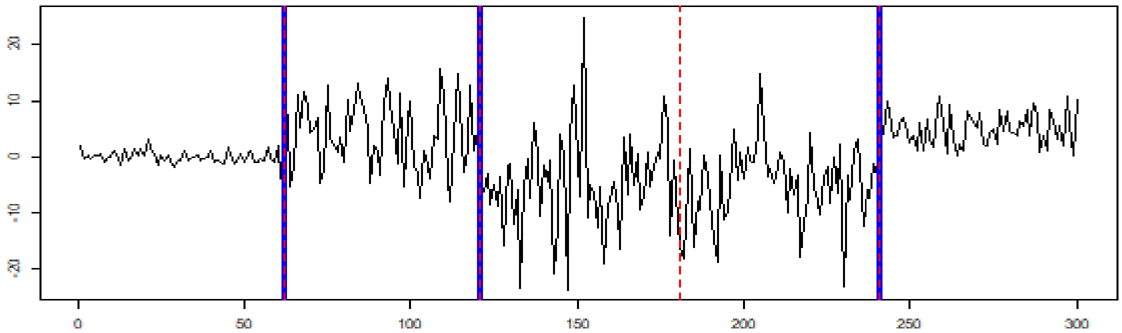
Figure:4.9 three change point detect in each series at different place through e-divisive simulation

We have selected the value of k=3, the three change points for each series such as doted lines have shown in figure 4.9 and after applying this method, we have detected the change point at different locations. Three change points have been detected in series one, two change points have detected in series two and three points have been estimated in series three but these change points exist at different places.

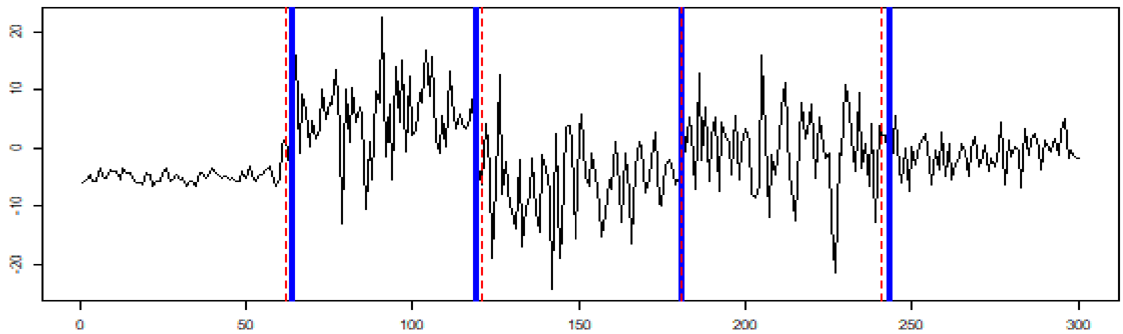
Figure:4.10 Detect the different location of change points when K=4



Four change points in generated (Y_1)



Four change points in generated (Y_2)

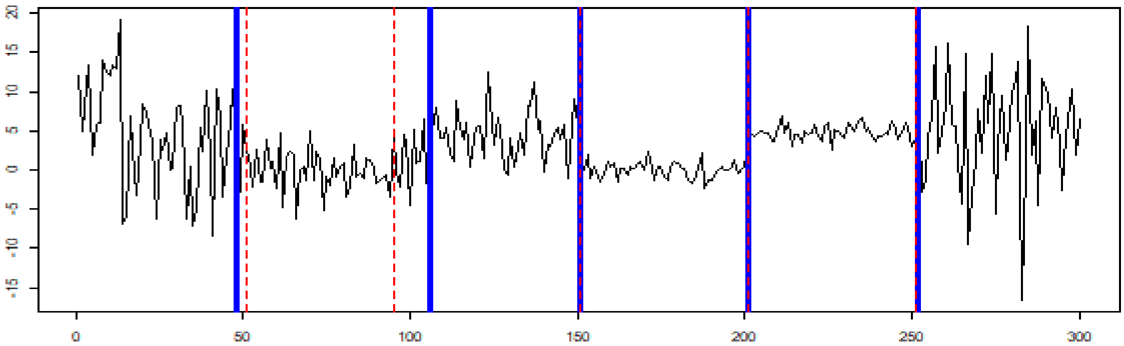


Four change points in generated (Y_3)

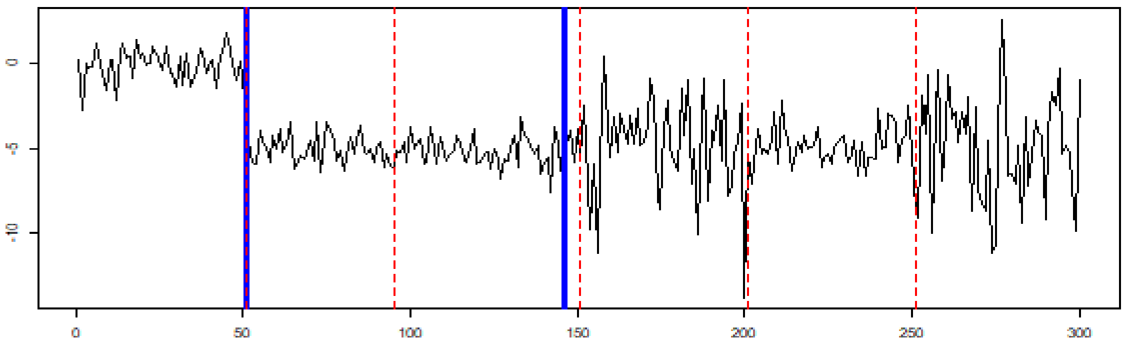
Figure:4.10 this graph has informed about change points detection

In figure 4.10, there are four change points in generated series. We have estimated the four change points for every series but at different locations. Solid lines have estimated change points after applying e-divisive simulations.

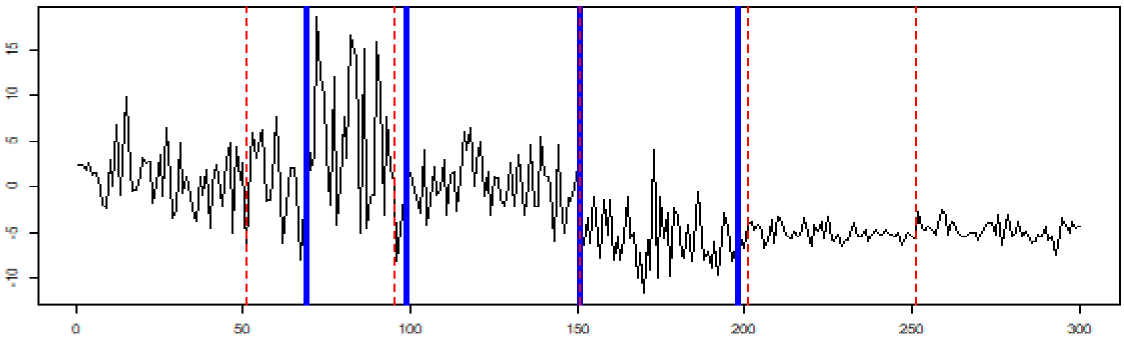
Figure:4.11 Detect the change point at different location when K=5



Five change points in generated (Y_1)



Five change points in generated (Y_2)



Five change points in generated (Y_3)

Different five change point detection at different locations

In fig 4.11 there are five change points detected at different places. In every simulation graph has given different observation and also detected the change points at different location.

4.1.3 Detect the change point in Co-Variance at same and different location

In this section, this study detect the change point in covariance structure. We generated three series and randomly select the value of variance and covariance matrix. By putting new value of covariance to detect the change point. We randomly select the value of covariance between -10000 to +10000.

$$\begin{bmatrix} 1 & 2835 & -2437 \\ 2835 & 3 & -2818 \\ -2437 & -2818 & 7 \end{bmatrix}$$

In the same matrix, we changed value of covariance to detect the change points

$$\begin{bmatrix} 1 & -4834 & 6269 \\ -4834 & 3 & 5107 \\ 6269 & 5107 & 7 \end{bmatrix}$$

4.1.3.1 Change point detection in co-variance at same location

Table 4.5 Empirical power calculation of E-Divisive method for 5000 iterations at same location

K (change points)	Power for multivariate case	Power for univariate cases		
		Series 1	Series 2	Series 3
1	0.58	0.19	0.18	0.12
2	0..25	0.03	0.01	0.05
3	0.1	0	0	0.01
4	0.01	0.01	0	0
5	0.01	0	0	0.

Table 4.5 showing that the power calculation for individual series and multivariate series, when we introduce the new change point k then the power is less but the power is more less in each series in case of covariance structure. E-Divisive method is better performance in level of mean and variance rather than covariance structure, because power of change point detection is high in case of mean and variance.

Table 4.6 Detected change points at same location

K	**Actual location of change points detected**		
	Series 1	**Series 2**	**Series 3**
K=1	49		
K=2	75 120	120	75 120
K=3			70 120
K=4		130 170	170
K=5			49 120 201 260

In table 4.6 is the same location of change points on each series. When we introduce the change point k=1 then only one change point detected in series 1 but there is no change points in series 2 and 3. Similarly, we introduced the different value of k (k= 2, 3, 4, 5) the E-Divisive method is detected the same change points on each series, just like show in table 4.6.

4.1.3.2 Change point detection in co-variance at different location

In this section, we detect the power calculation of change point with the help of different k value.

Table 4.7 Empirical power calculation of E-Divisive method for 5000 iterations at different location

K (change points)	Power for multivariate case	Power for univariate cases		
		Series 1	Series 2	Series 3
1	0.52	0.17	0.14	0.17
2	0..24	0.06	0.04	0.03
3	0.1	0	0.02	0
4	0.03	0	0	0
5	0.01	0	0	0.

Table 4.7 represents the power calculation of detected change points on different location. Power for multivariate case is much better then power for univariate casesbecause E-Divisive method is detecting change points in case of multivariate, whenever in case of univariate its performance is poor.

Table 4.8 Detected change points at same location

K	Actual location of change points detected		
	Series 1	Series 2	Series 3
K=1	144	145	
K=2		160	46 175
K=3	60 220	91 160	120
K=4		180 230 260	
K=5	160 260	120 145 270	

Table 4.8 shows the location of change points like: if we put the value of k=1 then the E-Divisive method detected the only one change point in each series at different location i.e series 1 shows the change point at 144 and series 2 shows change point at 145, while there is no change point in series 3. Similarly, while assigning the value of k=2 , this method detected two change points. There is no change point in series 1 whenever one change point at 160 in series 2 and two change point 46 and 175 in series 3. Likewise, when we assign the k value 3, 4, 5 then this method estimate the change points at different location. All these change points are significant because p-value is less the 0.05.

CHAPTER 5

RESULTS AND DISCUSSION

This section contained the empirically conducted exercise for Pakistan, using variables measured on monthly frequency. For the identified number of change points and location of change points, we have run the change points detection, and adjustment procedure suggested by James and Matteson (2013).

Remaining chapter have, section 5.1 illustrated data description. 5.2 consisting analytical framework and compress graphical Analysis of Raw series.

5.1 Data Description

The empirical study conducted for Pakistan is considered by examining nine monthly measured time series. We have taken two sectors of Pakistan economy: banking sector and energy sector. For multivariate change point detection in correlation. We have considered 3 cases from these sectors: Cash in Pakistan ,Balance with SBP , Borrowing from SBP, Electricity, Natural Gas, Crude Oil, Exchange rate, Real interest rate and Money Supply (M2) (1990M1 to 2016M12). The data is taken from State Bank of Pakistan andInternational Financial Statistic (IFS). The descriptive statistics of the data are estimated and is given in the appendix. Descriptive statistics helps us understanding the nature of data.

5.2 Analytical Framework and compress graphical Analysis of macro series

E-Divisive method is used to identify multiple change points in Macro Economic time series data.

E-Divisive method is used for performing hierarchical divisive estimation of multiple change points. E-Divisive method consists of following steps:

- Multiple change point are estimated by iteratively applying procedure for locating single change point.
- At each iteration a new change point location is estimated so that it divides on existing segments.
- The statistical significance of an estimated change point is determined through a permutation test. Since the distribution of the test statistic depends upon the distributions of the observations which is unknown in general.

Hypothesis:

The hypothesis testing is done to check the possibility of single change point in the given series especially using e-divisive method. This method is used to check the homogeneity between two parts of the data set. It exist then no change point is faced otherwise some change points can be detected in its correlation structure. So, thehypothesis developed are:

Ho:No change point

Ha:A single change point

Before we perform e-divisive method to identify change point in the series, we first locate the correlation structure among these series.

As identified in fig (2.1) the three cases are discussed as follows:

Case 1: Borrowing from SBP, Cash in Pakistan and Balances with SBP.

Case 2: Electricity, Crude oil and Natural gas.

Case 3: Real interest rate, Exchange rate and Money Supply (M2).

Table 5.1:Correlation Metrix (Energy sector)

	Natural gas	Electricity	Crude Oil
Natural gas	1		
Electricity	0.44	1	
Crude Oil	0.60	0.49	1

Table 5.1 : The table is correlation structure among the series

Table 5.1 is representing the correlation structure between three series of energy sector. The correlation on off diagonals and variance on principal diagonal. This segments shows that the variables are possibility correlated. The Correlation between natural gas and electricity is 0.44. whereas the correlation structure between natural gas and crude oil is 0.60. Whenever the correlation between crude oil and electricity is 0.49 and vice versa for the upper diagonal same elements. This means that relation between crude oil and natural gas is more than any other relation among the series.

The analysis was carried out by using $\alpha = 1$ (when $\alpha = 1$is used the e-divisive method identify the changes in mean and variance. If $\alpha = 2$ is selected that means the e-divisive method can only identify change in mean. We are interested in finding changes, both in mean and variance, therefore we set $\alpha = 1$) minimum cluster size of 30, and used R=199 (where R is the maximum number or random permutation to use in each repetition of the permutation test) permutations with the level of significance $\alpha = 0.05$. The permutation is the act of arranging the members of a set into a sequence or order, or if the set is already ordered, rearranging its elements a process is called permutation. Based on this information the estimated change point using the E-Divisive method are located as following,

Table 5.2: Grouping data (change point references)

Groups	No of observations
1	176
2	159

Table 5.2: The grouping data of the series

In table 5.2, group have been designed based on e-divisive , where it means each is representative of some data around that cluster. Here group one consists of 176 data points centered around the first cluster. Similarly, group two consists of 159 data points.

Table 5.3: Location of multiple change points of time series data

Order form	**Change point location**			
consider. Last			103	
order. Found	1	324	177	136
estimate	136	177		103
p-value	0.025*	0.005*		0.165

Table 5.3: The table is the location and number of significance change points

We have used three series in our analysis but found only two change point in the series. One change point is 136 and second change point is 177. In table 5.3 is shows that, here the estimate refers to the location of the statistically significant change point. Considered. last is the location of the last change point that was not found to be statistically significant at the given significance level. Here permutation shows the number of permutation performed by each of the sequential permutation test, the order in which the change points are estimated and lastly p-value is estimated from each permutation test. The calculated p-value for each change point is noted. Here results indicate that change points are significant.

The calculated P- Value for each change point is given significance levelat $\alpha = 0.05$. It means that these change points are significant. These change points are detected and located for changes in mean and variance corresponding to the $\alpha = 1$. On the basis of these results , the null hypothesis is rejected with indication that there is heterogeneity[6] in our series and we accept the alternative hypothesis. It means that there are multiple significant change points in our series. All these values are compared by using P-value. Estimated change point with their locations in individual series under this method are shown in Figure 5.1, 5.2, 5.3.

Figure 5.1: Detected change point in Electricity

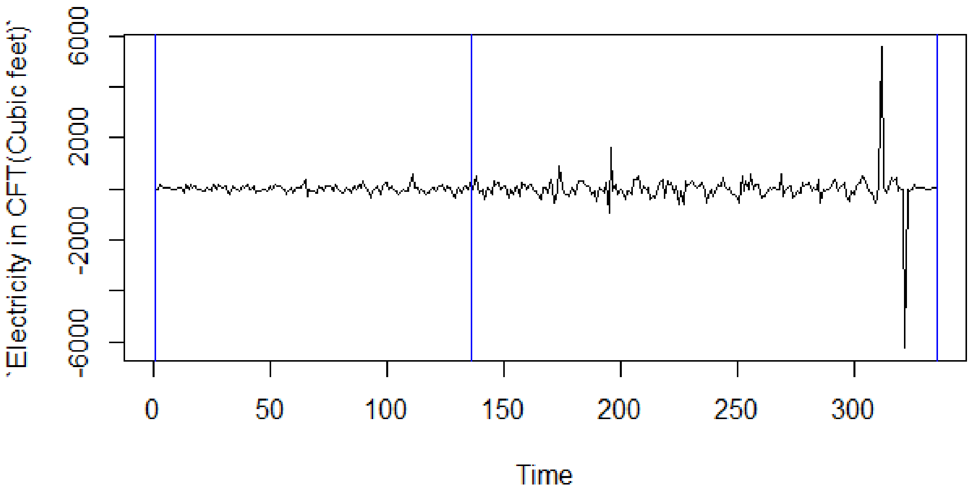


Figure 5.1: *Figure indicates the change points in the series of electricity. On y-axis thereis electricity and on x-axis there is time. Series is stationary*

Figure 5.1 represents the detection of change point in electricity series. Along Y-axis we have electricity production in cubic feet. On X-axis we have taken which time is counted in years. Total time span in our study is 1990- 2016.Total sample comprised of 27 years. Figure 5.1 shows that at time t=136 (March 2001), we have a single change point in our series. According to graph before break the variance of the data looks small

[6] If the variance is not constant (i.e. dependent on X's), then the linear regression model has heteroscedastic errors and likely to give incorrect estimates.

and after the break the variance is more. According to figure in the 2014 there are two outliers one is positive and one is negative. This change point is significant having P-value < (0.05). There are several possible reasons for the change point to occur at this point. The most common is, as given in the literature and economic survey of the country was under the electricity crisis along with others. Such as Electricity shortage/crisis, demand and supply gap, delay in adopting energy mix policy, inefficiency of managing authorities. Due to the increase in price of crude oil, decline in oil consumption between 2001-2006 was observed. This led to serious electricity shortages. In this time period, year of Pervez Musharraf electricity was also generated through oil. Oil was imported from other countries. After electricity was transferred to common. As a result, the price of electricity become higher. Demand of electricity increases while supply decreases due to higher prices.

Figure 5.2: Detected change point series in natural gas

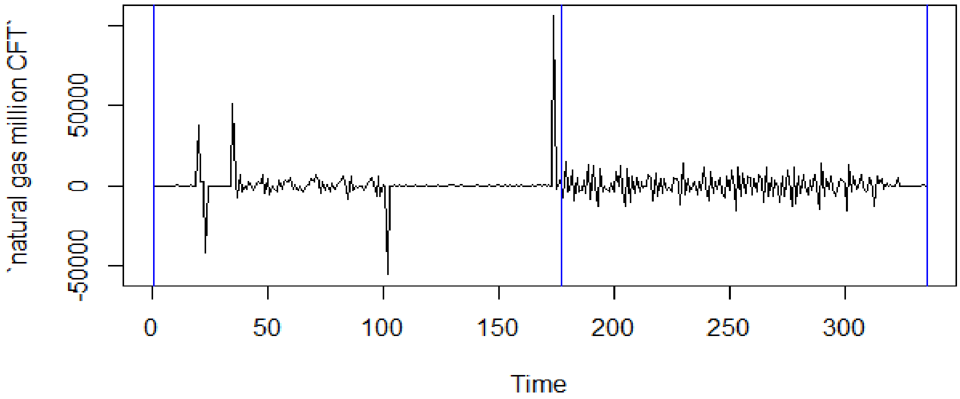


Figure 5.2 : *Figure indicates the change points in the series of natural gas. On y-axis thereis natural gas and on x-axis there is time. Series is stationary at I(1).*

The Figure 5.2 above represents the detected change points in the natural gas. The variable natural gas is in the first difference form thus the variable is stationary. The visuals from the graph indicates that there is only one change point in the series over the time. The change point occurs at August 2004. The series indicates the stationary process beyond and before the August 2004. There may be several reasons to incorporate this change point, but, the most important as determined by the economic history of Pakistan is the demand and supply gap, import of expansive fuel, energy consumption increases faster the economic growth process in Pakistan during 2004-05. Therefore Pakistan real GDP as high as 8.4% in this period. Also oil prices like during 2007-08 reduces. Pakistan export there by rising current account deficit at 8.4% of GDP as compare to 1.8% of GDP in 2003-04. Thus, based on these arguments the change in the behavior of the series over the time process can be mainly attributed to the shocks in the electricity due to the demand and supply mis-match.

Figure 5.3: Detected change point in crude oil

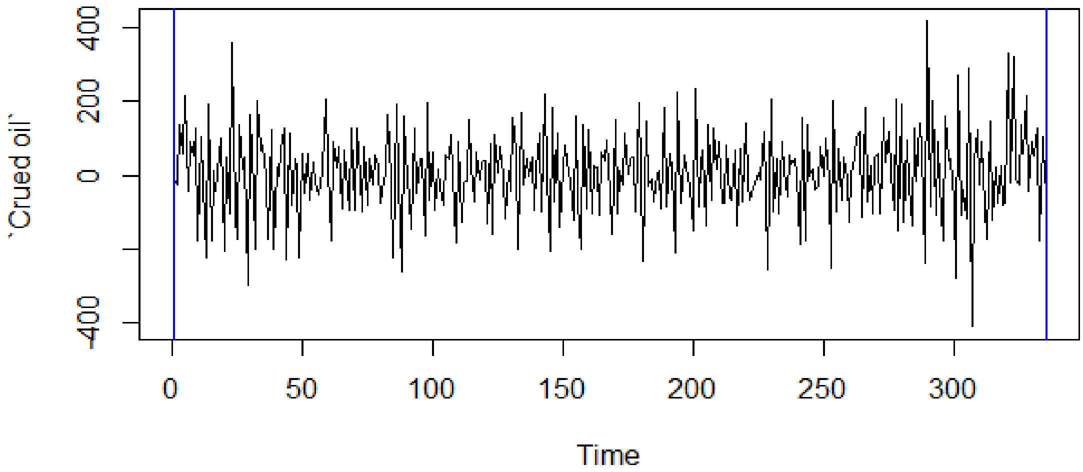


Figure 5.3: *The estimated change point locations for the E-Divisive method are indicated by the vertical lines*.

The fig 5.3 shown above indicates that there is no significant change point in the crude oil. The reason is that the supply and demand of the crude oil remains more or less table. It is the prices that fluctuate considerably instead the crude oil.

Financial Data series and the prevalence of change points.

Similarly the E-Divisive method to the time series data is applied on other different data series to identify the multiple change points. These series are related to banking sector like Cash in Pakistan, Borrowing from State Bank of Pakistan (SBP) , and balances with State Bank of Pakistan (SBP) in(millions) . The data comprises of 324 monthly observations on each series spanning 1990-2016.

Table 5.4 Correlation matrix of Financial series

	Cash in Pakistan	**Balances with SBP**	**Borrowing from SBP**
Cash in Pakistan	1		
Balances with SBP	0.93	1	
Borrowing from SBP	0.8971	0.89124	1

The table 5.4 indicates the correlation among the series. It indicates that the series are highly correlated.

The correlation between cash in Pakistan Balances with SBP is 0.93.The correlation between Cash in Pakistan and Borrowing from SBP is 0.8971.The correlation between Balances with SBP and borrowing from SBP is 0.89124. Thus, it can be regarded as the sufficient evidence for the application of the E-Divisive method.

The results and possible grouping is given in the table below. The groups or clusters are based on the presence of the significance of the change points. Thus, for each change point there is a single cluster. It indicates that there is a cluster for which there appeared a single change point in the data series. There are five clusters and of which there are 5 significance change points in the multivariate analysis.

Table 5.5 Clustering Based on E-Divisive

Groups	Size of observations
1	145
2	74
3	34
4	30
5	41

There are five groups based on the change points, each group assumes a single change point.

There are five clusters or groups in the multivariate analysis. The first cluster corresponds to the first group having 145 observations or data points. The second groups consists of 74 and there are 34, 30 and 41 observations in 3rd , 4th and 5th cluster or change point respectively. Thus, the cluster can also be regarded as indicating the frequency of each cluster. Thus it indicates that the 1 has appeared 145 times and 2 appeared 74 times and 3,4,5 appeared 34,30 and 41 times respectively. The detected change points based on thisinitial observed clusters are given below.

Table 5.6: Report about the location of multiple change points in time series data.

Order form	Change point location						
consider. Last	124						
order. Found	1	324	260	186	209	294	274
Estimate	186	209	260	274	294	124	
p-value	0.005*	0.005*	0.050*	0.050*	0.050*	0.335	

The table gives the information about number and location of the change points

The table 5.6 gives the information of the location and significance of the detected change points. It represents that there are six change points. The order of change points is such that it appeared first change point on the observation number 186 which is significant because its probability value is far below to the conventional level of probability required to reject the null hypothesis.

Thus, the second change point lies at 209 which is also significant based on the p-value criterion. Similarly, the other change points are also significant and their location with their order of occurrence is given in the table. Only, the last change point is insignificant, but this cannot be disregarded, as for the methodology of process is concerned it may continue to look for the change points until the change becomes insignificant. Thus, the information is equally valuable. The data series on financial sectors are thus subjected to various change points as detected by the e-divisive. One possibility is that they are correlated among each as the correlation between cash balance and borrowing from the State Bank is 0.93 , 0.8971 between borrowing from SBP and cash in Pakistan , 0.89124 between borrowing from SBP and balances with SBP.

Further, the analysis is concluded in the graphs which are prepared based on the detection of change points, their location and order their appearance. The information contained in the graphs is given below.

Figure 5.4: Change point detection series in cash in Pakistan

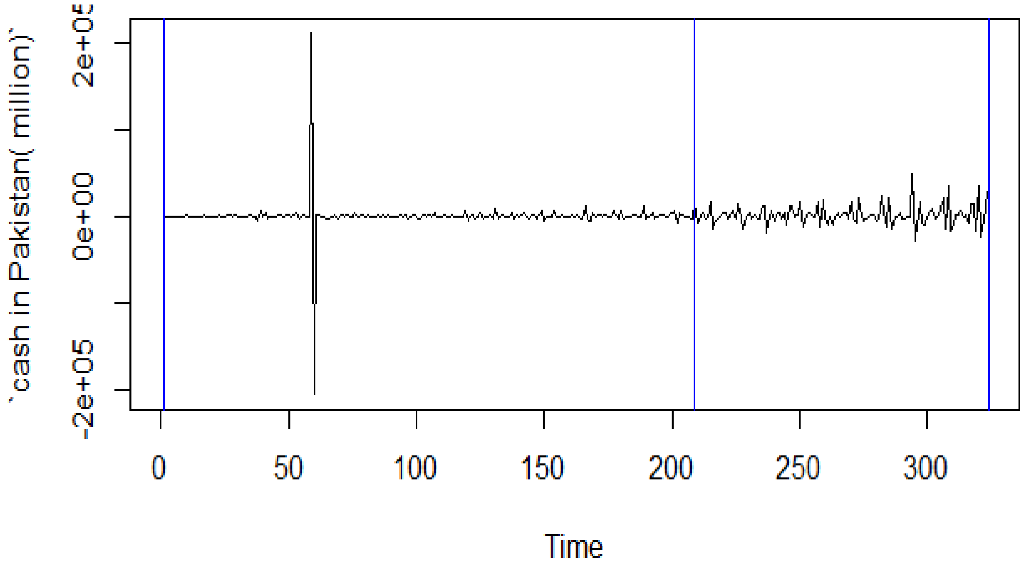


Figure 5.4 indicates the change points in the series of Cash in Pakistan. On y-axis there is Cash in Pakistan (million) and on x-axis there is time. Series is stationary.

The figure 5.4 is showing that detected change points in the time series data. This series is also stationary at first difference that means autocorrelation is constant. Each point is calculated on the basis of p-value($\alpha = 0.05$). The graph is showing that, there is a significance change point in series of cash in Pakistan that is on 209 (p-value < 0.05). Estimated change point locations for individual three series under this method are shown in Figure 5.4. Financial crisis: Pakistan's stock market showed considerable immunity to the global turbulence and boosted the index to peak highs in the mentioned year. There were many factors which contributed to the booming condition. They are improvement in the country's economic fundamentals, stability in exchange rate, reduction in interest rates by banks, recovery of outstanding/overdue loans, rescheduling of foreign debts and prepayment of the expensive foreign loans(ministry of finance).

Figure 5.5: Change point detection in series of Balances with SBP

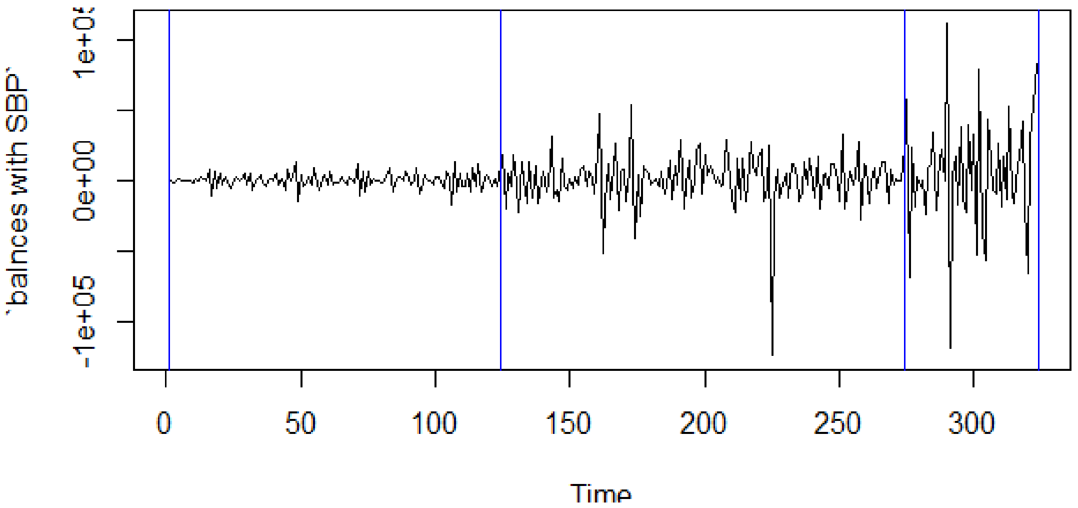


Figure 5.5: *The estimated change point locations for the E-Divisive method are indicated by the vertical lines.*

The figure above represents that, there is one change point that is 274. On this change point p-value is 0.005 . These values are less than level of significance. The fluctuation in the mentioned year was due to heavy repayments to the IMF, net outflows to other IFIs, and anemic foreign investments. Forex reserves were also exhausted (Economic Survey of Pakistan).

The series was first differenced stationary. Thus, it represents the pure behavior of the series over the time and subjected to various change points. The series do not show any such trend or irregular movement that could be regarded an indication that it would add to the presence or occurrence of the change points. This is the essence of making the series stationary. Thus, the change points detected by the methodology in practice indicates that these are true behavior of the series over time. Thus, the reasons for these changes or simply the breaks in the pattern of the series are more economic than any other change.

Figure 5.6: Change point detection series in Borrowing from SBP

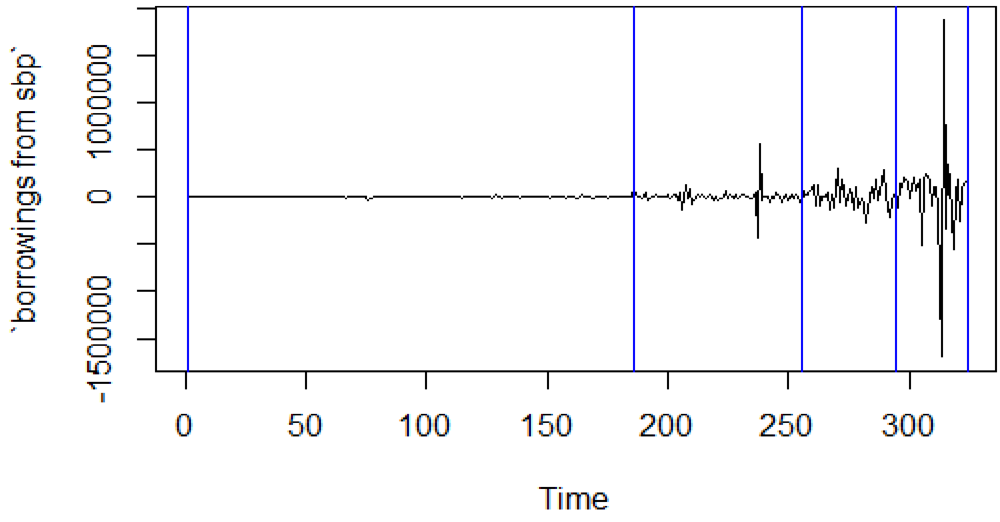


Figure 5.6 *The vertical lines describe the change points detected in the borrowing from the state bank of Pakistan. There two significant change points in the graph above denoted by the vertical bars.*

There are three change points one is 186, second is 260 and third is 294. The probability value of these change points is below the 5% level of significance which indicates that these are significant points on which there occurs change over the time span of the data series. The possible reason for these change points is attributed to the economic and financial issues. The series are highly correlated to each other which causes the fluctuation in one series as a consequence of the change in the other series. Therefore, this intra series correlation causes the changes in the behavior of the series in the time span over which the data points are subjected to the E-Divisive method. The data was subjected to first difference to make it stationary. The analysis thus corresponds to the stable and stationary data. The financial reports and economic survey of Pakistan indicates that Higher bank borrowing occurred(fromFY2010 and onwards) due to a pernicious combination of rising fiscal spending, and lower availability of external financing[7] as a result these changes have taken place.

Table 5.7 correlation between Real interest rate, Exchange rate and M2.

	Real interest rate	Exchange rate	M2
Real interest rate	1		
Exchange rate	0.12621	1	
M2	0.17855	0.93872	1

The table indicates the correlation among the real interest rate, exchange rate and M2 money supply. There are 1 on the main diagonal entries and off diagonals are different from 1 thus, indicating the correlation between the data series.

There is a high correlation between money supply and exchange rate with the value of multiple coefficient of correlation 0.9387. The correlation between the exchange rate and real interest rate is not much high with the correlation value of 0.126. The correlation coefficient between the money supply (M2) and the real interest rate is about 0.1785. thus indicating the weak association between these variables. The expected change points and the clusters are given below which indicates that based on this correlation among the series there are two clusters around which the data is centered. The first group consists of 163 data points and second consists of 137 .

The clustering indicates that there are two possible change points in the data series. The change points have been detected by the E-Divisive method. The results and possible significance level along with the ordering of the change points have been given in the table 5.8

[7] Within the banking sector, borrowings from the commercial banks fell October onwards. I happened because the government largely adhered to its pre-auction targets that were set lower for Q2-FY10 in anticipation of revenue receipts such as coalition support funds. Hence, reliance on borrowings (from State Bank of Pakistan) increased during October-April 2010. The government increased resources and SBP borrowing was made. However, in the absence of sufficient commercial bank borrowings, government borrowings from the central bank had exceeded its quarterly limits by the end of third week of December 2009.

Table 5.8 Clusters/Groups in the data series.

Groups	No of observations
1	163
2	137

The table 5.8 has shown that there are two groups or clusters in the data series. The series have shown low correlation among them as a result this can be attributed to the reason why they have low groups and that they have less change points in them.

The possible groups and the number of observations in each group has given in the above table but their location and probability values are given in the table 5.9.

Table 5.9: Reports the location of multiple change points of time series data

Order form	Change point location		
consider. Last	141		
order. Found	1	300	204
Estimate	204	141	
p-value	0.005*	0.085	

Tables gives information about location of the change points in series

Table 5.9 reports the location of multiple change points of the series. Here considered last is the insignificant change point that appears at 141thobservation at the given level of significance. Here order found two range for change point detection, one between 1-300 and second is between 300-204. Likewise for the estimates but here the range is 1-204 and 204-300. The p-value at 0.005 level of significance and 0.085 is insignificant.

Figure 5.7: Change point detection in Real interest rate

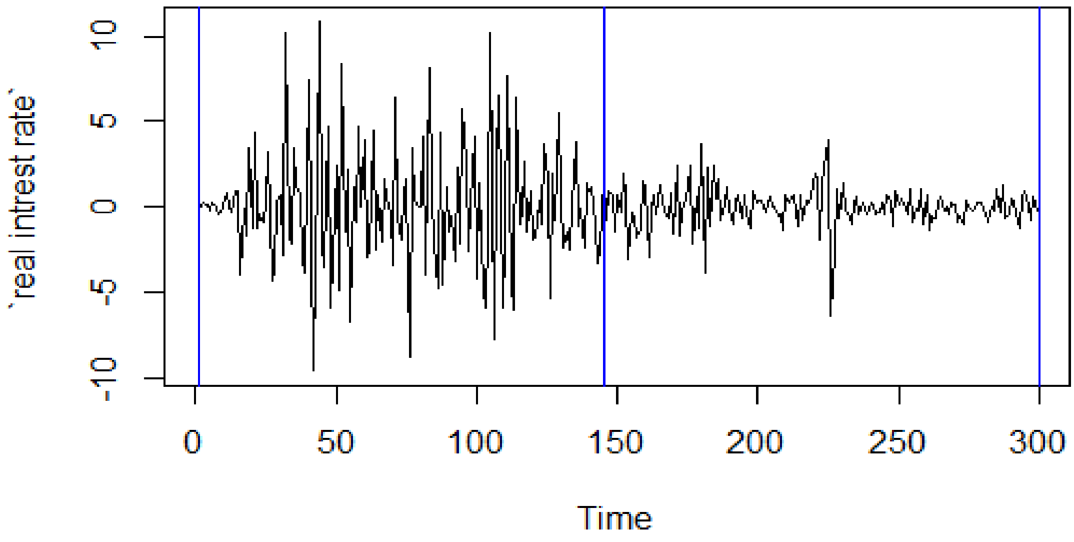


Figure 5.7:Change point detection in real interest rate

The fig has reported that there is one change points at 145. The probability value of the change points is below the 5% level of significance which indicates the significant point on which the changes occur over the time span of the data. The possible reason for this change point is attributed to the economic and financial issues. The series are highly correlated to each other which causes the fluctuation in one series as a consequence of the change in the other series. Therefore, this intra series correlation causes the changes in the behavior of the series in the time span over which the data points are subjected to the E-Divisive method. The data was subjected to first difference to make it stationary. The analysis thus corresponds to the stable and stationary data. This change point is due to inflationary pressures and liquidity constraints. These three series are related to banking sector. E-Divisive method is applied in case of multivariate series and it has detected the one change point in case of individual series, but it has detected many change points in collectives series.Therefore in this series, the change point is 145.

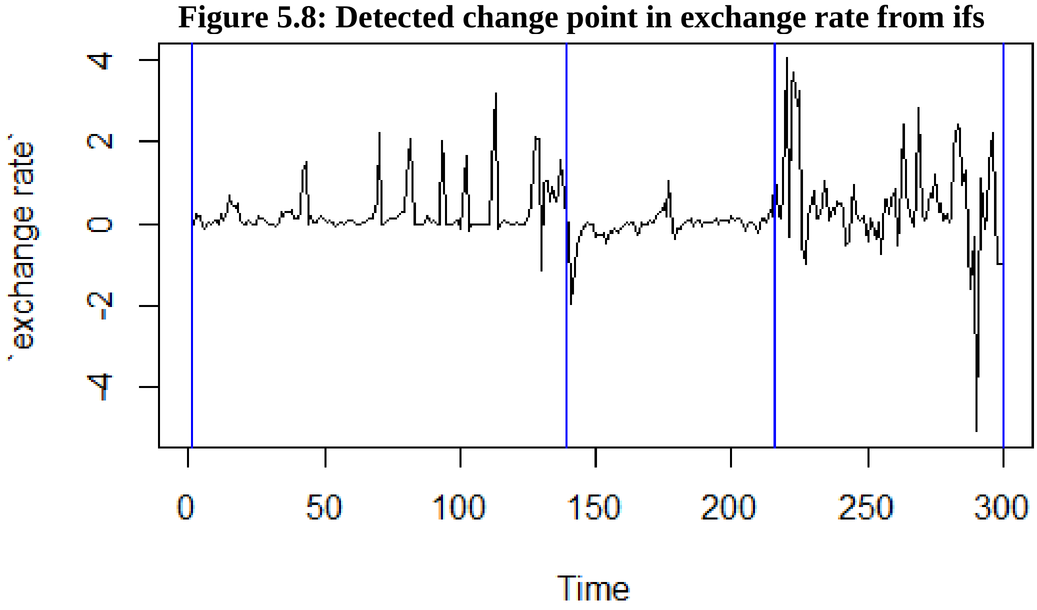


Figure 5.8:Different location of change points in exchange rate.

Here figure 4.8 reports the two change points, one is t=139(June 2001), second is t= 216 (November 2007). On these two change points p-value is 0.005 and 0.0150. In 1998 Pakistan conducted nuclear test which resulted the main indicators of macroeconomics highly volatile due to external pressure. State Bank of Pakistan reported in 2011 that Current account also turned to deficit due to remarkable decrease in exports, and exchange rate was also volatile.

Figure 5.9: Change point detection in M2

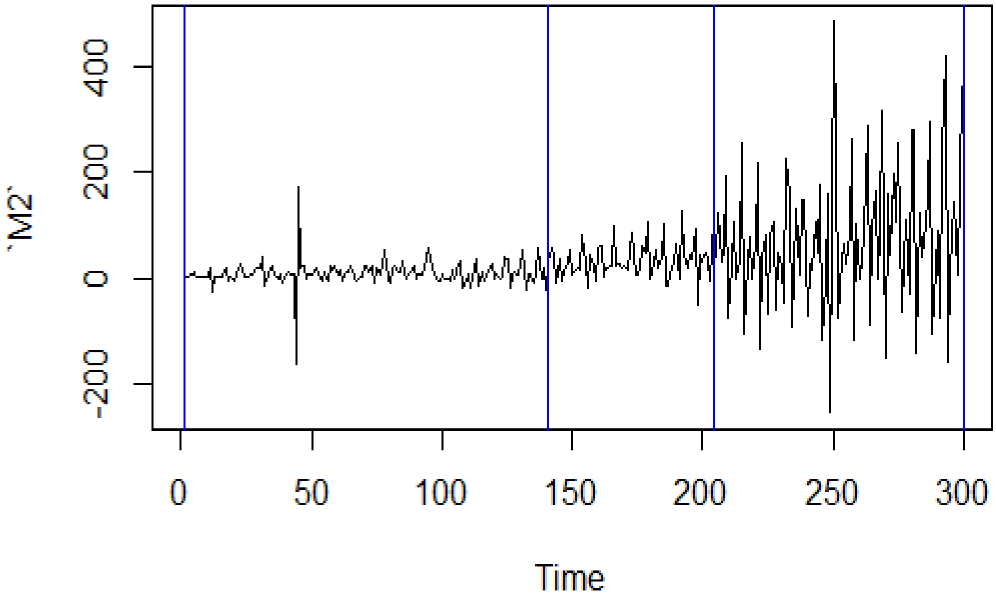


Figure 5.9 reports the estimated location of change points in M2 series.

Above figure 5.9 shows that there is one significant change point at t=204(December 2006). On this change point the p-value is 0.005. In 2006 all macroeconomic indicators were flourishing. Targeted economic growth was achieved which resulted in the growth of M2. (see table in appendix). Although there is another change point in the series as detected by the vertical bars in the above figure but, its probability value is greater than the 5% critical value that makes it insignificant. As a result this cannot be regarded as the significant change point. The series further adds to the analysis that over the time period of the change point there was a major macroeconomic activity in the economy that ultimately increased the supply of money in the economy. Therefore, this era can be regarded as the major activity era as compared to the previous government. The GDP growth recorded at its highest in this era.

CHAPTER 6

CONCLUSION

We have presented a method to perform multiple change point analysis of an independent sequence of multivariate observations. The proposed methods are able to estimate both the number of change points and their locations. The divisive version hierarchically tests the statistical significance of each hierarchically estimated change point. Because we have established consistency for the divisive procedure we prefer it in practice, even though its computation is dependent on the number of change points that are estimated. We prefer to use the E-Divisive method, even though its running time is output-sensitive and depends on the number of estimated change points. Through applications to real data and simulations, we observe that the E-Divisive approach obtains reasonable estimates for the locations of change points. We have used two sectors Energy and Banking sector with three different cases. We have used three series of energy sector in case one three series of banking sector in case two and three more series from banking sector as third case. When we apply the e-divisive algorithm on these three cases, the method has detected the number of change point and location of change points correctly. In banking sector (cash in Pakistan, balances with SBP and borrowing from SBP), e-divisive method is more appropriate for location of change points and number of change points. But in energy sector, e-divisive method is able to number and location of change points for two series: electricity prices and natural gas prices but not in third series i.e. crude oil where this method has not detected the number of change points and location of change points correctly. We have checked the performance and power of e-divisive algorithm in multivariate analysis through simulation. We have divided the simulation into two parts, the first one has simulation based on detecting the change points at same location for each series and second simulation experimentally based on change point detection when changes are at different locations in each series. In first simulation, when we have kept different k values 0,1,2,3,4,5. In first scenario, the best performance of e-divisive method to correct change point detection is in case of multivariate analysis related to individual analysis. We increase the k values, the power of calculation in case of multivariate is low but the power calculation of individual analysis is lower. We simulate 5000 times and keep the value of k=1, the e-divisive method detect the 97 time in multivariate analysis. Whenever other values of k is small, detection is better than individual analysis. In second simulation, we calculate the power of change point detection at different locations. The results of simulations indicate that the E-Divisive method performs well in both scenario. It detects all the change points with high power in case of multivariate structure. Finally we conclude that e-divisive method is best performing in multivariate structure because its power is high rather than individual case. We have established consistency for the divisive procedure, we prefer it though its computation is dependent on the number and location of change points that are estimated.

RECOMMENDATIONS

Economy may face economic issues, ups and downs which can be tackled by adopting appropriate policies. To avoid future economic problems the policy makers should have former information of structural breaks while selecting the series for further analysis. If the policy makers with prior knowledge take the corrected series (banking sector and energy sector), it will enable them to do careful analysis, so that the next policy will be accurately formulated on the basis of data.

Appendix

Summary Statistics

V. name	Mean	Medium	Std. Dev	Skewness	Kurtosis
Cash In Pakistan	677.6279	24.10000	17922.33	0.273363	114.6567
Balances With SBP	1657.729	613.6000	20326.45	- 0.585779	14.68747
Borrowings from SBP	5798.105	912.0000	166490.1	0.931653	84.80922
Crude Oil	4.517028	8.000000	120.5626	0.020654	3.421026
Electricity	5.117907	0.879300	522.3166	- -1.530397	103.9226
Natural Gas	389.4211	15.00000	9295.919	4.228729	60.92747
Exchange Rate	0.265384	0.074564	0.828872	0.618210	12.82683
M2	33.75545	16.28100	83.06402	1.684050	9.418449
Real Interest Rate	0.007940	0.000780	2.686787	0.454684	6.250838

V. name	Maximum	Minimum	Q1	Q2	Q3
Cash In Pakistan	210522.6	-206637.6	-1502.5	24.1	1769
Balances With SBP	110998	-123484	-4870.8	613.6	839.3
Borrowings from SBP	1871127	-1697702	-2295	912	9762
Crude Oil	417	-410	-77	8	78
Electricity	5549.92	-6245.744	-99.94	0.889	119.58
Natural Gas	105781	-55792	-2000	15	2311
Exchange Rate	4.045	-5.08	-3.80	0.074	0.31
M2	485.5	-256.3	1.621	16.281	45.34
Real Interest Rate	10.8387	-9.6211	-1.068	0.00078	0.896

Variable and reasons of detected change points

Sr. no.	Macro economic variable	Time	Event	Source
1	Electricity	March(2001)	Electricity shortage/crisis, demand and supply gap, delay in adopting energy mix policy, inefficiency of managing authorities. Due to the increase in price of crude oil, decline in oil consumption between 2001-2006 was observed. This led to serious electricity shortage. In this time period, year of Pervez Musharraf electricity was made through oil. Oil was import from other countries. After electricity transference to people. The price become higher. Demand increases supply decreases due to price become higher.	Pakistan Economic Survey (2005-2006) Uqaili et al.(2017)
2	Natural gas	August(2004)	Demand and Supply gap, Import of expansive fuel, Energy consumption increases faster the economic growth process in Pakistan during 2004-05. Therefore Pakistan real GDP as high as 8.4% in this period. Also oil prices like during 2007-08 reduces. Pakistan export thereby rising current account deficit at 8.4% of GDP as compare to 1.8% of GDP in 2003-04.	Pakistan economic survey (2004)
3	Cash in Pakistan	April(2007)	Financial crisis: Pakistan's stock market showed considerable immunity to the global turbulence and boosted the index to peak highs in the mention year. There were many factors which contributed to the booming condition. They are improvement in the country's economic fundamentals, stability in exchange rate, reduction in interest rates by banks,	Ministry of finance

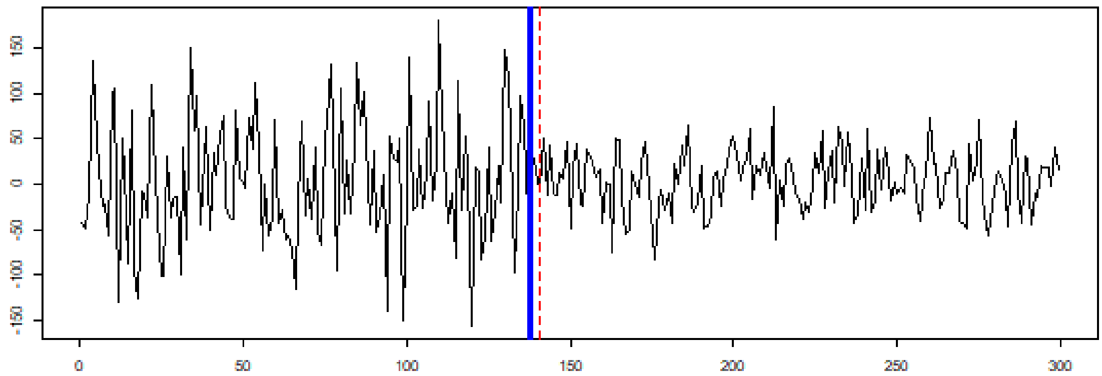
			recovery of outstanding/over due loans, rescheduling of foreign debts and prepayment of the expensive foreign loans.	
5	Borrowing from SBP	May (2005), July (2011)	Higher bank borrowing occurred(fromFY2010 and onwards) due to a pernicious combination of rising fiscal spending, and lower availability of external financing[8]	Pakistan economic survey(2009-2010)
6	Real interest rate	December (2001)	Inflationary pressures and liquidity constraints	Pakistan economic survey
7	Exchange rate	June (2001), Nov (2007)	Financial crisis: In 1998 Pakistan conducted nuclear test which resulted the main indicators of macroeconomics highly volatile due to external pressure. State Bank reported in 2011 that Current account also turned to deficit due to remarkable decrease in exports and exchange rate was volatile.	Hafeez(2015)
8	M2 (money supply)	Dec(2006)	In 2006 all macroeconomic indicators were flourishing. Targeted economic growth was achieved which resulted in the growth of M2.	SBP(2006)

[8] Within the banking sector, borrowings from the commercial banks fell October onwards. I happened because the government largely adhered to its pre-auction targets that were set lower for Q2-FY10 in anticipation of revenue receipts such as coalition support funds. Hence, reliance on borrowings (from State Bank of Pakistan) increased during October-April 2010. The government increased resources and SBP borrowing was made. However, in the absence of sufficient commercial bank borrowings, government borrowings from the central bank had exceeded its quarterly limits by the end of third week of December 2009.

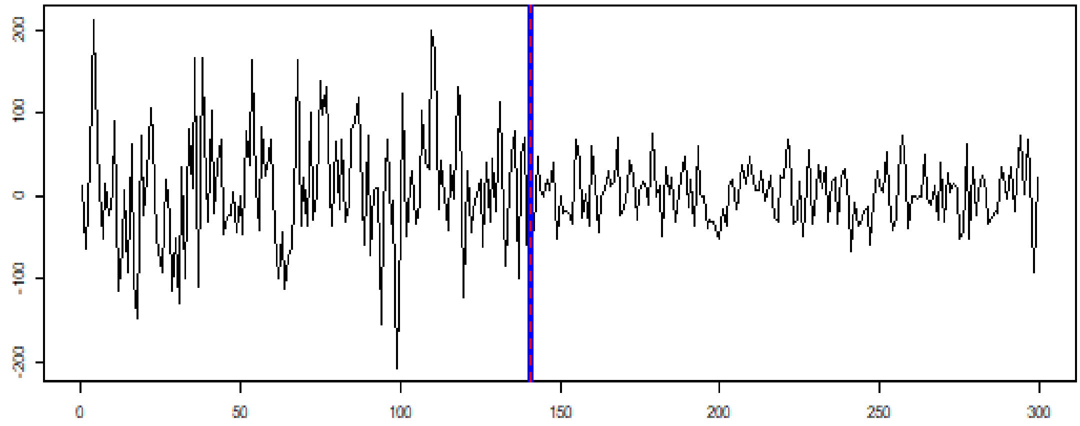
Performance of E-Divisive in case of multivariate change points in multivariate time series					
Introduce cps	**Detect cps In multivariate**		**Detected cps**	**Detected cps**	**Detected cps**
k	Correct detection	Power calculate	Series 1	Series 2	Series 3
0	95	0.95	92	96	96
1	100	1	90	90	91
2	90	0.9	65	64	68
3	94	0.94	47	54	57
4	94	0.94	46	52	55
5	95	0.95	41	44	37

Graphs in covariance structure at same and different location

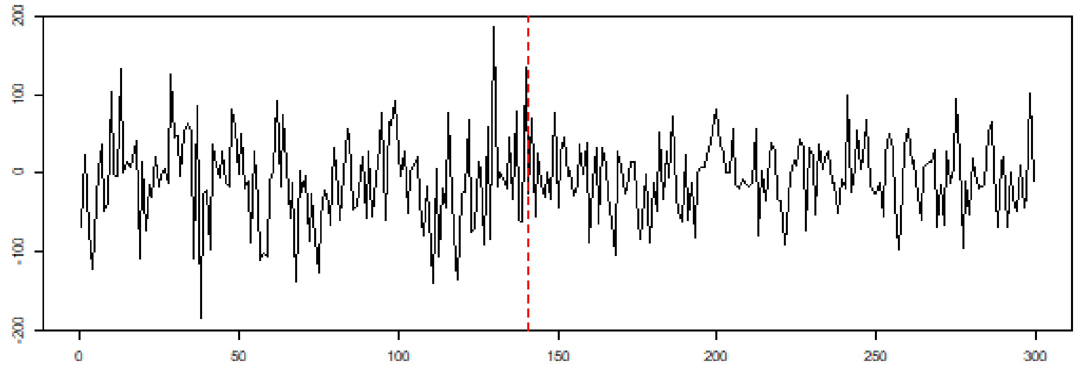
K=1



One change point in generated series one

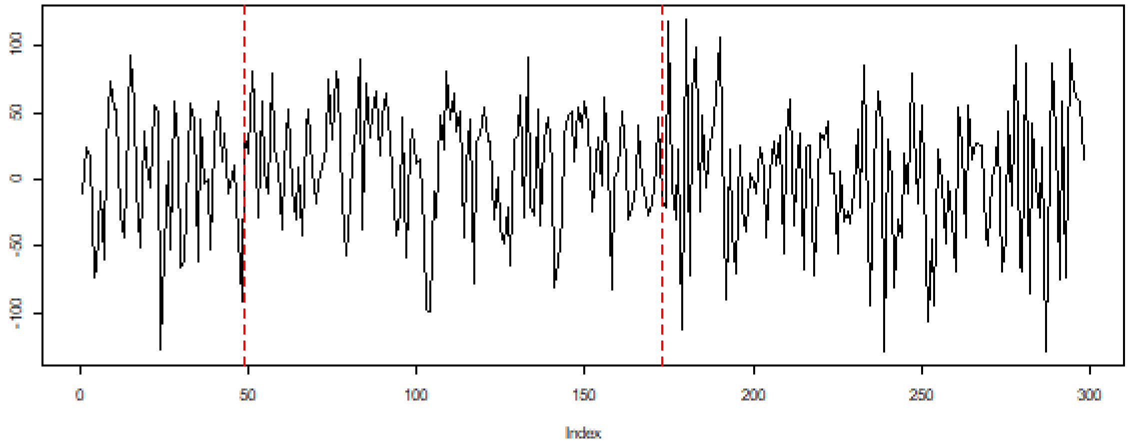


One change point in generated series two

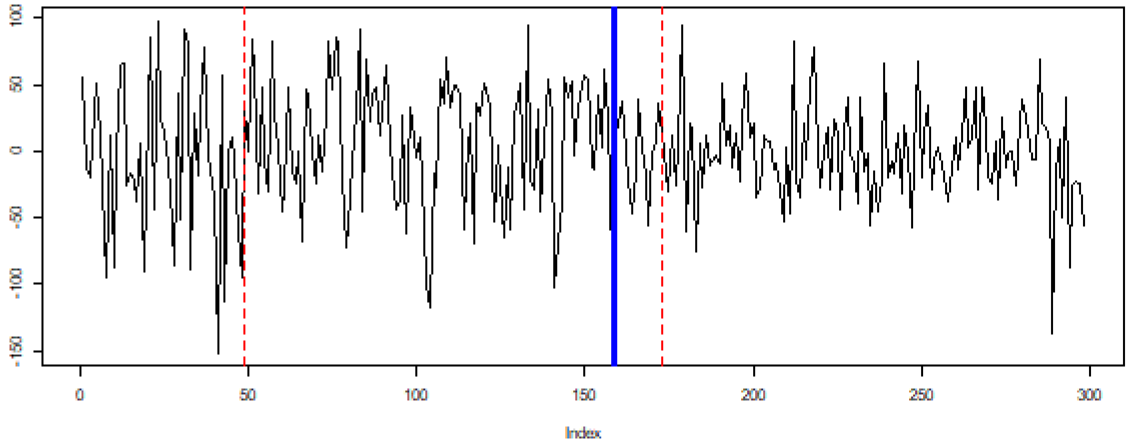


No change point in generated series three

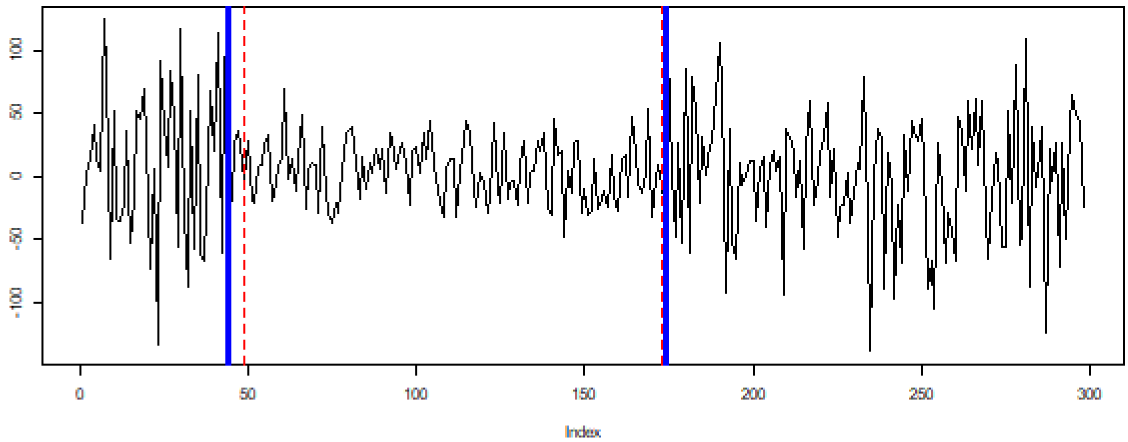
K=2



CPs:2 CPs Detected:0

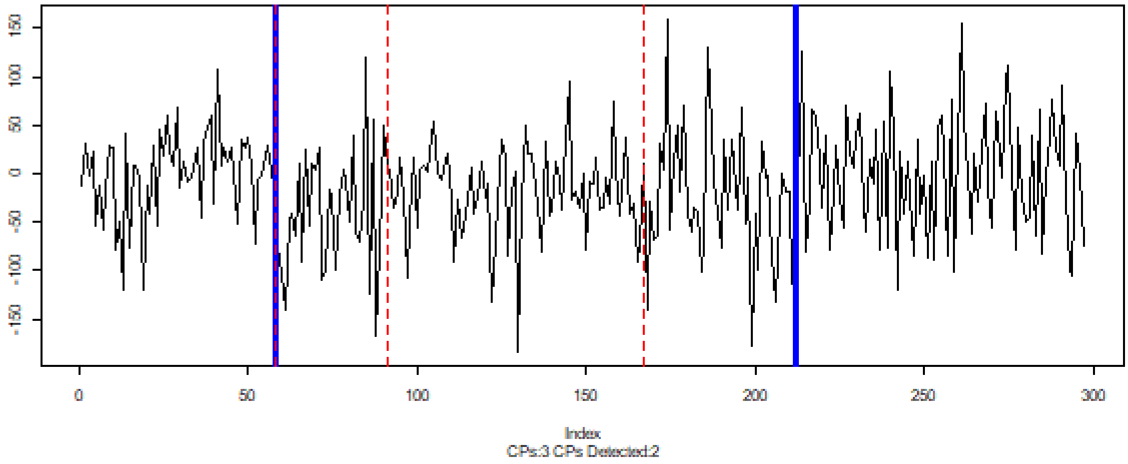


CPs:2 CPs Detected:1

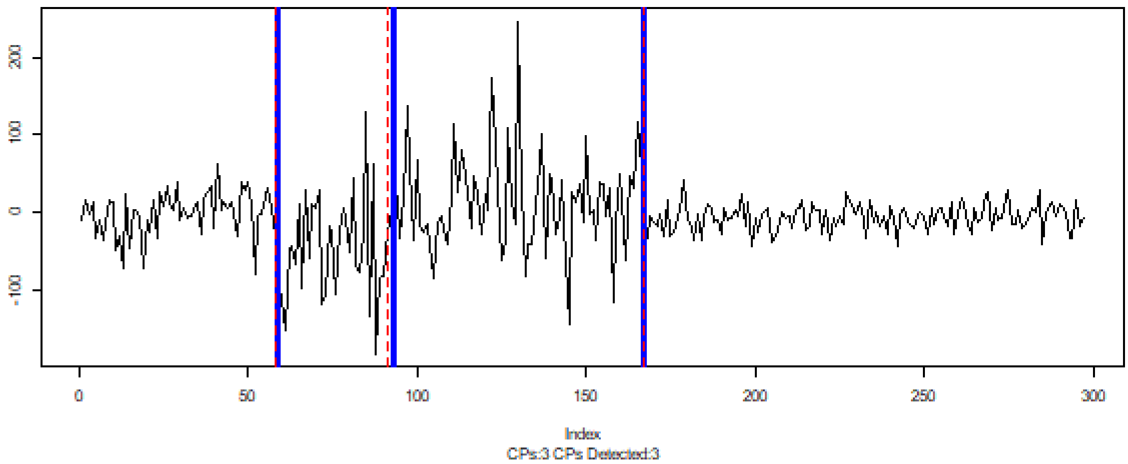


CPs:2 CPs Detected:2

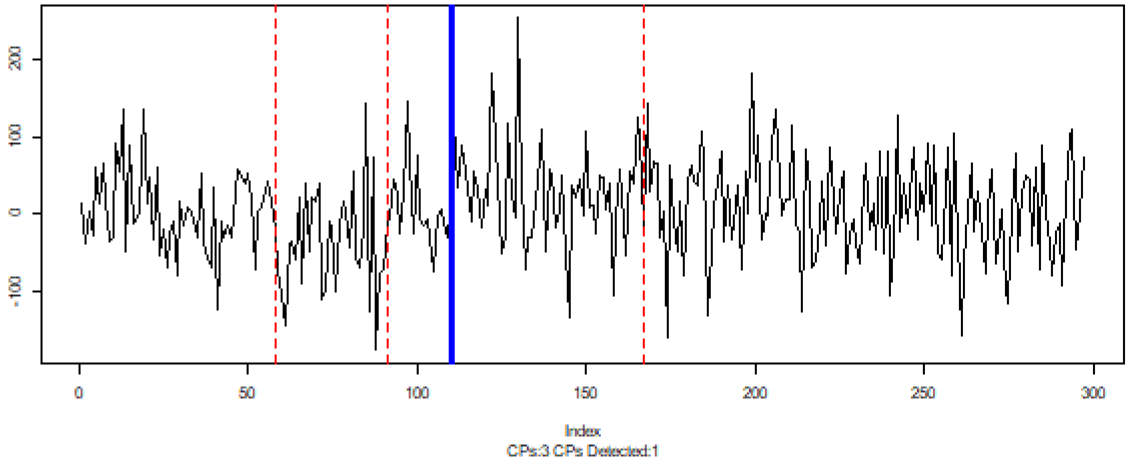
K=3



Index
CPs:3 CPs Detected:2

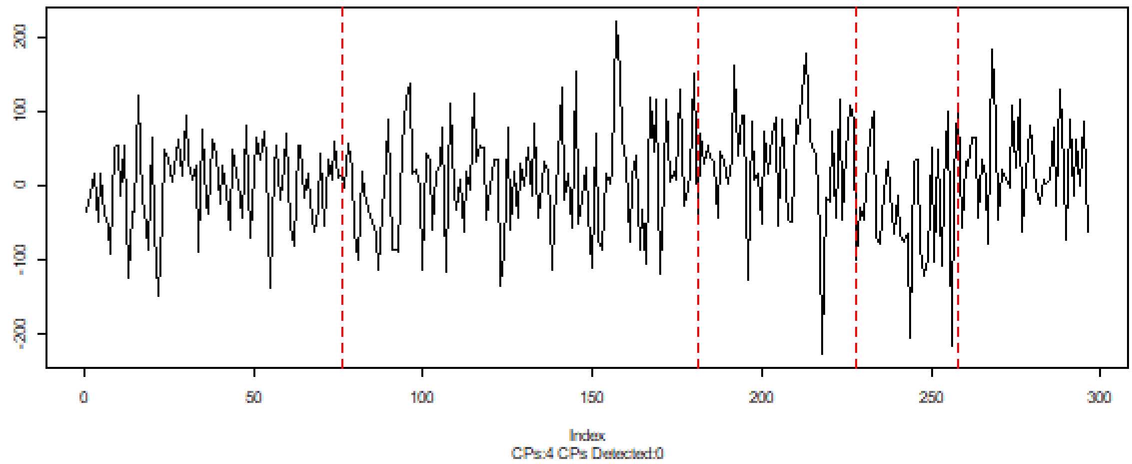


Index
CPs:3 CPs Detected:3

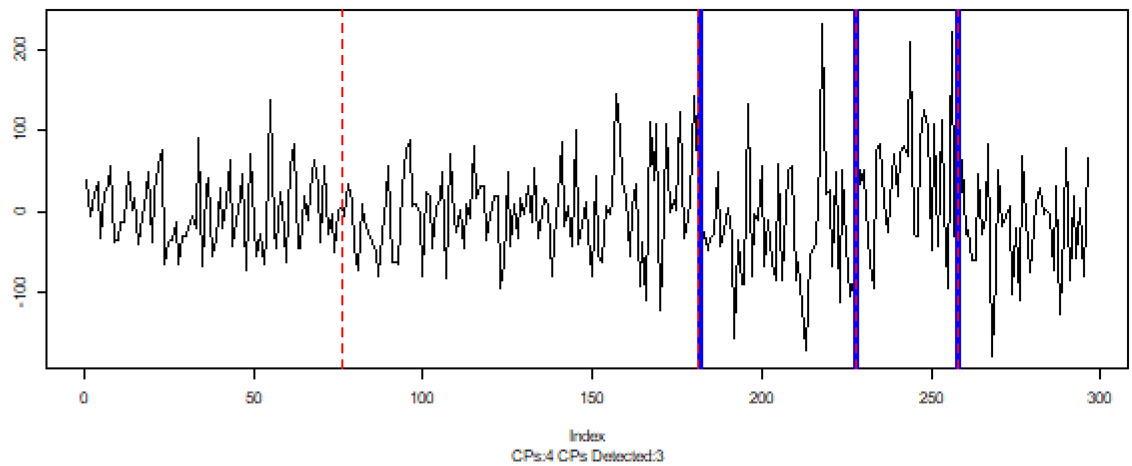


Index
CPs:3 CPs Detected:1

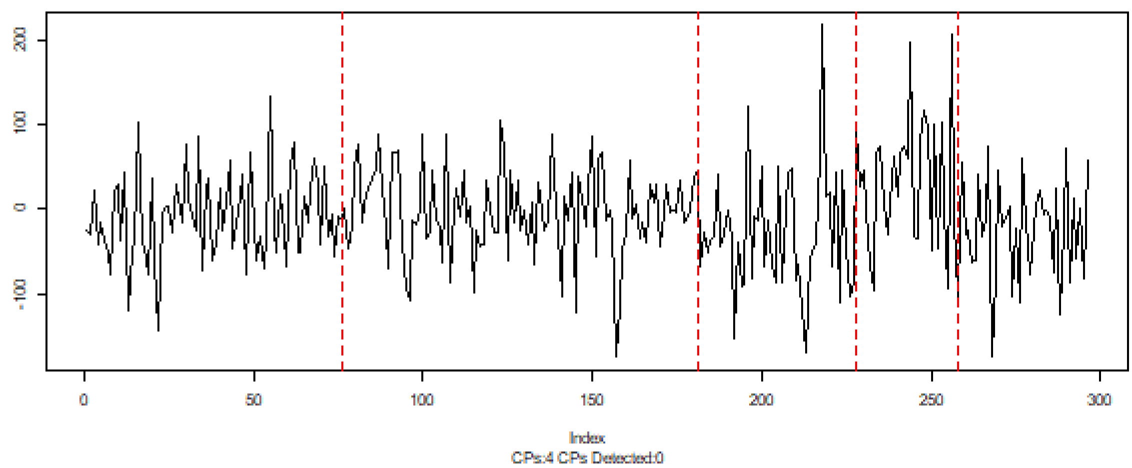
K=4



200
100
0
-100
-200
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:0

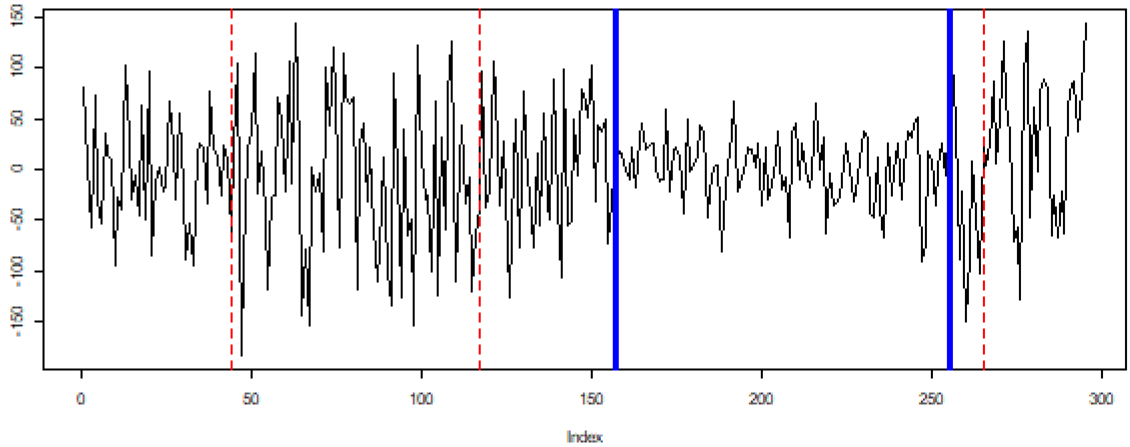


200
100
0
-100
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:3

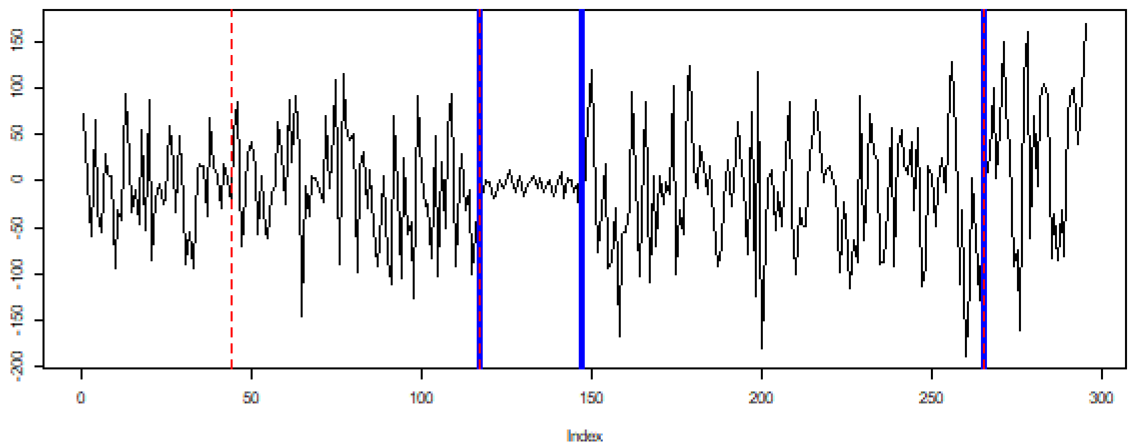


200
100
0
-100
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:0

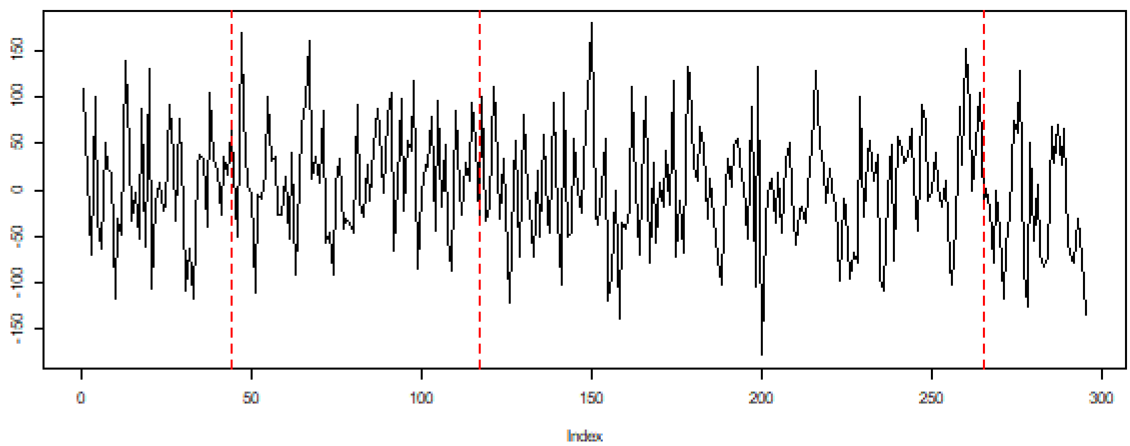
K=5



CPs:5 CPs Detected:2



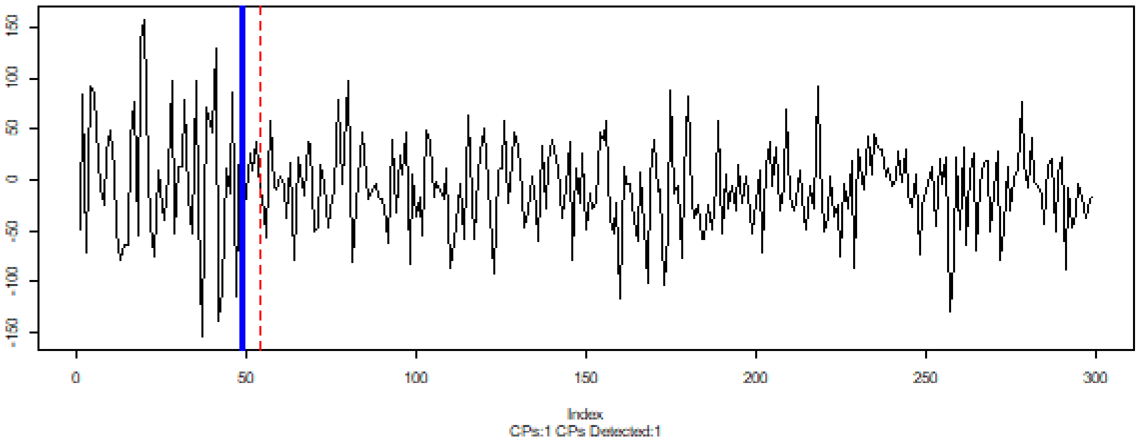
CPs:5 CPs Detected:3

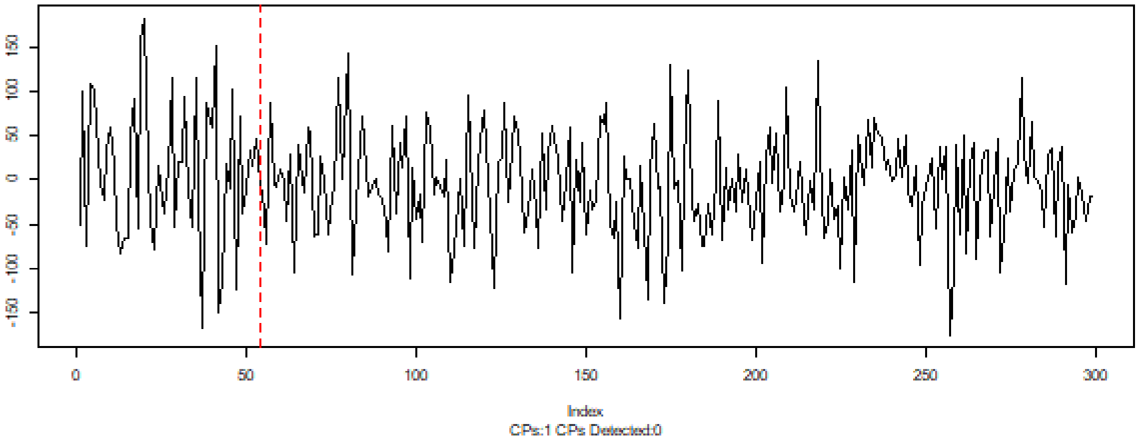


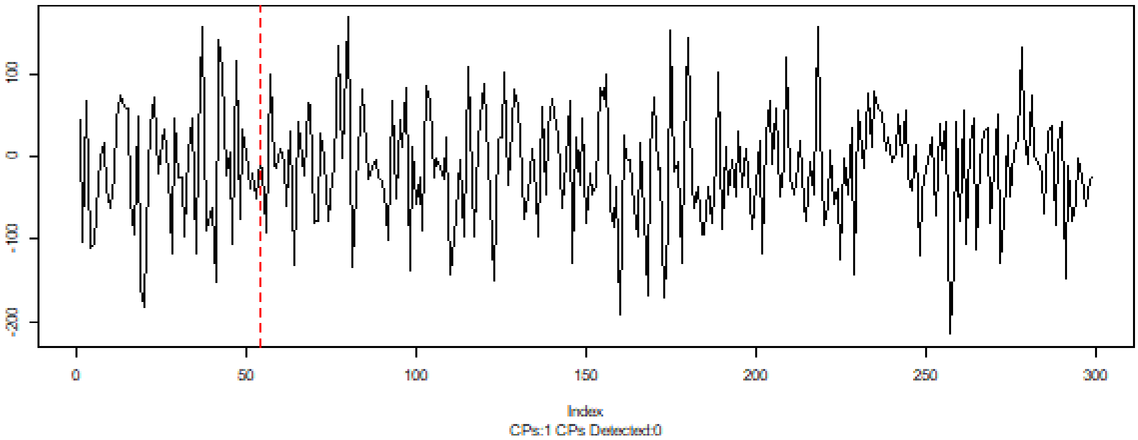
CPs:5 CPs Detected:0

At same location

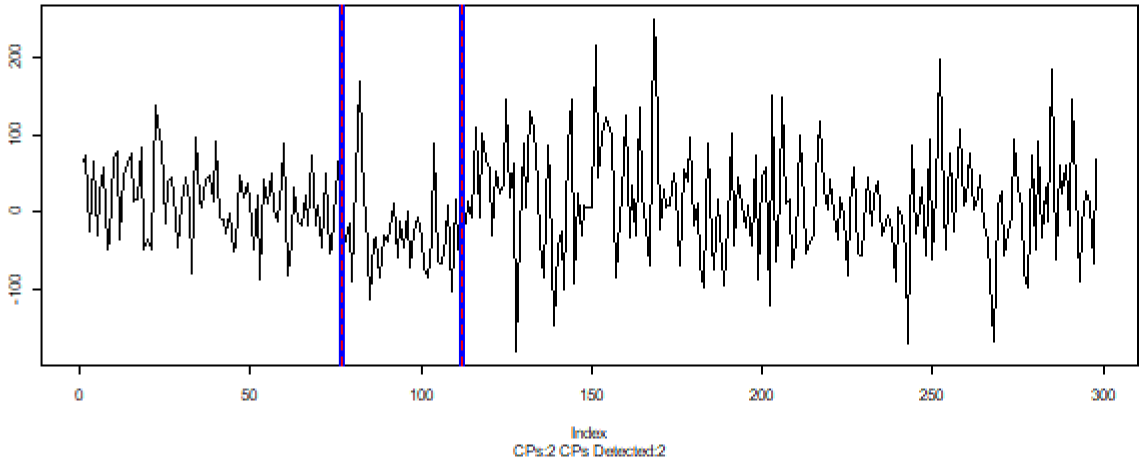
K=1

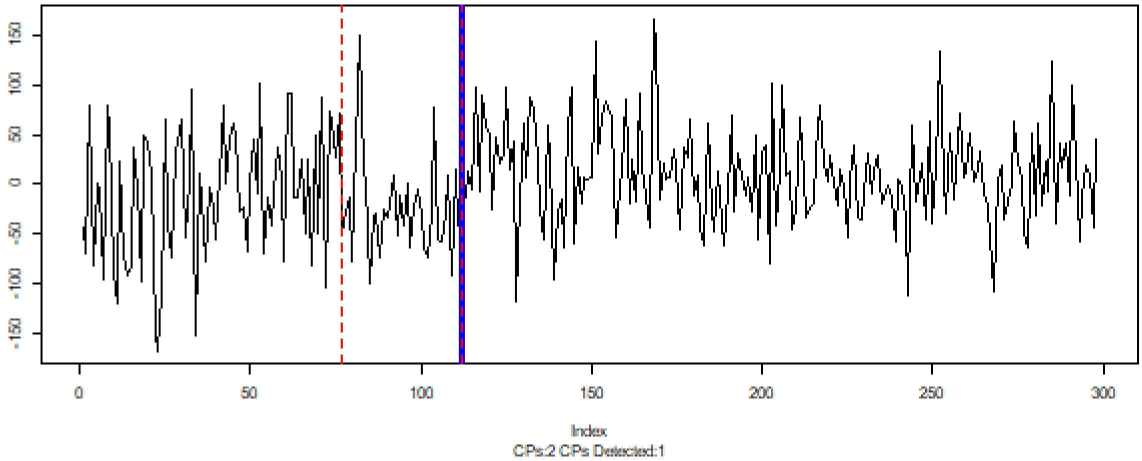


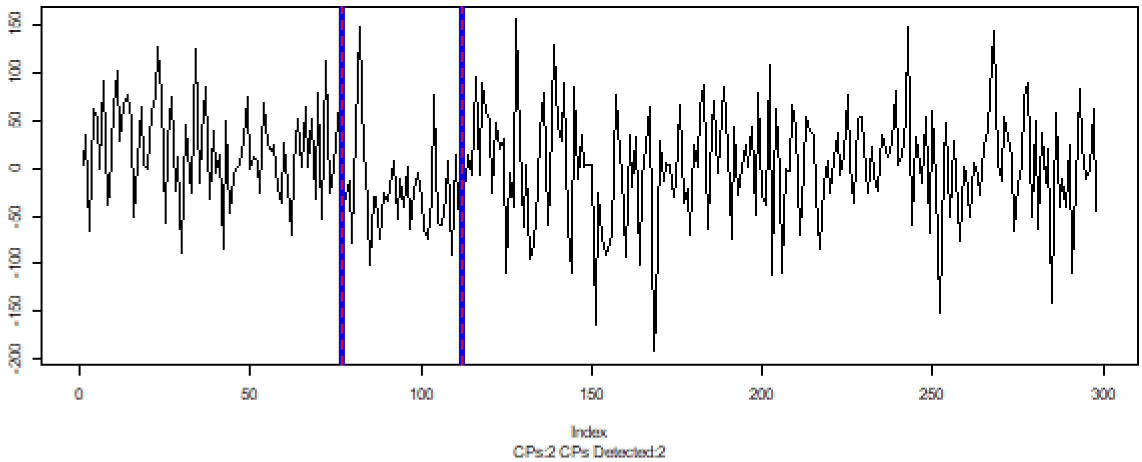




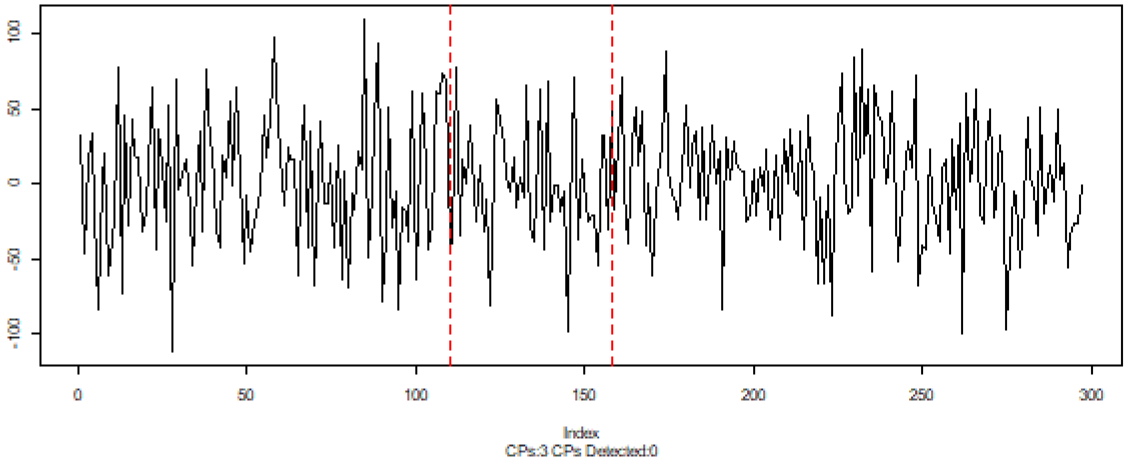
K=2

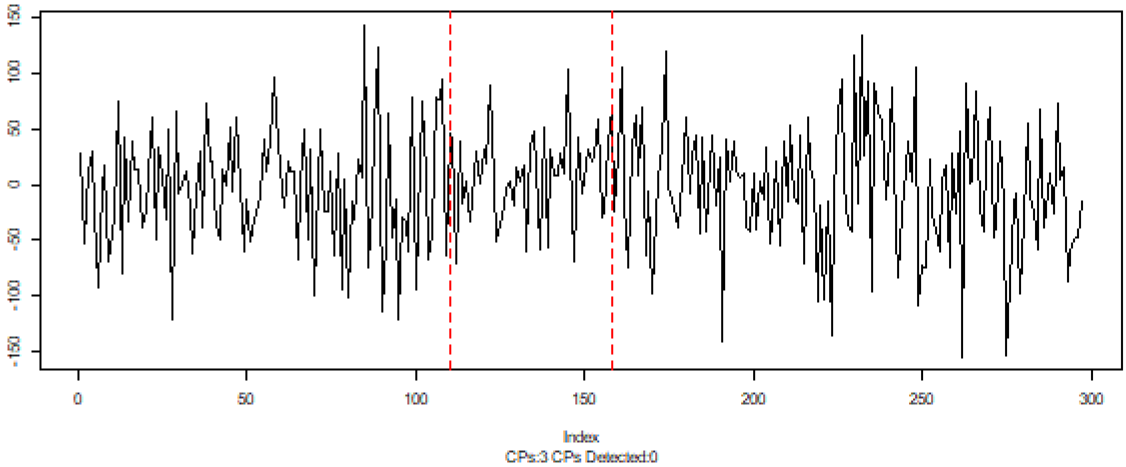

200
100
0
-100
0
50
100
150
200
250
300
Index
CPs:2 CPs Detected:2

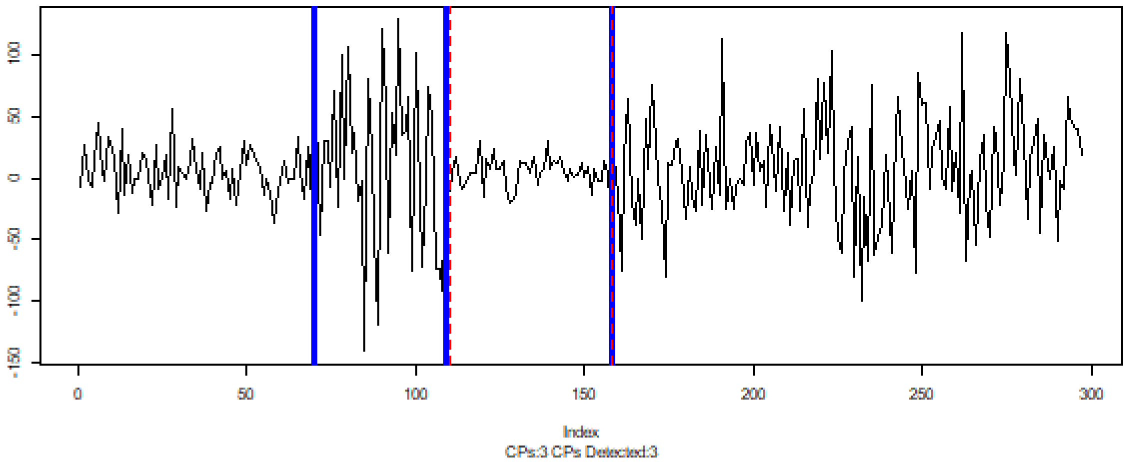

150
100
50
0
-50
-100
-150
0
50
100
150
200
250
300
Index
CPs:2 CPs Detected:1


150
100
50
0
-50
-100
-150
-200
0
50
100
150
200
250
300
Index
CPs:2 CPs Detected:2

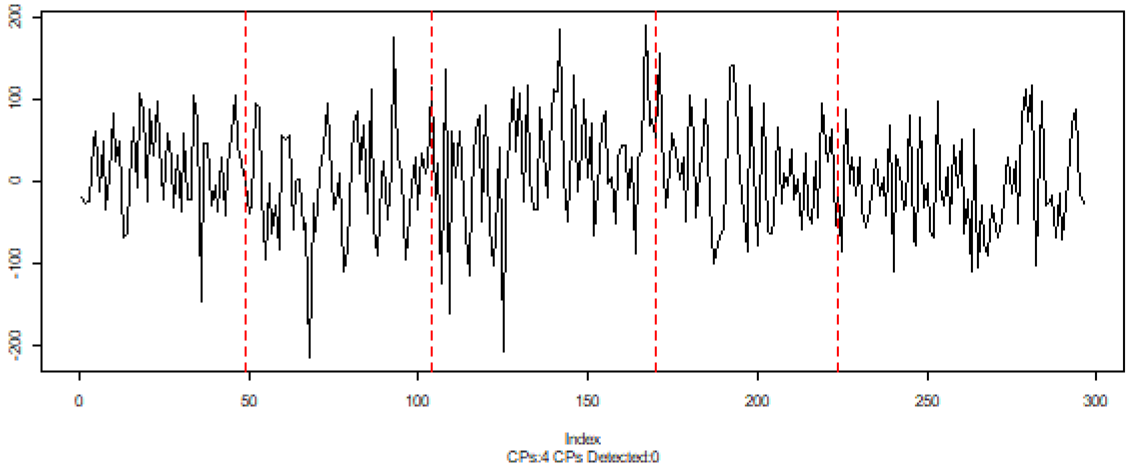
K=3


100
50
0
-50
-100
0
50
100
150
200
250
300
Index
CPs:3 CPs Detected:0

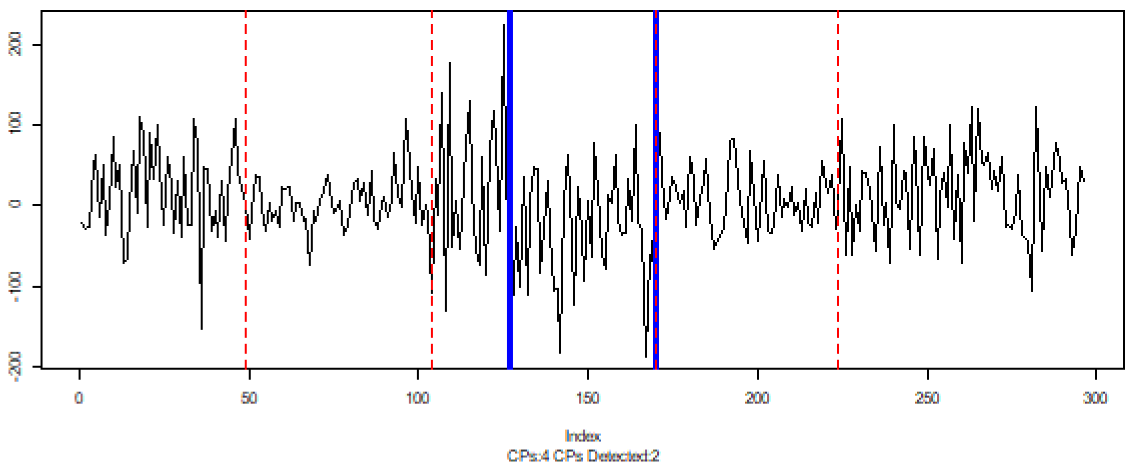

150
100
50
0
-50
-100
-150
0
50
100
150
200
250
300
Index
CPs:3 CPs Detected:0


100
50
0
-50
-100
-150
0
50
100
150
200
250
300
Index
CPs:3 CPs Detected:3

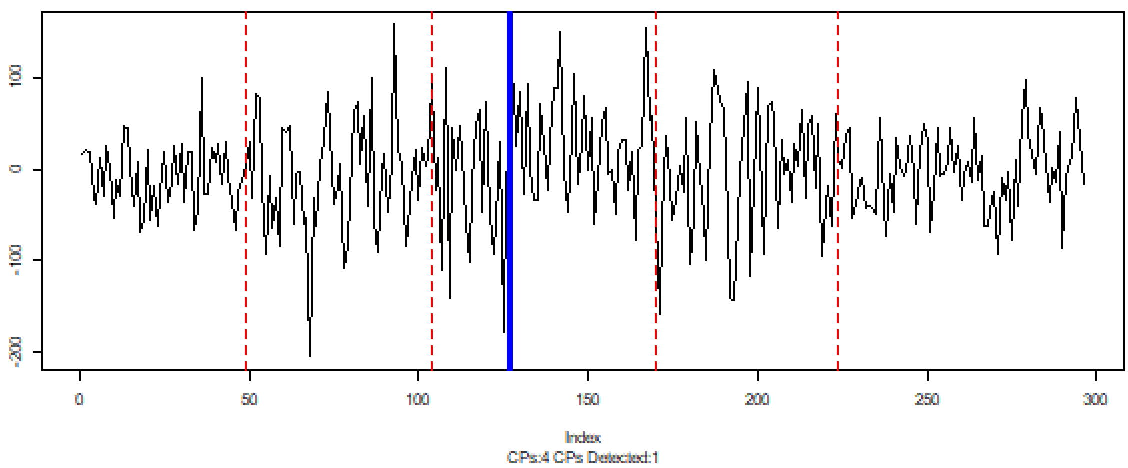
K=4



200
100
0
-100
-200
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:0

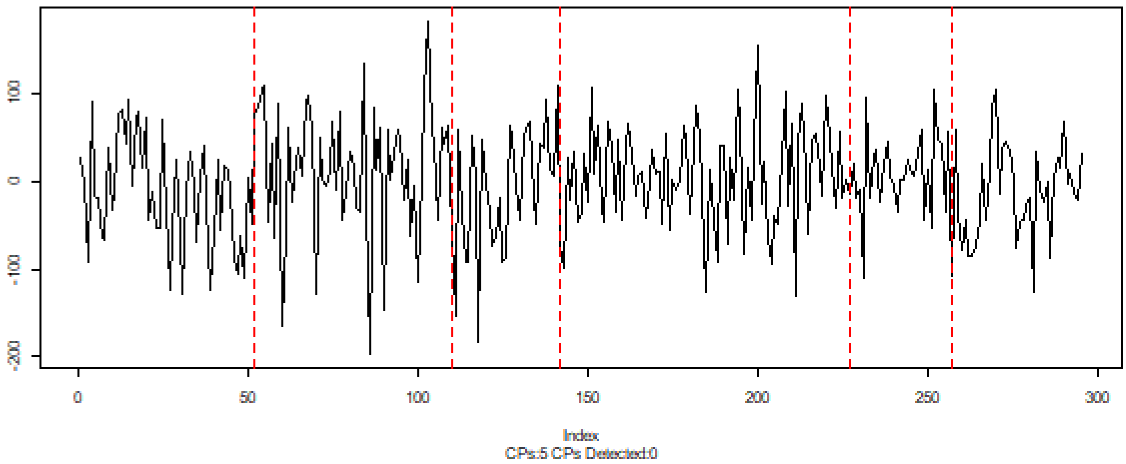


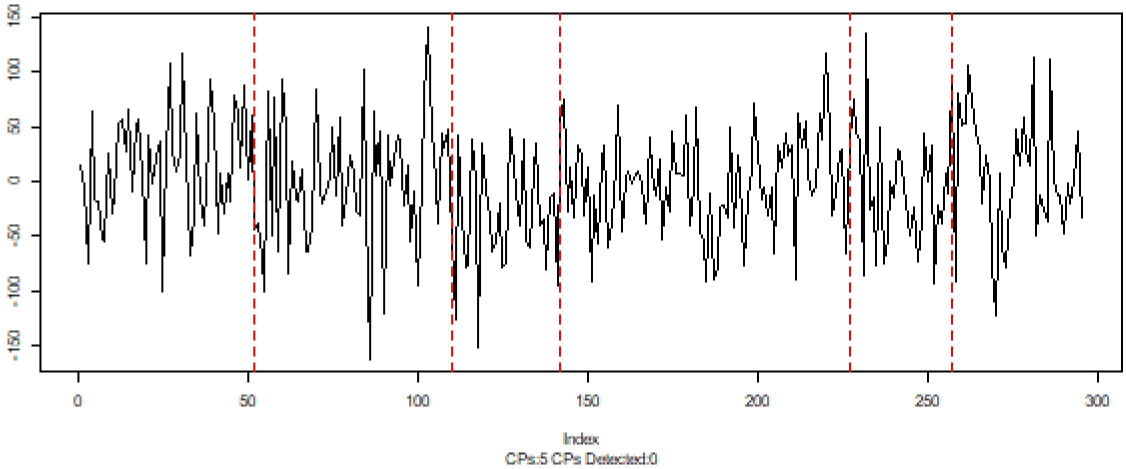
200
100
0
-100
-200
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:2

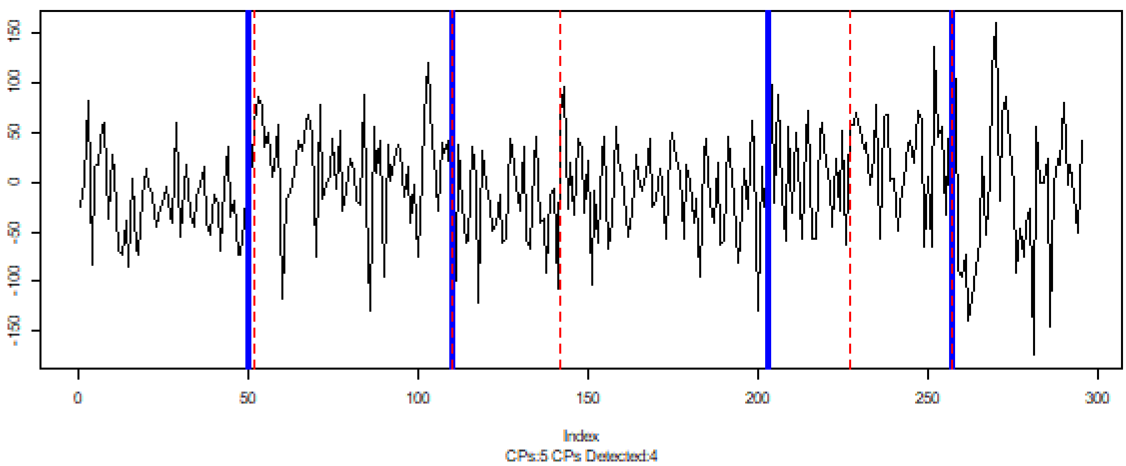


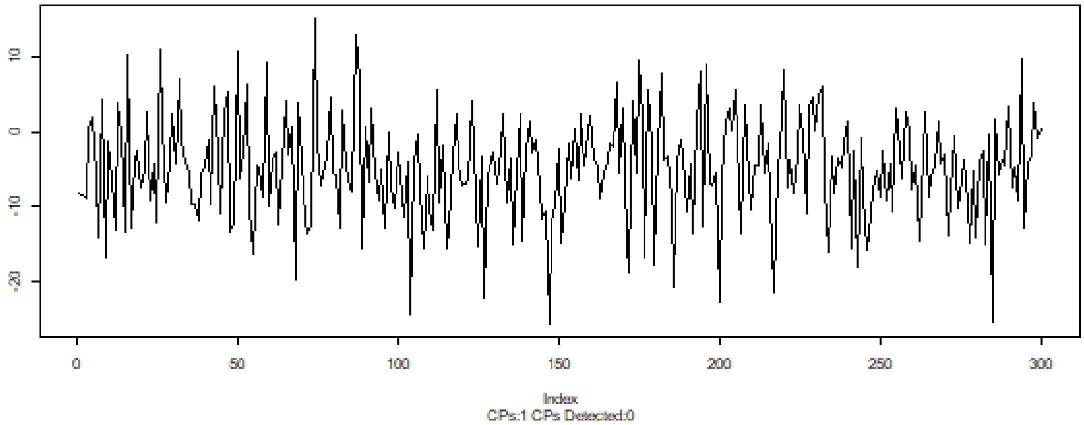
100
0
-100
-200
0
50
100
150
200
250
300
Index
CPs:4 CPs Detected:1

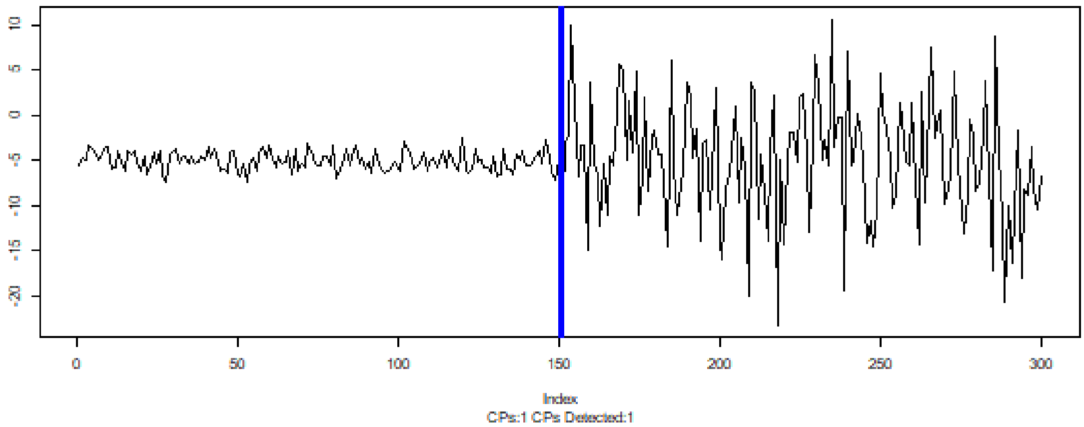
K=5

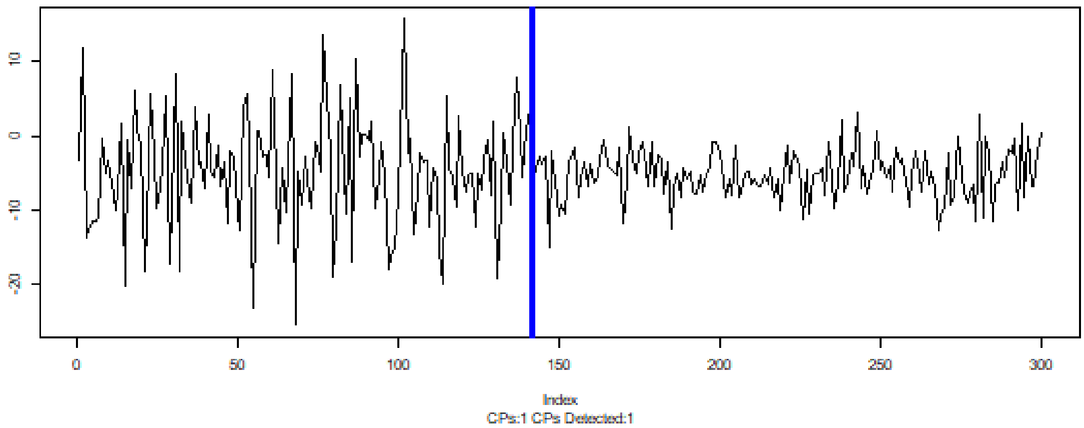

Index
CPs:5 CPs Detected:0

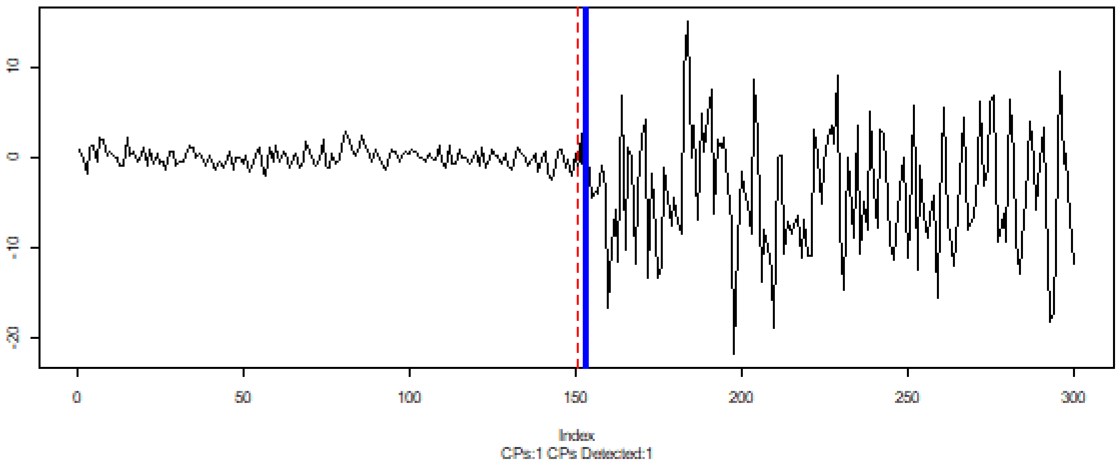

Index
CPs:5 CPs Detected:0

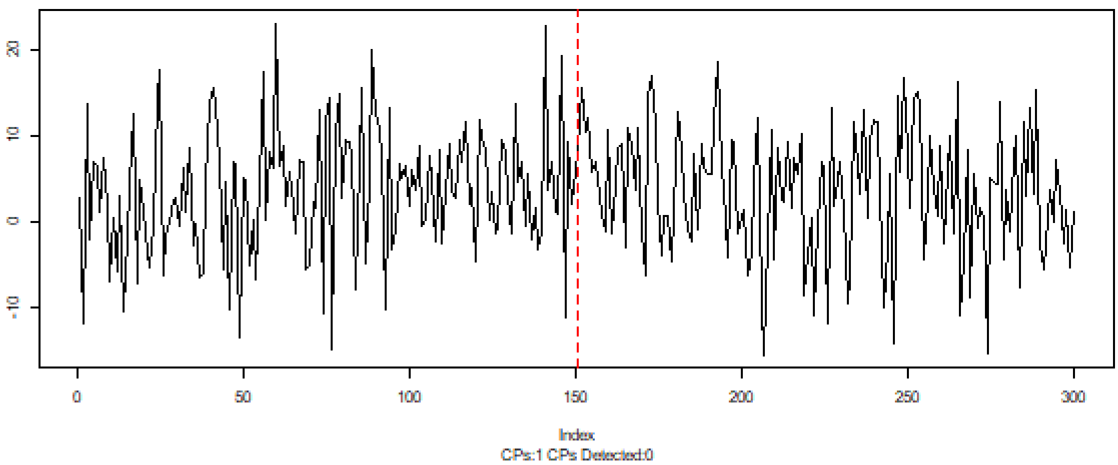

Index
CPs:5 CPs Detected:4

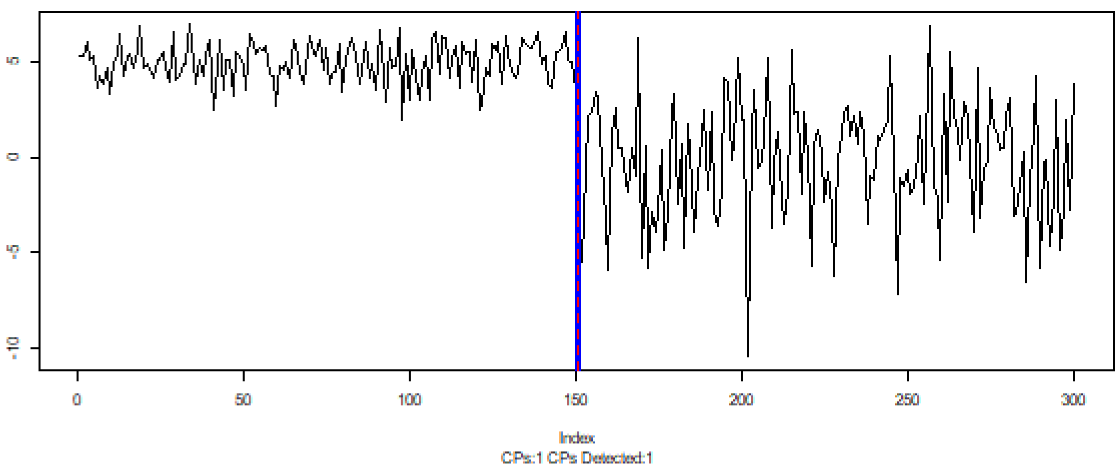

10
0
-10
-20
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:0

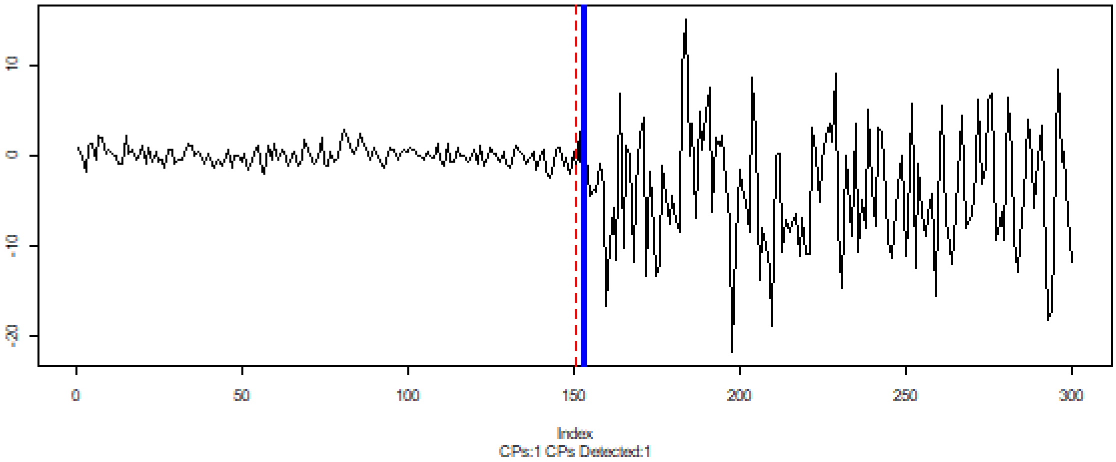

10
5
0
-5
-10
-15
-20
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1

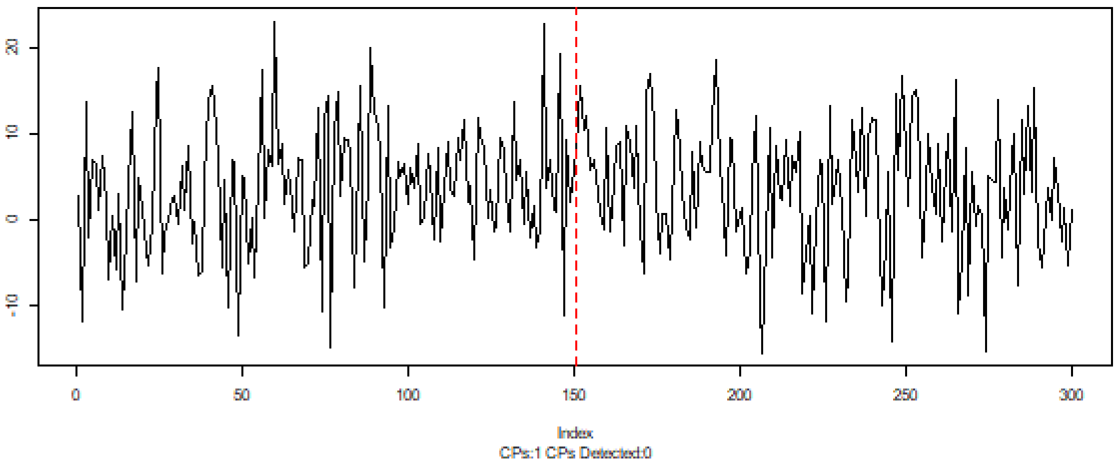

10
0
-10
-20
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1

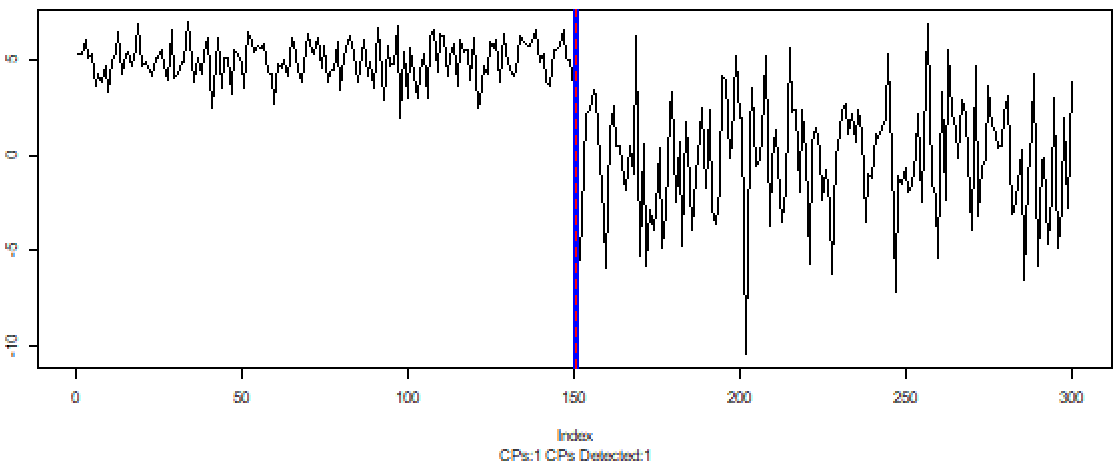

10
0
-10
-20
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1


20
10
0
-10
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:0


5
0
-5
-10
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1


10
0
-10
-20
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1


20
10
0
-10
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:0


5
0
-5
-10
0
50
100
150
200
250
300
Index
CPs:1 CPs Detected:1

REFERENCE

Adams, R. P., & MacKay, D. J. (2007). Bayesian Online Changepoint Detection. *stat*, *1050*, 19.

Alippi, C., Boracchi, G., Carrera, D., & Roveri, M. (2016). Change Detection in Multivariate Data streams: Likelihood and Detectability Loss. In *Joint Conference on Artificial Intelligence (IJCAI-16)* (pp. 1368-1374).

Aminikhanghahi, S., & Cook, D. J. (2017). A survey of methods for time series change point detection. *Knowledge and information systems*, *51*(2), 339-367.

Arlot, S., Celisse, A., & Harchaoui, Z. (2012). Kernel change-point detection. *arXiv preprint arXiv:1202.3878*, *6*.

Bai, J. (1997). Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, *79*(4), 551-563.

Bhattacharya, P. K. (1994). Some aspects of change-point analysis. In *Change-point problems* (pp. 28-56). Institute of Mathematical Statistics.

Biau, G., Bleakley, K., & Mason, D. M. (2016). Long signal change-point detection. *Electronic journal of statistics*, *10*(2), 2097-2123.

Bulteel, K., Ceulemans, E., Thompson, R. J., Waugh, C. E., Gotlib, I. H., Tuerlinckx, F., & Kuppens, P. (2014). DeCon: A tool to detect emotional concordance in multivariate time series data of emotional responding. *Biological Psychology*, *98*, 29-42.

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Grassmann, M., & Ceulemans, E. (2017). Detecting correlation changes in multivariate time series: A comparison of four non-parametric change point detection methods. *Behavior Research Methods*, *49*(3), 988-1005.

Cabrieto, J., Tuerlinckx, F., Kuppens, P., Hunyadi, B., & Ceulemans, E. (2018). Testing for the Presence of Correlation Changes in a Multivariate Time Series: A Permutation Based Approach. *Scientific REPORTS*, *8*, 769.

Cho, H., & Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society*, *77*(2), 475-507.

Galeano, P., & Wied, D. (2014). Multiple break detection in the correlation structure of random variables. *Computational Statistics & Data Analysis*, *76*, 262-282.

Garreau, D., & Arlot, S. (2018). Consistent change-point detection with kernels. *Electronic Journal of Statistics*, *12*(2), 4440-4486.

Guralnik, V., & Srivastava, J. (1999, August). Event detection from time series data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 33-42).

Hafeez, M., & Hussain, H. (2015, November). Monetary Policy and Exchange Rate Volatility in Pakistan: A VAR Approach. In *Hafeez, Muhammad and Hammad Hussain," Monetary Policy and Exchange Rate Volatility in Pakistan: A VAR Approach" 4th ICETM, 2015 (International Conference Emerging Trends in Management) at Mohammad Ali Jinnah University, Islamabad, Pakistan*.

Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, *37*(3), 323-341.

Haynes, K., Fearnhead, P., & Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, *27*(5), 1293-1305.

James, N. A., & Matteson, D. S. (2013). ecp: An R package for nonparametric multiple change point analysis of multivariate data. *arXiv preprint arXiv:1309.3295*.

Jandhyala, B. K., & Fotopoulos, S. B. (1999). Capturing the distributional behaviour of the maximum likelihood estimator of a changepoint. *Biometrika*, *86*(1), 129-140.

Jérémy, T., Guy, G., Brynjar, K., & Catherine, M. (2009). *Sources of bias in synchronization measures and how to minimize their effects on the estimation of synchronicity: application to the uterine electromyogram*. INTECH Open Access Publisher. Chapter(5), 73-99.

Kawahara, Y., & Sugiyama, M. (2009, April). Change-point detection in time-series data by direct density-ratio estimation. In *Proceedings of the 2009 SIAM International Conference on Data Mining* (pp. 389-400). Society for Industrial and Applied Mathematics.

Lavielle, M., & Teyssiere, G. (2006). Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, *46*(3), 287-306.

Liu, S., Yamada, M., Collier, N., & Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. *Neural Networks*, *43*, 72-83.

Lung-Yut-Fong, A., Lévy-Leduc, C., & Cappé, O. (2015). Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la Société Française de Statistique*, *156*(4), 133-162.

Matteson, D. S., & James, N. A. (2014). A nonparametric approach for multiple change point analysis of multivariate data. *Journal of the American Statistical Association*, *109*(505), 334-345.

Olshen, A. B., Venkatraman, E. S., Lucito, R., & Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, *5*(4), 557-572.

Peel, L., & Clauset, A. (2014, March). Detecting change points in the large-scale structure of evolving networks. In *th International Conference on Artificial Intelligence* (Vol. 291, No. arXiv: 1403.0989, pp. 4-2920).

Reeves, J., Chen, J., Wang, X. L., Lund, R., & 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.

Roy, S., Atchadé, Y., & Michailidis, G. (2017). Change point estimation in high dimensional Markov random-field models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *79*(4), 1187-1206.

Schröder, A. L. (2016). *Methods for change-point detection with additional interpretability* (Doctoral dissertation, The London School of Economics and Political Science (LSE)).

Sharkey, P., & Killick, R. (2014). Nonparametric methods for online changepoint detection. *STOR601 Research Topic II*.

Truong, C., Oudre, L., & Vayatis, N. (2018). A review of change point detection methods. *arXiv preprint arXiv:1801.00718*, 1-31.

Valasai, G. D., Uqaili, M. A., Memon, H. R., Samoo, S. R., Mirjat, N. H., & Harijan, K. (2017). Overcoming electricity crisis in Pakistan: A review of sustainable electricity options. *Renewable and Sustainable Energy Reviews*, *100*(72), 734-745.

Zhu, X., Li, Y., Liang, C., Chen, J., & Wu, D. (2013, January). Copula based change point detection for financial contagion in Chinese banking. In *ITQM* (pp. 619-626).