University College London.

M.Phil Thesis.

Change-point Detection in Time Series with Hydrological Applications.

Olalekan Kazeem Obisesan.

Department of Statistical Science, University College London.

October 7, 2011

Certification.

I Olalekan Kazeem Obisesan confirm that the work presented in this thesis is my own. Where information has been derived from other sources I confirm that this has been indicated in the thesis.

Dedication.

This work is dedicated to my Late father, Chief Y Obidairo Obisesan who died on Monday 18th July 2011. I will never forget my father for all his prayers, guidance and advises during the course of this research even at difficult times. May his gentle soul rest in peace(Amen).

Acknowledgements.

I write to acknowledge the Almighty God, the creator of the universe, men, and knowledge for providing me the opportunity of studying at University College London. I thank God for sparing my life and for providing me the strength of writing up this thesis.

I must thank Dr Richard Chandler, my principal supervisor for providing first-class supervision and training during the course of writing this thesis. Richard did everything within his capacity to make sure that this work is of the highest quality. He made many personal sacrifices to support this research. I THANK YOU VERY MUCH.

I also appreciate the fantastic efforts of Professor Trevor Sweeting, a former Head of Department of Statistical Science (first known Department of Statistics in the world!) at UCL for offering me a postgraduate-teaching-assistantship. This funding has opened my eyes wider and it has improved my teaching and research careers. I met students from many parts of the world and learnt lessons of my life even while teaching at UCL. I learnt a lot of the British culture while studying the environment and the people for a better understanding of their ways of life. I have been able to interact with many Statisticians both within and outside UCL while attending many training courses, conferences and seminars.

I appreciate all members of staff of the Department of Statistical Science at UCL including Professor Tom Fearn, Dr Serge Guillas, Dr Paul Northrop, and Dr Afzal Siddiqui for their helps in various forms. Dr Richard Taylor of UCL Department of Geography is appreciated for his professional assistance during the course of this work. Administrative staff in the Department of Statistical Science like Marion, Karen, Sue, and of-course Steve played very helpful roles in making things work for all within the Department.

I thank all research fellows, postdoctoral researchers, teaching assistants, postgraduate students both new and old in the department of Statistical Science at UCL particularly those that have worked with me in the past. I must mention Simon and Shafiq (a good friend). Simon particularly provided urgent helps which afforded me another opportunity of gaining from his wide experience.

The UCL Environment and Water institutes and The Royal Statistical Society are also

appreciated for providing time-to-time training in many relevant courses from many parts of the world. I also thank the UCL graduate school and the London Taught Course Center (LTCC) for their training-intervention at various times.

All members of staff of my home University, University of Ibadan, are appreciated for their positive efforts towards my research. I must mention Dr A. A. Adelakun, the first honourable commissioner for Environment and Water Resources in Oyo State, Nigeria who was the first to introduce Environmental Statistics to me. Others to be appreciated for various positive roles include Dr G. N. Amahia, Dr A. A. Sodipo, Dr O. I. Shittu, Dr O. E. Olubusoye, Dr Olaomi, Dr Oladiti and Dr Adekambi. Also worth of mentioning are Professors Bamiduro, Iyaniwura, Bamiro, Agbaje, Osibanjo, Ekhaguere, Ayoola, and Olayinka. The University administration under the leadership of the vice-chancellor, Professor I. F. Adewole is appreciated for its interventions at times of need.

However, it will be unfair not to mention two people who were with me when I was completely alone during the course of the research. These two people particularly made huge financial sacrifices during the course of this work. The first is my wife, Mrs Omolola Obisesan. The second is my mother, Mrs Titilola Obisesan. This work would never have been possible without these people. They both fed me with words of wisdom that kept me moving on till today. Omolola told me that not all that glitters is gold and Titilola told me that the greatest degree of inner tranquility comes from the development of love and compassion and that the greater we care for the happiness of others, the greater is our own sense of well-being.

I also appreciate the endurance of my immediate family members, the Obisesans and the Omotosos including my daughter, Oyinkansola, and my son, Oladoyin. Others include Dolapo, Kolade, Rukayat, uncle Ismail, Dr Nurudeen, Idris, and Sakiru. I must not forget Muhideen Ibrahim Ayede and Dr Sogo Adediran for their encouragements. There are many other people with whom I have worked in the past years so numerous that I cannot mention them here. I say THANK YOU ALL and God bless. E SE PUPO !

Contents

List of Tables

List of Figures

Abstract

Water pollution is a global problem that is more serious in developing countries. This work was motivated by the need to understand water pollution and quality in Nigeria and specifically in the city of Ibadan. Initially, data were obtained consisting of multivariate monthly water chemistry measurements between 2003 and 2007 from two reservoirs in the city, and admissions to three nearby hospitals for several water-borne diseases between 1997 and 2008. Also used are monthly rainfall data over a fifteen-year period from a nearby weather station. An initial analysis revealed that the quality of the water chemistry data was poor. In particular, many of the monthly time series contained abrupt changes and breaks. This thesis therefore focuses on the detection of change-points with particular application to hydrological time series. Additional data-sets with a simpler structure, United Kingdom Blackwater rainfall proportion and Haemolytic Uraemic Syndrome (HUS) disease data are used to estimate and explain positions of change-points. An extensive literature review is given including a theoretical demonstration showing why some standard statistical theories fail some regularity conditions for this problem. A wide variety of techniques to detect change-points are considered, with illustrations on different hydrological data-sets using profile likelihood and Markov Chain Monte-Carlo (MCMC) to estimate change-points. A conclusion is that although the sample size may hinder the accurate change-point detection in some series, state-space models provide a promising framework for the analysis of complex series in which multiple change-points may be present. This will help in detecting the potential presence of changes indicating data quality problems and signifying the direction of solutions towards future work.

Keywords: Water pollution, Change-points, Regularity, Markov Chain Monte Carlo, Profile likelihood, State-space model.

Chapter 1

Introduction

This research is motivated by an analysis of water quality data from Ibadan, Nigeria with suspicion of abrupt breaks in their behaviour. Water pollution is a global problem that is very serious worldwide with devastating consequences in developing countries especially due to its health-related effects. Adequate access to safe drinking water and sanitation is important worldwide and this necessity is part of the Millennium Development Goals (MDGs) of the United Nations (UN). It is listed in Target 7 of the United Nations' report, UN (2010). Efforts at reducing the proportion of people without access to safe drinking water and basic sanitation has brought some improvements but unfortunately the rural-urban gap access to safe drinking water is still wide in some developing countries. The improvements to halving the people without access to safe drinking water is still too low to reach this objective by the year 2015 (UN, 2008, 2010).

1.1 Background of the Research

The main background of the research can be linked to the fact that water pollution is a major problem in Nigeria and that there seem to be no clear incentives for good adoption of pollution abatement regulations and little disincentives for polluting the environment (Adelegan, 2004). Ibadan is the third largest city in Nigeria with a population of about 3 million people (NPC, 2006). Eleyele and Asejire reservoirs are the main sources of drinking water in Ibadan while 60% of households in Ibadan municipality are linked to piped water supply from reservoirs and other

sources. Decomposed solid waste from open dump sites run through the reservoirs (Omoleke, 2004). There exist high levels of Total Coliform (E.coli) in lakes. Water-borne diseases such as dysentery, are leading causes of morbidity (Adelegan, 2004). It is known that water supply and sanitation monitoring are vulnerable to climate change. In particular, increased precipitation intensity and high temperatures contribute to water pollution (Kundzerwicz, 2007). However, inadequate and poor water quality data with very weak collaboration (leading to duplication of efforts and waste of resources) among relevant agencies such as, the Standards Organisation of Nigeria (SON) and the World Health Organisation (WHO) are other problems of concern. There is therefore the need to have appropriate scientific regulations for monitoring the effects of water pollution. It includes the need to build statistical models for evaluating water pollution and the need to determine water quality and its pollution levels in Nigeria. Moreover we need to determine pollutants with the strongest and weakest effects.

1.2 Datasets

In this study, we investigate the quality of water supplied for human consumption in Ibadan, Nigeria. Data-sets to be discussed consist of multivariate monthly water chemistry data from two reservoirs, Asejire and Eleyele, in Ibadan Nigeria over a 5-year period (obtained from the Water Corporation of Oyo State Nigeria). We also have data for Eleyele reservoir from the Oyo State Ministry of Environment and Water Resources Ibadan for about a 5-year period (50 months between May 2003 and June 2007).

In subsequent exploratory analysis on the water chemistry data, Tur is used to stand for Turbidity measured in Nephelometric Turbidity Units (NTU), Col for Colour measured in Hazen Units (HU), PH for pH measured in Logarithmic Units (LU), DO for Dissolved Oxygen measured in milligram per litre(mg/l), Alk stands for Alkalinity in mg/l, TH for Total Hardness in mg/l, CaH for Calcium Hardness in mg/l, Cl for Chloride in mg/l, Fe for Iron in mg/l. Also used here is Si for Silica measured in mg/l, Sol for Total Solid in mg/l, DS for Total Dissolved Solids in mg/l, SS for Total Suspended Solids in mg/l, NITRT for Nitrate measured in parts-per-million (ppm), COD for Chemical Oxygen Demand in mg/l, and BOD for Biochemical Oxygen Demand measured in mg/l.

For clarity purposes, Table 1.1 lists the variables together with their units of measurement

and corresponding abbreviations. Table 1.2 tabulates the maximum concentration levels for some selected variables as listed in SON (2007) and World Health Organisation (2008) and in using the listed standards, Table 1.3 shows some indication of pollution in Eleyele reservoir. The values on Table 1.3 refer to months and it can be seen that there exists some indication of pollution since Turbidity, Colour and Iron show that some of the sixty monthly samples contain substances that go above standards and can therefore render the water polluted or contaminated.

Variables	**Units of Measurements**
Turbidity (Tur)	Nephelometric Turbidity Units (NTU)
Colour (Col)	Hazen Units (HU)
PH (pH)	Logarithmic Units (LU)
Dissolved Oxygen (DO)	milligram per litre (mg/l)
Alkalinity (Alk)	milligram per litre (mg/l)
Total Hardness (TH)	milligram per litre (mg/l)
Calcium Hardness (CaH)	milligram per litre (mg/l)
Chloride (Cl)	milligram per litre (mg/l)
Iron (Fe)	milligram per litre (mg/l)
Silica (Si)	milligram per litre (mg/l)
Total Solids (Sol)	milligram per litre (mg/l)
Dissolved Solids (DS)	milligram per litre (mg/l)
Total Suspended Solids (SS)	milligram per litre (mg/l)
Chemical Oxygen Demand (COD)	milligram per litre (mg/l)
Biochemical Oxygen Demand (BOD)	milligram per litre (mg/l)
Nitrate (NITRT)	parts per million (ppm)

Table 1.1: Variables and corresponding units of measurement.

The graphs in Figure 1.1 to Figure 1.4 display the data. Figure 1.1 shows data from Eleyele Reservoir emanating from the Ministry of Environment and Water Resources (MEWR) between May 2003 and June 2007. Figure 1.1 show some unexpected and irregular behaviours for some variables. In particular, the BOD and COD series show some seasonality behaviour over the same period and the Nitrates series exhibit a significant change towards the 25th month. Figure

Attribute	SON	WHO
Colour	15HU	15HU
Turbidity	5NTU	5NTU
Chloride	250mg/l	250mg/l
Total Hardness	150mg/l	200mg/l
Iron	0.3mg/l	0.3mg/l
pH	(6.5-8.5)LU	(6.5-8.0)LU
Total Dissolved Solids	500mg/l	600mg/l

Table 1.2: Maximum chemical concentrations based on SON (2007) and WHO (2008) reports.

Variables	Safe	Polluted	Contaminated
Turbidity	18	31	11
Colour	57	2	1
Chloride	60	0	0
Total Hardness	60	0	0
Iron	0	0	60
Total Dissolved Solids	60	0	0

Table 1.3: Chemical concentrations from Eleyele: The values refer to month(s) out of a total of sixty months considered where the determinand levels fall within safe, polluted or contaminated. Turbidity and Iron show indications of pollution based on SON and WHO reports.

1.2 also shows patterns for variables from Asejire but noticeably repeated colour concentrations of 5HU. In Figure 1.3 we see irregular patterns for variables from Eleyele and occasional extreme values of Dissolved Oxygen. Although there are extreme values evident in other variables, Colour and DO are very important indicators of pollution. Colour indicates the physical appearance and DO supports aquatic organisms in water. In Figure 1.4 we see a bar-plot comparing the average chemical profile from the two reservoirs. The average has been obtained from the raw and final water samples taken from the two reservoirs over the same period of sixty months. This shows that Eleyele reservoir, which is in the city centre, is more polluted than the Asejire reservoir. Therefore, Eleyele is prone to receiving more pollution from various major sources

such as industries in the city. We therefore need to understand these behaviours as as to build appropriate statistical model to capture and explain the structure of the series.

Moreover, based on SON (2007) and World Health Organisation (2008) we can check for conformance using some of the pollutants discussed. Due to the available data selecting Turbidity, Colour, Chloride, Total Hardness, Iron and Total Dissolved Solids we can understand the thresholds set for pollutants. For example, levels of Turbidity below 5 NTU, between 5 NTU and 10 NTU and above 10 NTU might be considered as safe, polluted and contaminated respectively. Also Colour concentrations below 15 HU, between 15.1HU and 20HU and 20.1HU upwards can be categorised as safe, polluted and contaminated respectively. Considering each units in each case, we can do the same for Chloride: below 250mg/l (safe), between 250.1mg/l and 300mg/l (polluted) and 300.1mg/l upwards (contaminated). In the same direction we might have Total Hardness below 150mg/l as safe, between 150.1mg/l and 200 mg/l as polluted and 200.1 upwards as contaminated. Iron concentrations below 0.3mg/l can be safe, between 0.31mg/l and 0.5mg/l polluted, and 0.51 upwards(contaminated) and Total Dissolved Solids below 600mg/l may be safe, between 600.1mg/l and 800mg/l may be polluted, and 800.1mg/l upwards contaminated. Considering these thresholds we tabulate in Table 1.3 concentration levels for Eleyele reservoir. The values in the table refer to month(s) out of a total of sixty months considered. The table shows some indication of pollution in Eleyele reservoir while the same comparison for Asejire reservoirs indicate standard conformance for many variables.

In summary, the exploratory analysis shows some irregular behaviour such as occasional suspect values of chemical data and the dissimilarity of reservoirs is also confirmed by different changes in values for the reservoirs. However due to irregular suspect behaviour of the data such as for BOD, COD and DO discussed, we can identify some priorities for the statistical analysis of such data. Specifically, methods for identifying change-points in time-series are required. Multivariate methods that are robust to data errors and outliers should be considered and methods involving the use of multivariate time-series models for non-normal data may also be required. The specific objectives of the thesis can therefore be stated as:

(a) Development of statistical methods for assessing and monitoring water quality data

(b) Use of non-standard statistical methods such as the identification of change-points in a time-series.

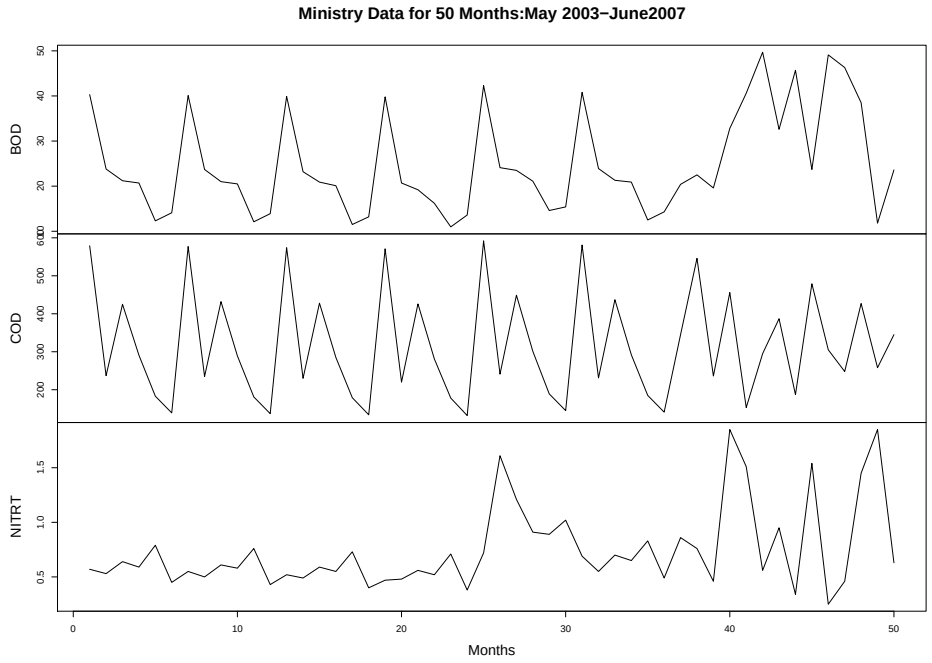

Figure 1.1: Time plots of monthly data from Eleyele Reservoir between 2003 and 2007, obtained from the Ministry of Environment and Water Resources in Oyo State Ibadan for BOD, COD, and Nitrate respectively.

For further illustration in subsequent work, two other data sets will be used to illustrate some of the techniques discussed. These data sets have been chosen because they seem to have simpler structure than the water quality data described above. The first is obtained from the United Kingdom Meteorological Office (UKMO) consisting of measurements from one of three sites (site 283424). Specifically, the data-sets refer to the annual proportions of wet-days from 1908 to 2000 from the Blackwater region of the United Kingdom. The data are displayed in Figure 1.5. From preliminary analysis noted in Yang et al. (2006), it can be viewed that after around 1970 the record becomes rather less complete and a substantial change in the resolution of the recordings can be observed. Yang et al. (2006) noted further that the resolution of values was $0.3mm$ and in the early 1970s the resolution was improved to $0.1mm$ for most sites except for some tipping bucket gauges for which the highest resolution was $0.2mm$. To detect the effect of change in resolution over a fairly long period, the authors selected the three longest sequences from the Blackwater data-set in the United Kingdom. Annual amounts and numbers of wet days are computed using a threshold of $0.3mm$ to compare with $0.1mm$. The authors calculated

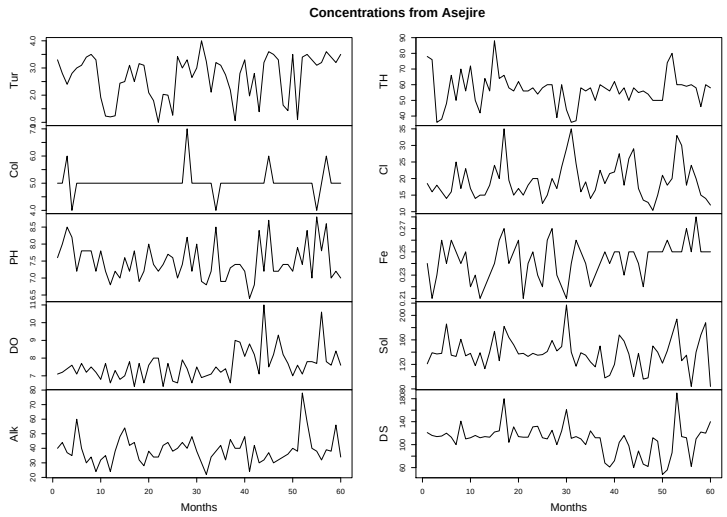

Figure 1.2: Time plots of monthly concentrations between 2003-2007 for Asejire Reservoir: Data supplied by the Water Corporation of Oyo State.

the annual amounts and numbers of wet days for the recorded series and then re-computed the quantities applying a threshold of $0.3mm$ so that the effect of recording to a resolution of $0.3mm$ could be known. Plotting values over the resolutions indicate that the improvement in recording resolution may introduce an illusory increasing trend in rainfall occurrences. Therefore, the rainfall proportion data could be used to illustrate suspect changes. It should be noted that Figure 1.5 shows the Black water data-set annual proportion over time with abrupt changes in the early 1970s.

The second data-set for illustration refers to the number of cases of Haemolytic Uraemic Syndrone (HUS) in Birmingham and Newcastle-upon-Tyne (1970-1989), United Kingdom. HUS is a life-threatening diarrhoea-related disease that causes severe illness primarily for infants and young children. It is linked with levels of E.coli in the environment (Lindsey, 2001). In Figure 1.6, cumulative counts of Haemolytic Uraemic Syndrome for two centers (Newcastle and Birmingham) are shown. The Newcastle data indicate an increase in slope around 1984 while the Birmingham series tends to change around 1980. There are some suspicion of changes that need to be understood in the data-sets and the times of changes in the HUS data sets could be interesting. Relevant applications of changes in the HUS data sets can also be found in Henderson and Mathews (1993) and Lindsey (2001).

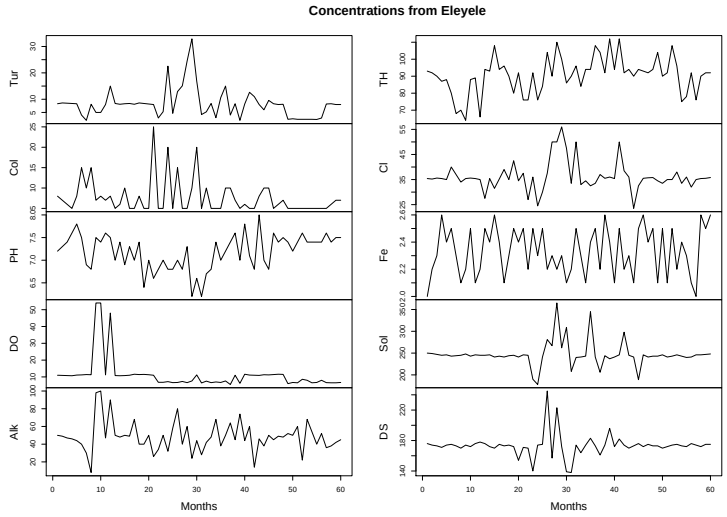

Figure 1.3: Time plots of monthly concentrations between 2003-2007 for Eleyele Reservoir: Data supplied by the Water Corporation of Oyo State.

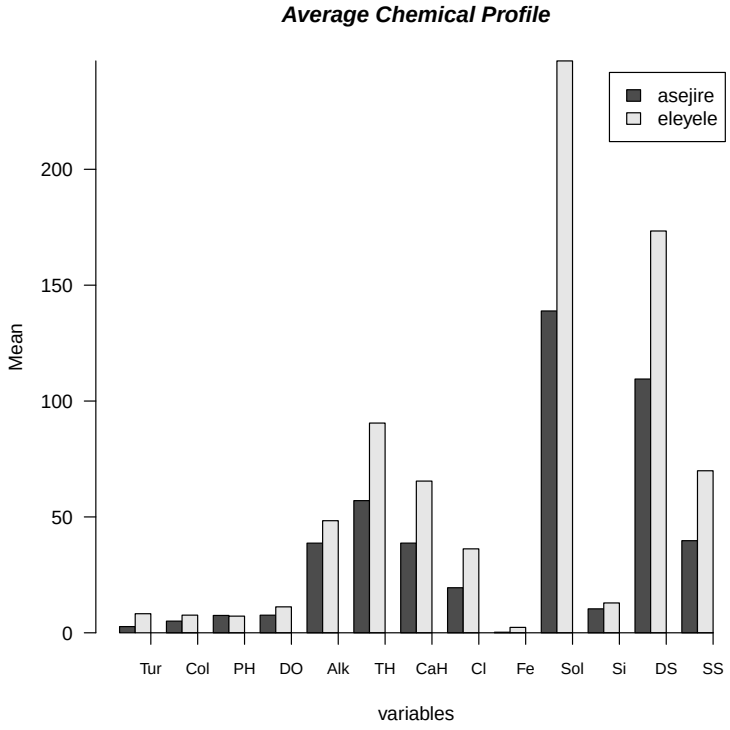

Figure 1.4: Barplot of average chemical profile for each variable taken between (2003-2007) for Asejire and Eleyele: The average chemical profile are derived from raw and final samples from the two reservoirs.

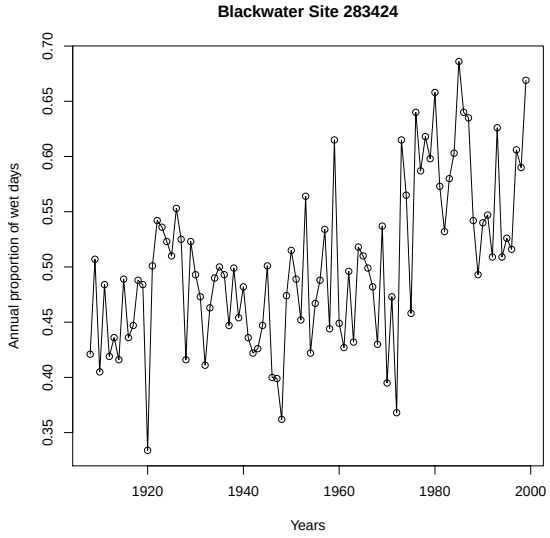

Figure 1.5: Time plot of annual proportion of wet days from Blackwater United Kingdom site 283424 as discussed in Yang et al. (2006).

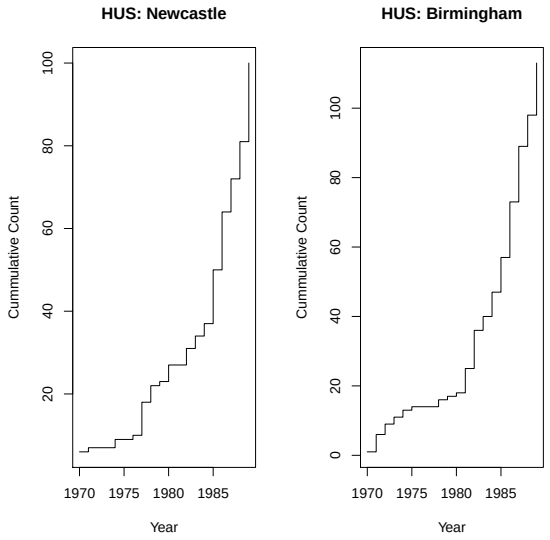

Figure 1.6: Cumulative counts of Haemolytic Uraemic Syndrome (HUS) in Newcastle-upon-Tyne and Birmingham: 1970-1989.

1.3 Structure of Dissertation

In this Chapter, we have described three different data-sets in order to study their characteristics thereby allowing the data-sets to suggest suitable models that may best explain their structures. The exploratory data analysis done in this Chapter has therefore motivated the need to investigate change-point problems. Chapter 2 discusses change-point problems and the implications of the failure of the standard theory of likelihood asymptotics. Also discussed is the consistency and distribution of the maximum likelihood estimators. Chapter 3 discusses off-line methods of change-point problems which have a change in the mean level. Change-point problems are discussed with examples using normally distributed random variables. Simulation is also used to show the change-point problem with respect to the likelihood function. Bayes factors for normal model were also computed with applications to the United Kingdom rainfall data. Computation of Bayes factors for a Poisson model are also computed with applications to the HUS data sets. We then discuss more literature change-point methods with applications to non-parametric methods, regression, and time series. Examples on multiple change-points are also discussed. In Chapter 4 on-line procedures are discussed with examples. Univariare and multivariate state-space models are explained using Kalman filtering equations. A multivariate state-space model is proposed for the water chemistry data. Chapter 5 gives a summary of findings and recommendations with respect to future work.

Chapter 2

Asymptotic Change-point Problem

As indicated in Section 1.2 of Chapter 1, the water chemistry series contain some abrupt changes in behaviour. Such abrupt changes or discontinuities are referred to as change-points (Chen and Gupta, 2000). Possible causes of change-points include changes in locations of observations, equipment, measurement methods, environmental effects, regulations, standards and so on. We sometimes need to investigate the potential presence of possible changes in the data set indicating data quality problems that should be resolved prior to any subsequent analysis. However, changes do occur even in the best regulated system as noted in Yang et al. (2006); these authors emphasise that discrepancies in records, occasional disagreement between documentation and data, abnormal data entry, changed units of measurement and other problems require adequate attention. Most of the time we are faced with the problems of detecting the number of change-points or jumps and their locations. The change-point problem is much easier if the point of change is known and this leads to intervention analysis (Manly, 2001).

The applications of change-point analysis are huge: we find relevant literature in many fields including: Biology, Chemistry, Environmental Sciences and Climate Change, Engineering, Econometrics, Medicine, Behavioral Sciences and also in Political Science, Finance, Image Analysis and Security. This list is not exhaustive in any way. The earliest work in this area seems to be that by Page (1954, 1955, 1957) who developed cumulative sum (CUSUM) methods. However since the 1950s many more papers have emerged. Reviews of many techniques are given by Jandhyala and MacNeil (1986) and Jandhyala et al. (1999). The reason for such a large body of literature on change-points is that the standard theory breaks down where the time of change is

unknown. Further details with respect to the standard theory on change-points can be found in Easterling and Peterson (1995), Chen and Gupta (2000), Lu et al. (2005), Hanesiak and Wang (2005) and Wang (2006). In the general case when a single change-point occurs at a time δ say we consider a sequence of random variable $X_1, ..., X_\delta$ with distribution function F_1 and another sequence of random variables $X_{\delta+1}, ..., X_n$ with distribution function F_2.

In this thesis, before reviewing the literature on change-point problems, the standard theory of likelihood asymptotics will be summarised paying attention to the role of assumptions and figuring out the consequences of their failure (especially **Assumptions** 1-4 in Section 2.1 and Section 2.2) in change-point problems. The likelihood theory will be linked to change-point problems to illustrate why change-point they are difficult.

2.1 Standard Techniques of Likelihood Asymptotics

To study inference in change-points and understand why inferences in change-point problems are non-standard, some properties of likelihood based inference will be reviewed. The likelihood function for a scalar parameter θ based on data $\mathbf{X} = X_1, ..., X_n$ as a collection of independent observations is defined to be

$$L(\theta|\mathbf{X}) = f(\mathbf{X};\theta) = \prod_{i=1}^{n} f(X_i;\theta)$$

which is simply the joint density of the data, regarded as a function of the parameter (Rice, 2007). For convenience we study the log-likelihood function $\ln L(\theta|\mathbf{X}) = l(\theta)$ and write

$$l(\theta) = \ln L(\theta|\mathbf{X}) = \ln f(X_1;\theta) + \ln f(X_2;\theta) + ... + \ln f(X_n;\theta) = \sum_{i=1}^{n} \ln f(X_i;\theta).$$

The maximum likelihood estimate of θ, is $\widehat{\theta}$ which is a value of θ that maximises the log-likelihood function. If the likelihood function is a differentiable function of θ then $\widehat{\theta}$ will be the root of $\frac{\partial l(\theta)}{\partial \theta} = 0$. Moreover, for a local maximum we need $\frac{\partial^2 l(\theta)}{\partial \theta} < 0$ at $\hat{\theta}$. The main assumptions here can be stated simply as

Assumption 1: The log-likelihood is a twice differentiable function.

Assumption 2: The second derivative $\frac{\partial^2 l(\theta)}{\partial \theta^2} < 0$ at $\widehat{\theta}$.

2.2 The Score Function

Under **Assumption 1** of Section 2.1, the first derivative is usually called the *score function*: $U(\theta, \mathbf{X}) = \frac{\partial \ln L(\theta)}{\partial \theta} = \frac{\partial [\sum_{i=1}^{n} \ln f(X_i;\theta)]}{\partial \theta}$ and is regarded as a function of θ for fixed $\mathbf{X}$. This function plays a central role in maximum likelihood theory. We can also define the observed information as

$$J(\theta) = -\frac{\partial^2 (l(\theta)}{\partial \theta^2} = -\frac{\partial^2}{\partial \theta^2} \sum_{i=1}^{n} l_i(\theta) = -\sum_{i=1}^{n} \frac{\partial^2 \ln f(X_i;\theta)}{\partial \theta^2}$$

which is a sum of n components. Also the Fisher information is defined as

$$I(\theta) = E\left\{-\frac{\partial^2 l(\theta)}{\partial \theta^2}\right\} = E\{J(\theta)\}$$

and

$$I(\theta) = E\left\{-\frac{\partial^2}{\partial \theta^2} \sum_{i=1}^{n} \ln f(X_i;\theta)\right\}$$

which can be written

$$\begin{aligned} I(\theta) &= \sum_{i=1}^{n} -E\left\{\frac{\partial^2 \ln f(X_i;\theta)}{\partial \theta^2}\right\} \\ &= n i_f(\theta), \ say \end{aligned}$$

where $i_f(\theta)$ refers to single observation information.

Now we show some characteristics of the *score function* when data are assumed generated from $f(.;\theta_0)$ so that θ_0 (assumed true value of θ) is the parameter to be estimated. If we have an independent and identically distributed sample of size n, the log-likelihood is written as

$$l(\theta) = \sum_{i=1}^{n} \ln f(X_i|\theta). \tag{2.1}$$

A careful illustration of the behaviour of the score function is given in Figure 2.1. This shows the sampling variation of score functions for different models (Normal, Poisson, Binomial and Cauchy) for samples of size n=10 (taken only for illustrative purposes). Figure 2.1(a) shows 20

score functions, each based on independent and identically distributed samples of size $n = 10$ from $N(4, 1)$. Each function is exactly linear and the score varies around 0 at the true parameter $\theta = \theta_0 = 4$. Figure 2.1(b) shows score functions for 20 independent samples of size 10 from a Poisson distribution with mean 4 (Near $\theta = 4$, each function looks approximately linear) and at the true parameter $\theta = 4$, the score function also varies around 0. Figure 2.1(c) shows score functions of 20 independent samples of size $n = 10$ from binomial(10,0.4) where $\theta = 0.4$. In Figure 2.1(d), the score functions for Cauchy($\theta = 4$) distributions are rather irregular and fail to behave as the previous models (although the score function also varies around 0 at $\theta = 4$, but there is the potential for multiple roots to the score equation). This case indicates problems with a complicated likelihood.

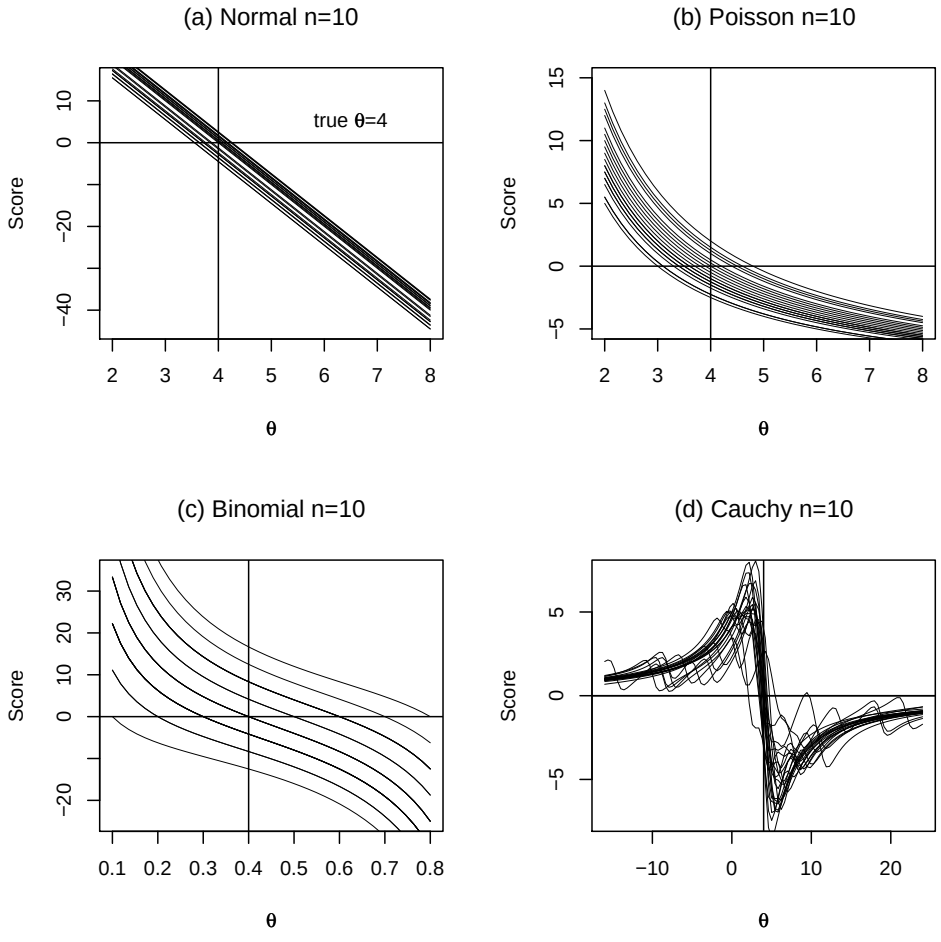

Figure 2.1: Sampling variation of score functions for different distributions.

In all examples in Figure 2.1, the score varies around zero at the true parameter value. We now show this is generally the case. Recall that the score is the first derivative of the log-likelihood function where we set $U(\theta, \mathbf{X}) = \frac{\partial l(\theta)}{\partial \theta}$, then at the true value of θ which is θ_0 we have

$$\begin{aligned} E(U(\theta_0, \mathbf{X})) &= \int_{-\infty}^{\infty} u(\theta_0, x).f(x|\theta_0)dx \\ &= \int_{x:f(x|\theta_0)>0} \frac{\partial l(\theta)}{\partial \theta_{|\theta=\theta_0}} f(x|\theta_0)dx \\ &= \int_{x:f(x|\theta_0)>0} \frac{1}{f(x|\theta_0)} \frac{\partial f(x|\theta)}{\partial \theta_{|\theta=\theta_0}} f(x|\theta_0)dx \\ &= \frac{\partial}{\partial \theta} \int f(x|\theta)dx = \frac{\partial}{\partial \theta}[1] = 0. \end{aligned}$$

A major assumption here is needed to justify interchanging the order of differentiation and integration and can be stated as the third assumption as in

Assumption 3: The range of integration of the variable does not depend on θ

Therefore using the stated assumptions we have $E(U(\theta_0, \mathbf{X})) = 0$ as required. We can also find the variance of the score function as

$$V(U(\theta_0; \mathbf{X})) = E(U^2(\theta_0; \mathbf{X})) - (E(U(\theta_0; \mathbf{X})))^2 = E[U^2(\theta_0, \mathbf{X})]$$

since $E(U(\theta_0, \mathbf{X})) = 0$ as seen above. We can rewrite $U(\theta_0, \mathbf{X})$ as say

$$U(\theta_0, \mathbf{X}) = \sum_{i=1}^{n} \left\{ \frac{\partial \ln f(x_i, \theta)}{\partial \theta_{|\theta=\theta_0}} \right\} = \sum_{i=1}^{n} S_i(\theta_0).$$

This implies that

$$E(U(\theta_0, \mathbf{X})) = E\left[\sum_{i=1}^{n} S_i(\theta_0) \right] = nE(S_i(\theta_0)).$$

Now

$$E(S_i(\theta)) = \int_{-\infty}^{\infty} \frac{\partial \ln f(x; \theta)}{\partial \theta} f(x; \theta_0)dx \tag{2.2}$$

and $E(S_i(\theta_0)) = 0$. Differentiating Equation 2.2 with respect to θ

we have

$$\frac{\partial}{\partial\theta}E(S_i(\theta)) = \int_{-\infty}^{\infty} \frac{\partial^2 \ln f}{\partial\theta^2} f(x;\theta_0)dx.$$

Therefore

$$\begin{aligned}
\frac{\partial}{\partial\theta}E(S_i(\theta)) &= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\left[\frac{\partial \ln f}{\partial\theta}\right] f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \frac{\partial}{\partial\theta}\left[\frac{1}{f}\frac{\partial f}{\partial\theta}\right] f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2} - \frac{1}{f^2}(\frac{\partial f}{\partial\theta})^2\right] f(x;\theta_0)dx.
\end{aligned}$$

At θ_0 therefore, we have

$$\begin{aligned}
\frac{\partial}{\partial\theta_{|\theta=\theta_0}}E(S_i(\theta)) &= \int_{-\infty}^{\infty} \left\{\frac{1}{f(x;\theta_0)}\frac{\partial^2 f}{\partial\theta^2_{|\theta=\theta_0}} - \frac{1}{f^2(x;\theta_0)}\left\{\frac{\partial f}{\partial\theta_{|\theta=\theta_0}}\right\}^2\right\} f(x;\theta_0)dx \\
&= \int_{-\infty}^{\infty} \frac{\partial^2 f}{\partial\theta^2_{|\theta=\theta_0}}dx - \int_{-\infty}^{\infty}\left[\frac{1}{f(x;\theta_0)}\frac{\partial f}{\partial\theta}_{|\theta=\theta_0}\right]^2 f(x;\theta_0)dx.
\end{aligned}$$

The score function is a sum of n independent random variables, the equation above shows that

$$V(U(\theta_0;\mathbf{X})) = E(U^2(\theta_0,\mathbf{X})) = -nE\left[\frac{\partial^2 \ln f(x;\theta)}{\partial\theta^2}\right]_{\theta=\theta_0}$$

since

$$E(U(\theta_0,\mathbf{X})) = E\left[\sum_{i=1}^{n} S_i(\theta_0)\right] = nE(S_i(\theta_0)).$$

Next we see how $U(\theta_0,\mathbf{X})$ behaves by studying $\frac{1}{n}U(\theta_0,\mathbf{X})$. As $n\to\infty$ we have,

$$E\left[\frac{U(\theta_0,\mathbf{X})}{n}\right] = \frac{1}{n}E(U(\theta_0,\mathbf{X})) = 0$$

and also that (assuming $\partial^2 \ln f/\partial\theta^2 < \infty$)

$$V\left[\frac{U(\theta_0,\mathbf{X})}{n}\right] = -\frac{1}{n}E\left[\frac{\partial^2 \ln(f(x;\theta))}{\partial\theta^2}\right]_{\theta=\theta_0} \to 0 \tag{2.3}$$

as $n\to\infty$. Hence, $n^{-1}U(\theta_0,\mathbf{X})\to 0$ in probability as $n\to\infty$.

The discussion so far has dealt with the behaviour of the score function at θ_0, the true parameter value. We now consider its behaviour at other values of θ. In general for $\theta \neq \theta_0$ we find that there may be need for another assumption.

Assumption 4: For $\theta \neq \theta_0$ the density $f(x, \theta)$ differs from $f(x, \theta_0)$ on a set of nonzero measure.

Note that $E(U(\theta, \mathbf{X})) \neq 0$ unless $\frac{\partial f}{\partial \theta} = 0$ for all x (which itself contradicts **Assumption 4**). Then for an arbitrary value of $\theta \neq \theta_0$

$$E(U(\theta, \mathbf{X})) = \int u(\theta, x) f(x|\theta_0) dx = \int \frac{1}{f(x|\theta)} \frac{\partial f(x, \theta)}{\partial \theta} f(x|\theta_0) dx \tag{2.4}$$

Now consider $V(U(\theta; \mathbf{X})) = V\left[\sum_{i=1}^{n} \frac{\partial \ln f(\mathbf{X}|\theta)}{\partial \theta}\right] = nV\left[\frac{\partial \ln f(\mathbf{X}|\theta)}{\partial \theta}\right] \propto n$ if $V\left[\frac{\partial \ln f(\mathbf{X}|\theta)}{\partial \theta}\right]$ is finite for all θ. Therefore for a case when $\theta \neq \theta_0$ we have $E(U(\theta, \mathbf{X})) = E(\frac{\partial \ln f(\mathbf{X};\theta)}{\partial \theta}) \neq 0$ and

$$V\left[n^{-1} U(\theta; \mathbf{X})\right] = \frac{1}{n^2} n V\left[\frac{\partial \ln f(\mathbf{X}, \theta)}{\partial \theta}\right] \to 0$$

and so $n^{-1}U(\theta, \mathbf{X}) \to E\left[\frac{\partial \ln f(\mathbf{X},\theta)}{\partial \theta}\right] \neq 0$. Therefore as $n \to \infty$ the function $U^*(\theta) = n^{-1}U(\theta, \mathbf{X})$ tends to a deterministic function with a root at θ_0.

2.3 Consistency of Maximum Likelihood Estimators

We now consider whether $\widehat{\theta}$ is a consistent estimator of θ_0. Using a Taylor expansion for $U(\theta)$ around θ_0 we have

$$U(\theta) = U(\theta_0) + (\theta - \theta_0) \frac{\partial U}{\partial \theta}_{|\theta=\theta^\dagger}$$

for some $\theta^\dagger \in (\theta_0, \theta)$ and so we can write

$$\theta - \theta_0 = \frac{U(\theta) - U(\theta_0)}{\frac{\partial U}{\partial \theta}_{|\theta^\dagger}}.$$

In particular, when $\theta = \hat{\theta}$ then we have (noting that $U(\hat{\theta}) = 0$)

$$\widehat{\theta} - \theta_0 = -\frac{U(\theta_0)}{\frac{\partial U}{\partial \theta}_{|\theta^\dagger}}, \tag{2.5}$$

which can be rewritten as

$$\widehat{\theta} - \theta_0 = -\frac{n^{-1}U(\theta_0)}{n^{-1}\frac{\partial U}{\partial \theta}_{|\theta^\dagger}}. \tag{2.6}$$

Note that the numerator of Equation (2.6) approaches 0 as $n \to \infty$. If we assume that the denominator is guaranteed nonzero, then Equation (2.6) implies that $\widehat{\theta} - \theta_0 \to 0$ and therefore $\widehat{\theta} \to \theta_0$. This requires the following assumption which can be seen as a strengthened version of **Assumption 2**.

Assumption 5: $\frac{\partial U}{\partial \theta}$ is non-zero in an interval containing θ_0

2.4 Limiting Distribution of $\widehat{\theta}$

As well as demonstrating the consistency of the maximum likelihood estimator $\hat{\theta}$, Equation (2.5) allows us to establish its distribution when n is large. Recall again that

$$V(U(\theta_0)) = E(U^2(\theta_0, \mathbf{X})) = -nE\left[\frac{\partial^2 \ln f(x;\theta)}{\partial \theta^2}\right]_{\theta=\theta_0}.$$

Moreover, $U(\theta_0)$ is a sum of independent and identically distributed contributions. Hence from the central limit theorem we have asymptotically,

$$\frac{U(\theta_0)}{\sqrt{-nE\left[\frac{\partial^2 \ln f}{\partial \theta^2_{|\theta_0}}\right]}} \sim N(0,1).$$

Now write $Z = \frac{U(\theta_0)}{\sqrt{-nE\left[\frac{\partial^2 \ln f}{\partial \theta^2_{|\theta_0}}\right]}}$. Then from Equation (2.6) we have

$$\hat{\theta} - \theta_0 = \frac{Z\sqrt{-nE\left[\partial^2 \ln f/\partial \theta^2_{|\theta_0}\right]}}{\partial U/\partial \theta_{|\theta^\dagger}}.$$

Hence we write,

$$(\hat{\theta} - \theta_0)\sqrt{-nE\left[\partial^2 \ln f/\partial \theta^2_{|\theta_0}\right]} = Z\frac{-nE\left[\partial^2 \ln f/\partial \theta^2_{|\theta_0}\right]}{\partial U/\partial \theta_{|\theta^\dagger}} = Z\frac{-E\left[\partial^2 \ln f/\partial \theta^2_{|\theta_0}\right]}{n^{-1}\partial U/\partial \theta_{|\theta^\dagger}}. \tag{2.7}$$

Now, as $n \to \infty$, $n^{-1}\partial U/\partial\theta_{|\theta^\dagger} \to E\left[\partial^2 \ln f/\partial\theta^2_{|\theta^\dagger}\right]$ and $\theta^\dagger \to \theta_0$ since it lies between $\hat{\theta}$ and θ_0. Also, if $\partial^2 \ln f/\partial\theta^2$ is continuous in θ then as $\theta^\dagger \to \theta_0$,

$$E\left[\partial^2 \ln f/\partial\theta^2_{|\theta^\dagger}\right] \to E\left[\partial^2 \ln f/\partial\theta^2_{|\theta_0}\right].$$

In this case, the final term in Equation (2.7) tends to 1 and we have

$$(\hat{\theta} - \theta_0)\sqrt{-nE\left[\frac{\partial^2 \ln f}{\partial\theta^2_{\theta_0}}\right]} \to Z$$

which has a standard normal distribution asymptotically.

2.5 Limiting Chi-square Distribution: Likelihood Ratio Statistic

Now we discuss the basic test statistic used for testing hypothesis using the principles of likelihood functions. Suppose that $l(.)$ is the log-likelihood established from the probability density f. Then the consistency of $\widehat{\theta}$ implies that we can write

$$l(\theta_0) = l(\widehat{\theta}) + (\theta_0 - \widehat{\theta})l'(\widehat{\theta}) + \frac{(\theta_0 - \widehat{\theta})^2}{2}l''(\theta^\dagger)$$

where $\theta^\dagger$ is between $\widehat{\theta}$ and θ_0.

Then representing the likelihood ratio statistic with $L_r = 2(l(\widehat{\theta}) - l(\theta_0))$ gives

$$L_r = 2(\widehat{\theta} - \theta_0)l'(\widehat{\theta}) - (\widehat{\theta} - \theta_0)^2 l''(\theta^\dagger)$$

and since $l'(\widehat{\theta}) = 0$ by definition we can write that

$$L_r = ni_f(\theta_0)(\widehat{\theta} - \theta_0)^2 \times \frac{l''(\theta^\dagger)}{l''(\theta_0)} \times \frac{l''(\theta_0)}{-ni_f(\theta_0)} = ni_f(\theta_0)(\widehat{\theta} - \theta_0)^2 \times \frac{l''(\theta^\dagger)}{l''(\theta_0)} \times \frac{l''(\theta_0)/n}{-i_f(\theta_0)} \qquad (2.8)$$

Referring to the standard normal derived in Section 2.4, it is clear that the first part of Equation (2.8) is asymptotically the square of a standard normal random variable and it is therefore a χ^2_1 distribution. In addition, the last two ratios $\frac{l''(\theta^\dagger)}{l''(\theta_0)}$ and $\frac{l''(\theta_0)/n}{-i_f(\theta_0)}$ tend to 1 using similar arguments to those applied in the previous subsection. In the same direction, we can obtain the χ^2 distribution for a case when θ is vector (without proof) in that as above we

write $L_r(\theta_0) = 2[l(\widehat{\theta}) - l(\theta_0)] = (\widehat{\theta} - \theta_0)^T i_f(\theta_0)(\widehat{\theta} - \theta_0)$. It is therefore noted that $L_r(\theta_0)$ has an approximate chi-square distribution on p (say) degree(s) of freedom for repeated sampling of data from the model. We can therefore write $L_r(\theta_0) \xrightarrow{D} \chi^2_p$. More can be found in Paul et al. (2002).

2.6 Likelihood Theory for Change-point Problems

In starting to review methodologies used in the analysis of change-point problems in the literature, it is helpful to consider a simple change-point problem and demonstrate that some of the standard asymptotic theory do not hold in this problem. Specifically, we show that **Assumption 1** and **Assumption 4** from the previous sections may fail. We discuss single change-points under different classifications and methods with respect to two-mean levels.

2.6.1 Simple examples: The two-mean model

David Hinkley's contribution to change-point analysis is a central research that should be reviewed in detail. Hinkley (1970) considered sequences of random variables and discussed the point at which the probability distribution changes using a normal distribution with changing mean. The asymptotic distribution of the maximum likelihood estimate discussed in this paper is particularly relevant to change-points in the study of the likelihood ratio statistic for testing hypotheses about the change-point. The author indicated the simplest model over a whole range of data as $X_t = \theta(t) + \epsilon_t$ for $t = 1, ...T$ where $\theta(t)$ is a mean function and ϵ_t refer to error terms. Hinkley (1970) computed the asymptotic distribution in the normal case when θ_0 and θ_1 are unknown. The asymptotic distribution is found to be the same when the mean levels are known. The two-mean model to be considered supposes that there exists a mean $\theta_0(t)$ and mean $\theta_1(t)$ for $t = 1, ..., \delta$ and $t = \delta + 1, ..., T$ respectively. He also computed the asymptotic distribution of the likelihood ratio statistic for testing hypothesis on the change-point δ. The maximum likelihood estimate of the change-point $\hat{\delta}$ (where θ_0 and θ_1 are known and δ is unknown) is obtained from a sample $x_1, ..., x_T$ by simply maximising the likelihood function of the form

$$L(\delta, \theta_0, \theta_1) = \prod_{i=1}^{\delta} f(x_i, \theta_0) \prod_{i=\delta+1}^{T} f(x_i, \theta_1)$$

which can be written in form of the log-likelihood as

$$l(\delta, \theta_0, \theta_1) = \sum_{i=1}^{\delta} \ln f(x_i, \theta_0) + \sum_{i=\delta+1}^{T} \ln f(x_i, \theta_1). \tag{2.9}$$

Given that the variance of the independent observations is $\widehat{\sigma^2}$, consider the case where the mean levels θ_0 and θ_1 are not known. The log-likelihood of the observed sequence $(x_1, ..., x_T)$ is

$$l(\delta, \theta_0, \theta_1, \sigma^2 | x_1, ..., x_T) = -\frac{T}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{\delta} (x_i - \theta_0)^2 + \sum_{i=\delta+1}^{T} (x_i - \theta_1)^2 \right\} \tag{2.10}$$

If we assume that δ is known, therefore the maximum likelihood estimators of θ_0 , θ_1 and σ^2 respectively are $\widehat{\theta}_0 = \frac{\sum_{i=1}^{\delta} x_i}{\delta}$, $\widehat{\theta}_1 = \frac{\sum_{i=\delta+1}^{T} x_i}{T-\delta}$ and $\widehat{\sigma^2} = \frac{\sum_{i=1}^{\delta}(x_i-\theta_0)^2+\sum_{i=\delta+1}^{T}(x_i-\theta_1)^2}{T}$.

Particularly for convenience Hinkley (1970) substituted $\sigma^2 = 1$ as known so that Equation (2.10) becomes

$$l(\delta, \theta_0, \theta_1, 1, | x_1, ..., x_T) = -\frac{1}{2} \left\{ \sum_{i=1}^{\delta} (x_i - \theta_0)^2 + \sum_{i=\delta+1}^{T} (x_i - \theta_1)^2. \right\} \tag{2.11}$$

Putting the maximum likelihood estimates of θ_0 and θ_1 back in to the log-likelihood in Equation (2.11) and re-arranging the emerging sums of squares conditional on t, Equation (2.11) becomes

$$l(t) = -\frac{1}{2} \left\{ \sum_{i=1}^{T} (x_i - \bar{x}_T)^2 - t(T-t)(\bar{x}_t - \bar{x}_t^{\star})^2 / T \right\}, \tag{2.12}$$

for which $\bar{x}_T = \frac{\sum_{i=1}^{T} x_i}{T}$, $\widehat{\theta}_0 = \bar{x}_t$, $\widehat{\theta}_1 = \bar{x}_t^{\star}$ and $\hat{\delta}$ is the value of t that maximises the observed value z_t^2 of Z_t^2 where

$$Z_t^2 = t(T-t)(\bar{X}_t - \bar{X}_t^{\star})^2 / T.$$

Now suppose that σ^2 is unknown then we can use Equation (2.10) above to obtain a profile log-likelihood function for δ by substituting all maximum likelihood estimates of the other parameters except the change-point into Equation(2.10) to obtain the profile log-likelihood function of the change-point $l_p(t)$ as in

$$l_p(t) = -\frac{T}{2} \ln \left[\frac{\sum_{i=1}^{T} (x_i - \bar{x}_T)^2}{T} - \frac{t(T-t)(\bar{x}_t - \bar{x}_t^{\star})^2}{T^2} \right] - \frac{T}{2} \tag{2.13}$$

For illustration of the change-point problem using the United Kingdom rainfall data sets described in Section (1.2), plots of the log-likelihood function in Equation (2.12) and that of profile log-likelihood in Equation (2.13) can be seen below in Figure 2.2. The (a) part of Figure (2.2) shows the log-likelihood function as used by Hinkley (1970) when $\sigma^2 = 1$ and the (b) part shows the profile log-likelihood function for the change-point δ when the variance is unknown. They both have similar rough shape and are both step functions. The (b) part also shows the maximum with a vertical line indicating the change-point at point 65 corresponding to the early 1970s as noted in Yang et al. (2006). The failure of **Assumption 1** discussed in Section (2.1) is indicated by the step function. Moreover by setting $l_p(t) = l_p(\delta)$ and plotting the profile log-likelihood function described in Equation (2.13) for which $f(x/\theta_0) = f(x/\theta_1)$ for changing δ produces a constant function for a reasonable range of the change-point. However checking all assumptions, the constant function will indicate the failure of **Assumption 4** in Section (2.2).

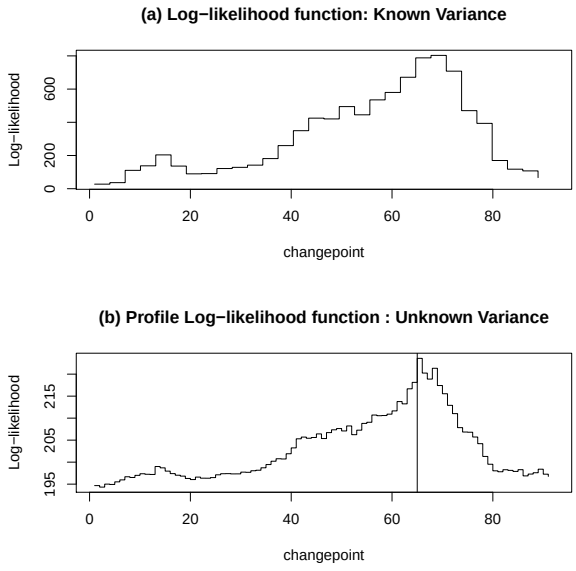

Figure 2.2: Log-likelihood functions: (a): Represents log-likelihood function in Equation (2.12) as used by Hinkley with known variance and (b): Represents the profile log-like-lihood function of δ in Equation (2.13) with unknown variance σ^2. The two are step functions showing the problem of change-points.

Moreover, in a bid to illustrate the change-in-mean level problem we may use simulation to study the distribution of the log-likelihood ratio statistic. To obtain a suitable likelihood ratio statistic under the hypotheses $H_0 : \theta_0 = \theta_1$ against $H_1 : \theta_0 \neq \theta_1$. Using Equation (2.10) and setting $\delta = t$, we can derive the log-likelihood ratio statistic as

$$L_h = -T \times \ln \left[\frac{\sum_{i=1}^{t}(x_i - \widehat{\theta}_0)^2 + \sum_{i=t+1}^{T}(x_i - \widehat{\theta}_1)^2}{\sum_{i=1}^{T}(x_i - \widehat{\theta})^2} \right] \tag{2.14}$$

or equivalently

$$L_h = -T \times \ln \left[\frac{\sum_{i=1}^{T}(x_i - \bar{x}_T)^2 - t(T-t)(\bar{x}_t - \bar{x}_t^{\star})^2/T}{\sum_{i=1}^{T}(x_i - \bar{x}_T)^2} \right]$$

Showing the quantile plot of likelihood ratio statistic L_h in Equation (2.14) observed in 10000 simulated normal samples each of size 1000 in comparison with the χ_1^2 distribution, Figure 2.3 on the left indicates very high discrepancy between the distribution of the likelihood ratio statistic L_h and that of a χ_1^2 distribution. On its right Figure (2.3) shows conformance with the distribution of χ_1^2 (regular case). This clearly shows that when the position of the change-point is known the problem is that of a regular case conforming to the χ_1^2 distribution. We can therefore conclude that the likelihood ratio method may not produce a test statistic with known probability distribution such as the square of a standard normal.

Additionally giving an example on testing hypothesis on the existence of change-point, we again use the Hinkley (1970) method described earlier on the United Kingdom rainfall data set. We use bootstrapping to test the hypothesis on the existence of a change-point in the data set. We can bootstrap the rainfall proportion 10000 times and obtain the likelihood ratio statistic L_h (to check the existence of a change in the rainfall series). The bootstrapping procedure is done considering the rainfall data as the population from which we draw samples. The procedure involves taking 10000 samples each of the same size as the rainfall data with replacement and computing the likelihood ratio statistic L_h in each iteration. Doing this a very small pvalue of 0 is obtained and this indicates strong evidence against the suggestion of no change-point hence the change-point needs to be investigated. At the maximum, point 65 is detected as the position of change-point corresponding to changes in 1973 where the mean level seem to have changed. This also supports the claim in Yang et al. (2006) as point 65 coincides with 1973. We can see the position of the change-point in Figure (2.4) where the likelihood ratio and the rainfall proportion are plotted on two different vertical axes. This is done by plotting the likelihood

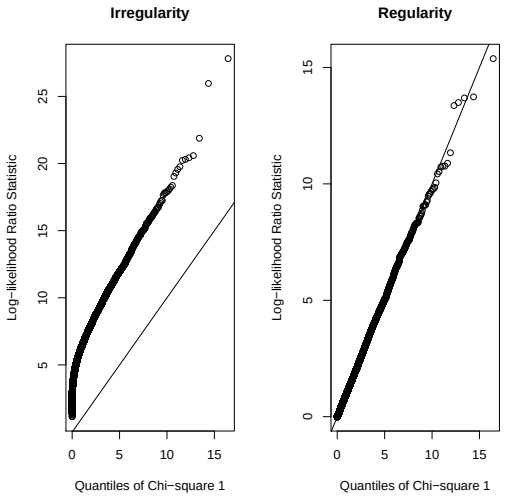

Figure 2.3: 10000 simulations of the Log-likelihood Ratio Statistic L_h with Quantiles of a χ_1^2 distribution. Left: Irregularity when the change-point is unknown then distribution of the Log-likelihood is not regular. Right: The distribution of the Log-likelihood Ratio Statistic L_h conforms to the chi-square 1 distribution when the position of the change is known.

$L_h = W$ on the right side (red line) and the rainfall proportion on the left side (blue line). Both are plotted against the Year and the vertical line indicates the position of change in 1973.

However Hinkley (1970) tested hypotheses directly on the change-point (and also investigated the differences between θ_0 and θ_1 when both are unknown and when one of them is known) instead of checking for the existence of the change-point in the first instance. Applying the Hinkley model to the Turbidity concentrations from Asejire (also Eleyele for comparison) with an underlying assumption of normality gives the estimates in Table 2.1. The year of change is estimated to be 2007. It is interesting that the data collection agencies witness a lot of changes in the year 2007. Government and personnel changes took place in 2007 and also the units of measurements are not standardised and the agency lack measurement facilities and qualified personnel. Suspicous values of concentration are also mixed up in records and the sample size is small for change detection. The Hinkley model suggests that there might be more changes in the series taking place between year 2006 and year 2007 rather than year 2005 suggested in Figure 2.5. For both reservoirs, the points of change are close.

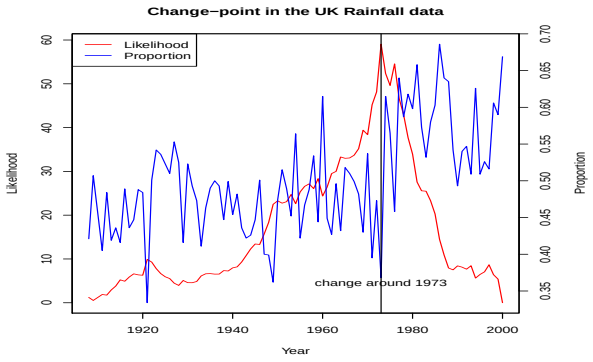

Figure 2.4: Fitted Model of change-point Using Hinkley (1970) method indicate change at point 65 around 1973: Proportion of Rainfall (blue line) and Likelihood Ratio Statistic: W(red-line) on different axes. The vertical lines shows the position of change in 1973.

Hinkley	δ	θ_0	θ_1	σ^2	$l_p(\delta)$	W	Year
Asejire	51	2.562	3.36	0.563	-12.77	7.97	2007
Eleyele	48	9.2	4.3	27.16	-129	7.77	2006

Table 2.1: Hinkley Model Application to Turbidity Concentration from Asejire and Eleyele Reservoirs. The points of change in the reservoirs are close.

Moreover, suppose that θ_0 and θ_1 are the means before and after δ where $l_p(\delta)$ is the profile log-likelihood function of the change-point. The Hinkley normal model can be fitted to the rainfall data set. Table 2.2 shows the same fit and hence the Hinkley model is appropriate for the rainfall data as the model detects the position of change at $\delta = 65$.

Model	δ	θ_0	θ_1	$l_p(\delta)$	Year of Change
Hinkley	65	0.47	0.58	223.59	1973

Table 2.2: Estimates of: Change-point(δ), Mean before the change(θ_0), Mean after the change(θ_1) and the Profile log likelihood for the change point($l_p(\delta)$) using Hinkley methods.

Important works on the asymptotic theory for inference of the change-point based on sampling theory can be found in Hinkley (1970, 1971, 1972). Also a similar example of change in parameter (proportion) of a binomial variable has also been investigated by Hinkley and Hinkley

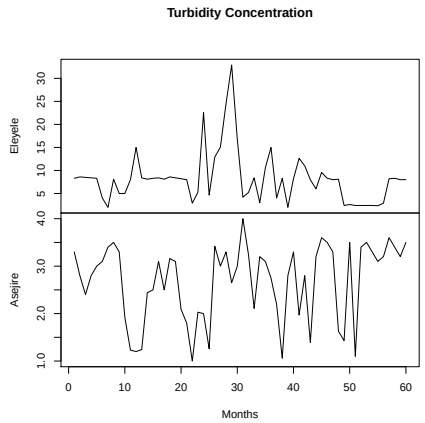

Figure 2.5: Eleyele shows significantly higher concentrations of Turbidity(NTU) and both reservoirs show changes close to the 30th month corresponding to June, 2005.

(1970). In this paper, the maximum likelihood estimate of the change-point was obtained by maximising the function R_q below as

$$R_q = \sum_{i=1}^{q} \left\{ X_i(\ln(\frac{\theta_0}{\theta_1})) + (1 - X_i)\ln(\frac{1-\theta_0}{1-\theta_1}) \right\}$$

where $q = 1, ..., n-1$ and

$$P(X_i = 1) = \begin{cases} \theta_0 & \text{for } i = 1, ..., \delta \\ \theta_1 & \text{for } i = \delta + 1, ..., n. \end{cases}$$

The estimate of δ, δ_m say is the value of q that maximises the sequence R_q. It is noted that here θ_0 and θ_1 both refer to proportions before and after δ. Moreover, changes in means in a sequence of normal random variables are also considered by Severini (2006) where the sequence of independent normally distributed random variables $X_1, ..., X_n$ is considered such that $X_1, ..., X_\delta$ have mean μ and variance σ^2 and $X_{\delta+1}, ..., X_n$ have mean $\mu + \beta$ and variance σ^2 where $\mu, \beta,\ \delta$ are all unknown parameters for $\beta \neq 0$ and $\delta \in \{1, 2, ..., n\}$. Severini (2006) therefore computed the likelihood ratio test statistic to test the hypotheses $H_0 : \delta = n$ against $H_1 : \delta < n$. This is basically the same model discussed by Hinkley (1970) apart from the different parameterisation.

2.7 Summary of Chapter Two

In this Chapter we discussed the regular asymptotic likelihood theory with connections to regular assumptions. We also discussed the likelihood asymptotics of change-point problems with emphasis on standard likelihood asymptotics. The score function is used to show the sampling variation for different models and simple examples of two-mean models are considered. Hypotheses are formulated to test the existence of change-points and simulation is used to study the distribution of the profile log-likelihood statistic function of the change-point model in Hinkley (1970) therefore providing links to the failure of **Assumption 1** and **Assumption 4** establishing the change-point problem. The main issue established in this Chapter is the failure of regular assumptions possed by change-point problems justifying why change-point problems are difficult. In Chapter 3, we will discuss the off-line methods for change-point problems . This will consider the change-point problem under different models in the literature with little or no facility for information update.

Chapter 3

Off-line Procedures

In this Chapter, we shall consider off-line methods of change-point detection in time series. The main feature of these methods is that they require processing of all available observations each time we need to detect a change-point. They can be applied many times to the same data-sets to detect multiple change-points. They can be applied to change detection in time series, regression, non-parametric models and so on.

3.1 Change in Mean levels: Bayes Factor Applications

In the previous section, we explained a simple change-point model involving changes in the mean level of independent observations. We now present an alternative strategy for the analysis of such models using Bayes factors which are very useful in comparing hypotheses and statistical models (Jim, 2008). We give a short description of the method of Bayes factors and thereafter provide its application in testing for a change-point. Suppose we have two hypothesis H_0 and H_1, for $i = 0, 1$, the prior probabilities for models under each of the hypotheses have posterior probabilities

$$Pr(H_i|x) = \frac{Pr(x|H_i)Pr(H_i)}{Pr(x|H_0)Pr(H_0) + Pr(x|H_1)Pr(H_1)}.$$

In this case, as against the frequentist approaches, the alternative hypothesis H_1 is restricted to be fully specified in a Bayesian setting. However for many applications such as testing hypothesis on the existence of change-point, it is valuable to use the odds in favour of H_1

(Davison, 2003). The odds in favour of H_1 can be written as

$$\frac{Pr(H_1|x)}{Pr(H_0|x)} = \frac{Pr(x|H_1)}{Pr(x|H_0)} \times \frac{Pr(H_1)}{Pr(H_0)}.$$

The change in prior to posterior odds for H_1 relative to H_0 will depend on data through the Bayes factor

$$B_{10} = \frac{Pr(x|H_1)}{Pr(x|H_0)}.$$

Davison (2003) further noted the simplest case when both H_0 and H_1 are simple hypotheses. He indicated that the hypotheses involve the mean levels θ_0 and θ_1 and computed for $i = 0, 1$, $P(x|H_i) = \int f(x|H_i, \theta_i)\pi(\theta_i|H_i)d\theta_i$ where $\pi(\theta_i|H_I)$ is the prior for θ_i under H_i. This integral is a combination of the likelihood function and the prior density. However interpreting Bayes factor B_{10} can be done in favour of H_0 over H_1 noting that computing B_{10} is the same as computing B_{01}^{-1}. Therefore $-2\ln B_{10}$ provides us with evidence against H_1. Analogously with the likelihood ratio statistics, $2\ln B_{10}$ is often used to summarize the evidence for H_1 compared to H_0. Some rough interpretations as given by Davison (2003) are shown in Table 3.1. This interpretation is done in contrast to that of the likelihood ratio statistic having a null χ^2 distribution for nested models depending on the difference in their degrees of freedom.

B_{10}	$2\ln B_{10}$	Evidence in support of H_1
$1-3$	$0-2$	Hardly worth mentioning
$3-20$	$2-6$	Positive
$20-150$	$6-10$	Strong
>150	>10	Very strong

Table 3.1: Rough interpretation of Bayes factor B_{10} in favour of H_1 over H_0 given by Davison (2003).

In this Section, we shall compute Bayes factors for the existence of a change-point in the mean levels of Poisson, Normal and Exponential models. In the context of the two-mean model, suppose that the hypotheses to be tested for example are $H_0 : \theta_1 = \theta_2 = \theta$ against H_δ of θ changing after a fixed time δ. If $x_1, ..., x_n$ are independent realisations of $\mathbf{X}$ under H_δ, then we can derive the Bayes factor for a change-point at time δ as the ratio $\frac{Pr(x|H_\delta)}{Pr(x|H_0)}$. Details on Bayesian analysis of change-point can be seen in Gelfand et al. (1990) and Rasmussen (2001).

3.1.1 Normal Model

Assuming we have normal realisations we may consider the computations of Bayes factors for a Normal change-point model. Under H_δ we may easily consider a situation where we suspect a fixed change-point δ that the series possess different means and variances. Considering some data vector $\mathbf{x}$ of length T being an independent and identical realisation of normal random variables we might suspect a change-point δ such that

$$x_i \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{for } 1 \leq i \leq \delta \\ N(\mu_2, \sigma_2^2) & \text{for } \delta < i \leq T. \end{cases}$$

Letting $Q_n(\mathbf{x}) = \int \pi(\theta)\pi_n(x|\theta)d\theta$, a combination of the prior density $\pi(\theta)$ and the likelihood function $\pi_n(x|\theta)$ denote the predictive posterior density for a sequence $\mathbf{x}$ of n independent observations from a normal distribution with parameters $\theta = (\mu, \sigma^2)'$. Then given a time series $\mathbf{x} = (x_1, ..., x_T)$ we can compute the Bayes factor for a change-point at δ as

$$B_F = Q_\delta(\mathbf{x}_{1:\delta})Q_{T-\delta}(\mathbf{x}_{\delta+1:T})/Q_T(\mathbf{x}). \tag{3.1}$$

Equation (3.1) is the product of the components for the two-parts of the series divided by the Bayes factor component for the full series. It should be noted here that we can only factorise the numerator of Equation (3.1) this way when both the mean and variance change. We therefore proceed to find a generalising Bayes factor for Normal change-point model. Finding a general expression for $Q_n(\mathbf{x})$ we write $\psi = \sigma^2$. For mathematical convenience Gelman et al. (2000) used the scaled Inverse Chi-square distribution for $\pi(\psi)$ and following their lead, we take π as scaled Inv-χ^2 as in Inv-$\chi^2(v, \sigma_0^2)$ and we also take $\pi(\mu|\psi) \sim N(\mu_0, \psi/k)$. Therefore we have

$$\pi(\theta) = \pi(\psi)\pi(\mu|\psi) = \frac{(v/2)^{v/2}}{\Gamma(v/2)}\sigma_0^v\psi^{-(v/2+1)}e^{-v\sigma_0^2/2\psi} \times \frac{\sqrt{k}}{\sigma\sqrt{2\pi}}e^{\frac{-k(\mu-\mu_0)^2}{2\psi}}$$

and then

$$\pi(\theta) = A\psi^{-(v/2+1)}\frac{e^{-\frac{[v\sigma_0^2+k(\mu-\mu_0)^2]}{2\psi}}}{\sigma}.$$

where $A = \frac{(v/2)^{v/2}\sigma_0^v\sqrt{k}}{\Gamma(v/2)\sqrt{2\pi}}$. The likelihood function for μ and ψ based on n independent observations is

$$\pi_n(\mathbf{x}|\theta) = (2\pi\psi)^{-n/2} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\psi}} = (2\pi\psi)^{-n/2} e^{-\frac{1}{2\psi}((n-1)s^2+n(\bar{x}-\mu)^2)},$$

where $\bar{x}$ and s^2 are the sample mean and variance respectively. On this basis to compute the components in Equation (3.1), we can write the integrand $\pi(\theta)\pi_n(\mathbf{x}|\theta)$ as

$$\pi(\theta)\pi_n(\mathbf{x}|\theta) = (2\pi)^{-n/2} A\psi^{-(v+n+3)/2} \times e^{-\frac{1}{2\psi}[v\sigma_0^2+k(\mu-\mu_0)^2+(n-1)s^2+n(\bar{x}-\mu)^2]}.$$

This integrand can be rewritten as

$$(2\pi)^{-n/2} A\psi^{-(v+n+3)/2} \times e^{-\frac{1}{2\psi}[(n-1)s^2+v\sigma_0^2)]} \times e^{-\frac{1}{2\psi}[k(\mu-\mu_0)^2+n(\bar{x}-\mu)^2]}.$$

After some rearrangement and completing the square, the exponent in the final term here can be written as

$$-\frac{k+n}{2\psi}[(\mu-\eta)^2 + \lambda(1-\lambda)(\bar{x}-\mu_0)^2]$$

where $\lambda = k/(k+n)$ and $\eta = \lambda\mu_0 + (1-\lambda)\bar{x}$. We thus write

$$Q_n(\mathbf{x}) = (2\pi)^{-n/2} A \int_{\psi} \psi^{(v+3+n)/2} e^{-\frac{1}{2\psi}[(n-1)s^2+v\sigma_0^2+(k+n)\lambda(1-\lambda)(\bar{x}-\mu_0)^2]} \times \int_{\mu} e^{[-\frac{k+n}{2\psi}(\mu-\eta)^2]} d\mu d\psi.$$

Clearly, the μ-integrand here is a scaled $N(\eta, \psi/(k+n))$ density therefore μ-integrand is equal to $\sqrt{2\pi\psi/(k+n)}$. Thus

$$Q_n(\mathbf{x}) = (2\pi)^{-(n-1)/2}(k+n)^{-1/2} A \int_{\psi} \psi^{-(\frac{v+n}{2}+1)} e^{-\frac{1}{2\psi}[(n-1)s^2+v\sigma_0^2+(k+n)\lambda(1-\lambda)(\bar{x}-\mu_0)^2]} d\psi.$$

or equivalently

$$Q_n(\mathbf{x}) = (2\pi)^{-(n-1)/2}(k+n)^{-1/2} A \int_{\psi} \psi^{-(\frac{v+n}{2}+1)} e^{-B_n(\mathbf{x})(v+n)/2\psi} d\psi$$

where $B_n(\mathbf{x}) = \frac{(n-1)s^2+v\sigma_0^2+(k+n)\lambda(1-\lambda)(\bar{x}-\mu_0)^2}{v+n}$. This is a scaled density of Inv-$\chi^2(v+n, B_n(\mathbf{x}))$. Hence the ψ-integral is

$$\frac{\Gamma(\frac{v+n}{2})}{(\frac{v+n}{2})^{(v+n)/2}[B_n(\mathbf{x})]^{(v+n)}},$$

and finally we obtain

$$Q_n(\mathbf{x}) = \frac{(2\pi)^{-(n-1)/2}(k+n)^{-1/2}A\Gamma((v+n)/2)}{(\frac{v+n}{2})^{(v+n)/2}[B_n(\mathbf{x})]^{(v+n)}}. \tag{3.2}$$

Further variations to Equation (3.1) can be investigated for example when the two parts of the series $\mathbf{x}_{1:\delta}$ and $\mathbf{x}_{(\delta+1):T}$ have different means μ_1 and μ_2 but the same variance σ^2. The model here is such that

$$x_i \sim \begin{cases} N(\mu_1, \sigma^2) & \text{for } 1 \leq i \leq \delta \\ N(\mu_2, \sigma^2) & \text{for } \delta < i \leq T. \end{cases}$$

In this case, following a similar setting and assuming the same prior distribution for the variance, the prior distribution for which $\theta = (\mu_1, \mu_2, \sigma^2)'$ then is

$$\pi(\theta) = \pi(\sigma^2)\pi(\mu_1|\sigma^2)\pi(\mu_2|\sigma^2).$$

We suppose that $\pi(\mu_1|\psi) \sim N(\mu_0, \psi/k)$ and that $\pi(\mu_2|\psi) \sim N(\mu_0^*, \psi/k)$. For a similar result as in Equation (3.2) we can obtain the prior distribution as

$$\pi(\theta) = \frac{(v/2)^{v/2}}{\Gamma(v/2)}\sigma_0^v\frac{k}{2\pi}\psi^{-(\frac{v+4}{2})}e^{-\frac{1}{2\psi}\left[k[(\mu_1-\mu_0)^2+(\mu_2-\mu_0^*)^2]+v\sigma_0^2\right]}.$$

Then by writing $A_2 = \frac{(v/2)^{v/2}\sigma_0^v k}{\Gamma(v/2)2\pi}$ we can write $\pi(\theta)$ in this case as

$$\pi(\theta) = A_2\psi^{-(v+4)/2} \times e^{-\frac{\left[v\sigma_0^2+k[(\mu_1-\mu_0)^2+(\mu_2-\mu_0^*)^2]\right]}{2\psi}}.$$

The complete likelihood to be combined with this prior is therefore

$$(2\pi\psi)^{-n/2}e^{-\left[\frac{\sum_{i=1}^{\delta}(x_i-\mu_1)^2+\sum_{i=\delta+1}^{n}(x_i-\mu_2)^2}{2\psi}\right]}.$$

Using $\bar{x}_1$, $\bar{x}_2$, s_δ^2, s_a^2 as the sample means and variances before and after the change δ we can rewrite the likelihood and combine it with the stated prior as in

$$(2\pi)^{-n/2}A_2\psi^{-(v+4+n)/2}e^{-\frac{\left[v\sigma_0^2+k[(\mu_1-\mu_0)^2+(\mu_2-\mu_0^*)^2]\right]}{2\psi}} \times e^{\frac{-\left[(\delta-1)s_\delta^2+\delta(\bar{x}_1-\mu_1)^2+(n-\delta-1)s_a^2+(n-\delta)(\bar{x}_2-\mu_1)^2\right]}{2\psi}}.$$

Re-arranging carefully and integrating with respect to μ_1, μ_2 and ψ we have

$\lambda_1 = k/(k+\delta)$, $\lambda_2 = k/(k+(n-\delta))$,$\eta_1 = \lambda_1\mu_0 + (1-\lambda_1)\bar{x}_1$ and $\eta_2 = \lambda_2\mu_0^* + (1-\lambda_2)\bar{x}_2$ and then the final component is

$$A_2(2\pi)^{-(n-2)/2}(k+\delta)^{-\frac{1}{2}}(k+n-\delta)^{-\frac{1}{2}}\int_{\psi}\psi^{-((v+n)/2+1)}e^{(B(v+n))/2\psi}d\psi$$

where $B = \frac{(\delta-1)s_{\delta}^2+(n-\delta-1)s_a^2+v\sigma_0^2+(k+\delta)\lambda_1(1-\lambda_1)(\bar{x}_1-\mu_0)^2+(k+n-\delta)\lambda_2(1-\lambda_2)(\bar{x}_2-\mu_0^*)^2}{v+n}$. We can therefore compute the numerator of B_F in Equation 3.1 as

$$\frac{(2\pi)^{-(n-2)/2}(k+\delta)^{-\frac{1}{2}}(k+n-\delta)^{-\frac{1}{2}}A_2\Gamma(v+n)/2}{(\frac{v+n}{2})^{(v+n)/2}B^{v+n}}. \tag{3.3}$$

These results can be implemented the same way as done with Equation (3.2) given an additional value of μ_0^*.

3.1.2 Normal Model: Application to Rainfall Data

To check how Equation (3.2) and Equation (3.3) work we use the rainfall data set to compute Bayes factors in each case. In the first case where the series experience a change in mean along with a change in variance we shall compute $2\ln B_F$ using Equation (3.2) with reference to Equation (3.1). Also for the second case where we expect change in mean levels but same variance we simply use the numerator of the Bayes factors B_F obtained in Equation (3.3) to compute $2\ln B_F$.

An important issue in computing Bayes factors involves the selection of an appropriate prior distribution for the parameter of interest. It is therefore necessary that we rightly choose the hyper-parameters of the prior distribution for example to reflect our prior beliefs about the parameter θ say. The rainfall observations are proportions whose distributions are being approximated as normal. It may seem reasonable to set the prior mean μ_0 as centred on 0.5 and with a variance that is as large as possible subject to the mass being concentrated in (0,1); and the prior for ψ (variance) to give an expectation of roughly 1/365 (the variance would be $1/(4 \times 365)$ if the daily rainfall occurrences were independent and identically distributed Ber(1/2), but in practice the variance will be inflated by seasonality). Since the expectation of Inv-$\chi^2(v, \sigma_0)$ is $(v \times \sigma_0)/(v-2)$, take $v = 3$ (smallest possible) and $\sigma_0 = 1/(365 \times 3)$. If we are taking the variance of $\pi(\mu|\psi) = \psi/k$ a priori then we want (say) 3 standard deviations of μ to be equal to 0.5. Since ψ is around 1/365, we therefore want something like $3 \times \sqrt{(}1/365 \times k) = 0.5$. Then $k = 365 \times 0.25/9 = 10$. In the first case where both mean and variance change, we

therefore select $\mu_0 = 0.5, v = 3, \sigma_0 = 1/(v \times 365), k = 10$. On the other hand we select an additional value μ_0^* by setting $\mu_0 = 0.5, \mu_0^* = 0.5, v = 3, \sigma_0 = 1/(v \times 365), k = 10$. What is plotted on the vertical axis in Figure 3.1 is $2 \ln B_F$ referred to as Bayes factors. The two cases as indicated by Equation 3.2 and Equation 3.3 show positive values of Bayes factors in favour of H_δ (see Table 3.1). There is therefore very strong evidence of change in the year 1973. This result is very similar to what has been obtained in Figure 2.4 Chapter 2, where scaled likelihood ratio statistic were plotted.

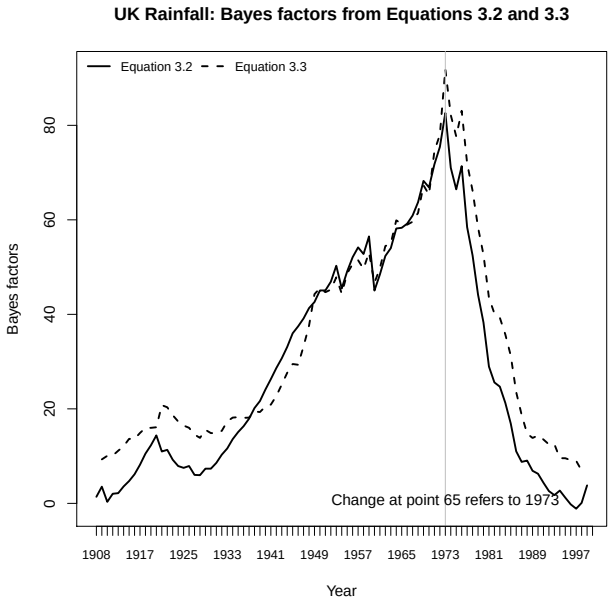

Figure 3.1: Strong evidence in favour of H_δ suggests changes in the year 1973 and vertical line shows the point of change since Bayes factors indicates positive values of $2 \ln B_F$ as shown on Table 3.1.

3.1.3 Poisson Model

In change-point detection, the underlying distribution of the data sets does not have to be normal. Davison (2003) treated change-in-mean levels with the Birmingham HUS counts data with an underlying Poisson distribution and computed Bayes factors to investigate the change. Suppose $x_1, ..., x_n$ are independent realisations of $\mathbf{X}$, then if under the H_δ, θ_1 and θ_2 are mean levels each treated as a variable possessing independent gamma prior densities with means γ/τ

and variance γ/τ^2 we can write

$$Pr(x|H_\delta) = \int_0^\infty \prod_{i=1}^{\delta} \frac{\theta_1^{x_i}}{x_i!} e^{-\theta_1} \times \frac{\tau^\gamma \theta_1^{\gamma-1}}{\Gamma(\gamma)} e^{-\tau\theta_1} d\theta_1 \int_0^\infty \prod_{i=\delta+1}^{n} \frac{\theta_2^{x_i}}{x_i!} e^{-\theta_2} \times \frac{\tau^\gamma \theta_2^{\gamma-1}}{\Gamma(\gamma)} e^{-\tau\theta_1} d\theta_2$$

as the conditional probability of x given the alternative hypothesis. Noting that

$$Pr(x|H_0) = \int_0^\infty \prod_{i=1}^{n} \frac{\theta^{x_i}}{x_i!} e^{-\theta} \times \frac{\tau^\gamma \theta^{\gamma-1}}{\Gamma(\gamma)} e^{-\tau\theta} d\theta,$$

then in a compact form we can write

$$Pr(x|H_\delta) = \frac{\tau^{2\gamma}}{\Gamma(\gamma)^2 \prod_{i=1}^{n} x_i!} \frac{\Gamma(\gamma + s_\delta)\Gamma(\gamma + s_n - s_\delta)}{(\tau + \delta)^{\gamma+S_\delta}(\tau + n - \delta)^{\gamma+s_n-s_\delta}}$$

where as before $s_\delta = x_1 + ... + x_\delta$. Accordingly we can then write the Bayes factor $B_{\delta 0}$ for a change-point in year δ as

$$B_{\delta 0} = \frac{\Gamma(\gamma + s_\delta)\Gamma(\gamma + s_n - s_\delta)\tau^\gamma(\tau + n)^{\gamma+s_n}}{\Gamma(\gamma)\Gamma(\gamma + s_n)(\tau + \gamma)^{\gamma+s_\delta}(\tau + n - \delta)^{\gamma+s_n-s_\delta}}$$

where $\delta = 1, ..., n-1$.

3.1.4 Poisson Model: Application to HUS data

For the HUS data-set from Birmingham, we compute $2\ln B_{\delta 0}$ to test the hypothesis of no change-point using four different priors. Following the lead in Davison (2003) the prior density for θ is chosen so that it has unit mean and the variances $1, 10^2, 10^3, 10^4$ indicating increase in prior uncertainty. The results obtained in Table 3.2 indicate negative values of $2\ln B_{\delta 0}$ corresponding to evidence in favour of H_0. There exists evidence of change in any year from 1976 to 1986 since values obtained in this range are positive values that cannot be ignored as shown in Table 3.1. Figure 3.2 shows $2\ln B_{\delta 0}$ as Bayes factors on the vertical axis using different priors on Table 3.2. They all have same shape indicating the same time of change occurring between 1976 and 1986.

Years	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
HUS Data	1	5	3	2	2	1	0	0	2	1
$2logB_{\delta 0}, \gamma = \tau = 1$	4.9	-0.5	0.6	3.9	7.5	13	24	35	41	51
$2logB_{\delta 0}, \gamma = \tau = 0.01$	-1.3	-5.9	-4.5	-1.0	3.0	9.7	20	32	39	51
$2logB_{\delta 0}, \gamma = \tau = 0.001$	-5.9	-10.4	-9.0	-5.5	-1.5	5.2	15.5	27.7	34.8	46.3
$2logB_{\delta 0}, \gamma = \tau = 0.0001$	-10	-15	-14	-10	-6.1	0.6	11	23	30	42
Years	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989
HUS Data	1	7	11	4	7	10	16	16	9	15
$2logB_{\delta 0}, \gamma = \tau = 1$	63	55	38	42	40	31	11	-2.9	-5.3	0
$2logB_{\delta 0}, \gamma = \tau = 0.01$	64	57	40	47	46	38	18	1.8	1.2	0
$2logB_{\delta 0}, \gamma = \tau = 0.001$	59.4	52.4	36.1	42.2	41.3	33.7	13.4	-2.5	-3.1	0
$2logB_{\delta 0}, \gamma = \tau = 0.0001$	55	48	31	38	37	29	8.8	-7.1	-7.7	0

Table 3.2: Bayes factors computed to detect change-points in Birmingham HUS data. The negative values suggest evidence of change between 1976 and 1986.

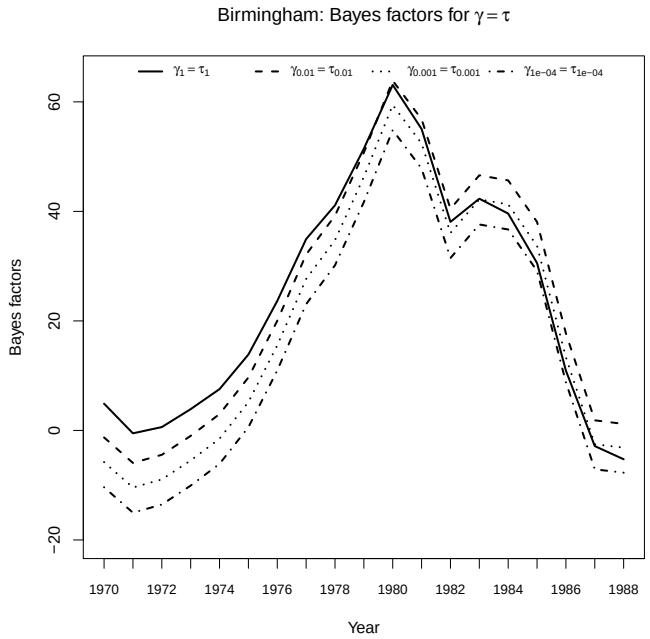

Figure 3.2: Bayes factors $2 \ln B_{\delta 0}$ for different priors. The subscripts on the legend equals the values for which $\gamma = \tau$ in each case as seen on Table 3.2. There is evidence of change between 1976 and 1986.

3.1.5 Exponential Model

Suppose now that in using the Bayes factor to test the hypothesis of change against no change we have $x_1, ..., x_n$ as realisations of other distributions rather than normal or Poisson we compute the Bayes factor using the one parameter exponential density $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ as the underlying distribution of the realisations $x_1, ..., x_n$ describing the random variables $X_1, ..., X_n$. We obtain the Bayes factor assuming independent gamma prior densities for λ_1 and λ_2 say. If the parameters of the gamma prior has mean γ/τ and variance γ/τ^2 then we write

$$Pr(x|H_\delta) = \int_0^\infty \prod_{i=1}^{\delta} \lambda_1 e^{-\lambda_1(x_i)} \times \frac{\tau^\gamma \lambda_1^{\gamma-1} e^{-\tau\lambda_1}}{\Gamma(\gamma)} d\lambda_1 \int_0^\infty \prod_{i=\delta+1}^{n} \lambda_2 e^{-\lambda_2(x_i)} \times \frac{\tau^\gamma \lambda_2^{\gamma-1} e^{-\tau\lambda_2} d\lambda_2}{\Gamma(\gamma)} d\lambda_2.$$

The λ_i integral can be written as

$$f_p = \frac{\tau^\gamma}{\Gamma(\gamma)} \int_0^\infty \lambda_1^{\delta+\gamma-1} e^{-\lambda_1(\sum_{i=1}^{\delta} x_i + \tau)} d\lambda_1.$$

Using a change of variable and recognising the appropriate gamma function we set $k = \lambda_1(s_\delta + \tau)$ where $s_\delta = \sum_{i=1}^{\delta} x_i$ to obtain

$$f_p = \frac{\tau^\gamma}{\Gamma(\gamma)} \frac{\Gamma(\delta+\gamma)}{(s_\delta+\tau)^{\delta+\gamma}}.$$

Then we can write for the full series

$$Pr(x|H_\delta) = \frac{\tau^\gamma}{\Gamma(\gamma)} \frac{\Gamma(\delta+\gamma)}{(s_\delta+\tau)^{\delta+\gamma}} \times \frac{\tau^\gamma}{\Gamma(\gamma)} \frac{\Gamma(n-\delta+\gamma)}{(s_n-s_\delta+\tau)^{n-\delta+\gamma}} \tag{3.4}$$

Noting that under the null hypothesis of no change we have $Pr(x|H_0) = \frac{\tau^\gamma}{\Gamma(\gamma)} \frac{\Gamma(n+\gamma)}{(s_n+\tau)^{n+\gamma}}$ so that we obtain the Bayes factor for testing a change-point in period δ to be

$$B_{\delta O} = \frac{\tau^\gamma}{\Gamma(\gamma)} \frac{\Gamma(\gamma+\delta)\Gamma(\gamma+n-\delta)(s_n+\tau)^{\gamma+n}}{\Gamma(n+\gamma)(\tau+s_\delta)^{\delta+\gamma}(s_n-s_\delta+\tau)^{n-\delta+\gamma}.} \tag{3.5}$$

Although the Bayes factor derivation for the exponential distribution could not be appropriate for detecting changes in the available data, it could be applied to several other environmental mechanisms.

3.2 Full Conditional Distribution for δ using Gibbs sampling.

In Section 3.1, we considered Bayes factors which involves posterior densities which may end up becoming a difficult expression such that we may not be able to integrate out some or all of the variables. We therefore need a way of understanding posterior densities which may not rely on being able to analytically integrate the posterior density. Wilkinson (2006) stated that given any complex non-standard multivariate probability distribution, we need ways to understand it so as to obtain measures such as moments and conditional marginal distributions. A well known Markov chain monte-carlo (MCMC) method such as the Gibbs sampling provides a possible solution. The Gibbs sampling procedure provides a way of simulating from multivariate distributions based on the ability to simulate from conditional distributions. It is more appropriate when sampling from marginal distributions seems difficult or impossible. Moreover, there also exist cases when it is difficult to simulate from the full conditional distribution. In this case the Gibbs sampler is combined with another MCMC method called the Metropolis-Hastings algorithm (Wilkinson, 2006). Praha and Praha (2008) combined both the Gibbs sampler and the Metropolis-Hastings algorithm to estimate change-points in temperature data using three models. The first model assumes a normal distribution with the same variance for the series before and after δ, the second model assumes a two-phase linear model with a jump at δ and the third assumes a two-phase linear model with a gradual change. The posterior distribution of the change-point and other parameters are found using the combination of the Gibbs sampler and the Metropolis Hastings algorithm.

In this section, we will consider the two-mean models using Gibbs sampling to obtain full conditional distributions for parameters. Assuming that δ is unknown, Davison (2003) suggested an example using Bayesian analysis for independent observations. It is assumed that we have consecutive independent normal random variables with variance σ^2 and means λ_1 for $i = 1, ..., \delta$ and λ_2 for $i = \delta + 1, ...n$. If we take uniform prior densities for λ_1 and λ_2 and that $\sigma^2 \sim IG(a, b)$ for specified a and b are the parameters of an inverse-gamma distribution such that δ is assumed uniform on $\{1, ..., n-1\}$ a priori, then the posterior density of the parameters can be represented as

$$\pi(\lambda_1, \lambda_2, \sigma^2, \delta|\mathbf{x}) \propto (\sigma^2)^{-n/2-a-1} \times e^{-\left[\frac{\sum_{i=1}^{\delta}(x_i-\lambda_1)^2+\sum_{i=\delta+1}^{n}(x_i-\lambda_2)^2}{2\sigma^2}+\frac{b}{\sigma^2}\right]} \tag{3.6}$$

and then for each parameter we can simply derive the full conditional distribution from Equation 3.6 as follows

$$\pi(\lambda_1|\mathbf{x}, \lambda_2, \sigma^2, \delta) \propto e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{\delta}(x_i-\lambda_1)^2} \propto e^{-\frac{1}{2\sigma^2}[\delta(\lambda_1-\bar{x}_1)^2]}$$

where $\delta = 1, ..., n-1$, $\delta\bar{x}_1 = \sum_{i=1}^{\delta} x_i$, $(n-\delta)\bar{x}_2 = \sum_{i=\delta+1}^{n} x_i$. Then we can write $\lambda_1|\mathbf{x}, \lambda_2, \sigma^2, \delta \sim N(\bar{x}_1, \sigma^2/\delta)$ and also $\lambda_2|\mathbf{x}, \lambda_1, \sigma^2, \delta \sim N(\bar{x}_2, \sigma^2/(n-\delta))$. From Equation 3.6 we derive the full conditional distribution of σ^2 in the following way. Start by writing

$$\pi(\sigma^2|\mathbf{x}, \lambda_1, \lambda_2, \delta) \propto (\sigma^2)^{-n/2-a-1} e^{-b/\sigma^2} e^{-1/2\sigma^2\sum_{i=1}^{\delta}(x_i-\lambda_1)^2} e^{-1/2\sigma^2\sum_{i=\delta+1}^{n}(x_i-\lambda_2)^2}$$

?? which can be re-arranged as

$$\pi(\sigma^2|\mathbf{x}, \lambda_1, \lambda_2, \delta) \propto \{\frac{1}{\sigma^2}\}^{a+n/2+1} e^{-\frac{1}{\sigma^2}[b+\frac{1}{2}\sum_{i=1}^{\delta}(x_i-\lambda_1)^2+\frac{1}{2}\sum_{i=\delta+1}^{n}(x_i-\lambda_1)^2]}$$

indicating that

$$\sigma^2|\mathbf{x}, \lambda_1, \lambda_2, \delta \sim IG(a+n/2, b+1/2[\sum_{i=1}^{\delta}(x_i-\lambda_1)^2 + \sum_{i=\delta+1}^{n}(x_i-\lambda_2)^2]).$$

Finally, we have

$$P(\delta = k|\lambda_1, \lambda_2, \sigma^2, \mathbf{x}) \propto e^{-\frac{1}{2\sigma^2}[\sum_{i=1}^{\delta}(x_i-\lambda_1)^2+\sum_{i=\delta+1}^{n}(x_i-\lambda_2)^2]}.$$

The Gibbs sampler will then simulate from each of these conditional distributions. The normal procedure is that a starting value is chosen and the sampler iteratively updates by sampling from each parameter in turn from its conditional distribution given the current values of the remaining parameters. An important issue in running the Gibbs sampler is the choice of initial values for the parameters. Wilkinson (2006) indicated that for a posterior distribution $\pi(\theta|x)$ say, simulation is possible from $\pi(\theta|x)$ by first initialising the sampler somewhere in the support of θ and then running the Gibbs sampler. Afterwards the resulting chain could be observed for convergence after which the 'burn-in' period is discarded for further analysis. We can then see after convergence that the simulated values are approximately from $\pi(\theta)$. Wilkinson (2006)

indicated further that the sampler can be initialised anywhere in the half-plane where the posterior and the prior has support. However convergence may be faster if the chain is not started in the tails of the distribution. He suggested the possibility of starting the sampler near the posterior mode or at a point simulated from the prior distribution or better still at the mean of the prior distribution.

3.3 Gibbs sampling: Application to rainfall data-sets

To illustrate the use of the Gibbs sampler in this context, it is now used to fit the two-mean normal model to the rainfall data discussed in Section 1.2. The data-sets as earlier indicated represent proportions of wet days. Assuming a normal distribution, we now apply the full conditional distribution derived in Section 3.2. As indicated in the last section suppose we specify that values $a = 2$ and $b = 0.06$ then the initial values for λ_1, λ_2 σ^2 and δ can be obtained using the means of their prior distributions which gives values $0.5, 0.5, 0.06, 46$ for λ_1, λ_2, σ^2 and δ respectively. Running Gibbs sampler and repeating the process for R iterations where R is $500, 1000, 2000, 5000$ and 10000, we find the results in Table 3.3 and in Figure 3.3. Table 3.3 shows for each set of iterations, the means, variances and the 95% credible intervals for δ. Figure 3.3 shows the Gibbs sampler for all parameters and noticeably column 4 shows concentration on point 65 (In each row the horizontal axis refers to the number of iterations performed while the vertical axis refers to the sampled values). There is overwhelming evidence that indicates changes in the early 1970s since all the credible intervals include points 65, 66, 67 and 68. This result supports the change in the early 1970s noted in Yang et al. (2006). Yang et al. (2006) indicated that some substantial change in the resolution of the rainfall recordings were observed by a careful check on the data. They noted that the resolution was 0.3mm before the 1970s and that in the early 1970s the resolution was improved to 0.1mm. The result therefore suggests that there are some changes reacting to the change in measurement resolution in the early 1970s.

We investigate convergence by running 11000 iterations on the Gibbs sampler for each parameter starting at five different initial points spanning the range of the priors shown in Table 3.4. Provided we discard the first 1000 iteration as burn-in in each case for each parameter all the 10000 iterations are represented with green, blue, red, yellow and black lines respectively in Figure 3.4 (In each row the horizontal axis refers to the number of iterations performed while

the vertical axis refers to the sampled values). We can see that there exist some fairly general overlaps suggesting general convergence for all parameters.

Iterations(R)	**Mean**(δ)	**Variance**(δ)	**Credible Interval** for δ
R=500	65.4960	11.5531	(64,68)
R=1000	65.0870	29.3748	(64,68)
R=2000	65.4915	8.6224	(65,68)
R=5000	65.5516	1.7689	(65,68)
R=10000	65.5589	2.6958	(65,68)

Table 3.3: UK rainfall Data: Gibbs Sampling Iterations for $R = 500, 1000, 2000, 5000$ and 10000 with means, variances and credible intervals for δ. The initial starting values used for the parameters are $\lambda_1 = 0.5, \lambda_2 = 0.5, \sigma^2 = 0.06$ and $\delta = 46$.

Initial Values	λ_1	λ_2	σ^2	δ
First:10000 iterations(Green line)	0.5	0.5	0.06	46
Second:10000 iterations(Blue line)	0.05	0.95	0.057	44
Third:10000 iterations(Red line)	0.95	0.05	0.003	2
Fourth:10000 iterations(Yellow line)	0.3	0.7	0.03	23
Fifth:10000 iterations(Black line)	0.7	0.3	0.036	28

Table 3.4: Five different initial values spanning the range of the priors for the Gibbs sampler in testing parameter convergence for $\lambda_1, \lambda_2, \sigma^2, \delta$. In each case 10000 iterations are represented by green, blue, red, yellow, and black lines as shown in Figure 3.4.

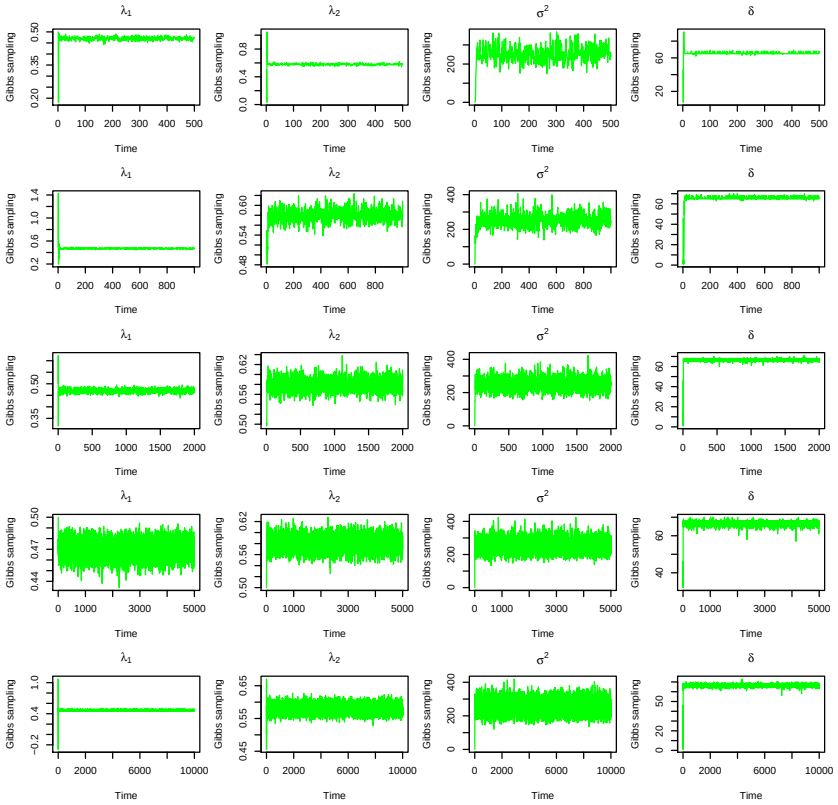

Figure 3.3: UK rainfall data: Matrix plot shows by row respectively Gibbs Sampling for parameters $\lambda_1,\ \lambda_2,\ \sigma^2, \delta$ with initial values $0.5, 0.5, 0.06, 46$ for $\lambda_1,\ \lambda_2,\ \sigma^2$ and δ respectively. Rows 1 to 5 each represents respectively iterations $R = 500, 1000, 2000, 5000$ and 10000. The distribution of δ in column 4 shows clearly the existence of a change-point clustering around point 65 in all cases.

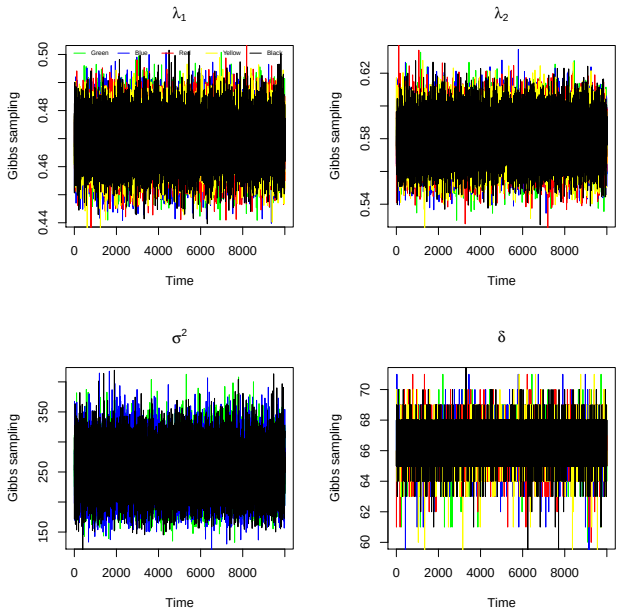

Figure 3.4: Convergence plots for λ_1 and λ_2 (first row) also σ^2 and δ (second row). Starting from five different initial values as observed in Table 3.3. Each plot is obtained from 10000 iterations each represented by green, blue, red, yellow and black lines after a burn-in of 1000 iterated values have been discarded. Each plot shows signs of overlapping suggesting convergence. .

3.4 Change-point in Regression

Another set of models common in the literature estimate change-points in a regression framework. Following Hinkley (1971) the change-point problem was formulated as a two-phase model written in the form

$$Y_i = \begin{cases} \theta + \beta_1(x_i - \delta) + \varepsilon_i & \text{for } i = 1, 2, ..., k \\ \theta + \beta_2(x_i - \delta) + \varepsilon_i & \text{for } i = k+1, ..., n \end{cases}$$

where Y_i is the response variable and ε refer to the random errors in each segment for $x_1 < \cdots < x_k \leq \delta < x_{k+1} < \cdots < x_n$, θ is the mean value and k is the unknown to be estimated. In this paper, maximum likelihood estimation was used to estimate δ, θ, β_1,β_2 and k. Relevant treatment can be found in Hudson (1966), Hinkley (1969a) and Hinkley (1969b). There is also a generalisation of Hudson's method in Williams (1970). Particularly, Hinkley provides approximate large sample confidence intervals for the parameters and also large sample

tests for the hypotheses $\beta_1 = \beta_2$(no change in slope) and $\beta_2 = 0$. Another method for testing $\beta_1 = \beta_2$ is given by Farley and Hinich (1970).

Julious (2001) formulates the two-line model in a slightly different way writing the regression function $f(x_i) = E[Y|X = x_i]$ as

$$f(x_i) = \begin{cases} \alpha_1 + \beta_1 x_i & \text{for } X_0 < x_i \leq \delta \\ \alpha_2 + \beta_2 x_i & \text{for } \delta \leq x_i < X_1 \end{cases}$$

In this paper the design matrix $\mathbf{X}$ for the estimation of the parameters $\beta = (\alpha_1, \beta_1, \beta_2)'$ is written as

$$\mathbf{X} = \begin{pmatrix} 1 & \delta & x_1 - \delta \\ 1 & \delta & x_2 - \delta \\ \vdots & \vdots & \vdots \\ 1 & \delta & x_t - \delta \\ 1 & x_{t+1} & 0 \\ 1 & x_{t+2} & 0 \\ \vdots & \vdots & \vdots \\ 1 & x_T & 0 \end{pmatrix}$$

where the estimate obtained here is the usual $\hat{\beta} = (\mathbf{X'X})^{-\mathbf{1}}\mathbf{X'Y}$ and $\delta = \frac{\alpha_1 - \alpha_2}{\beta_2 - \beta_1}$ indicating a constant slope until a point along the x-axis, δ. The function $f(x)$ is continuous but not differentiable at δ. For the interval (X_0, X_1), each half of the model is treated such that a sum of squares Q needs to be minimised where

$$Q = (\mathbf{Y_1} - \mathbf{X_1}\beta_\mathbf{1})'(\mathbf{Y_1} - \mathbf{X_1}\beta_\mathbf{1}) + (\mathbf{Y_2} - \mathbf{X_2}\beta_\mathbf{2})'(\mathbf{Y_2} - \mathbf{X_2}\beta_\mathbf{2}) \tag{3.7}$$

Q is minimised subject to another linear constraint of the form

$$g(\hat{\beta}_\mathbf{1}', \hat{\beta}_\mathbf{2}') = (\hat{\beta}_\mathbf{1}', \hat{\beta}_\mathbf{2}')q = d \tag{3.8}$$

and here for two straight lines $q = (1, \delta, -1, -\delta)'$ and $g(\hat{\beta}_\mathbf{1}', \hat{\beta}_\mathbf{2}') = \alpha_1 + \hat{\beta}_1\delta - \alpha_2 - \hat{\beta}_2\delta = d$. Note that the implication of the constraint is that we have a continuous regression with no break when $d = 0$ although in the general case $d \neq 0$.

To minimise Equation 3.7 subject to the contraint in Equation 3.8, introduce a Lagrange multiplier k and minimize $Q(\beta) + kg(\beta)$. This leads to the following three equations, the last of which is clearly the constraint

$$-\mathbf{X_1'Y_1} - \mathbf{Y_1'X_1} + (\mathbf{X_1'X_1} + \mathbf{X_1X_1'})\beta_1 + k\delta \tag{3.9}$$

$$-\mathbf{X_2'Y_2} - \mathbf{Y_2'X_2} + (\mathbf{X_2'X_2} + \mathbf{X_2X_2'})\beta_2 + k\delta \tag{3.10}$$

$$\alpha_1 + \hat{\beta}_1\delta - \alpha_2 - \hat{\beta}_2\delta - d = 0. \tag{3.11}$$

The resulting estimates of $\hat{\beta}_\mathbf{1}$ and $\hat{\beta}_\mathbf{2}$ are obtained from

$$\begin{pmatrix} \hat{\beta}_\mathbf{1} \\ \hat{\beta}_\mathbf{2} \end{pmatrix} = \begin{pmatrix} \hat{\beta}^*_\mathbf{1} \\ \hat{\beta}^*_\mathbf{2} \end{pmatrix} + \frac{d-s}{t}(C^{-1}q)$$

where $\beta^{*\prime}_\mathbf{1} = (\mathbf{X_1'X_1})^{-\mathbf{1}}\mathbf{X_1Y_1}$ and $\beta^{*\prime}_\mathbf{2} = (\mathbf{X_2'X_2})^{-\mathbf{1}}\mathbf{X_2Y_2}$ are the unrestricted or unconstrained respectively. The matrix C is derived as

$$\mathbf{C^{-1}} = \begin{pmatrix} C_1^{-1} & 0 \\ 0 & C_2^{-1} \end{pmatrix}$$

having $\mathbf{C_1} = \mathbf{X_1'X_1}$ and $\mathbf{C_2} = \mathbf{X_2'X_2}$, $s = (\beta^{*\prime}_\mathbf{1}, \beta^{*\prime}_\mathbf{2})q$ and $t = q'C^{-1}q$. It is noted that s and t follow the notation of Hudson (1966). Then the least squares estimates in the general case where $(\hat{\beta}'_\mathbf{1}, \hat{\beta}'_\mathbf{2})q = d$ as given in Seber and Wild (1989) is used by setting $d = 0$ for which $(\hat{\beta}'_\mathbf{1}, \hat{\beta}'_\mathbf{2})q = 0$. Estimates for which $\alpha_1 + \hat{\beta}_\mathbf{1}\delta = \alpha_2 + \hat{\beta}_\mathbf{2}\delta$ are therefore derived as

$$\begin{pmatrix} \hat{\beta}_\mathbf{1} \\ \hat{\beta}_\mathbf{2} \end{pmatrix} = \begin{pmatrix} \beta^*_\mathbf{1} \\ \beta^*_\mathbf{2} \end{pmatrix} - \frac{s}{t}(C^{-1}q)$$

which allows the estimation of the parameters in the two-line regression model by using the unconstrained estimates for the slopes. The regression can then be adjusted to ensure that the two-lines meet at a particular point. Julious (2001) proposes an algorithm to overcome non-linearity when δ is unknown. It leads to some numerical optimisation which attempts to select the best model. The protocol suggests that all possible unconstrained models should be fitted

first. Then if two lines meet such that $x_t \leq \delta \leq x_{t+1}$ we compute the total restricted residual sums of squares otherwise we compute the total unrestricted residuals sums of squares. The next step is to check for the minimum residual sums of squares. This algorithm takes the model with the smallest restricted residual sums of squares as the best so as to obtain the change-point. An important contribution in this work is that a non-parametric bootstrapping method was used to overcome the problem posed by the failure of the F-test to conform to its expected parametric form as against the problems in Hinkley (1970). In Hinkley (1970), change-point problem led to highly intractable function leading to the failure of the F distribution.

For the same two-line regression model, Seber and Lee (2003) computed a confidence interval for δ while Chen (1998) used the Schwartz Information Criterion(SIC)to estimate δ. The method of Chen (1998) uses the minimum information principle to select a model which minimises the SIC in the regression framework. The SIC is written as

$$SIC = -2logL(\hat{\theta}) + \lambda log(n) \tag{3.12}$$

where $L(\hat{\theta})$ refers to the maximum likelihood function for the model, λ refers to the number of free parameters of the model and n is the sample size. The decision rule used is to select the model with no change-point (accept the null hypothesis) if $SIC(n) \leq SIC(\delta)$ for all δ. That is we select a model with a change at $\hat{\delta}$ if $SIC(\hat{\delta}) = \min\{SIC(\delta) : 1 < \delta < n-1\} < SIC(n)$ where $\hat{\delta} = 2, ..., n-2$. Enhancing the concepts of single change point problem Chen (1998) further estimated the change point δ in a multivariate sense. He simply repeated and extended the same method described in a determined number of times. This however may involve estimating many parameters and also takes more time when the sample size n is large. It is therefore necessary to develop methods that can simultaneously detect changes. Further details on the applications of change-point in regression models with some fairly different applications can be found in Chia-Shang and White (1992), Bang et al. (2006), Bellera et al. (2008a), Bellera et al. (2008b) and Bellera et al. (2009).

3.4.1 Changepoint in Regression: Application to Ministry of Environment Data

In Figure 1.1 of Chapter 1, we presented Biochemical Oxygen Demand(BOD), Chemical Oxygen Demand(COD) and Nitrates(NITRT) obtained from the Ministry of Environment and Water Resources in Ibadan over fifty months. A critical look at Figure 1.1 gives an impression of some seasonality behaviour for the concentrations.

We now discuss an amendment of the methods in Julious (2001) as applied to data sets from the Ministry of Environment and Water Resources in Ibadan. To estimate and test hypotheses on the existence of a change-point we suppose that each concentration is treated as a response variable Y and months as factors (so that we can investigate possible differences in the levels of the factor. In this case, for each concentration from the ministry we simply write

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{3.13}$$

where α_i refers to a factor occurring at $i = 1, 2, ...I$ levels with $j = 1, ..., J_i$ observations per level and μ and ε refer respectively to the overall mean and the error process respectively. We then suspect a change point δ such that Equation 3.13 becomes

$$Y_{ij} = \begin{cases} \mu^{(0)} + \alpha_{ij}^{(0)} + \varepsilon_{ij}^{(0)} & \text{for } y \leq \delta \\ \mu^{(1)} + \alpha_{ij}^{(1)} + \varepsilon_{ij}^{(1)} & \text{if } y > \delta \end{cases} \tag{3.14}$$

The position of change both before and after δ is indicated in Equation 3.14. The superscripts in brackets (0) and (1) refer to effects before and after δ respectively. Julious (2001) applied bootstrap methods to identify change-points by first obtaining the best fitting two-line model and then sampling with replacement from the set of residuals from the best fitting line (with the smallest residual sum of squares). Using Equation 3.14 and following the lead in Julious (2001) as applied to the COD data and obtaining a 1000 bootstrap distribution of the F-statistic indicates a small p-value of 0.003 at point 39 as evidence for the existence of change in the series. Point 39 supports a change in COD concentration in the year 2006. An advantage in using this computing intensive method is that it overcomes the failure of the F-statistic noted by Hinkley (1970) as the distribution takes its appropriate form after several iterations.

3.5 Change-point in Time Series

We have so far been concerned with attempts to model independent series by a sequence of random variables described by a number of structural parameters which can change abruptly at known and unknown change-points, we now concentrate on time-dependent series. Change-points mostly refer to time as we always plot series in order to look for change-points, trend, seasonality, outliers, extreme values or erroneous values that may cause some abrupt changes in series. Considering the time of change, Chandler and Scott (2011) defined a change-point as a point in time at which the properties of a process change abruptly. This might involve abrupt shifts in means, variances or the autocorrelation structure of the process. This might lead to erroneous implications arising from serious data errors as noted in Yang et al. (2006). Such errors among others might arise from faulty equipment, observer practise and unstandardised measurement resolutions. There is a huge literature on change-point problems in time series both for known and unknown change-point.

Intervention analysis is usually performed when the position of a change-point is known. Commandeur and Koopman (2007) used a local linear trend model with intervention variable δ_t and wrote the response $Y_t = \mu_t + \beta X_t + w\delta_t + \varepsilon_t$ for ε_t is white noise, and estimated the parameters $\theta = (\beta, w, \mu_t)'$. The effects of intervention may lead to recovery taking a bit of time as against instantaneous and permanent effects. Representing these effects with an intervention model, Chandler and Scott (2011) proposes a plausible parametric intervention model (where X_t refers to abundance of some affected species due to a case of an oil-spill intervention)

$$X_t = \begin{cases} \mu + \varepsilon_t & \text{if } t \leq \delta, \\ \mu + \zeta + \xi e^{[-\lambda(t-\delta)]} + \varepsilon_t & \text{if } t > \delta, \end{cases}$$

where ε_t is random white noise, μ is the mean level before intervention at δ. Also ζ is the long-term change in the mean level, ξ is the additional immediate change at time δ and $\lambda > 0$ is the rate of recovery after the intervention at δ. This model can be fitted with non-linear least square methods since it is intrinsically non-linear in some of the parameters.

On the other hand, cases of unknown change-points are more complicated as the standard F test may not be valid as shown in Chapter 2. However Hinkley (1970) derived approximate methods to be used in this case for which the process seems highly intractable and as indicated in

section 3.4, Julious (2001) used modern computing intensive methods to overcome the difficulties noted in Hinkley (1970).

3.5.1 Structural Changes in Time Series

It is important to consider the structure of time-series so that we detect the time of changes. Many papers on structural changes in time-series can be found in the literature. McCulloch and Tsay (1993) used a likelihood based approaches and simulation techniques for detecting changes. These methods are later extended by Ray and Tsay (2002) using autoregressive fractionally integrated moving average models (ARFIMA) to study structural breaks using the Nile River minima data for illustration. Pfaff (2006) considered the model $X_t = \mu + \tau D_t + X_{t-1} + \varepsilon_t$ where $D_t = 1$ when $t = \delta$ (δ assigns the structural break date and it will be indicated as a subscript to indicate the change time) and $D = 0$ otherwise. He then presented a drift equation as $X_t = \tau S_t + X_0 + \sum_{i=1}^{t} \varepsilon_t$ where $S_t = 1$ for $t \geq \delta$, $S_t = 0$ otherwise and τ quantifies both D and S. Relevant works on this include the papers by Perron (1989, 1990); Perron and Vogelsang (1992). Particularly in Perron (1989), three models were used and the structural break has been assumed known. The models respectively are

$$\begin{aligned} X_t &= \mu + d^{\star} D(T_\delta) + X_{t-1} + \varepsilon_t \\ X_t &= \mu_1 + (\mu_2 - \mu_1) DU_t + X_{t-1} + \varepsilon_t \\ X_t &= \mu_1 + d^{\star} D(T_\delta) + (\mu_2 - \mu_1) DU_t + X_{t-1} + \varepsilon_t \end{aligned}$$

where $1 < T_\delta < T$ refers to a priori known break point; $D(T_\delta) = 1$ if $t = T_\delta + 1$ and 0 otherwise and $DU_t = 1$ for $t > T_\delta$ and 0 otherwise. Also $d^{\star}$ is the coefficient of $D(T_\delta)$. The author in the first of the three models above allowed a one-time shift in the levels of the series and in the second model he allowed a change in the rate of growth and combined the effects of the previous two models in the third model. He later used a test statistic that is dependent on fraction of the structural break point with respect to the total sample size T, $W^* = \frac{T_\delta}{T}$ and provided the critical values for the test. Zivot and Andrews (1992) used the same data-set as used in Perron (1989) and found less evidence of rejecting the assumption of a unit root process. The approach chooses the date of the structural shift for that point with the least favorable result for the null

hypothesis of a random walk with drift.

Also Lutkepohi et al. (2004) proposed a procedure for estimating a vectorised error correction model (VECM) for a case of a simple shift. The paper assumes that the $N \times 1$ vector process $\{\mathbf{Y}_t\}$ is generated such that

$$\mathbf{Y}_t = \mu_0 + \mu_1 t + \tau d_{t\delta} + \mathbf{X}_t,$$

for $d_{t\delta}$ is a dummy variable and $d_{t\delta} = 0$ for $t < \delta$ and $d_{t\delta} = 1$ for $t \geq \delta$. Here the assumption is that δ is unknown and it is fixed fraction of the sample size such that $\delta = [T\lambda]$ with $0 < \underline{\lambda} \leq \lambda \leq \overline{\lambda} < 1$ where $\underline{\lambda}$ and $\overline{\lambda}$ are real numbers and $[\cdot]$ refers to integer part and the shift might neither occur at the beginning or the end of the sample. Another assumption here is that the process $\{\mathbf{X_t}\}$ is represented as a vector autoregression $VAR(p)$. Then the estimation of the break point is based on the following equations.

$$\mathbf{Y}_t = \mathbf{v}_0 + \mathbf{v}_1 t + \tau d_{t\delta} + \mathbf{A}_1 \mathbf{Y}_{t-1} + ... + \mathbf{A}_p \mathbf{Y}_{t-p} + \varepsilon_{t\delta},$$

for $t = p+1, ..., T$ and $\mathbf{A}_i$ with $i = 1, ..., p$ assign the $N \times N$ coefficient matrices and ε_t is the N-dimensional error process so that the estimator of δ can be defined as

$$\hat{\delta} = \arg \ min_{\delta \in \mathfrak{T}} \ det(\sum_{t=p+1}^{T} \hat{\varepsilon}_{t\delta} \hat{\varepsilon}'_{t\delta})$$

for $\mathfrak{T} = [T\underline{\lambda}, T\overline{\lambda}]$ and $\hat{\varepsilon}'_{t\delta}$ are least-square errors. Once the least squares estimator $\hat{\delta}$ is estimated the data are adjusted using the relationship $\hat{\mathbf{X}}_t = \mathbf{Y}_t - \hat{\mu}_0 - \hat{\mu}_1 t - \hat{\tau} d_{t\hat{\delta}}$. Instead of using the test in Lutkepohi et al. (2004) a new test $L_r(r) = T \sum_{j=r+1}^{N} \ln(1 + \hat{\lambda}_j)$ with critical values supplied in Trenkler (2003) could be used. The advantage here is that the critical values are extensive and more precise.

3.6 Multiple change-points

There exists situations where the changes occur at many points in the series. In this case it may not be practicable to repeat the previous methods in a number of times especially when the sample size is large. This means we have to think of methods on multiple change detection that seek to estimate multiple changes simultaneously. There are many authors interested

in multiple change-point analysis. Inclan (1993) and Stephen (1994) focused on detection of multiple changes using Bayesian methods. Also in Chib (1998) multiple change-point were estimated using a Bayesian approach by formulating latent discrete state variables indicating the regimes from which a particular observation has been drawn. He used a discrete-state Markov process and estimated the model by Markov chain monte carlo methods (MCMC). His method is useful in that it can fit many complex change-point models. He proposes $X_t = \{x_1, x_2, ..., x_n\}$ time series and that the density of X_t given X_{t-1} depends on a parameter ξ changing at unknown times $\Upsilon_m = \{\delta_1, \delta_2, \delta_3, ...\delta_m\}$ and remain constant otherwise. Also if $\delta_1 > 1$ and $\delta < n$ then the model to capture many changes is

$$\xi_t = \begin{cases} \theta_1 & \text{if } t \leq \delta_1 \\ \theta_2 & \text{if } \delta_1 < t \leq \delta_2 \\ \vdots & \vdots\vdots \\ \theta_m & \text{if } \delta_{m-1} < t \leq \delta_m \\ \theta_{m+1} & \text{if } \delta_m < t \leq n, \end{cases}$$

However the problem here is that we need to estimate the parameter vector $\Theta = \{\theta_1, ...\theta_{m+1}\}$ along with the detection of unknown multiple change-points $\Upsilon_m = \{\delta_1, \delta_2, \delta_3, ...\delta_m\}$. Also we may have to compare models with different numbers of change-points. The approach formulates the change-point problem in terms of a latent discrete state variable that indicates the regime from which a particular observation has been drawn. The state variable is specified to evolve according to a discrete-time discrete-state Markov process with transition probabilities constrained so that the state-variable can either stay at the current value or jump to the next higher value. This process is based on the MCMC method discussed in Chib and Greenberg (1996) and it is valuable since it fits more complex models.

3.7 Nonparametric Methods in Change-point

In non-parametric analysis it is recognised that assumption of a particular parametric distribution may not hold for regular independent observations. These methods are sometimes called distribution free methods. In Bhattacharyya and Johnson (1968) a linear rank statistic with score function was used in a Bayesian framework to estimate shifts at an unknown time-point.

The use of ranks in testing hypothesis of change normally assumes that X_t has unknown distribution functions

$$F_0(x), t = 1, ..., \delta$$

and

$$F_1(x), t = \delta + 1, ..., n.$$

The null hypothesis of no-change corresponds to $\delta = n$ or $F_0(x) = F_1(x)$ and for change $F_0(x) \neq F_1(x)$ where $1 \leq \delta \leq n$. To study change in mean levels under non-parametric methods let W_t represent the common Mann-Whitney-Wilcoxon two sample test statistic for $X_1, ..., X_t$ and $X_{t+1}, ..., X_n$ and μ_t and σ_t^2 are the mean and variance of W_t. Sen and Srivastava (1975) used $S_s = Max_{1 \leq t < n} |(W_t - \mu_t)|/\sigma_t$ while Pettitt (1979) also used $P_t = Max_{1 \leq t < n} |W_t - \mu_t|$ as test statistics for no-change null hypothesis and the rejection is based on large values. Again papers by the same authors Pettitt (1979, 1981) discussed the exact and approximate results for testing the null hypothesis of no change. Pettitt defined for each $\delta = 1, 2, ..., n$, $U_{\delta,n} = \sum_{i=1}^{\delta} \sum_{j=\delta+1}^{n} sgn(X_i - X_j)$ and used Bernoulli random variables to obtain a Kolmogorov-Smirnov two-sample statistic. A follow-up to this can be seen in Schechman (1982). In this paper, a normalised type of the statistic used by Pettitt was used.

Also for a shift at an unknown point, a linear rank statistic similar to the one used by Bhattacharyya and Johnson (1968) was used by Hsieh (1984). Moreover Lombard (1987) used the statistic S_t (for which R is a variable)

$$S_t = \sum_{i=1}^{n} \left[\sum_{i=1}^{\delta} S(R_i) \right]^2$$

for a single abrupt change where $S(R_i)$ for $i = 1, ..., n$ are normalised rank scores and similar rank statistics were also suggested for multiple change model and smooth change model.

Parameter estimation problems are treated among others in Pettitt (1980), Darkhovskii (1980) and Carlstein (1988). Particularly Pettitt (1980) used a Mann-Whitney type statistic to obtain an estimate of δ and Darkhovskii (1976) had earlier used a Mann-Whitney statistic by dividing a sequence of n independent random variables into two samples of sizes δ and $n - \delta$. He then considered the values of δ that minimise or maximise the statistic as the possible

change-points and showed that the estimator is consistent. Similar to the work of Darkhovskii (1976) is that of Carlstein (1988) where strongly consistent estimators were proposed for a single change-point. Many other authors include Eastwood (1993) who used a test statistic that converges in distribution to a corresponding Gaussian function under the assumption of no change in distribution (null hypothesis H_0). He used the test under a single change-point detection framework. Also Brodsky and Darkhovsky (1993) integrated most change-point estimation techniques with focus on non-parametric techniques by investigating segments of homogeneity and non-homogeneity of sample observations. Details of non-parametric analysis of change-point problems can be found in the book by Chen and Gupta (2000). The book considered methods that do not possess heavy reliance on assumptions for the parameters in the model. In addition, Chu (2002) used non-parametric bootstrap methods to determine the frequency distribution of the mean annual cyclones in the Central North Pacific (CNP) from 1966 to 2000. This is based on the application of change-point analysis which shows the existence of shifts in 1982 and 1995. The shifts favour some cyclone incidences and indicate that the tropical cyclones in the Eastern North Pacific possess a better chance to enter CNP. Correct estimation of such shifts is therefore important for safety. This is because tropical cyclones are one of the most dangerous natural disasters that can cause loss of lives and properties around the world.

A report under the World Climate Programme on hydrological data which summarised the contributions of a conference held in the United Kingdom in 1998 is published in WMO (2000). This report indicated that most hydrological series are non-normally distributed and that it makes sense to employ more non-parametric methods. The report discussed rank-based tests, normal scores transformation and resampling methods with emphasis that the series are independent.

In a regression framework, Dempfle and Stute (2002) proposed a nonparametric method to estimate discontinuity in regression. They obtained the optimal rate of convergence n^{-1} under minimal assumptions when no smoothing is needed. Reeves et al. (2006) proposed nonparametric methods to detect undocumented change-points in climate data series. They used the standard normal homogeneity (SNH) test, Wilcoxon's nonparametric test, two-phase regression (TPR) methods, information criteria and inhomogeneity tests. They compared all the methods and showed that TPR and a Bayes criteria method seem to be optimal for most climate time series and submitted that the SNH and its non-parametric variant are best when trend and periodic

effects are diminished by using a homogeneous reference series. Most recently, Hoffman et al. (2010) used a joint-point two-regime regression of a broken line model composed of one regression line and a horizontal ray. The method used does not depend on the assumption of normality, independence and identically distributed error terms. Hoffman et al. (2010) also compared their methods with those of Quandt (1958) in a simulation study showing that the new model performs adequately in the normal error term and is therefore preferable in small samples where error terms are correlated with non constant variance.

3.8 Summary of Chapter Three

In this Chapter, we looked deeper into the change-point literature considering various off-line methods with applications to the available data-sets. Bayes factors for normal, poisson and exponential models were shown for testing hypothesis of change. Using the rainfall data-sets, the Bayes factors assuming a normal model were computed based on cases when the variances of the series before and after δ are different and when the variance is constant. A full Bayesian conditional distribution of change-points is also derived using Gibbs sampling and there seem to be convergence for the change-point parameter estimated to be 65 corresponding to 1973 as indicated in Yang et al. (2006) using the rainfall datasets. Evidence of change is also noted in the COD data from the Ministry of Environment. This is done by following the lead in Julious (2001) to overcome the problem of the failure of the F-statistic noted in Hinkley (1970). The main categories of models discussed are regression and time-series change-point models with special application to structural changes. We also discussed change-point detection in non-parametric models and concluded with discussion on multiple change-point estimation. However, it could be noted that most models discussed could not be applicable to the Ministry of Environment water chemistry data, the bootstrapping method detected a change in the COD data around the year 2006. Another difficulty that can militate against change-point detection in these data is the small sample size worsened by unregulated units of measurements. We therefore require other sophisticated methods that could detect changes in data sets that have the characteristics discussed and can estimate changes as soon as they occur. In the next Chapter, we discuss on-line methods with respect to quality control techniques and state-space modeling.

Chapter 4

On-line Procedures

These methods, also called real-time methods, involve the development of models that allow data to be processed as soon as they are collected. They allow for rapid processing by not requiring a complete reanalysis of the entire data set each time a new observation is collected. Common applications of online methods include sequential analysis and recursive Bayesian methods able to take care of false alarm and wrong-detection rate (Poor and Hadjiliadis, 2009). Further applications of on-line methods can also be found in Lai (1995a) and Moreno et al. (2005) and Mei (2010). In this Chapter we give some brief discussion on quality control methods, and recursive state space modelling for detecting on-line changes in parameter levels.

4.1 Quality Control Methods

Suppose there is a continuous production procedure. The quality of a product is usually monitored and often should ideally be maintained as constant for some time. Unfortunately the quality can drop or deteriorate to an unsatisfying level at a particular time. This may be due to a problem at a stage of the process yielding defective products. Identification of such a change-point is extremely critical so that instant solutions at the right time can be provided. Page (1954, 1955, 1957) considered a normal mean μ_0 changing to μ_1 and it is suggested that the whole process is not in statistical control when $S_n - Min_{1\leq\delta\leq n}S_\delta$ becomes large for $S_\delta = \sum_{i=1}^{\delta}(X_i - \mu^*)$ where $\mu_0 < \mu^* < \mu_1$. Thereafter more explanations on the sequential control methods were discussed by Khan (1978). Since the attempts of the 1950s several models of change-points

have been developed in the framework of statistical quality control. Lai (1995b) conducted a survey of a large variety of sequential detection procedures on fault detection and signal processing using sequential techniques for on-line implementation. Also Lorden and Pollak (2005) considered independent observations with a known parameter value θ_0 when the process seems to be in statistical quality control and the parameter changes when the process is out of control. This is then applied to the problem of detecting changes in the shape parameter of a Gamma distribution in both univariate and multivariate cases. Furthermore, Chicken et al. (2009) noted that many industrial processes are only capable of generating rich and complex data records that could not allow the use of standard quality control methods. A semi-parametric wavelet method is proposed for monitoring changes in sequences of non-linear profiles. Monte Carlo simulation is used in this paper to indicate quick changes in in-control profiles.

4.2 State-Space Modeling: Recursive Bayesian Methods

We now describe state space models as applied to continuous changes in model parameters using recursive Bayesian methods. State space models are general linear models that can capture the structural behaviour of most time series. The real interest is that attempts are made to capture the underlying structure of time series' behaviour as the parameters change (Chandler and Scott, 2011). These models can capture sudden changes in time series and have been used in Engineering since the 1950's. The applications of these models can be traced back to the works of Plackett (1950) and Kalman and Bucy (1961). The latter used filtering methods to estimate parameters while the former used adaptive regression methods. The main objective of the Kalman filter used in state space modelling is to infer values for the state variables from noisy measurements so as to use the estimated values to evaluate the system.

In a recursive context we suppose a two-equation linear model, adopting the model described by Pole et al. (1994) as

$$X_t = F_t'\theta_t + v_t \tag{4.1}$$

$$\theta_t = G_t\theta_{t-1} + w_t. \tag{4.2}$$

where $\theta_0 \sim N(m_0, C_0)$, $v_t \sim N(0, V_t)$ and $w_t \sim N(0, W_t)$ and we describe F_t' and G_t below.

Values of the state at time t are in a column matrix θ_t and are linear combinations of the values of the state at time $t-1$ together with a random variation w_t for the system from a multivariate normal distribution. In this two-equation linear model, G_t defines the linear combination of values of the state at time at $t-1$. The variance-covariance matrix of the system noise will be represented by W_t. The observation at time t is a column vector X_t which is a linear combination of the states computed by a matrix F_t. The random variation or the measurement error v_t are assumed to have a normal distribution with mean 0 and variance-covariance matrix V_t and be uncorrelated over time. In this model all matrices can vary with time but G_t mostly is constant. The two-equation model in Equations 4.1 and 4.2 is often referred to as a Dynamic Linear Model(DLM). The DLM involves states as unknown coefficients in the linear model and other equations allowing the change. The main feature of dynamic linear models is that they allow parameters to vary rather than being fixed and therefore generalise linear models for varying parameters. The system is said to be observable if it is possible to infer the values of all components of the state from the noisy observations.

For illustration, suppose an Environmental Statistician is interested in knowing the influence of general pollution level α and Total Solid P on Turbidity pollution concentration X_t' in a reservoir. A simple model for his interest is proposed as

$$\begin{aligned} X_t' &= \alpha_t + \beta_t P_t + e_t \\ \alpha_t &= \alpha_{t-1} + \Delta\alpha_t \\ \beta_t &= \beta_{t-1} + \Delta\beta_t \end{aligned}$$

It is easy to see that these equations can be written as

$$X_t' = \begin{pmatrix} 1 & P_t \end{pmatrix} \begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} + e_t$$

and

$$\begin{pmatrix} \alpha_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{pmatrix} + \begin{pmatrix} \Delta\alpha_t \\ \Delta\beta_t \end{pmatrix}$$

where e_t, $\Delta\alpha_t$ and $\Delta\beta_t$ are all independent random deviations with mean 0. It is clear that

the intercept or the general pollution level α_t and the coefficient of total solid β_t are allowed to vary over time. In the notations of Equations (4.1) and (4.2), $w_t = \begin{pmatrix} \Delta\alpha_t \\ \Delta\beta_t \end{pmatrix}$ and $F_t = \begin{pmatrix} 1 \\ P_t \end{pmatrix}$, also $G_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, and $X_t = X_t'$.

Furthermore, multivariate DLMs can be used to handle changes in levels of many variables at a time. For example, if we have to observe two variables $Y_t^{(1)}$ and $Y_t^{(2)}$ then using a bivariate model with the same explanatory variable X in each case we can write (μ,v_t, β refer to the parameters and the superscripts indicate the series)

$$\begin{aligned} Y_t^{(1)} &= \mu_t^{(1)} + \beta_t^{(1)} X_t + \varepsilon_t^{(1)} \\ Y_t^{(2)} &= \mu_t^{(2)} + \beta_t^{(2)} X_t + \varepsilon_t^{(2)} \end{aligned}$$

and the state equations reflecting the changes in parameters can be written as follows

$$\begin{aligned} \mu_t^{(1)} &= \mu_{t-1}^{(1)} + v_{t-1}^{(1)} + e_t^{(1)} \\ v_t^{(1)} &= v_{t-1}^{(1)} + \zeta_t^{(1)} \\ \beta_t^{(1)} &= \beta_{t-1}^{(1)} \\ \mu_t^{(2)} &= \mu_{t-1}^{(2)} + v_{t-1}^{(2)} + e_t^{(2)} \\ v_t^{(2)} &= v_{t-1}^{(2)} + \zeta_t^{(2)} \\ \beta_t^{(2)} &= \beta_{t-1}^{(2).} \end{aligned}$$

In this case we then have a state-space form for the observed equation as

$$\begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 0 & X_t & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & X_t \end{pmatrix} \begin{pmatrix} \mu_t^{(1)} \\ v_t^{(1)} \\ \beta_t^{(1)} \\ \mu_t^{(2)} \\ v_t^{(2)} \\ \beta_t^{(2)} \end{pmatrix} + \begin{pmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{pmatrix}$$

and the state-space form for the states-equation can be written as

$$\begin{pmatrix} \mu_t^{(1)} \\ v_t^{(1)} \\ \beta_t^{(1)} \\ \mu_t^{(2)} \\ v_t^{(2)} \\ \beta_t^{(2)} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{t-1}^{(1)} \\ v_{t-1}^{(1)} \\ \beta_{t-1}^{(1)} \\ \mu_{t-1}^{(2)} \\ v_{t-1}^{(2)} \\ \beta_{t-1}^{(2)} \end{pmatrix} + \begin{pmatrix} e_t^{(1)} \\ \zeta_t^{(1)} \\ 0 \\ e_t^{(2)} \\ \zeta_t^{(2)} \\ 0 \end{pmatrix}$$

The final columns of the last two matrix equations refer to the matrices of uncorrelated errors in each case.

4.2.1 Dynamic Linear Model: Kalman Filtering

Applications of DLMs exist heavily in science and engineering, cruise control in automobiles, autopilots in aircraft and so on. In the chemical sciences for example, Papadopoulous et al. (1991), Tiwari and Dienes (1994) and Cai and Tiwari (2000) applied DLMs to chemical concentrations with interests on time varying parameters to capture seasonality. More theoretically extensive application of DLMs can be found in Pole et al. (1994), West and Harrison (1997), Durbin and Koopman (2001) and Commandeur and Koopman (2007).

To apply the DLMs, we can distinguish prediction from filtering and smoothing. Basically prediction has to do with forecasting future values of the state. Filtering makes the best estimate of the current values of the state from the record of observation including the current observation while smoothing is making the best estimates of past values of the state given the record of observations. Explaining the on-line nature of DLMs we use the Kalman filtering equations to obtain useful estimates. Suppose data D_t obtained till time t, is represented as $D_t = (D_{t-1}, X_t)$ meaning combination of data till time $t-1$ and the observation at time t. Using the Bayes's formula we can write the following supposing the parameter of interest (state) is θ

$$\pi(\theta_t|D_{t-1}, X_t) = \frac{\pi(X_t|\theta_t, D_{t-1})\pi(\theta_t|D_{t-1})}{\pi(X_t|D_{t-1})}.$$

From the two-equation model identified in Equations 4.1 and 4.2, it is clear that X_t is independent of D_{t-1} given θ_t. Hence $\pi(X_t|\theta_t, D_{t-1}) = \pi(X_t|\theta_t)$ and then

$$\pi(\theta_t|D_{t-1}, X_t) \quad \propto \quad \pi(X_t|\theta_t)\pi(\theta_t|D_{t-1}).$$

Supposing the prior distribution of the parameter θ and the likelihood (obtained from the data) are both normally distributed, Bayes Theorem can be used to see the analytic form of the posterior distribution. In this case the mean of the posterior distribution represents a weighted mean of the prior distribution and the observation with weights proportional to their precisions (reciprocal of the variance). It can also be shown that the precision of the posterior distribution is the sum of the precision of the prior distribution and the precision of the observations (see Cowpertwait and Metcalfe (2009) and Petris et al. (2009)). Therefore for a multivariate normal distribution, we can write

$$\theta_t|D_t \sim N(m_t, C_t)$$

where m_t and C_t are iteratively computed for $t = 1, ..., n$ from the following algorithm called the Kalman filter algorithm. The initial values m_0 and C_0 are specified as part of the model. The Kalman filter algorithm indicates the prior distribution for θ_t, the likelihood and the posterior distribution for θ_t respectively as follows $\theta_t|D_{t-1} \sim N(a_t, R_t)$, the likelihood $X_t|\theta_t \sim N(F_t'\theta_t, V_t)$ and the posterior distribution for θ_t as $\theta_t|D_t \sim N(m_t, C_t)$. Then for $t = 1, ..., n$, the Kalman filter equations are written as

$$\begin{aligned}
a_t &= G_t m_{t-1}, \quad f_t = F_t' a_t \\
R_t &= G_t C_{t-1} G_t' + W_t, \quad Q_t = F_t' R_t F_t + V_t \\
e_t &= X_t - f_t, \quad A_t = R_t F_t Q_t^{-1} \\
m_t &= a_t + A_t e_t, \quad C_t = R_t - A_t Q_t A_t'.
\end{aligned}$$

From the algorithm f_t is the forecast value of observation at time t, the forecast being made at time $t-1$. In the algorithm, e_t represents the forecast error, variance of the posterior is a weighted sum of the prior mean and the forecast error, and the variance of the posterior C_t is less than the variance of the prior distribution R_t. The equations above are called Kalman Filter equations and for more information, see Cowpertwait and Metcalfe (2009) and Chandler and Scott (2011). To start predicting from the posterior estimate of the state from the Kalman filter on time (day) t on which the forecast is made a one-step-ahead forecast can be done as follows

$$E[X_{t+1}|D_t] = E[F'_{t+1}\theta_{t+1} + v_{t+1}|D_t] = F'_{t+1}E[\theta_{t+1}|D_t] = F'_{t+1}a_{t+1},$$

which is equal to f_{t+1}

and also

$$V[X_{t+1}|D_t] = V[F'_{t+1}\theta_{t+1} + v_{t+1}|D_t] = F'_{t+1}V[\theta_{t+1}|D_t]F_{t+1} + V_{t+1} = F'_{t+1}R_{t+1}F_{t+1} + V_{t+1},$$

which is equal to Q_{t+1}. We can then forecast for g-steps ahead generally using

$$X_{t+g}|D_t \sim N(f_{t+g|t}, Q_{t+g|t}),$$

where

$$\begin{aligned} f_{t+g|t} &= F'_{t+g}G^{g-1}a_{t+1} \\ Q_{t+g|t} &= F'_{t+g}R_{t+g|t}F_{t+g} + V_{t+g} \\ R_{t+g|t} &= G^{g-1}R_{t+1}(G^{g-1})' + \sum_{j=2}^{g} G^{g-j}W_{t+j}(G^{g-j})'. \end{aligned}$$

For smoothing, we need to make the best estimates of past values of the state given the record of observations. Therefore a one-step back procedure can be used using the rule of total probability as

$$\pi(\theta_{t-1}|D_t) = \int \pi(\theta_{t-1}|\theta_t, D_t)\pi(\theta_t|D_t)d\theta_t$$

and also

$$\pi(\theta_{t-1}|\theta_t, D_t) = \pi(\theta_{t-1}|\theta_t, D_{t-1}).$$

Logically X_t will provide no further information once θ_t is known so that we can write

$$\pi(\theta_{t-1}|\theta_t, D_{t-1}) = \frac{\pi(\theta_t|\theta_{t-1}, D_{t-1})\pi(\theta_{t-1}|D_{t-1})}{\pi(\theta_t)|D_{t-1})}.$$

In this case given θ_t and D_{t-1} the numerator includes the posterior density at time $t-1$ following from the Kalman filter. Assuming normal distribution we have

$$\theta_{t-1}|D_t \sim N(a_t(-1), R_t(-1))$$

where

$$\begin{aligned} a_t(-1) &= m_{t-1} + B_{t-1}(m_t - a_t), \\ R_t(-1) &= C_{t-1} - B_{t-1}(R_t - C_t)B'_{t-1}, \\ B_{t-1} &= C_{t-1}G'R_t^{-1}. \end{aligned}$$

4.2.2 Dynamic Local Linear Models: Kalman filtering Applications to Rainfall Data.

Dynamic linear models have an advantage in that model parameters can change over time and that a change in mean level can be allowed. Many packages exist in the **R** software for fitting state space models via the Kalman filtering equations but the informed choice that the user has to make for Kalman filtering is in terms of the prior distribution. Fernando (2011) reviewed many of the packages in **R** including **dse**, **sspir**, **dlm**, **FKF** and **KFAS** and assessed the relative speed of the packages. He estimated parameters in Gausssian models of different sizes by maximum likelihood. In each package, optimisation of the likelihood is performed using the **R** function **optim**. Therefore the choice of packages to be used depends sometimes on what is needed. In this Section, we will use the **R** function **StructTS** to study the structure of the rainfall data by fitting a local level model to the data. Studying the change in mean level the model to be fitted is

$$\begin{aligned} X_t &= \theta_t + v_t \\ \theta_t &= \theta_{t-1} + w_t. \end{aligned}$$

where $v_t \sim N(0, V)$ and $w_t \sim N(0, W)$ such that V and W are estimated by maximum likelihood.

Obtaining some smoothing for the rainfall time series via the local state-space model gives the best estimate of the state θ_t at each time point based on the whole observed series. The smoothing on Figure 4.1 shows the level position as it changes. It shows signs of changes at

points 15, 40 and 65 indicating multiple changes. Also we can use the residuals obtained from the local linear model above to study the time when the error process seems to go out of statistical quality control. Figure 4.2 shows a Cumulative Sum Chart for the error terms obtained from the local level model and a total of five points are out of the upper decision boundary(UDB) and the lower decision boundary(LDB) suggesting that the process is out of statistical quality control at around points 6 and 68. We can also investigate the exact position of the change deeper if we estimate the variances V and W from the first 40 data points that come in and use the fitted model to carry out on-line change-point detection in the remainder of the series. Figure 4.2 shows that 13 points fall outside UDB indicating that the process is outside control and that the changes occurred almost after point 68. It is therefore reasonable in an on-line setting to experience such changes since we do not need to consider all data set each time a new observation is observed.

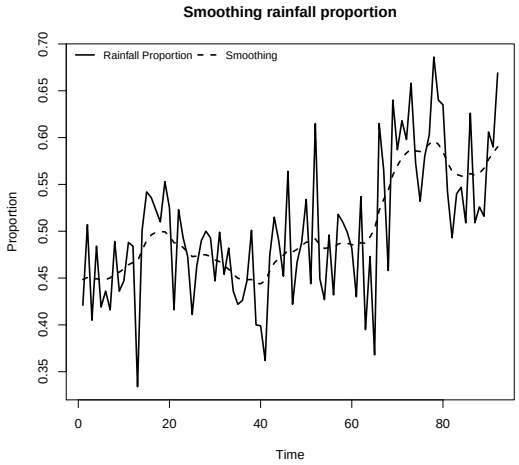

Figure 4.1: Comparing the local level smoothing (dashed line) with the United Kingdom rainfall proportion (solid line).

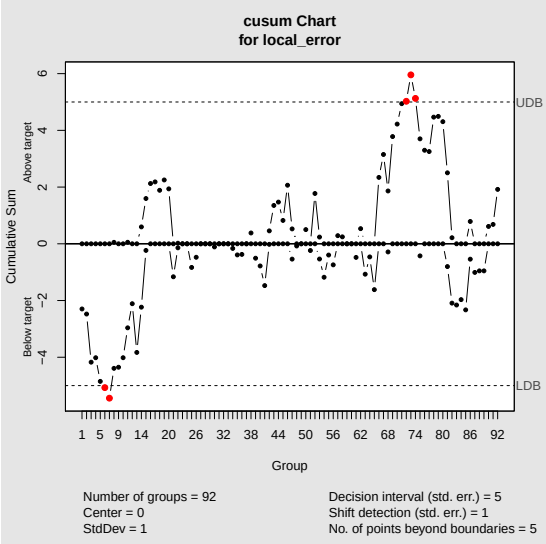

Figure 4.2: Cumulative Sum Chart for the error terms obtained from the local level model: A total of five points are out of the upper and lower boundaries suggesting that the process is out of statistical quality control at around points 6 and 68

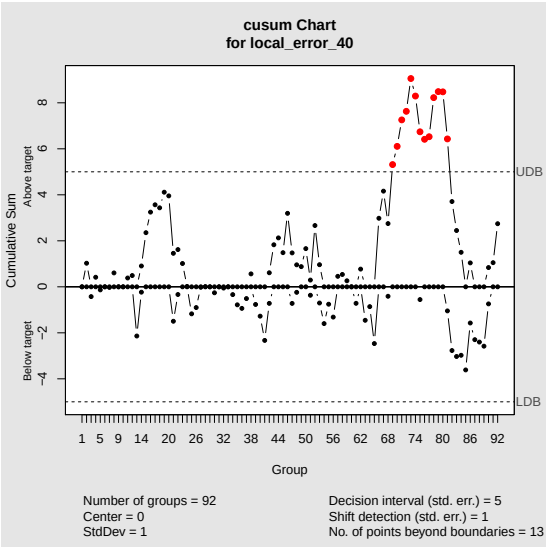

Figure 4.3: Cumulative Sum Chart for residuals from local level model using the rainfall data: The plot shows the model for the first 40 data points and the remaining data after an initial model is fitted to the first 40 data points.

4.2.3 Bivariate Linear Models: Proposed Model for Water Chemistry Data

Eleyele and Asejire reservoirs discussed in Chapter 1 are the main sources of urban and rural water needs in Ibadan. Increasing population and low rainfall jointly make efficient management of water resources an important priority. Although both reservoirs receive heavy pollution from many sources, we can justify from the exploratory analysis done in Chapter 1 that Eleyele reservoir is more polluted with respect to the set standards of the World Health Organisation (2008) and SON (2007). Envisaging how to evaluate pollution in the reservoirs we can use the monthly data reported from Eleyele and Asejire on Turbidity (T_t) and Total Solids (T_t). Turbidity is chosen due to its importance as a pollution variable and Total solid has been chosen due perceived heavy solid pollution in the reservoirs.

We intend to set up a simple state space model that can adapt to the changing Turbidity and Total Solid levels. (Harvey, 1989, p 156) discussed non-linearity and non-normality in state-space models where the system matrix G_t can be stochastic such that they can depend on information that is available at time $t-1$. These models can be susceptible to treatment by the Kalman filter. He also discussed another kind of non-linearity when the observations in the measurement equation are no longer a linear function of the state vector.

Allowing the coefficients $\theta_1, ..., \theta_{10}$ to change over time and letting T_t and S_t stand for mean adjusted Turbidity and Total Solid concentration for month t the model to be fitted is

$$\begin{aligned} T_t &= \theta_1 + \theta_2 T_{t-1} + \theta_3 S_{t-1} + \theta_4 \sin(\frac{2\pi t}{12}) + \theta_5 \cos(\frac{2\pi t}{12}) + v_{T,t} \\ S_t &= \theta_6 + \theta_7 T_{t-1} + \theta_8 S_{t-1} + \theta_9 \sin(\frac{2\pi t}{12}) + \theta_{10} \cos(\frac{2\pi t}{12}) + v_{S,t} \end{aligned}$$

The state-space form therefore can be written as

$$\begin{pmatrix} T_t \\ S_t \end{pmatrix} = \begin{pmatrix} 1 & T_{t-1} & S_{t-1} & \sin(2\pi t/12) & \cos(2\pi t/12) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & T_{t-1} & S_{t-1} & \sin(2\pi t/12) & \cos(2\pi t/12) \end{pmatrix} \times \mathbf{K}$$

where $\mathbf{K} = \begin{pmatrix} \theta_{1,t} & \theta_{2,t} & \theta_{3,t} & \theta_{4,t} & \theta_{5,t} & \theta_{6,t} & \theta_{7,t} & \theta_{8,t} & \theta_{9,t} & \theta_{10,t} \end{pmatrix}'$ and then for the varying parameters we write $\theta_{i,t} = \theta_{i,t-1} + w_{i,t}$. It is expected that $v_{T,t}$, $v_{S,t}$ and $w_{i,t}$ will capture the random components with respect to time in each case.

4.3 Summary of Chapter Four

In this Chapter, we have discussed quality control methods as regards monitoring and conformance to standards. We discussed dynamic linear models with respect to recursive Bayesian techniques. Examples of dynamic linear models are given both for univariate and bivariate models. Kalman filter equations were derived to explain the recursive nature of state-space models and to explain the change-in-mean levels with filtering showing positions of change. A seasonality-inclined model is proposed for handling changes in the water chemistry data using Turbidity and Total Solids. The model will be developed to monitor the abrupt changes in the Water Chemistry and other data sets with such characteristics.

Chapter 5

Findings and Recommendations

Data suspected with change-point problems need to be carefully studied. When the change-point is unknown, the problem is complex in that the failure of the regularity conditions leads to more complicated cases where we have to estimate parameters by iterations and non-linear algorithms. In this thesis, we have followed the lead in Hinkley (1970) to derive a test statistic used in testing hypothesis of change. The test-statistic is derived against Hinkleys assumption of known variance and simulation is used to show the effect of the regularity conditions not holding when the position of the change is unknown. The profile likelihood function for the change-point is also derived. Furthermore review applications of frequentist and Bayesian methods for estimating single change-point distributions are discussed while hypotheses of change are also tested using Bayes factors for the normal change-point models. State-space models are also fitted to the available data and changes in levels are noted. Significant changes however have been noted in the water pollution data and the United Kingdom rainfall data. It seems that there exist multiple changes in the rainfall data sets. However, the sample size of the turbidity data may make it difficult to detect changes. It is recommended to build more complex state-space models that can capture the complete behaviour of the data sets.

5.1 Limitations of the Study

There seem to be a number of limitations to this study. One for example is the small sample size of the water chemistry data-sets which were part of the motivations leading to the research. The

small sample size can limit the amount of information that can be derived from the observations. Another is the suspicion of the presence of some abrupt changes in the data sets which may make it a complex task detecting appropriate changes. The limitations include the presence of some extreme values and unexpected seasonality behaviours in the water chemistry data-sets. Therefore considering these disadvantages, most of the off-line methods could not appropriately explain the water chemistry data.

5.2 Future Work

Change-point problems considered in this dissertation have involved the estimation of single change-points using different frequentist and Bayesian methods. We intend to extend these methods to estimate multiple change-points. This will be done by extending both the frequentist methods and the Bayesian methods of Gibbs Sampling and Metropolis Hastings algorithm to estimating multiple changes.

Other volatile structures in the data set will be investigated and the existence of change investigated. For example, we might follow our discussion in Section 3.1.2 to compute Bayes factors to test hypotheses of change when the variance changes. In this case, we suppose $\mathbf{x}_{1:\delta}$ and $\mathbf{x}_{(\delta+1):T}$ have different variances σ_1^2 and σ_2^2. Then for $x_1, ..., x_T$ independent variables we have the model

$$x_i \sim \begin{cases} N(\mu, \sigma_1^2) & \text{for } 1 \leq i \leq \delta \\ N(\mu, \sigma_2^2) & \text{for } \delta < i \leq T. \end{cases}$$

More importantly, we will extend the state-space models to further understand the nature and structure of the available noisy data sets from Nigeria. This will be done by monitoring change in variable levels in the direction of our discussion in Section 4.2.3. Furthermore attempts will be made to incorporate interventions in to the state-space models to estimate change-points in many variables at a time. We shall then use multivariate methods that are robust to outliers and data errors (for example in Eleyele: DO concentrations have some strange values). This will be done with the use of multivariate time-series models for non-normal data (most observations have abrupt changes) in mind.

It is also intended to gather more data sets for longer period (to give better insights in

to the data structure for better and easy detection of significant changes) so as to link water pollution data sets to other hydrological data and public heath. This will help to investigate the connections with environmental systems and climate change.

Bibliography

Adelegan, J. A. (2004). The History of Environmental Policy and Pollution of Sources in Nigeria. *BCHD Global Envt.*

Bang, H., Mazumdar, M., and Spence, J. D. (2006). Tutorial in biostatistics: Analysing associations between total plasma homocysteine and b vitamins using optimal categorisation and segmented regression. *Nueroepidemiology*, 27:188–200.

Bellera, C. A., Hanley, J. A., Joseph, L., and Albersen, P. C. (2008a). Hierarchical changepoint models for biochemical markers illustrated by tracking postradiotheraphy prostate specific antigen series in men with prostate cancer. *Elsevier*, 18(4):270–282.

Bellera, C. A., Hanley, J. A., Joseph, L., and Albertsen, P. C. (2008b). Detecting trends in noisy data series: Application to biomarker series. *American Journal of Epidemiology*, 167(9):1131–1339.

Bellera, C. A., Hanley, J. A., Joseph, L., and Albertsen, P. C. (2009). A statistical evaluation of rules for biochemical failure after radiotheraphy in men treated for prostrate cancer. *Int.J.Radiation Oncology Biol.Phys.*, 75(5):1357–1363.

Bhattacharyya, G. K. and Johnson, R. (1968). Nonparametric Tests for Shifts at an Unknown Time Point. *Ann.Math.Statist.*, 39(1):1731–43.

Brodsky, B. E. and Darkhovsky, B. S. (1993). *Nonparametric Methods in Changepoint Problems*. Kluwer Academic Publishers.

Cai, Z. and Tiwari, R. C. (2000). Application of a Local Linear Autoregressive Model to BOD Time Series. *Environmetrics*, 11:341–350.

Carlstein, E. (1988). Nonparametric Changepoint Estimation. *Ann. Statist*, 16(1):188–197.

Chandler, R. E. and Scott, E. M. (2011). *Statistical Methods for Trend Detection and Analysis in the Environmental Sciences*. John Wiley Ltd.

Chen, J. (1998). Testing for a Change Point in Linear Regression Models. *Commun.Statist.*, 27(10):2481–2493.

Chen, J. and Gupta, A. L. (2000). *Parametric Statistical Change Point Analysis*. Birkhauser Boston.

Chia-Shang, J. C. and White, H. (1992). A Direct Test for Changing Trend. *American Statistical Association*, 10(3):289–299.

Chib, S. (1998). Estimation and Comparison of Multiple Change-point Models. *Journal of Econometrics*, 86:221—241.

Chib, S. and Greenberg, E. (1996). Markov chain montecarlo simulation methods in econimetrics. *Econometric Theory*, 12:409–431.

Chicken, E., Pignatiello, J., and Simpson, J. R. (2009). Statistical Process Monitoring of Nonlinear Profiles using Wavelets. *Journal of Quality Technology*, 41(2):198–212.

Chu, P. (2002). Large-scale Circulation Features Associated with Decadal Variation of Tropical Cyclone Activity over the Central North Pacific. *Journal of Climate*, 15:2679–2689.

Commandeur, J. J. F. and Koopman, S. J. (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press.

Cowpertwait, P. S. P. and Metcalfe, A. V. (2009). *Introductory Time-Series*. Springer.

Darkhovskii, B. S. (1976). A Nonparametric Method for the a Posteriori Detection of the Disorder Time of a Sequence of Independent Random Variable. *SIAM, Theory of Prob. and Appl.*, 21:178–183.

Darkhovskii, B. S. (1980). A Posteriori Detection of the Disorder Time in a Random Sequence. *SIAM, Theory of Prob. and Appl.*, 25:624–628.

Davison, A. C. (2003). *Statistical Models*. Cambridge University Press.

Dempfle, A. and Stute, W. (2002). Nonparametric Estimation of a Discontinuity in Regression. *Statistica Neerlandica*, 56(2):233–242.

Durbin, J. and Koopman, S. J. (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Easterling, D. R. and Peterson, T. C. (1995). A New Method for Detecting Undocumented Discontinuities in Climatological Time Series. *Int.J. Climatol*, 15:369–377.

Eastwood, V. R. (1993). Some Nonparametric Methods for Changepoint Problems. *The Canadian Journal of Statistics*, 21(2):209–222.

Farley, J. U. and Hinich, M. J. (1970). A Test for Shifting Slope Coefficient in a Linear Model. *J.Am.Stat.Assoc*, 65:1320–1329.

Fernando, T. (2011). Kalman filtering in R. *Journal of Statistical Software*, 39(2):1–24.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990). Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling. *J.Am.Stat.Assoc*, 85:972–985.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2000). *Bayesian Data Analysis*. Chapman/Hall/CRC.

Hanesiak, J. M. and Wang, X. L. (2005). Adverse Weather Trends in the Canadian Arctic. *J.Climate*, 18(2):3140–3156.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman filter*. Cambridge University Press.

Henderson, R. and Mathews, J. N. S. (1993). An investigation of Changepoints in the Annual Number of Cases of Haemolytic Uraemic Syndrome. *Appl.Statist*, 42(3):461–471.

Hinkley, D. V. (1969a). Inference about the Intersection in Two-phase Regression. *Biometrika*, 56:495–504.

Hinkley, D. V. (1969b). On the Ratio of Two Correlated Normal Random Variables. *Biometrika*, 56:635–639.

Hinkley, D. V. (1970). Inference about the Change-point in a Sequence of Random Variables. *Biometrika*, 57(1):1–17.

Hinkley, D. V. (1971). Inference in Two-phase Regression. *J.Am. Stat. Assoc.*, 66:736–743.

Hinkley, D. V. (1972). Time-ordered classification. *Biometrika*, 59:509–523.

Hinkley, D. V. and Hinkley, E. A. (1970). Inference about the change-point in a sequence of binomial random variables. *Biometrika*, 57:477–488.

Hoffman, L. R., Knowfczynski, G., and Clark, S. (2010). Estimation of Parameters in the Joint-point Two-regime Regression Model. *Communication in Statistics- Simulation and Computation*, 39:1562–1576.

Hsieh, H. K. (1984). Nonparametric tests for scale shift at an unknown time point. *Commun. Statist. Theor.Meth.*, 13(11):1335–1355.

Hudson, D. J. (1966). Fitting Segmented Curves whose Join Points have to be Estimated. *J.Am. Stat.Assoc*, 61:1097–1129.

Inclan, C. (1993). Detection of Multiple Changes of Variance Using Posterior Odds. *Journal of Busines and Economic Statistics*, 11:289—300.

Jandhyala, V. K., Fotopoulus, S. B., and Evaggelopoulos, N. (1999). Change-point Methods for Weibull Models with Applications to Detection of Trends in Extreme Temperature. *Environmetrics*, 10:547–564.

Jandhyala, V. K. and MacNeil, I. B. (1986). The Changepoint Problem: A Review of Applications. *Elsevier: In Statistical Aspects of Water Quality Monitoring(Eds. A.H. ElShaarawi and R.E. Kwiatkowski)*, pages 381–387.

Jim, A. (2008). *Bayesian Computation with R*. Springer.

Julious, S. A. (2001). Inference and Estimation in Changepoint Regression Problem. *The Statistician*, 50(1):51–61.

Kalman, R. and Bucy, R. (1961). New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83:95–108.

Khan, R. A. (1978). A Sequential Detection Procedure and the Related Cusum Procedure. *Sankhya,Ser.B*, 40(3-4):146–162.

Kundzerwicz, Z. W. (2007). Fresh Water Resources and their Management: Climate Change, Impacts, Adaptation and Vulnerability. *IPCC:Cambridge*, pages 173–210.

Lai, T. L. (1995a). Sequential change point detection in quality control and dynamic systems. *J R StatSoc Ser B*, 57:613–658.

Lai, T. L. (1995b). Sequential Changepoint Detection in Quality Control and Dynamical Systems. *J.R.Statist.Soc.B*, 57(4):613–658.

Lindsey, J. K. (2001). *Nonlinear Models in Medical Statistics*. Oxford University Press.

Lombard, F. (1987). Rank tests for changepoint problems. *Biometrika*, 74(3):615–624.

Lorden, G. and Pollak, M. (2005). Nonanticipating Estimation Applied to Sequential Analysis and Changepoint Detection. *The Annals of Statistics*, 33(3):1422–1454.

Lu, Q., Lund, R., and Seymour, L. (2005). An update of US temperature trends. *J. Climate*, 18:4906–4919.

Lutkepohi, H., Saikkonen, P., and Trenkler, C. (2004). Testing for the cointegrating rank of VAR with level shift at unknown time. *Econometrica*, 72(2):647–662.

Manly, B. F. L. (2001). *Statistics for Environmental Science and Management*. Chapman Hall/CRC,Boca Raton.

McCulloch, R. and Tsay, R. (1993). Bayesian Inference and Prediction for mean and variance shifts in Autoregressive Time Series. *Journal of American Statistical Association*, 88:968–978.

Mei, Y. (2010). Efficient Scalable Schemes for Monitoring a Large Number of Data Streams. *Biometrical*, 97(2):419–433.

Moreno, E., Casella, G., and Garcia-Ferrer, A. (2005). An Objective Bayesian Analysis of the Change-point Problem. *Stochastic Environ.Res Risk Assess*, 19(3):191–204.

NPC (2006). *Population Results for Oyo State*. National Population Commision.

Omoleke, I. I. (2004). Management of Environmental Pollution in Ibadan, An African City: The Challenges of Health Hazard Facing Government and the People. *Jhum. Ecol*, 14(4):1–26.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41:100–114.

Page, E. S. (1955). A test for a Change in a Parameter Occurring at an Unknown Time Point. *Biometrika*, 42:523–526.

Page, E. S. (1957). On Problems in which a Change in Parameters occurs at an Unknown Point. *Biometrika*, 44:248–252.

Papadopoulous, A. S., Tiwari, R. C., and Muha, J. (1991). Bootstrapping Procedures for Time Series Analysis of BOD Data. *Ecological Modelling*, 55:57–65.

Paul, G., Jolliffe, I., and Jones, B. (2002). *Statistical Inference*. Oxford University Press.

Perron, P. (1989). The great crash, the oil price shock and the unit root hypothesis. *Econometrica*, 57(6):1361–1401.

Perron, P. (1990). Testing for a unit root in a time series with a changing mean. *Journal of Business Economic Statistics*, 8(2):153–162.

Perron, P. and Vogelsang, T. J. (1992). Testing for a unit root in a time series with a changing mean: Corrections and extensions. *Journal of Business Economic Statistics*, 10:467–470.

Petris, G., Petrone, S., and Campagnoli, P. (2009). *Dynamic Linear Model with R*. Springer.

Pettitt, A. N. (1979). A Non-parametric Approach to the Change-point Problem. *Applied Statist.*, 28(2):126–135.

Pettitt, A. N. (1980). Estimation of Changepoint Using Nonparametric Type Statistics. *J. Statist.Comput.Simul.*, 11:261–274.

Pettitt, A. N. (1981). Posterior Probabilities for a Change-point using Ranks. *Biometrika.*, 68(2):443–450.

Pfaff, B. (2006). *Analysis of Integrated and Cointegrated Time Series with R.* Springer.

Plackett, R. (1950). Some Theorems in Least Squares. *Biometrika*, 37:149–157.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting.* Chapman and Hall.

Poor, H. V. and Hadjiliadis, O. (2009). *Quickest Detection.* Cambridge University Press.

Praha, J. M. and Praha, D. L. (2008). Application of MCMC to Change Point Detection. *Application of Mathematics*, 4:281–296.

Quandt, R. E. (1958). The Estimation of the Parameters of a Linear Regression System Obeys Two Seperate Regimes. *J.American Statistical Asssociation*, 53:873–880.

Rasmussen, P. (2001). Bayesian estimation of change-point using the general linear model. *Water Resource Res*, 37:2723–2731.

Ray, K. and Tsay, R. S. (2002). Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, 23(6):687–705.

Reeves, J., Chien, J., Wang, X. L., Lund, R., and Lu, Q. (2006). A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatatology*, 46:900–915.

Rice, J. A. (2007). *Mathematical Statistics and Data Analysis.* Duxbury.

Schechman, E. (1982). A Nonparametric Test for Detecting Changes in Location. *Comm.Statist.Theor.Meth*, 11(13):1475–1482.

Seber, G. A. F. and Lee, A. J. (2003). *Linear Regression Analysis.* John Wiley.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression.* John Wiley. Inc.

Sen, A. and Srivastava, M. S. (1975). On Tests for Detecting Change in Mean. *The Annals of Statistics*, 3(1):98–108.

Severini, T. A. (2006). *Likelihood Methods in Statistics*. Oxford University Press.

SON (2007). *Nigerian Standards for Drinking Water Quality*. Standards Organisation of Nigeria Governing Council, Lagos,Nigeria. .

Stephen, D. A. (1994). Bayesian Retrospective Multiple-change Identification. *Journal of Econometrics*, 43:159–178.

Tiwari, R. C. and Dienes, T. P. (1994). The Kalman Filter Model and Bayesian Outlier Detection for Time Series Analysis of BOD Data. *Ecological Modelling*, 73:159–165.

Trenkler, C. (2003). A new set of critical values for systems cointegration tests with a prior adjustment for deterministic terms. *Economics Bulletin*, 3(11):1–9.

UN (2008). *Assessing Progress in Africa towards the Millenium Development Goals Report*. United Nations. .

UN (2010). *The Millenium Development Goals Report*. United Nations. .

Wang, X. L. (2006). Climatology and Trends in Some Adverse and Fair Weather Conditions in Canada,1953-2004. *J.Geophys. Res*, 111.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer.

Wilkinson, D. J. (2006). *Stochastic Modelling for Systems Biology*. CRC Press.

Williams, D. A. (1970). Discrimination Between Regression Models to Determine the Pattern of Enzyme Synthesis in Synchroneous Cell Cultures. *Biometrics*, pages 23–32.

WMO (2000). Detecting trend and other changes in hydrological data. In Kundzewicz, Z. W. and Robson, A., editors, *World Climate Programme-Water*, Geneva. World Meteorological Organisation, World Meteorological Organisation.

World Health Organisation (2008). *Guidelines for Drinking-water Quality: Incorporating the first and second addenda*. WHO, Geneva. Volume 1 Recommendations.

Yang, C., Chandler, R. E., Isham, V. S., and Wheater, H. S. (2006). Quality Control for Daily Observational Rainfall Series in the UK. *Water and Environment Journal*, 20:185–193.

Zivot, E. and Andrews, D. (1992). Further evidence of the great crash, the oil-price shock and the unit-root hypothesis. *Journal of Business Economic Staistics*, 10(3):251–270.