UNIVERSITY OF CAPE COAST

ROBUST VARIATIONAL BAYES ANALYSIS OF LINEAR CHANGE-POINT PROBLEM

SETH ASARE

2021

UNIVERSITY OF CAPE COAST

ROBUST VARIATIONAL BAYES ANALYSIS OF LINEAR CHANGE-POINT PROBLEM

BY

SETH ASARE

Thesis submitted to the Department of Statistics of the School of Physical Sciences, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Master of Philosophy degree in Statistics

DECEMBER 2021

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original work and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature: ………………………… Date:…………………

Name: Seth Asare

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature ……………………… Date………………

Name: Dr. David Kwamena Mensah

Co-Supervisor's Signature: ……………………….…… Date………………

Name: Prof. Bismark Kwao Nkansah

ABSTRACT

The deterioration of the condition of a physical system that produces output with linear relationship with the input can manifest in the data generated by such system via change-points. As a result, timely detection and analysis of a change-point in such systems form a significant element in providing pragmatic solutions towards the smooth operation of the system. In this regard, the thesis considered novel Variational Bayes methods for modeling, detection, and inference of change-point in linear systems. In particular, Variational Lower Bound Difference(VLBD), Variational Bayes Information Criteria (VBIC), and Variational Akaike Information Criteria (VAIC) ratio-based change-point detectors are developed for a single change-point detection in linear systems. The methods are assessed with linear change-point datasets in both simulation and real data of a refinery process, and their utility is soundly illustrated. Interestingly, the Variational lower bound difference-based detector shows robustness over its VBIC and VAIC counterparts in situations where there exist multiple change-points. This was evidenced by the real-data application.

KEY WORDS

Change-Point Problem

Switching and Non-Switching Linear Models

Variational Akaike Information Criterion

Variational Bayesian Information Criterion

Variational Lower Bound

Variational Lower Bound Difference

Variational Marginal Likelihood

ACKNOWLEDGEMENT

I wish to express my profound gratitude to my supervisors, Dr. David Kwamena Mensah and Prof. Bismark Kwao Nkansah for their expert guidance, advice, encouragement and immense contributions to making this study successful. I am very grateful to Dr. Francis Eyiah Bediako, the Head of Statistics Department of the University of Cape Coast, for his seasoned advice and encouragement through my study.

I am indebted to all the academic staff and supporting staff of the Departments of Statistics and Mathematics of the University of Cape Coast for their various contributions towards the completion of my programme. I would like to express my sincere gratitude to my beloved wife, Margaret Gyasiwaa and my children, Joel Nana Kofi Adams Asare and Beatrice Asare for their undaunted support and contribution to my education. Finally, I would like to thank all those who contributed in diverse ways to make this work a success. I say may the Almighty God bless you.

.

DEDICATION

To my lovely family for their love, encouragement and prayers

TABLE OF CONTENTS

Page

DECLARATION ii

ABSTRACT iii

KEY WORDS iv

ACKNOWLEDGEMENT v

DEDICATION vi

TABLE OF CONTENTS vii

LIST OF TABLES xi

LIST OF FIGURES xii

LIST OF ABBREVIATIONS xiv

CHAPTER ONE: INTRODUCTION

Background to the Study 1

Statement of Problem 11

Significance of the Study 14

Definition of Terms 16

Organization of the Study 16

Chapter Summary 17

CHAPTER TWO: LITERATURE REVIEW

Introduction 18

Change-Point 18

Change-Point Detection Approaches 19

Probabilistic Approach to Inference 21

Bayesian Inference 21

Prior Parameter Modeling 23

Likelihood Information Modeling 27

Posterior Information about a Parameter 29

Posterior Inference via Summary Statistics 32

Variational Probability Models as Approximation to True Models 34

Variational Inference for Bayesian Models 37

The Variational Lower Bound as Approximate Information 38

The Variational Bayes Perspective of Model Selection 41

Variational Information Criteria 42

Variational Akaike Information Criterion 44

Variational Bayesian Information Criterion 46

Empirical Reviews 47

Empirical Studies on Change-Point Problem in the Bayesian Paradigm 48

Empirical Studies on Change-Point Problem for Linear Regression Models 49

Empirical Studies on Change-Point Models as Model Selection Problem 54

Chapter Summary 59

CHAPTER THREE: RESEARCH METHODS

Introduction 60

Bayesian Switching Model 61

Variational Inference for Bayesian Model Switching 63

Variational Bayes Marginal Likelihood for Switching Models 65

Application of Variational Inference to Developed Models 68

Algorithm 1: Variational Algorithm for Null Model (1) 70

Algorithm 2: Variational Algorithm for Switching Model (2) 70

Variational Bayes Change-Point Detector 71

Variational Lower Bound-Based Approach 72

Algorithm 3: Variational Lower Bound-Based Detector Algorithm 74

Variational Information-Based Approach 75

Algorithm 4: VAIC and VBIC Ratio-Based Detection Algorithm 77

Performance Evaluation 77

Implementation 79

Application of Methods to Datasets 79

Example 1: Simulated Study Based on Fixed *k* 80

Example 2: Real Data Application: Manufacturing Process in Refinery 81

Chapter Summary 82

CHAPTER FOUR: RESULTS AND DISCUSSION

Introduction 84

Application to Simulated Data 84

Example 1: Simulated Based on Fixed *k* 87

Real Data Application: Manufacturing Process in Refinery 97

Chapter Summary 107

CHAPTER FIVE: SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction 109

Summary 109

Conclusions 111

Recommendations 112

Suggestions for Further Studies 113

REFERENCES 114

APPENDICES 128

APPENDIX A: Derivations for the Null Model 128

APPENDIX B: Derivations for the Switching Model 132

LIST OF TABLES

Page

1 Comparison of Change-Point Statistic Estimates Based on VLBD, VAIC,and VBIC using the $L_{(q)}^{0}=-200.7394$, $VAIC^{0}=668.7677$, $VBIC^{0}=360.5402$ 96

2 Comparison of Change-Point Parameter Estimates Based on VLBD, VAIC, and VBIC 97

3 Real Data. Comparison of Statistic Estimates for VLBD, VAIC, and $,L^{0}(q)=-200.8703$, $VAIC^{0}=537.1086$, $VBIC^{0}=311.448$ 104

4 Real Data. Comparison of Change-Point Parameter Estimates Based on VLBD, VAIC, and VBIC 105

LIST OF FIGURES

Page

1 Randomly Selected Sample Simulated Datasets for Switching Linear System 86

2 Randomly Selected Sample Simulated Linear System 86

3 Switching Model Plot of Lower Bound Attained at Convergence for 6 Randomly Selected k Candidates for Dataset 11 87

4 Non-Switching Model Plot of Lower Bound Attained at Convergence for 6 Randomly Selected Datasets 88

5 Nature of Parameter Estimates $\left(\beta_0, \beta_1, \beta_2, \theta_0, \theta_1, \theta_2\right)$ of Switching Model for a Randomly Selected Dataset over k Values 90

6 Nature of Parameter Estimates $\left(\beta_3, \beta_4, \theta_3, \theta_4\right)$ of Switching Model for a Randomly Selected Dataset over k Values 91

7 Mean Absolute Error (MAE) Pattern of Switching Model Parameters over k Values 92

8 Plot of RMSE Switching Model Parameters Against k Values 92

9 Dynamics of Variational Bayes Information Criterion and the Lower Bound over Switching Positions k 93

10 Nature of Linear System Generated by Refinery Manufacturing Process 98

11 Plot of Data Obtained from Refinery Manufacturing Process 99

12 Lower Bound Attained at Convergence for 6 Randomly Selected k Candidates for Real Data Application 100

13 Pattern of Switching Model Parameter Estimates over k 101

14 Dynamics of Change-Point Detectors (LB Difference, VAIC Ratio

and VBIC Ratio Methods) 106

15 Performance of Change-Point Detectors (LB Difference, VAIC Ratio and VBIC Ratio Methods) 107

LIST OF ABBREVIATIONS

AIC	:	Akaike Information Criterion
BIC	:	Bayesian Information Criterion
DLB	:	Difference of Lower Bound
LB	:	Lower Bound
MCMC	:	Markov Chain Monte Carlo
VAIC	:	Variational Akaike Information Criterion
VB	:	Variational Bayes
VBIC	:	Variational Bayesian Information Criterion
VLB	:	Variational Lower Bound
VLBD	:	Variational Lower Bound Difference

CHAPTER ONE

INTRODUCTION

Background to the Study

The Problem of detection and estimation of change-points or switch points in linear regression models has received a lot of attention in the literature, particularly, in statistical computations and analysis due to the awareness of its importance in many applications. The change-point problem was pioneered by (Page, 1957, 1955, 1954), who proposed a search procedure for detecting a parameter change at an unknown point, and thus prove the existence of a switching point (Chen, Gupta and Pan, 2006; Chen and Gupta, 2001). His goal was to detect a shift in the mean of Gaussian variables that were independently and identically distributed (iid) for the purposes of industrial quality control. Thereafter, the problem of detection of change-points related to changing sequences of random variables has been explored extensively. Also, one particular version of a change-point problem termed the two-phase and, or switching regression problem has received enormous attention in many fields including economics, finance, bioinformatics, medical research, genomics research, signal processing, psychology, geology, econometrics, etc., and even in our daily lives (Chen & Gupta, 2001). According to Adams and MacKay, (2007) change-points are abrupt changes in a generative process of a sequence of random variables. In other words, shift point in statistical applications generally defines a location or time point of a data generative system and that the observations before and after that change-point follow different distributions (Basalamah, Said, Ning, & Tian, 2021; Chen & Gupta, 2001; Kang, 2015).

The process of making an informed and precise statistical statement about an unknown population based on a representative random sample is paramount in Statistical inference. The process involves modeling, estimation, hypothesis testing, and decision making. The inference is key in any modeling framework. In particular, in the change-point modeling framework, statistical inference about change-point is usually in two folds, exhibiting process patterns like following a two-stage modeling framework. However, all the two folds happen within one modeling framework. The first component of the above two folds' process involves the detection of whether or not there exists a change in the observed sequence of random variables. The second focused on the estimation of the number of changes and their respective locations if there exists a change-point. (Chen & Gupta, 2001; Holbert, 1982).

Several studies have been conducted over the years on the identification of change-points issues related to; change-in-mean, change-in-variability, or both, and also change in regression parameters from a prospective and retrospective view point. The prospective change-point issue, also known as the online change-point detection problem, seeks to identify changes in the distributional features of model parameters in a real-time scenario sequentially or as soon as they occur (Adams & MacKay, 2007; Truong, Oudre, & Vayatis, 2020). In contrast, for the retrospective change-point issue, inferences about changes in model parameters are formed based on the entire dataset or necessitate the participation of all sample points received or seen, and it is also known as offline or posterior change-point detection (Truong et al., 2020).

The existence of a change-point in a system might cause a structural change or a switch in regression parameters, affecting any or all of the model parameters. In many applications, these conditions have different theoretical and practical ramifications (Chaturvedi & Shrivastava, 2016). A structural change is considered to have occurred if at least one parameter of the linear regression model changed at a specific position (referred as change-point) across the whole sample period. A change-point could represent a transition between states, outliers, or anomaly of a system. These changes (states) may be informative on some interesting features or phenomena such as early signs of an emergency situations. Thus, timely detection of such changes can serve as a benchmark for saving unforeseen accidents. One practical example is in industry. Such a change in a production system, could be an indication of the process being out of control that might lead to defective products for which health implications are enormous. The existence of change in the generative process of systems manifests in the data generated by such systems. Another level of challenge is also seen in the probabilistic modeling of such data as the underlying distribution of the data may be changed in such a manner that a single distribution may not be fit such dataset. Thus, an intrinsic transfer of challenge to a statistician in terms of modeling and decision making.

In many theoretical and practical circumstances, a statistician or researcher faced the challenge of determining the number of jumps or change-points and their corresponding locations (referred as change-point problem) in a generative process or system (Chen, 1998). The impact of change-point problem is seen in a wide range of practical situations from several disciplines, and understanding of these changes and their implications is essential for avoiding wasteful losses and finding alternative solutions for development and constructive transition (Chen & Gupta, 2012).

The intrinsic nature of physical systems of varied types and operations due to the recent technological advancements has resulted in an increasing generation of change-point datasets of different dimensions, sizes, and covariate information, in many modern scientific applications. The availability of change- point datasets coupled with their impact in varied forms has led to the revival of an unparalleled surge of interest in inferential methods for change-point analysis within the statistical community recently. Nevertheless, the application of standard statistical methods to change-point datasets is highly challenged, leading to a new learning paradigm termed change-point modeling framework. In this framework, standard statistical methods are modified or adapted in order to develop model and inferential schemes for making informed decisions about change-points. For instance, in regression analysis, it is appropriate to present more than a single statistical model to fit the observations if the structure of the dataset changes at a certain point in time. The use of a single statistical model with no account of the change-point in such situations obviously leaves the data poorly explained (Chen & Gupta, 2012; Chen, 1998). The introduction of change-point hypothesis in statistical analyses has sparked study of modeling and inference of switching regression models and has taken center stage in regression analysis. The study of change-point problems has expanded the The process of making an informed and precise statistical statement about an unknown population based on a representative random sample is paramount in Statistical inference. The process involves modeling, estimation, hypothesis testing, and decision making. The inference is key in any modeling framework. In particular, in the change-point modeling framework, statistical inference about change-point is usually in two folds, exhibiting process patterns like following a two-stage modeling framework. However, all the two folds happen within one modeling framework. The first component of the above two folds' process involves the detection of whether or not there exists a change in the observed sequence of random variables. The second focused on the estimation of the number of changes and their respective locations if there exists a change-point. (Chen & Gupta, 2001; Holbert, 1982).

Several studies have been conducted over the years on the identification of change-points issues related to; change-in-mean, change-in-variability, or both, and also change in regression parameters from a prospective and retrospective view point. The prospective change-point issue, also known as the online change-point detection problem, seeks to identify changes in the distributional features of model parameters in a real-time scenario sequentially or as soon as they occur (Adams & MacKay, 2007; Truong, Oudre, & Vayatis, 2020). In contrast, for the retrospective change-point issue, inferences about changes in model parameters are formed based on the entire datasets or necessitate the participation of all sample points received or seen, and it is also known as offline or posterior change-point detection (Truong et al., 2020).

The existence of a change-point in a system might cause a structural change or a switch in regression parameters, affecting any or all of the model parameters. In many applications, these conditions have different theoretical and practical ramifications (Chaturvedi & Shrivastava, 2016). A structural change is considered to have occurred if at least one parameter of the linear regression model changed at a specific position (referred as change-point) across the whole sample period. A change-point could represent a transition between states, outliers, or anomaly of a system. These changes (states) may be informative on some interesting features or phenomena such as early signs of an emergency situations. Thus, timely detection of such changes can serve as a benchmark for saving unforeseen accidents. One practical example is in industry. Such a change in a production system, could be an indication of the process being out of control that might lead to defective products for which health implications are enormous. The existence of change in the generative process of systems manifests in the data generated by such systems. Another level of challenge is also seen in the probabilistic modeling of such data as the underlying distribution of the data may be changed in such a manner that a single distribution may not be fit such datasets. Thus, an intrinsic transfer of challenge to a statistician in terms of modeling and decision making.

In many theoretical and practical circumstances, a statistician or researcher faced the challenge of determining the number of jumps or change-points and their corresponding locations (referred as change-point problem) in a generative process or system (Chen, 1998). The impact of change-point problem is seen in a wide range of practical situations from several disciplines, and understanding of these changes and their implications is essential for avoiding wasteful losses and finding alternative solutions for development and constructive transition (Chen & Gupta, 2012).

The intrinsic nature of physical systems of varied types and operations due to the recent technological advancements has resulted in an increasing generation of change-point datasets of different dimensions, sizes, and covariate information, in many modern scientific applications. The availability of change- point datasets coupled with their impact in varied forms has led to the revival of an unparalleled surge of interest in inferential methods for change-point analysis within the statistical community recently. Nevertheless, the application of standard statistical methods to change-point datasets is highly challenged, leading to a new learning paradigm termed change-point modeling framework. In this framework, standard statistical methods are modified or adapted in order to develop model and inferential schemes for making informed decisions about change-points. For instance, in regression analysis, it is appropriate to present more than a single statistical model to fit the observations if the structure of the datasets changes at a certain point in time. The use of a single statistical model with no account of the change-point in such situations obviously leaves the data poorly explained (Chen & Gupta, 2012; Chen, 1998). The introduction of change-point hypothesis in statistical analyses has sparked study of modeling and inference of switching regression models and has taken center stage in regression analysis. The study of change-point problems has expanded the scope of fitting regression models and generating predictions. Following the detection and location of the change-point in the regression models, some previously poorly fitting regression models were appropriately fitted to the datasets when the change-point has been detected and identified (Chen & Gupta, 2012). The significance of the study of switching regression models in the wake of change-point problem is self-evident. detection and location of the change-point in the regression models, some previously poorly fitting regression models were appropriately fitted to the datasets when the change-point has been detected and identified (Chen & Gupta, 2012). The significance of the study of switching regression models in the wake of change-point problem is self-evident. In practice, regression relationships may vary at unknown change-points (epochs), resulting in multiple regression regimes that must be detected and identified. A change-point model is considered a two-or multiple-phase regression, switching regression, segmentation regression, two-stage least squares regression, or broken-line regression in the regression literature (Hahn, Banerjee, & Sen, 2017; Khodadadi & Asgharian, 2008; Shaban, 1980).

Considerable work has been done in the past on the various problems of estimation and inference associated with single and multiple change-points and also the switching linear regressions for both univariate and multivariate cases and the references given in this work are by no means exhaustive (Carlin et al, 1992). Most of these researched works are also related to both abrupt and gradual changes in the literature. The conventional approaches to solving change-point problems in the literature have been the usual Bayesian and Non-Bayesian (Classical) methodologies based on different modeling contexts, for example, the parametric and non-parametric. There exists a rich literature on the above approaches for change-point problems. For details on the parametric change-point approaches see, for example, (Chernoff & Zacks, 1964; Hinkley, Chapman, & Runger, 1980; Hinkley, 1970; James, James, & Siegmund, 1987; Worsley, 1979; Worsley, 1986). On the nonparametric approaches, readers are referred to, for example, (Brodsky & Darkhovsky, 1993; Lombard, 1987; Pettitt, 1979).

Appealing alternative methods to handle the change-point problems are the information criterion and the decision-theoretic methods. However, several authors in most of the recent studies have proposed solution schemes to handle change-point problems from the Bayesian framework. In the literature, the standard regression modeling approaches to tackling change-point issues and its effects have been extensively studied, employing both Bayesian and classical methodologies. Many researchers have explored the change-point problems related to regression models using both classical and Bayesian approaches in this regard. In particular, (Acitas & Senoglu, 2020; Hinkley, 1969, 1971; Quandt, 1958, 1960; Sprent, 1961), have discussed two-phase regression in a classical point of view.

Also, Ferreira (1975) used the Bayesian paradigm to explore the sampling features of various Bayesian estimates. However, several studies of change-point related problems in the context of fitting regression models in the Bayesian framework used the exact Bayesian measures, which requires a high level of integration except for simple models with appreciable complexity, generating tractable posterior distributions. Choy and Broemeling (1980), for example, utilized Bayesian inference techniques to investigate switching linear models. Holbert (1982) also examined the switching simple linear regression models and switching multiple regression models using an exact Bayesian technique. Basically, the integrations associated with the exact Bayesian methods are extremely challenging to perform, either numerically or analytically in the case of a complex and intractable posterior distribution (Pandya & Sheth, 2016). As a result, a growing interest in the usage of the approximate Bayesian approaches to handle complex and intractable posterior distribution inference for change-point problems.

A considerable study of change-point problems in the context of regression models in the Bayesian viewpoint has been carried out using the two main approximate Bayesian methodologies namely, the Markov Chain Monte Carlo (MCMC) sampling methods and the Variational Bayes methods. The MCMC (Barry & Hartigan, 1993; Chib, 1998; Green, 1995; Lavielle & Lebarbier, 2001) sampling techniques using Metropolis-Hasting and, Gibbs sampling has dominated the probabilistic solutions in the literature with regards to change-point detection problem.

Moreover, very little work can be found treating the change-point detection in a sequence of random variables and even in the linear regression model for both univariate and multivariate systems by the Variational Bayes approach than the MCMC counterpart. In particular, there exists only single work done on Variational Bayes application to change-point detection in speaker change detection in signal processing. The development of Variational Bayes solution for a change-point problem is not straightforward owing to the nature of the change-point problem and its associated changes, especially, the appropriate choice of prior distributions for the change-point random variable. This study will approach the change-point problem from the Bayesian framework, particularly, the Variational arm of the Bayesian methods. The variational Bayes methods will be explored to develop appropriate Bayesian switching linear regression models for the univariate systems, with Variational Bayes inferential schemes model parameters. Furthermore, information-based criteria in the context of the Variational Bayes methods will be developed using Variational statistics of the developed models for the detection of the change- point and its location.

Statement of Problem

Statistical approaches to finding solutions to change-point problems in the linear systems based on MCMC sampling methods have been extensively explored in theory and practice, for example, (Barry & Hartigan, 1992; Blei, Kucukelbir, & McAuliffe, 2017; Chen, 1998; Choy & Broemeling, 1980; Holbert, 1982; Kang, 2015). However, it is well known that MCMC does not scale well in terms of computation for complex problems generating large datasets, which is the case of change-point problems. Though MCMC is known to yield exact solutions, the development of MCMC algorithms may result in intrinsic challenges inherited from the complexity of the change-point problem in terms of model calibration since the MCMC adopts the Bayesian framework based on a choice of prior models. An appealing fast approximate yet deterministic alternative to the MCMC technique is the Variational Bayes method.

Nevertheless, the application of the Variational Bayes methods to change-point problems has not been as extensively explored in the literature. This may be due to the underlying complexity associated with the change-point problems in terms of modeling random parameters such as the change-point location variable. A recent attempt to apply the Variational Bayes techniques to detect change-points in a sound in signal processing has been demonstrated by (Valente & Wellekens, 2005a). Valente & Wellekens, (2005b) developed a novel Variational Bayes speaker change detector based on the lower bound difference between a switching(change-point) model and a non-switching (non change-point) model assumed for speaker data. The Variational lower bound also known as the free energy is a by-product generated by an assumed Variational fitting algorithm. For models of the same complexity, the Variational lower bound, naturally serves the purpose of a modeling selection tool, in choosing an optimal model among competing models.

In this regard, it is straightforward to see the applicability of the difference in lower bounds mentioned above in detecting change-point. The work of (Valente & Wellekens, 2005a) based inference on the decision rule as the positive differences and applied a window approach in the detection because a threshold on the positive difference was needed to make an inference. Basically, inference as such that a change-point exists if the lower bound difference is positive within a given window, then a change-point is detected otherwise no change has occurred in such window, otherwise, the window varied and the process continues until a change-point is detected and identified. However, the use of a window and its associated variation contribute extra computational or detection complexity, especially for large change-point datasets. Also, there exists another issue of choice of window, starting window as well as the choice of window length.

Furthermore, it appears that there are other intrinsic drawbacks. For a linear system, if it is of interest to detect and locate a single change-point, there might be a situation with the following possibilities. The difference in the lower bound could

1. all be positive;
2. all be negative;

3. result in a positive and a negative and, or all difference being zero.

These suggest that the threshold adopted by them may not work for all situations. Another detection issue is for example, in the situation where the complexity in the null model is almost the same as that of the alternative model the difference does not work, it is negligible and this may result in no change-point detected in the system, on the basis of their detection rule, although a change-point might exist. As a result, the detector or searching scheme may not be able to detect and locate a change-point accurately in the linear system even though it may exist. This will render the detector or search algorithm ineffective for the detection problem.

Alternatively, information criteria can be applied if the change-point problem is considered in the context of model selection problem. The application of information criteria such as the Akaike Information Criterion (AIC) and its counterpart Bayesian Information Criterion (BIC) as a conventional approach to solving a change-point problem is widely established in the literature. See, for example, (Chen & Gupta, 2012) for details on theory and its application. On the other hand, their corresponding Variational Bayes versions namely the Variational Akaike Information Criteria (VAIC) and Variational Bayesian Information Criteria (VBIC) have been used for typical modeling selection in many modeling contexts, but not in change-point analysis. For details, see, for example, (You, Ormerod, & Mueller, 2014) for model selection in linear regression models.

For change-point problem analysis within the Variational Bayes framework, the application of VAIC and VBIC are not trivial and they have not been explored, due to the unknown switching nature of the problem. In the light of these, there is the need for a solution system that is robust and takes into account all the possibilities aforementioned, in particular, a Variational Bayes detector that can find the optimum change-point. This study seeks to develop Bayesian switching linear system with inferential methods in the Variational Bayes framework for parameter inference. In addition, the robust decision rules based on VLBD, VAIC, and VBIC as well as their differences will be developed for detection single change-point and its location in a linear system oriented in either positive plane or negative plane.

Objectives of the Study

The main objective of this study is to develop an appropriate Bayesian approach for modeling and inference for change-point datasets generated by linear systems in which the linear relationship existing among the response variable and predictors are oriented in either a positive or negative plane. In order to facilitate this objective, the following specific objectives are outlined:

1. Development of appropriate Bayesian switching and non-switching linear regression models
2. Development of Variational Bayes fitting algorithms for the developed models
3. Development of Variational Bayes change-point detection schemes.
4. Application of developed schemes for change-point analysis in both simulated and real datasets.

Significance of the Study

The significance of this study can be categorized into three main areas. Firstly, on methodology. This study throws more light on the applicability of the Variational Bayes technique to change-point problems of varied types based on the successful application to detection of single change-point in multiple linear regression models. This will motivate interest in change-point research in different scientific fields. Secondly, this study exhibits the potential to motivate the development of lightweight computational methods and their integration into prototype devices used in many physical systems such as medical devices for detecting changes in medical (health) conditions. A practical application can be seen in the detection of breaks (changes) in blood pressure, an early sign of hypertension, and strokes. This is possible because of the fast and deterministic nature of the associated Variational algorithms. Thirdly, this study can inform fast industrial quality control procedures if adopted. In particular, for industrial processes where the quality of finished products can change at any time during the manufacturing process due to machine or system error or generation of an unknown compound based on blending of two or more raw materials.

Delimitations

The study is based on the detection of a single change-point in univariate linear systems in which the data generated allows the fitting of multiple linear regression models with error models being normal. For such data, there exists only change in the mean process with somewhat fixed variability in the variance process. The study focussed on full Bayesian methods specifically Variational Bayes but not exact Bayesian methods via the MCMC for inference. Further, illustrative examples were tailored toward both simulations and real data application using industrial systems in line with the objectives of the study.

Limitations

The study explored the use of a Variational lower bound for the detection and identification of change-point in linear systems where there exists one switch in the mean process. As result, linear systems with changes in both mean and variances were not considered. In addition, inferences for multiple change-points within a given dataset for linear systems were not treated. In terms of computation methods, approximate inference in the MCMC framework was not considered due to the challenges associated with the specification of change-point distributional properties that will allow the application of the above methods. Finally, issues of time since the thesis are bound to be completed within two years with one year of research.

Definition of Terms

Change-Point Problem: This is the task of identifying and locating a change-point or step when the probability distribution of a data generative process changes.

Variational Marginal Likelihood: It is interpreted as the approximation of the true posterior distribution with a Variational distribution.

Variational Lower Bound: It is a lower bound on the probability of the observed data under a model.

Organization of the Study

This thesis is organized into five chapters: Introduction, Literature Review, Methodology, Analysis and Summary, Conclusions, and Recommendations.

Chapter 1 introduces the thesis. It provides the study's background, problem statement, objectives, significance of the study, delimitation,

limitation, definition of terms, and organization of the study. Chapter 2 reviews literature related to the study. It performs an extensive review of existing literature on change-point detection and its application in linear systems in the Bayesian paradigm. Chapter 3 focuses on methodological development. In particular, the fundamental theoretical aspects of the statistical and computational methods that were developed. This entails the modeling framework, inferential methods, and analysis of data. Chapter 4 focuses on the implementation of the developed computational methods in simulation and real data and the findings from the application of the various developed models. Finally, Chapter 5 summarizes the work, presents the conclusions, recommendations and suggesting for further studies.

Chapter Summary

The study aims at developing regression methods using Variational Bayes computation approaches that incorporate switching information for modeling and estimation. The primary goal of this chapter, however, has been to address the problem under study. This chapter provided an introduction to this study report highlighting issues, covering aspects such as the background to the study, problem statement, objectives, and proposition guiding the study, as well as the significance of the study. In addition to these, is the thesis organization. The chapter concludes with this summary.

CHAPTER TWO

LITERATURE REVIEW

Introduction

This chapter provides a brief review of relevant literature on change-point issues in linear regression settings within the Bayesian framework. Bayesian switching multiple linear regression model, Bayesian approach to inference and its characterization as well as approximate Bayesian computations will also be introduced in brief. A further review of the approximate Bayesian inference approach for complex Bayesian models in the context of Variational Bayesian methods and relevant empirical evaluations is carried out in the work.

Change-Point

The change-point analysis and its connected issues is an ever-growing field, resulting in a colossal literature discussing numerous aspects of such change-point problems. The change-point problem spans the detection of existing change-point(s) and identification of the location of such a change if it exists. This development features a wide range of applications, from industry, finance, medicine, and biological sciences, so on, and additionally provides rise to multiple methodologies several of the change-point problems. There are many studies that have explained or outlined change-points from totally different fields.

According to Adams and MacKay (2007), change -points are considered to be abrupt changes in a process generating a sequence of random variables. A change-point can also be thought of as the date or place where at least one parameter of a statistical model (e.g, mean, variance, intersection, trend) undergoes an abrupt change (Seidou, Asselin & Ouarda, 2007). Furthermore, a change-point is defined as a location or point in time at which the distributions of observations before and after this point differ (Chen and Gupta, 2001; Kang, 2015). These and many other definitions of change-points have been considered and explored in various aspects of change-point analysis. Additionally, the change-points could represent different conditions in a wide variety of application areas and deserve the special attention of a researcher and/or practitioner when it comes to modeling and inference. For example, in medicine and health, a change-point could represent a system anomaly and is frequently encountered in medical and health research (Zhou & Liang, 2008). According to MacNeil and Mao (1993), young people have relatively stable cancer incidence rates but change dramatically after a certain age. Further, in economic theory, fluctuations in any stock price are normal, but many of these shifts or changes are abnormal and deserve the special attention of an investor (Chen and Gupta, 2012). These emphasize the need to detect and locate changes (change points), anomalies, or thresholds in a system at the right time to avoid its effects on the system and also to find alternative remedies for it advancement and the benefit of the transition. They can also involve positive results, for example in quality control, when an intervention is implemented, the result is expected to evolve into a desired positive result.

Change-Point Detection Approaches

Change-point detection or change detection has become a crucial part of regression studies because the presence of a change-point can indicate a significant change in the regression model or the data generation process.

Change-point detection approaches are used to detect and or identify abrupt and gradual changes in the process of generating sequential data due to distributive or structural changes (Sharma, Swayne, & Obimbo, 2016). The ability to detect and react accurately and in a timely manner to sudden changes (change-points) is extremely desirable and also crucial for practical applications in most applied science and real-life scenarios. For example, validation of an untested scientific hypothesis (Henderson & Matthews, 1993), monitoring and evaluation of safety-critical processes (Elsner, Niu, & Jagger, 2004), and validation of hypothesis modeling (Fryzlewicz & Rao, 2011) among others are studies carried out on the issue of change-points.

Basically, the change-point detection approaches have been divided into two main branches, namely the offline change-point detection problem, and the online change-point detection problem. Offline change-point detection methods (called retrospective detection) require the full or fixed dataset for statistical analysis and detect changes when all datasets are accounted for and processed simultaneously (Truong et al. al., 2020). Detecting the offline change- point is sometimes referred to as segmentation and is generally considered to be more accurate because it involves all of the data in the analysis. In contrast, online change-point detection is used on live or sequential broadcast datasets and tries to detect changes when they occur in a real-time frame. Online change- point detection is also known as event or anomaly detection and is typically used for constant monitoring or immediate anomaly detection (Truong et al., 2020). These are very common phenomena in a wide range of disciplines. In the literature, it is possible to find a range of disciplines and a large number of underlying techniques or approaches to check whether or not there has been a change and also find where the change-point is located, if it exists. In this study, the offline change-point detection approach is used.

Probabilistic Approach to Inference

Statistical inference mainly concerns making valid and precise statements about the parameters of a given population, mainly represented by a random variable. This process requires the development of statistical models and estimation of uncertainty of the unknown parameters that characterize the given family of distribution of the data or population. Modeling the random variable implies modeling the population and after an appropriate model has been accepted, an appropriate approach to estimate the parameters is adapted to initiate the inference process. A broad range of inference methodologies is available in the literature of computational statistics. This section gives a brief exposition on the statistical inference method in the Bayesian Paradigm also referred to as the Bayesian inference method.

Bayesian Inference

The Bayesian paradigm is a probabilistic method based on a rigorous theory and which essentially represents uncertainty using probability distributions. In the Bayesian framework, probabilities provide a quantification of uncertainty of the unknown. Probability theory seeks to combine uncertain information from different sources to make optimal decisions under conditions of uncertainty. This philosophy is based on probability as logic and provides a powerful basis for the application of Bayesian inference (Beck, 2010; Cox, 1946, 1961; Jaynes, 2003) as well as Bayesian formulation and estimation(Bishop, 2006). The rationale of any

Bayesian analysis is mainly a probabilistic model which represents the uncertainty on a parameter of a population. In the Bayesian framework, the probability distributions represent precisely, at all stages of statistical analysis, what is known and unknown on each variable of interest. This probabilistic framework allows a consistent quantification of uncertainties by taking due account of all available information and is interpreted as subjective degrees of belief(Chick, 2006). The probability distribution often referred to as the degree of conviction expresses the degree of the expert belief in the truth of a certain proposition in the light of new information, which is the plausibility of the truth of a certain proposition, and is always conditional to our background knowledge or information.

Bayesian inference is a way to modify or update one's belief given the observed data using the Bayes theorem. The Bayesian framework allows the formal quantification of current beliefs as the prior distribution of the parameter. This expresses the uncertainty about the parameter value before the data is observed and the information provided by the new data or the data model, also called probability which reflects our beliefs about the data. Given a particular parameter value and applying Bayes theorem, one is able to update the beliefs and form the posterior distribution. The prior and posterior beliefs can be quantified as prior and posterior distributions respectively, and represent the uncertainty of the unknowns before and after the analysis of the information.

In the Bayesian paradigm, all statements of estimation and statistical inference depend on the posterior distribution. A detailed introduction on Bayesian inference can be found in (Box & Tiao, 2011; Bernardo & Smith, 2009; Lee, 2004). The parameters of a prior distribution are often referred to as hyper-parameters and are distinguished from that of the model parameters (Θ). In the applicability of the Bayes' theorem, the product of the likelihood and prior distribution is normalized to provide the conditional density (posterior probability distribution) of the data. Moreover, the prior distribution to large extent may have a great impact on the Bayesian estimates (for example, posterior distribution) under certain conditions (Du, Edwards, & Zhang, 2019). Regarding Bayesian inference, Prior probability distributions have been divided into two major types: informative priors and uninformative priors. In this study, four classes of prior distributions, namely, informative, weakly informative, less informative, and uninformative are presented according to information and the purpose of using the prior.

Prior Parameter Modeling

The prior distribution is used in Bayesian inference to describe knowledge about an unknown parameter. The posterior distribution (a probability distribution updated) is the optimal inference and decision-making component within the Bayesian paradigm. It is the product of the prior distribution and the probability of new data (likelihood function) (Berger, 2006). The prior distribution (often called the prior) presents the researcher's subjective belief about the unknown parameter(s) before the data is observed. The pre-data knowledge or belief about the parameter is quantified to provide a probabilistic statement about the parameter θ and is denoted as, $p(\theta)$, called the prior probability distribution. The prior distribution as mentioned earlier plays a defining role in Bayesian analysis and yet remains the most debatable component of Bayesian measurements unless there is a physical sampling mechanism to justify a choice.

Basically, the prior ought to be found by contemplation and thought of all accessible accurate knowledge about the unknown parameters before considering the information on observations (data). However, due to the subjective nature of prior beliefs, an important question in Bayesian analysis is how to choose or define a prior density in order to make an accurate and optimal inference about an unknown for reliable decision making. Studies have shown that, the prior has a considerable impact on the resulting inference, and that for accurate and reliable statistical inference, the choice or selection of the prior must be conducted with the utmost care (Du et al., 2019).According to Ghaderinezhad and Ley (2020), the effect of the prior diminishes as the sample size grows under certain regularity conditions. As a result, if the data size is large enough, especially for identical independent distributions, prior distribution has little influence on inference results. Conversely, when the size of the data is small, the type of inference is significantly more affected by the prior choice. Furthermore, in a situation where the data distribution and prior information are significantly different from each other, with a potential conflict between the two sources of information, the conditional density for the model under certain conditions may be strongly influenced by the prior information (O'Hagan & Forster, 2005; Rahman, Gao, D'Este, & Ahmed, 2016). The selection or specification of priors is obviously a key element of the Bayesian framework, as it may have an impact on inference results. As with many aspects of Statistics, there are several ways, and reasons for choosing different prior distributions. As mentioned early in this work, informative priors and uninformative priors are the two main types of prior distributions that have historically been used. Under these two headings, there are a number of important groups that describe various prior properties.

To facilitate the computation, the priors are often chosen such that the resulting posterior distribution and the prior probability distribution belong to the same family of distributions. In statistical models, when a posteriori and a priori distributions are from the same family of distributions, the prior distribution choice is called conjugate prior (Seber & Lee, 2003). In other words, a prior conjugate is one for which the application of the Bayes' theorem results in posterior with the same family of distribution as the prior (Rahman et al., 2016).

Practically, conjugate prior distributions have the advantage of being easy to handle both from a computational and interpretative point of view. It may not fully represent its belief accurately, but it is chosen for the analytical tractability of the corresponding posterior distribution. That is, it allows the results to be derived in a closed conjugate form (Seber & Lee, 2003). In the Bayesian framework, Conjugate priors yields analytically tractable Bayesian integrals. It has an intuitive interpretation as an expression of the results of previous (indeed imaginary) observations under the model. Given a standard form of distributions, an a priori conjugate is proper, however, it may belong to an a priori non-informative family, depending on the hyper-parameters values.

In defining a prior, the choice of hyper-parameters (prior parameters) is crucial. In the case of sufficient prior information about an unknown parameter, the prior option should strongly reflect one's prior belief. According to Son and Kim (2005), default priors such as non-informative priors or uniform priors are acceptable in Bayesian experiments with no prior knowledge of model parameters and/or when just a limited amount of information about an unknown is provided. The hyper-parameters in such a situation should be well specified so that a non-informative prior can be presented (Bernardo & Smith, 2009). For example, using a prior distribution with large uncertainty (variance), a Gaussian or a normal prior is proper but reflects little knowledge about an unknown parameter.

A proper prior with a strong prior belief about an unknown, not necessarily being conjugate, can be referred to as an informative prior. It should match one’s belief, as with insufficient data a wrong selection of prior may lead to an inappropriate posterior distribution. In a situation where little prior knowledge is available or the researcher tries to impart as little information as possible in order to allow the data to carry as much weight as possible in the posterior distribution, a prior should be designed to express the ignorance about an unknown. Reference priors, locally uniform priors, Jeffery’s prior also called Jeffrey’s rule, and Flat, diffuse or vague priors are some examples of non-informative priors (Bernardo & Smith, 2009). These priors do not belong to the family of standard distributions. They are improper densities, i.e. they do not sum or integrate to one ($\int P(\theta)d\theta=\infty$). A proper prior is one in which the prior probability distribution integrates to one and, an improper prior is one in which the prior probability distribution does not converge to one.

According to Son and Kim (2005), the majority of non-informative priors are improper and so it is instructive to pay attention to whether they are proper or improper. This is due to the fact that improper priors can result in logical inconsistencies as well as unspecified constants being incorporated into the Bayesian framework (Dawid, Stone, & Zidek, 1973). A typical example of such improper priors could be a uniform density over an infinite range, reflecting no specific, definite prior knowledge about the parameter. Improper priors may yield proper posterior distributions i.e. must integrate to 1, with a sufficiently informative data likelihood. A detailed discussion on such non-informative and improper priors can be found in (Box & Tiao, 2011).

Moreover, for a vector of parameters, θ (vector) i.e. $\theta=\{\theta_1,\theta_2,\theta_3,\ldots,\theta_{r-1,}\theta_r\}$, the prior distribution can be specified as independently on each component on θ or jointly on the entire vector or decomposing the joint distribution as product of conditionals as marginal as $P(\theta_1,\theta_2,\ldots,\theta_r)=P(\theta_1)P(\theta_2/\theta_1)\ldots P(\theta_r/\theta_1,\theta_2,\ldots\theta_{r-1})$ so that, it starts with a marginal prior for θ_1, conditional prior for θ_2 given θ_1, conditional prior for θ_3 given (θ_1,θ_2) and so on.

Likelihood Information Modeling

The likelihood function is a function of parameters and describes the joint probability density of the sample data. It plays a significant role in Bayesian analysis in the formulation of posterior distributions. Let y denote a random variable and $f(y;\theta)$, the probability density function of an unknown parameter, θ. This parameter identifies the population or family given the restriction that it must satisfy. Let consider a random sample of size n drawn independently from $f(y;\theta)$. Let $y_1, y_2, y_3, \ldots y_n$ be the sample. That is, $y_1, y_2, y_3, \ldots y_n \sim f(y;\theta)$. The model that is defining a relationship between a parameter, θ and a set of observed data denoted as y = { $y_1, y_2, y_3, \ldots y_n$ } is expressed through a conditional probabilistic statement and written as, $P(y;\theta)$, or $P(y|\theta)$. The term $P(y|\theta)$ is a function of data, y given a fixed (unknown) value of θ. The likelihood function of the parameter θ of a statistical model, given observed data y, denoted by $L(\theta|y)$.

It can be expressed as

$$L(\theta|y) = P(y|\theta) \tag{2.1}$$

Assumed independent and identically distributed (iid) random variables,

$$y = (y_1, y_2, \ldots, y_n),$$

we have,

$$L(\theta|y) = \prod_{i=1}^{n} P(y|\theta) \tag{2.2}$$

The likelihood function is a tool for summarizing the evidence of the data for unknown parameters, and it encapsulates all of the information about the parameter offered by the data and draws inferences from it so as to extract all the potential information contained in the data. The likelihood function $L(\theta|y)$,is the function by which the data y alters prior knowledge of, θ and it can thus be seen as conveying information about the parameter derived from the data (Box & Tiao, 2011). The basic information it provides about the parameter, θ is how likely the observed sample is generated by the possible parameter values. The likelihood $L(\theta|y)$ is not a probabilistic statement over parameter, θ given data, y. Unlike $P(y/\theta)$, it is not a p.d.f. of θ. It should not be expected to integrate to one (Vatsa, 2011).

Posterior Information about a Parameter

All inference regarding unknown parameters in the Bayesian framework is based on the posterior distribution (Bernardo & Smith, 2009). The posterior may be thought of as a belief or knowledge about an unknown parameter based on seen data, as well as a probability density, denoted by $p(\theta|y)$ that summarizes what is known about uncertain variables following data observation. It combines the prior probability density and the likelihood function to determine information contained in the observed data, which is referred to as "new evidence" (Bernardo & Smith, 2009; Box & Tiao, 2011). Furthermore, all valid Bayesian inferential claims about parameter values are included in the posterior probability density $p(\theta|y)$, which integrates the data information with any additional information contained in the prior density $P(\theta)$.

We assumed y and θ as continuous random vectors.

Given the likelihood $L(\theta|y)$ and a prior distribution $P(\theta)$, via the Bayes' theorem, conditional probability $p(\theta|y)$ can be obtained as:

$$p(\theta|y)=\frac{p(y|\theta)\,p(\theta)}{p(y)} \tag{2.3}$$

$$p(\theta|y)=\frac{p(y|\theta)\,p(\theta)}{\int p(y|\theta)\,p(\theta)d\theta} \tag{2.4}$$

$$=\frac{likelihood\times prior}{integrated\,likelihood}$$

For a given set of data, y, the posterior distribution of parameters can be represented proportionally to the product of prior distribution and data probability(likelihood). i.e.

$$p(\theta|y) \propto P(\theta)P(y|\theta) \tag{2.5}$$

Posterior ∝ Likelihood ∗ Prior

The term $P(y)$ in Equation (2.3) is called the marginal likelihood or model evidence and obtained as;

$$P(y)=\int_{\theta} p(y|\theta)p(\theta)d\theta \tag{2.6}$$

The integrated likelihood $\int_{\theta} p(y|\theta)p(\theta)d\theta$, form the normalising constant of the true posterior distribution and the integral is taken over the admissible range of the parameter θ. And it is important to ensure that the posterior $p(\theta|y)$ is a valid probability densities and integrates or sums (for discrete case) to one (Bishop, 2006; Box & Tiao, 2011).

Marginal likelihood provides an empirical basis for choosing suitable prior models or is often used in model checking, for example, the Bayes factor.

The posterior distribution $p(\theta|y)$ in Equation (2.4) is normalized to express it as a probability density function with a total probability of one and the marginal likelihood $\int_{\theta} p(y|\theta)p(\theta)d\theta$, serves as the normalization constant for a valid posterior. The term in Equation (2.5) is called the posterior normalized. This form is known as the posterior non-normalized joint distribution, and it may be used to make inferences. In most statistical problems, the parameter θ, is multidimensional; $\theta = \{\theta_1, \theta_2, \ldots, \theta_r\}$,with some nuisance parameters. In other words, a nuisance parameter is one that is included in a model's joint posterior distribution but is not of primary concern. One method for removing these bothersome factors is to integrate out or minimize the joint posterior distribution with regard to them. Define $\theta=(\varphi,\omega)$ where φ represents the major parameters of interest and where ω represents the nuisance parameters. The marginal posterior density of the major parameters of interest, φ is obtained as;

$$P(\varphi|y)=\int P(\varphi,\omega|y)\,d\omega \qquad (2.7)$$

As a result, in Bayesian inference, the marginal unnormalized joint posterior distribution is used to estimate unknown parameters. To estimate the parameters, for example, one may use the MAP technique to calculate the mode of the marginal unnormalized joint posterior distribution. These estimates are some function of posterior distributions which again require the computation of some multidimensional integrals. Solving integrals is therefore a necessary task in Bayesian computation. More often, the multidimensional integrals are either computationally intensive or intractable (Vatsa, 2011). In such cases, we rely on distributional or simulation type approximations for example Markov Chain Monte Carlo (MCMC) sampling technique and Variational Bayesian approximation discussed later in the chapter. It is critical to recognize that some types of conflicts may be created by the data (for example, outliers) and prior information. In such cases, the disagreement may have a significant impact on the posterior distribution, potentially leading to inconsistency/inappropriate statistical conclusions (Andrade & O'Hagan, 2006).

Posterior Inference via Summary Statistics

In Bayesian statistics, the posterior density function contains all the information about the parameter we already know and that we would like to know. As a result, several summary statistics can be computed from the posterior for inference about the parameter. A plot of the posterior density is frequently helpful; that is, the nature of the posterior density may be visually evaluated using graphs such as scatter plots, box plots, histograms, and contour plots, among others. Some of the numerical summary statistics representing the center of the posterior density that are commonly used are its mean, median, and mode. Each of these might be used as a Bayesian point estimate for parameter inference.

The posterior mean is computed as:

$$\mu_\theta = E\left[P(\theta/y)\right]$$

$$= \int \theta \;\; P(\theta/y)\, d\theta , \qquad (2.8)$$

The posterior mean μ_θ is defined as the parameter's expectation in relation to the posterior probability. On the other hand, the posterior mode is obtained as:

$$\theta_{\text{mod}} = \text{argmax}_\theta \left\{P(\theta/y)\right\} \qquad (2.9)$$

The value of the parameter of interest that maximizes the posterior density is defined as the posterior mode. For continuous posterior distributions, the estimate of the modality can be obtained using the principle of differentiation. The process can be difficult if the posterior has a complex structure. However, in such a situation, the maximum posterior estimate (MAP) is used which is the overall model of a posterior distributions. It is computed as:

$$\theta_{MAP}=\operatorname{argmax}_{\theta} P(\theta)P(y/\theta) \tag{2.10}$$

A MAP estimate plays a vital role in Bayesian analysis. Even if the posterior distribution is intractable, a MAP estimate can be found by some optimization methods, Newton’s optimization method. Also, if the prior distribution is non-informative the MAP reduces to the ML estimation of the parameter. A MAP estimate facilitates the approximation of complex or intractable posterior distributions, such as a Laplace approximation or a Gaussian approximation. Moreover, the measures of dispersion (spread) of the posterior distribution might be summarised, for example, by its variance or standard deviation. Not only that but also the posterior interquartile range and other quantiles are applicable (Mensah, 2010).

In the Bayesian framework, posterior measures of spread are used to represent the variability or spread of the parameter's posterior distribution.

The most common and frequently used are the posterior variance and standard deviation defined as follows: The posterior variance is

$$Var(\theta/y)=\int(\theta-\mu_{\theta})^{2}p(\theta/y)d\theta, \tag{2.11}$$

Where μ_{θ} represents the posterior mean and V also represents posterior variance.

The posterior standard deviation is given by

$$V_{\theta}=\sqrt{\int(\theta-\mu_{\theta})^{2}p(\theta/\text{¿}y)d\theta,\text{¿}} \tag{2.12}$$

The posterior knowledge of a parameter contained in its posterior distribution can be summarized not only by point estimates but also interval estimates. An interval estimate summarizes the posterior uncertainty of the parameter.

The interest may lie in specifying an interval that includes most of the posterior density and for the interval estimates for parameters, an analog of the confidence interval in the classical approach, called Bayesian credible intervals are utilized. The credible interval defines the domain of the posterior probability or predictive distribution.

As previously mentioned, the purpose of Bayesian inference is to infer the posterior probability distribution of a collection of parameters given observed data. However, in most circumstances, these posteriors are intractable, making direct quantifications of marginal distributions of parameters or other variables (estimates) of interest impracticable. Given the challenges of directly utilizing these posterior distributions, statistics literature has utilized two main methods for approximating the intractable posterior probability density and fitting Bayesian models: Markov Chain method, especially MCMC (Gelfand, Hills, Racine-Poon, & Smith, 1990; Geman & Geman, 1984) and Variational Bayesian approximations methods (Dempster, Laird, & Rubin, 1977) to accomplish the inference task. However, in this study, we consider a Variational Bayesian approximation technique particularly the Variational Bayesian method (referred to as Variational Bayes) for all our inferences.

Variational Probability Models as Approximation to True Models

Variational approximations are deterministic approaches for drawing conclusions for parameters in complex statistical models. Variational Bayes approaches have long been employed in control theory, physics and mathematical applications (Jaakkola & Jordan, 2000). According to Jordan (2004), Variational approximation has now become an important aspect of traditional computational approach, mostly utilized to address issues including, document retrieval, voice recognition, and genetic link analysis, to mention a few examples. Variational approximation approaches have lately benefited from increased application and development in statistical issues (Jordan, Ghahramani, Jaakkola, & Saul, 1999) by authors such as (Bishop, 2006; Jordan ,2004 & Titterington, 2004). In the year 2008, a software program based on Variational Bayes approximation called Infer.NET was developed under the pretext that it has the capability to deal with extensive issues (Minka, Winn, Guiver & Kannan, 2008). Variational Bayes techniques for approximating complicated calculi have their origins in calculus of variations and contain a wide variety of tools for evaluation of integrals and functions.

In general, calculus of variations fundamentally deals with optimization problems. It is the case of optimizing a given function on a particular class of functions on which this function is dependent. When a set of functions is constrained in some manner, generally to increase tractability, approximate solutions emerge (Ormerod & Wand, 2010). In spite of their enormous significance, Variational approximations methods are rarely employed in the statistical community. For making approximate inference, Markov Chain Monte Carlo methods have dominated and are more significant than Variational approximation methods and Laplace approximation methods (Gelfand & Smith, 1990; Hastings, 1970). Despite the fact that Variational techniques are routinely utilized in the field of machine learning in the statistical community, Bishop (2006) demonstrated that Variational Bayes solutions are efficient and effective alternatives to MCMC solutions for Bayesian computation and inference. In the case of large models, for example, Variational approaches provide a quick alternative to MCMC and has lately acquired prominence in the literature. Variational Bayes approximations, in particular, are significantly faster deterministic alternative to the MCMC sampling approach for Bayesian computation and facilitating approximate intractable posterior inference of complex statistical models (see, for example, Attias, 2000; Jordan et al., 1999; Waterhouse, MacKay, Robinson, & others, 1996) for early developments of the method and (Jordan, 2004; Ormerod & Wand, 2010; Titterington, 2004) for non-technical overviews and also seen as a productive sets of methods and ranked high than the Laplace approach. The Variational approximations however are restricted in their approximations accuracy as compared to the Markov chain Monte Carlo (MCMC) counterpart which can result in high accuracy if the Monte Carlo sample sizes are increased (Robert & Casella, 2004; Ormerod & Wand, 2010).

In the machine learning and statistics literature, Variational approximations methods have recently gained traction (Corduneanu & Bishop, 2001; Jordan et al., 1999; Ueda & Ghahramani, 2002). Some of these studies developed and published a new Variational approximations methodology for specific applications. Examples include the works (McGrory & Titterington, 2007; McGrory, Titterington, Reeves, & Pettitt, 2009). Titterington and Wang, (2006) also discussed the statistical properties of estimation obtained through Variational computations. The use of Variational approximations is enormous for statistical inference, in particular Bayesian inferences. According to Ormerod and Wand (2010), the application of Variational Bayes approximations is far more crucial for Bayesian inference in situations with intractable computing challenges. As a result, the majority of Variational Bayes approximation descriptions are used for Bayesian analysis. It is important to note that, Variational Bayes approximations solutions are beneficial and closely correlate to MCMC solutions.

However, the MCMC approach is inefficient for some issues.

These are useful when we need an estimated conditional probability faster than a typical MCMC method can generate it, such as when the data sets are large or the models are complex.

In these circumstances, Variational Bayes inference offers a fast and powerful alternative for Bayesian solutions (Blei et al., 2017). It could be a cost-effective and reliable method for analyzing larger datasets (You et al., 2014).

Variational Inference for Bayesian Models

A critical component of Variational inference is the measures of the approximate posterior density also referred to as the conditional density. The essential concept is to handle this problem through optimization. The optimization seeks the member of a group of densities, as an estimation to the conditional density (posterior distribution) of interest based on certain distance measures like Kullback-Leibler divergence. This measure quantifies the difference present in the approximate posterior density referenced to the conditional density (true posterior). In Bayesian solutions, the closet or the minimum Kullback-Leibler divergence measure is preferred. That is to find an approximate density that is maximally similar to the true posterior distribution, $p(\theta|y)$.

Blei et al. (2017) provided Variational evidence suggesting that the fitted Variational density derived through Variational Bayesian methods, for example ‘‘Variational Bayes’’, acts as a proxy for the true conditional densities. Variational Bayesian inference presents a class of approaches with inference optimization duality (Jang, 2016). Statistical inference problems, such as determining the values of parameters that minimize particular objective functions may be viewed as optimization problems. Furthermore, the optimal quantity in the Variational Bayesian framework is the Varioational Lower Bound (VLB, also known as evidence lower bound (ELBO), and its relevance pervades Variational Bayesian derivations and inference.

The Variational Lower Bound as Approximate Information

Using a general Bayesian model of parameter vector $\theta \in \upsilon$ and observed data vector y, for the purposes of Bayesian inference, the posterior distribution is generated as:

$$p(\theta|y)=\frac{p(y,\theta)}{p(y)}=\frac{p(y|\theta)p(\theta)}{\int_{\theta} p(y,\theta)d\theta} \qquad (2.13)$$

The phrase is marginal likelihood, and (Kass & Raftery, 1995). We will assume that and are continuous random vectors. The discrete case is handled in the same way, but with additions rather than integrals.

The term $p(y)$ marginal likelihood it is the basis for comparing competing models using Bayes factors (Kass & Raftery, 1995). We assume both parameters y and θ as continuous random vectors. The discrete case is handled similarly, but using summation but not integrals.

Assumed an arbitrary density function, q, over υ. As a result, the Variational lower bound (i.e., evidence lower bound) to the log (marginal) probability of the observations determined by (Beal ,2003) using Jensen's inequality.

as;

$$\log p(y)=\log\int_{\theta} p(y|\theta)p(\theta)d\theta=\log\int_{\theta} p(y,\theta)d\theta \tag{2.14}$$

$$=\log\int_{\theta} p(y,\theta)\frac{q(\theta)}{q(\theta)}d\theta \tag{2.15}$$

$$=\log\left(E_q\left[\frac{p(y,\theta)}{q(\theta)}\right]\right) \tag{2.16}$$

$$¿E_q\left[\log\frac{p(y,\theta)}{q(\theta)}\right] \tag{2.17}$$

$$¿E_q[\log p(y,\theta)]-E_q[\log q(\theta)] \tag{2.18}$$

The Variational lower bound, often called evidence lower bound, is represented in Equation (2.18). In Variational Bayes framework, we employ the distribution $q(\theta)$ in Equation (2.16) to estimate the posterior distribution $p(\theta|y)$. The Jensen's inequality for the concave log function is given by Equation (2.17).

The lower bound of the evidence is less than or equal to the marginal logarithmic probability of the observations. Altogether, the Evidence Lower Bound (ELBO) for a probability model $p(y,\theta)$ and approximation $q(\theta)$ to the posterior denoted as L is:

$$L=E_q[\log p(y,\theta)]-E_q[\log q(\theta)] \tag{2.19}$$

where $-E_q[\log q(\theta)]$ is termed the Shannon entropy.

As a result, instead of maximizing the marginal probability, we can maximize its Variational lower bound in some instances. With each update in the VB approximations, the lower bound on the marginal probability increases or remains constant. When the Variational approximation distribution is extremely close to the conditional density, a tight bound is established.

As a result, the VB technique can approximate the posterior distribution as: $p(\theta|y) \approx q(\theta|y)$. That is, approximation distributions that are as close to the true posterior distribution as practicable, and we achieve this by determining the parameter settings that get q close to the required posterior. Obviously, approximation distributions should be somewhat easy and tractable for inference. The Kullback-Leibler (KL) divergence is a standard statistic used to quantify the closeness of the two distributions. This metric computes the differences between the approximated distribution $q(\theta)$ and the true posterior distribution $p(\theta|y)$. The lower bound can alternatively be described in terms of a Kullback-Leibler divergence for Variational inference as;

$$KL[q(\theta)||p(\theta|y)] = \int_\theta q(\theta)\log\frac{q(\theta)}{p(\theta|y)}d\theta \qquad (2.20)$$

$$= -\left(\int_\theta q(\theta)\log\frac{p(y,\theta)}{q(\theta)} - \int_\theta q(\theta)\log p(y)\right)d\theta \qquad (2.21)$$

$$= -\int_\theta q(\theta)\log\frac{p(y,\theta)}{q(\theta)}d\theta + \log p(y)\int_\theta q(\theta)d\theta \qquad (2.22)$$

$$= -L + \log p(y) \qquad (2.23)$$

where L denotes the previously defined Variational lower bound. It is worth noting that the KL divergence from the true posterior is equal to the negative ELBO plus a constant. The normalization constraint $\int_{\theta} q(\theta)\, d\theta = 1$ yields as in Equation (2.23)

we rearrange the equations to obtain;

$$L = \log p(y) - KL\left[q(\theta) \| p(\theta|y)\right] \qquad (2.24)$$

KL divergence is always a nonnegative quantity (i.e., KL divergence $¿0.$ and zero when $p=q)$ (Kullback, 1997; Kullback & Leibler, 1951).

In Variational inference, the approximated posterior density, $p(y;q)$ is to a large extent tractable than the marginal likelihood $p(y)$,when density transform approach is employed. Tractability is achieved when q is restricted to a manageable class of densities and then maximized over that specific class. Considering Equation (2.24), the optimization of the Variational lower bound becomes achievable when maximizes $p(y;q)$ or minimizes the KL-divergence of the two quantities; that is *q* and $p(\theta|y)$ (Ormerod & Wand, 2010).

The Variational Bayes Perspective of Model Selection

A fundamental practical issue in many statistical investigations is the problem of model selection and has become an important part of regression analysis and inference. Model selection is a technique that statistically selects an appropriate model from a set of competing models, for a given dataset (Skonishi, Ando & Imoto, 2004). However, for several candidate models, the model that best fits a dataset is often not easily determined. For example, in the case of many complex observable datasets, and no one/single model explains the data-generating process for all datasets. The application of a model selection criteria is one of the most extensively utilized methods to handling change-point related problems. The number of components to include in a switch model is a challenging task to solve.

Fully Bayesian solutions are critical for model selection due to the fact that they include an implicit penalization term for the model complexity. Variational Bayesian methods for model selection directly aim at estimating the Bayesian integral even though in an approximated form. As a result, real or true posterior distributions over model parameters are substituted with approximated distributions referred to as Variational distributions that allow for tractable approximation.

Variational Information Criteria

The Variational Information Criteria is a developed approximation technique for marginal log-likelihood that offers an automatic model selection framework with tractable inference algorithms. Valente and Wellekens (2005a); You et al.(2014) are perhaps among the few major studies on Variational information criteria construction and applications for the purposes of model selection problem. The Variational Bayes approximation of the frequentist information criteria, that is, the deviance information criterion (DIC), Bayesian information criterion (BIC) (Spiegelhalter, Best, Carlin, & Linde, 2002), and its variants are of particular interest to us. These approximate approaches have become a key component of Bayesian solutions for model selection criterion and its related problems.

Approximation of difficult-to-compute probability densities is not trivial in modern statistics. This phenomenon is common in Bayesian statistics, where all inference about unknown quantities is framed as calculations with the posterior density. Most of the time, the intractability of some classes of complex posterior distributions has made it difficult to maximally extract information and put a sort of restrictions for Bayesian inference. These limitations seemed to be overcome by the advent of Variational Bayes approximate approach, which yields more tractable posterior distributions and enhances maximally extraction of useful information for effective and broader Bayesian inference.

In the context of selecting a linear regression model, You et al. (2014) proposed a Variational Bayes version of Bayesian Information Criterion (BIC) (Schwarz, 1978) and Akaike information criterion (AIC) (Akaike, 1973), referred to as Variational Akaike Information Criterion (VAIC) and Variational Bayes Information Criterion (VBIC). They asserted that the VAIC and VBIC have a high-performance rate and accuracy level and converge with the classical AIC and BIC solutions respectively, under a mild regularity condition (Yang, 2005).

Increasingly, an appreciable number of researchers and practitioners have modified existing methods and applied novel model selection techniques based on information-theoretic- approaches to the analysis of their change-point problem. One of the most extensively used strategies for detecting change-points is the model selection criteria (Jiang, 2015). According to Ninomiya (2015), A change-point model must contain an irregularity that calls for a different approach from asymptotic theory as a whole. According to McGrory and Titterington (2007), the DIC and BIC for Bayesian model selection can be extended to the change-point model framework, in our case, for linear systems, by using a Variational approximation. The researchers noted that estimates of these information criteria, particularly VAIC and VBIC, provide an additional model selection tool that can be used as alternative to MCMC sampling approaches for change-point detection in linear systems. It is important to note that VAIC and VBIC have never been used to solve change-point detection problems in the context of model selection in linear systems.

Variational Akaike Information Criterion

In recent years, research has been expanded to include complex structural models for better representation of real-world datasets, necessitating the development of some appropriate criterion to facilitate comparison of competing models. In model comparison, model complexity is a vital fundamental issue and trade-off between model fits and model complexity is required. Model selection has become a critical component in Bayesian solutions, and with increasingly complex models, a robust searching criterion is required. Spiegelhalter et al. (2002) proposed the Deviance Information Criterion(DIC), a model selection criterion that utilized Bayesian estimates of model complexity and model fit. This is arguably one of the most remarkable progress in the Bayesian framework for the Model Selection problem. From a Bayesian viewpoint, deviance information criterion (DIC) viewed as an approximate model selection method aims to explicitly balance the model complexity with fit to the data and server as a Bayesian version and a Multilevel modeling generalization of the AIC.

The Akaike Information Criteria (AIC) (Akaike, 1973) is a well-known criterion that defines the best model in a competing set as the one with the lowest AIC and is defined as:

AIC = -2Iog (maximum likelihood) + 2 (number of parameters)

The log-likelihood has been found to favor models with more parameters and to be penalized by the inclusion of the number of parameters term. In this perspective, the AIC is viewed as a criterion that balances model fit with model complexity. However, it has been claimed that the AIC selects a model with more free parameters than is required (Shibata, 1976). Spiegelhalter et al. (2002) present a model selection criterion that combines Bayesian evaluations of model fit or adequacy and the model's complexity.

The DIC is defined as

$$\text{DIC} = 2P_D - 2\log p(y|\tilde{\theta}) \qquad (2.25)$$

where *y* represents data and θ, represents unknown parameters vector on the parametric density $p(\cdot|\theta)$. Having $\tilde{\theta}$ as a Bayesian estimator for θ, That is, $\tilde{\theta}$ is the expectation of θ, $\mathbb{E}(\theta|y)$ and $p(y|\theta)$ is the likelihood function and

$$P_D = \{\mathbb{E}_{(\theta/y)}\left[-2\log p(y|\theta)\right]\} + 2\log p(y|\theta) . \qquad (2.26)$$

The model that minimizes the DIC is a logical choice for an appropriate model; that is, smaller values of DIC are preferred than large values of DIC.

The difference between the posterior mean of the deviation and the deviance evaluated at the posterior mean or mode, $\tilde{\theta}$, say, of the relevant parameters, is defined as the measure P_D. It represents estimate of the effective number of parameters in a model and is seen as a penalty term for model complexity, with the goal of preventing overfitting. In the DIC formulations, the term $-2\log p(y|\tilde{\theta})$ in Equation (2.25) represents a measure of the model's goodness of fit. When non-information or improper priors are used, the DIC might be helpful for model comparison. Explicit computation of the DIC necessitates knowledge about the model's posterior distribution, which is sometimes difficult to get precisely, especially in the case of a complex and intractable posterior distribution.

In practice, numerous studies have discussed the MCMC simulation-based approach to approximate the intractable distributions to obtain tractable posterior distributions for easily computation of DIC (Gelfand et al., 1990; Green, 1995; Robert, Ryden, & Titterington, 2000). You et al., (2014) on the other hand, provided a Variational Bayes approximation based on the work of (McGrory & Titterington, 2007) to approximate DIC. Their proposed procedure was presented substituting true posterior distribution $p(\theta|y)$ by the approximation distributions $q(\theta)$ and yielded to the Variational Akaike information criterion(VAIC) as an approximation of the DIC.

VAIC defined as;

$$\text{VAIC} \quad ¿-2\log p(y|\theta^{¿})+2P^{¿D} \qquad (2.27)$$

Where $\theta^{¿}= E_q(\theta)$ and $P^{¿D}=2\log p(y|\theta^{¿})-2E_q[\log p(y|\theta)]$. As pertaining to information criterion, the smaller values of VAIC are preferable.

Variational Bayesian Information Criterion

According to (You et al., 2014), the variational version of the Bayesian information criterion (BIC) of (Schwarz, 1978) is a valuable tool for modern Bayesian model selection. The Schwarz, (1978) Criterion, is a well-known model selection approach that selects more parsimonious models over more complex models. It is consistent in selecting an optimal model from a set of candidate models. This was accomplished by including a penalty term dependent on the number of free parameters evaluated in the model (Raftery, 1995; Schwarz, 1978). It is obtained as;

$$\text{BIC}\ \text{¿}[-2\log p(y|\hat{\theta}_{ML})+P\log(n)] \qquad (2.28)$$

Where the term $p(y|\hat{\theta}_{ML})$ in Equation (2.28) denotes maximum likelihood, P , the number of model parameters and n , the number of observations. According to Raftery (1995), the Bayesian information criterion (BIC) offers less assistance for additional effects or parameters and that, for effective and efficient applications of this popular criterion, it is necessary to determine the number of free parameters in advance, which is a challenging task where complex hierarchical models abound with intractable posterior probability. (You et al., 2014) considered Variational estimates updates for BIC for high dimensional distributions and complex hierarchical models by means of the Variational Bayes algorithms and called it Variational Bayesian information criterion (VBIC). It is obtained as:

$$\text{VBIC}\ \text{¿}[2(-E_q\log p_{-q}(y)+E_q\log p(\theta))] \qquad (2.29)$$

The goal is to update the approximation distributions all through the process to approximate the true posterior distributions while still functioning effectively when free parameters and observations are not clearly defined. According to You et al., (2014), the smaller values of the information criteria, specifically VAIC and VBIC, are desirable in the application of Variational Bayesian model selection.

Empirical Reviews

Detection and estimation of change-points or switch points in linear regression models has long been a concern for statisticians. A significant variety of strategies have been investigated to identify and estimate the change-point and its location in systems. The most commonly used include the Bayesian analysis test, maximum likelihood ratio test, stochastic process test, and non-parametric test, among others.

Empirical Studies on Change-Point Problem in the Bayesian Paradigm

Chernoff and Zacks (1964) investigated the first change-point issues in the Bayesian paradigm and estimated the real current mean of a normal distribution subjected to temporal changes. They used Bayesian inference as a technological tool to gain insight into the change-point problem, which led to simple robust procedures. The Bayesian estimate of a current mean for an a priori probability density over the entire real line was performed with a normal prior distribution for the mean and amount of change but, with a uniform prior distribution for the change-points. Broemeling (1972) studied Bayesian mean change-point in relation to a sequence of normally distributed random variables with fixed variability. For all of the parameters, he used a non-informative prior. It is important to note that in the Bayesian technique, everyone, without exception, has utilized just the uniform prior distribution for the switch and this study is no exception. In practice, detecting change-points is a complex process since number of change-points in the system is unknown. Smith (1975) proposed a Bayesian method to solving a change-point problem in a sequence of random variables when the underlying distribution changes. The posterior probability density of the possible change-points was used to make the inferences. With a numerical demonstration, he offered a comprehensive study on change-point detection and estimation for binomial and normal distributions.

Bayesian techniques have been used with single and multiple changes with known or unknown number of change-points. Gelfand, Hills, Racine-Poon and Smith (1990) investigated Bayesian analysis of change-points for a range of normal data models, including regression with unequal variances and a fixed number of change-points.

Stephens (2000) based his Bayesian analysis on multiple change-point issues where the number of the change-points remained uncertain. Examples of such techniques that assumed a fixed number of change-points are as follows: (Carlin, Gelfand, & Smith, 1992). Other writers approached the topic in such a manner that the data series comprised just one change-point.

Empirical Studies on Change-Point Problem for Linear Regression Models

A considerable number of studies in the literature examined the topic of change- point related with regression models, and several ways for verifying whether a change exist or not in the real values of a linear system have been extensively investigated. In the literature, two problems are encountered in detecting and identifying change-points in the linear regression model: jump discontinuity (change-point) and linked lines or continuous lines, both with known or unknown change-points (Chen, Chan, Gerlach & Hsieh, 2011). Quandt (1958) introduced the idea of abrupt switch in the simple linear regression models for the first time. He considered a number of pairs of observation $\left(y_i, x_i\right);\ i=1,2,\ldots n$ such that

$$Y_i=\alpha_1+\beta_1 x_i+e_i\ ;\ i=1,2,\ldots k \quad (2.30)$$

$$Y_i=\alpha_2+\beta_2 x_i+e_i\ ;\ i=k+1,k+2,\ldots,n \quad (2.31)$$

Where $\alpha_1,\alpha_2,\beta_1$ and β_2 represents the r model parameters, and *k* is the change-point at which the second structure of the model differs from the first, and is referred to in the literature as two-phase simple linear regression model. He used an iterative maximum likelihood estimation (MLE) approach to solve jump discontinuity change-point problem. The scheme selected rule was based on the maximum likelihood measures of *k* which corresponded to the maximum likelihood function. Thus, first, the switch point in the system is estimated and then, based on the switch point, the ML estimators of the regression parameters were derived. He also developed a test procedure based on an asymptotic (LR method) and a small sample test to investigate a change-point linear regression model, and he also used the methodology to analyze some real-world data meaningfully.

Two years later, Quandt (1960) with generated data, empirically investigated the distribution of $-2\log\lambda,$ where λ is the LR statistics and demonstrated that the term $-2\log\lambda$ is not a chi square distribution. On the basis, he also derived an empirical table of percentage points for $-2\log\lambda,$ to detect a single change-point in the system. Based on the numerical study, he observed that the logarithm likelihood ratio λ was independent on the length of the series. In his studies, he described a small sample property of tests like, a t-test and F-tests, based on the ‘’crossed residuals” which depend upon dividing the observations in to two non-overlapping groups. In this context, he remarked that it is preferable to divide the two groups arbitrarily than to use the MLE for the switching point as the dividing limit.

Finally, he recommended that the F-test be used to solving change-point problem in system changes. The two-phase simple linear model of Quandt (1958) was later extended to multi-phase multiple linear regression models by (Chow, 1960) on the assumption that the number of switch points is known (multiple change-points problems).

Sprent (1961) outlined a hierarchy of potential hypotheses of interest in relation to the two-phase simple linear regression model, and he also indicated that the outcome of an initial investigation should indicate the hypothesis to be considered. He utilized the F-test procedure to detect the change-point and explore least square estimates (LSE) of the parameters. He applied his proposed methods to real-life data consisting of three sets of data with each one relating to a two-phase regression or two-line regression model and discussed the problem of identifying the terminal point of buds of spur of the apple and concluded that the results correctly identified the original situation of the problem.

Farley and Hinich (1970) explored the abrupt change-point problem in a linear regression model by developing an LMPT for the stable and the switch linear regression models. They proposed the use of MCMC approach and observed that a shift that occurs at extreme ends of the data are challenged to be detected than a change-point that occurs in the neighbourhood of the center of the data. These works reviewed above all belong to the classical (non-Bayesian) approach to solving the change-point problems in linear systems.

Bacon and Watts (1971) presented an empirical model that may support both a smooth and abrupt change from one linear model to another. They defined a set of transition functions and considered the joining point and nature of the transition. They utilized the Bayesian technique and non-informative prior distribution for all model parameters. They obtained a joint and marginal posterior distribution of the parameters utilizing numerical integration techniques or exact Bayesian computations. They also applied their techniques to certain experimental data from R. A. Cook. They stated that their model is not sensitive enough to identify changes in slope. They also said that the analysis of their approach is applicable to multiple joint points and linear intersection functions.

Ferreira (1975) studied the abrupt change problem corresponding to two-phase linear model from the Bayesian viewpoint. He employed three different forms of priors for the change-point and a uniform prior for all of the other parameters. Mean biases and MES of the Bayesian estimates were computed and compared with those of the ML, estimates through Monte Carlo studies. The MSE of the Bayesian estimates was found to be uniformly smaller than those of ML estimates. He also analysed Quandt's data using the Bayesian methodology employing his three types of priors and got results which were almost identical.

Choy and Broemeling (1980) generalized the results of (Ferreira,1975) using Bayesian Methodology and a uniform prior for the change-point and a conjugate prior for all other parameters. They obtained the posterior distributions of all the parameters and demonstrated how to obtain point and interval estimators for the parameters. These techniques were applied to

(Quandt, 1958) data and they arrived at results that converged to that of Quandt's. They also derived the HPD region for the regression parameters whose posterior distributions were a mixture of t distributions. They conducted a number of numerical studies and noted that the estimates were quite close to the actual values. In addition to the work of Tsurumi (1980), Bacon and Watts (1971) investigated the gradual change problem in multiple linear regression system, with the assumption that the changes occurred in only one of the regression parameters. He considered non-informative priors for all parameters and a uniform prior distribution for the change-point location. The joint marginal posterior density of the change-point and the transition parameter were obtained using three types of transition functions. This proposed method was applied to U.S. petroleum data analysis to study the parameter shift in the supply and demand functions. He remarked that when data obey two different regimes and are treated as one homogeneous group would produce an erroneous result. This technique was also applied to the simultaneous equations model. As stated earlier in the work, approximate Bayesian methodologies are used for Bayesian inference in many situations where the conditional density, which is the key component for Bayesian inference, is intractable.

Holbert (1982) also used Bayesian methodology to solve the problem of two-phase or switching simple linear regression for both discrete and continuous cases, and he considered non-informative priors for all parameters and derived the parameter posterior distributions. He also obtained the highest posterior density (HPD) region for the point where the two regression lines intersected and described a test procedure for testing the hypothesis relating to the intersecting points and utilized the ratio of the posterior distribution's ordinates. He also illustrated the procedure by using some data from (Pool & Borchgrevink (1964) which was used earlier by (Hinkley, 1971) to illustrate the MLE method. On the basis of his numerical study, Holbert, (1982) concluded that his results compared quite favourably with those of Hinkley's. It should be remarked here that Hinkley's results were based on a certain chi-square approximation whereas no such approximation was involved in the result of Holbert. Again, Holbert (1982) investigated the change-point detection in multiple linear regression models from a Bayesian perspective.

The online detection of change-points in a linear regression model has been of much interest in many applications in recent years. Geng, Zhang, Huie, and Lai (2019) discussed the online change-point detection in linear regression settings and assumed a known pre-change coefficient of a linear model but an unknown post-change coefficient of the linear regression system. An efficient online scheme was considered to identified a change-point using both Classical and Bayesian formulations. They proposed a novel technique, the parallel-sum algorithm which, despite its modest computational complexity, ranked high in terms of the performance indicators of the corresponding parameter estimates. They concentrated on the impact of detection delay on the likelihood of detecting the true change-point. Liu, Zou, and Zhang (2008) discussed a nonparametric technique and empirical likelihood for detecting a change-point in a linear regression model's coefficient. They examined the effectiveness of empirical likelihood ratio test statistics versus the usual parametric likelihood technique. As a consequence, the two converge on an asymptotic null distribution. They evaluated the maximum empirical likelihood change-point estimator and discovered it to be reliable. The results also show that their proposed approach is sensitive and robust.

Empirical Studies on Change-Point Models as Model Selection Problem

In addition to the classic (likelihood ratio test, LRT) and Bayesian tests that dominated change-point analysis, the change-point problem has recently been viewed as a model selection problem, with the most commonly used methods being the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and their variants. These criteria-based information techniques and the variants serve as an additional tool for handling the change-point problem. In this regard, the change-point problem is considered to be a model selection problem. Changes in the dataset necessitate the selection of a better model to match the data from among the models given null and alternate assumptions. These approaches are primarily based on the parametric distribution of the parameters of interest, and the departure from specified parametric model may result in contradictions (Zhang & Siegmund, 2007).

Chen (1998) proposed an information criterion method such as the Schwarz Information Criterion (SIC) (Schwarz, 1978) to identify the location of the change in the switching of single and multiple linear regression models. Chen (1998) argued that the switch in the model was caused by changes in the coefficients of the linear regression parameters. It was instructive to use maximum likelihood estimates to compute the SIC values of the null model and alternative models for all free parameters, including the change-point location. The minimal information content concept has been extensively explored in order to identify the location of the change-point k in linear systems, if it exists. In the context of model selection, the Bayesian information criterion presented by (Schwarz, 1978) is defined as follows:

$$BIC_t = -2\log L(\hat{\Theta}) + t\log n \quad , \quad t = 1,2......T \qquad (2.32)$$

Where t represents the number of free parameters, L represents the maximum likelihood of the relevant model, and T's represents the number of all potential parameters in the model. The single model that minimizes the BIC is deemed the best model based on the principles of the minimal information content. That is, we accept H_0 if

$$BIC_{H_0}(n) < \min_{1<k<n} BIC_{H_1}(k) \qquad (2.33)$$

It thus implies that there is no change-point in the system. On the other hand, we reject H_1 if

$$BIC_{H_0}(n) > BIC_{H_1}(k) \qquad (2.34)$$

for some k ,and this signifies the existence of at least a single change-point in the model. As a result, the change-point location can be computed by $\hat{k}$ such that

$$BIC_{H_0}(n) = \min_{1<k<n} BIC_{H_1}(k) \qquad (2.35)$$

Some research explored the use of SIC approaches for single and multiple change-point detection in switching regression models, and they used the binary segmentation techniques proposed by (Vostrikova, 1981) and implemented by (Chen & Gupta, 2012).

In terms of statistical inference for model selection, with regards to detection and estimation of change-point, the SIC, and or other information criteria, such as MIC, (Basalamah et al., 2021) has proven to be a very efficient and effective methods for detection and identification of change-point and its location in linear systems. They utilize the same data as Holbert (1982) to demonstrate the SIC approach for determining the switching transition point in linear regression. They noted that their finding is consistent with Holbert's utilizing his technique. They noted that, while they and Holbert (1982) discovered the same change- point, Holbert's result is less affirmative, and that there is a propensity for a relative maximum at the endpoints when applying his Bayesian posterior density. They supported the argument that the SIC technique is an appealing approach that is simple to use, as well as assessing the change locations concurrently, which reduces the number of computations significantly when compared to the classic likelihood-ratio technique.

Furthermore, Cai, Said, and Ning (2016) and Ngunkeng and Ning (2014) examined the change-point with bathtub shape exponential model utilizing Schwarz information criteria (SIC). However, Chen, Gupta, and Pan (2006) pointed out that in the context of the change-point problem, Schwarz information criterion (BIC) techniques do not fully account for the contribution of the change location in the penalty term. That is, when there is a change at the extreme ends of the data, one of the two sets of the model's parameters based on SIC becomes entirely redundant. The reason for this is that in such cases, the complexity of the null and the alternative model is almost identical. Furthermore, search algorithms or detectors based on the SIC may fail to detect a change-point if it occurs at the model's extremes. They suggested a unique modified information criterion (MIC), which tries to improve on the traditional SIC-based criterion by refining model complexity as a function of change location in the context of the change-point problem.

Ngunkeng and Ning (2014) also said that the SIC approach may fail to identify changes at the model's extreme ends, and that there is a need for enough data to generate MLEs of the parameters, which interns is used to compute the values for the BIC statistics. In light of these findings, they developed a SIC-based (modified SIC) approach for detecting change-points in the skew-normal distribution's characteristics. Basalamah et al. (2021) presented a modified information criterion (MIC)-based test strategy for detecting points of change in a linear regression model with normal errors. This modified information criterion was considered as a modified version of Schwartz's information criterion (SIC) (Chen et al., 2006). The location of the change was included into the alternative model's penalty term during the derivation and calculation of the MIC values. They also applied the minimum MIC concept to determine the location of the change-point k in the simple linear regression model. Based on simulation experiments, the suggested procedure's performance was compared to that of a conventional SIC technique. Regarding performance metrics, they discovered that the powers of the two tests rise as sample size grows, and that the two powers are stronger when the change happens around the center of the data compared to the changes at the ends of the model. They found that the MIC technique outperformed the classic SIC procedure with different sample sizes and modification positions. They recommended the application of MIC approach as a highly competitive method for change-points detection. Finally, the proposed technique was successfully applied to NASA data to identify the change-point. Utilizing the NASA real-world calibration dataset, Mahmoud, Parker, Woodall, and Hawkins (2007) used parametric approaches to study the detection and estimation of change-points in simple linear models. They assumed a normal distribution of error. They observed the results indicated that there was a change in the interception. The application of Variational Bayes methods to change-point problems, particularly the linear change-point issue, has garnered considerable attention in recent times in the literature. Valente and Wellekens (2005a) discussed Bayes' Variational methods for detecting the point of changing speaker with an approximate learning algorithm. They developed a novel Variational Bayes lower bound difference statistic (VBLD) and compared its performance to the usual Bayesian information criteria (BIC). They observed that using the 1996 Hub4 experimental data set and the window approach, the proposed Bayes variation scheme improved detection performance by 7% compared to the usual BIC. This study will adapt and modify the VBLD proposed by (Valente & Wellekens, 2005a) and will explore its applications and those of a developed Bayes Variational information criterion, VAIC and VBIC to linear change-point systems.

Chapter Summary

This chapter presents procedures for detecting change-points in the Bayesian switching linear regression model with Variational Bayes computational approaches which take into account the location information for modeling change-point detection schemes. The chapter addresses the change-point problem which contains a known and unknown location of the changes. It also reviews the various Variational information criteria and the lower bound. Empirical review of works done in the Variational Bayes Paradigm are presented.

CHAPTER THREE

RESEARCH METHODS

Introduction

Data generating models can provide insight into the processes associated with how a given system performs in terms of the desired output based on the required inputs. As a result, the ability to model and calibrate the system using observed data is key to the discovery of important features underlying the system as well as its monitoring. This will enable the detection of mal-functioning component(s) of the system and provide pragmatic solutions. In this chapter, the statistical foundation in the Bayesian framework for developing the data generative models for performing change-point analysis of linear systems is formally introduced.

The chapter is organized as follows. First, Bayesian data generative models for switching and non-switching systems exhibiting linear patterns are formally introduced. Second, an introduction to Variational inference for Bayesian switching and non-switching is provided in brief. Variational Bays approach for obtaining tractable Bayesian marginal likelihoods for switching models are introduced. Application of the Variational Bayes ideas to models considered is outlined. Next, change-point detectors based on the Variational Bayes lower bound as well as Variational Bayes information and its Akaike information are developed. Finally, the chapter ends with the implementation of the developed Bayesian models and their illustrations using both simulation and refinery manufacturing process change-point dataset.

Bayesian Switching Model

Consider a sequence of response predictor pair (y_1,x_1), (y_2,x_2), . . ., (y_n,x_n) observed from some system. Suppose the response and the predictors are linearly dependent and that the nature of the underlying data pattern changes beyond a given point, say, k. The challenge is determining the true location of change-point in a data generative system. Assuming that the switch affects the mean but not the variance. We consider a Bayesian switching model.

$$H_0: y_i = X^{'}\beta + \varepsilon_i\,,\ i=1,2,\ldots,n \qquad (3.1)$$

$$\varepsilon \sim N(0,\sigma_{\varepsilon}^{2})\,, \qquad \sigma_{\varepsilon}^{2} \sim IG(a_{\varepsilon}^{0}, b_{\varepsilon}^{0})$$

$$\beta_0 \sim N_{(P+1)}(\mu_{\beta_0}, \textstyle\sum_{\beta_0})$$

where

$\beta^0 = [\beta_0^0\,,\ \beta_1^0,\ \ldots,\ \beta_p^0]\,,\ x_i = [1\,,\ x_{i1},\ x_{i2},\ \ldots,\ x_{ip}]$ against the alternative

$$H_1: y_i = x_i^{'}\beta + \varepsilon_i\,,\ i=1\,,2\,,\ldots,\ k.$$

$$y_i = x_i^{'}\theta + \varepsilon_i\,,\ i=(k+1)\,,(k+2)\,,\ldots,n \qquad (3.2)$$

$$\varepsilon \sim N(0,\sigma_{\varepsilon}^{2}), \qquad \sigma_{\varepsilon}^{2} \sim IG(a_{\varepsilon}, b_{\varepsilon})$$

$$\beta \sim N_r(\mu_\beta, \textstyle\sum_\beta), \qquad \theta \sim N_r(\mu_\theta, \textstyle\sum_\theta), \quad r=(p+1)$$

where $\beta^1 = [\beta_0^1,\ \beta_1^1,\ \ldots,\ \beta_p^1]\,,\ \theta = [\theta_0\,,\theta_1\,,\ldots,\ \theta_p]$

To complete the model specification, k is given a random treatment with uniform distribution

$$p(k)=\frac{1}{n-2p}, \quad k=p+1, \ldots, n-p.$$

Stacking the observation level models based on appropriate convention for notation, both the null and the alternative models can be expressed in vector notation. Now, write the design matrices corresponding to the switching and its non-switching counterpart as follows. Write X_k and X_{nk} for the design matrices associated with the switching model so that,

$X_k^1=[a, X_1, X_2, \ldots, X_k], X_j=[x_{i1}, x_{i2}, \ldots, x_{ip}], i=1, 2, \ldots, k$

$X_{nk}^1=[b, X_{(k+1)}, X_{(k+2)}, \ldots, X_{(n)}], X_j=[x_{i1}, x_{i2}, \ldots, x_{ip}],$

$i=(k+1), (k+2), \ldots, n$. Where *a* and b are column vectors of lengths k and $(n-k)$ respectively containing 1s.

Also let, $X=[c, X_1, X_2, \ldots, X_p], X_i=[x_{i1}, x_{i2}, \ldots, x_{ip}],$ $i=1, 2, \ldots, n$. Denotes the design matrix for the non-switching model defined by H_0 with c being a column vector of length *n*. The corresponding responses are also written in the following forms:

$Y_k=[y_1, y_2, \ldots, y_k], Y_{nk}=[y_{(k+1)}, y_{(k+2)}, \ldots, y_{(n)}]$ and

$Y=[y_1, y_2, \ldots, y_n]$

For H_1 and H_0, respectively. With the above convention for notation, the non-switching model (3.1) and the switching model (3.2) can be expressed as follows,

$$H_0: Y=X\beta^0+\varepsilon \tag{3.3}$$

$$H_1 : Y_k = X_k^1 \beta^a + \varepsilon_k \quad , \quad Y_{nk}^1 = X_{nk}^1 \theta + \varepsilon_{nk} \tag{3.4}$$

The compact forms of the models defined in Equations (3.3) and (3.4) were employed for the development of computational inference framework.

Variational Inference for Bayesian Model Switching

Bayesian inference is completely determined by the availability of the updated prior distribution based on conditional probabilities termed the posterior distribution. The number of free parameters for a given Bayesian model determines the model complexity given that the data is observed. Complex models usually have many free parameters; thus generates complex likelihood functions that are not easy to marginalize in terms of some parameters or all the parameters. This leads to intractable likelihood functions. As the likelihood function is one of the core elements of the Bayesian inference, its nature (i.e. tractable or intractable) directly influences the posterior distribution. Variational Bayes inference methods offer standard ways to handle intractable posterior distribution by varying the posterior using an easy to handle class of distributions and determining the departure from the true via optimizing Kullback-Leibler divergence (Attias, 2000; Bishop, 2006; Ormerod & Wand, 2010; Waterhouse et al., 1996).

Change-point problems present another area in which Variational Bayes methods can be explored due to complex likelihood functions generated by such problems. Considering Bayesian inference for change-point data, demands building comparative Bayesian solutions for both the change-point data and its non-change point counterpart. This is because the difference in the complexities of the two models can provide essential information for the existence of a change in the data under consideration. For a change-point data,

$\{Z,\omega\}$, where Z defines a response vector, which is linearly dependent on the design matrix, ω. Suppose the generative model underlying the data is of the form,

$$Z_k=\omega_k\gamma+u, Z_{k^{¿}}=\omega_{k^{¿}}\phi+u, \qquad (3.5)$$

Where$u$ $N(0,\sigma_u^2),\gamma,\phi,$ are set of regression parameters and u are measurement errors, $k^{¿}=n-k,k$ denotes the location of existing change-point. The corresponding change-point design matrices are defined as follows,

$$\omega_k=\begin{bmatrix} 1 & \omega_{11} & \omega_{12} & L & \omega_{1r} \\ 1 & \omega_{21} & \omega_{22} & L & \omega_{2r} \\ 1 & M & M & M & M \\ 1 & \omega_{k1} & \omega_{k2} & L & \omega_{kr} \end{bmatrix}, \quad \omega_{k^*}=\begin{bmatrix} 1 & \omega_{(k+1)1} & \omega_{(k+1)2} & L & \omega_{(k+1)r} \\ 1 & \omega_{(k+2)1} & \omega_{(k+2)2} & L & \omega_{(k+2)r} \\ 1 & M & M & M & M \\ 1 & \omega_{(n)1} & \omega_{(n)2} & L & \omega_{(n)r} \end{bmatrix} \qquad (3.6)$$

Because the location of the change-point is uncertain, it is treated as a parameter assumed random. Suppose the $k_{,}$ is discrete uniformly distributed between the second data point and the $(n-2)$ data point, so that it contributes a constant value over the interval $[2,n-2]$, otherwise 0. Let the set of unknown parameters be denoted by $\vartheta_c=(\gamma,\phi,\sigma_u^2,k)$. Furthermore, if there exist no break at position, k, the data can be modeled with single Bayesian model of the form;

$$Z=\omega\beta+u \qquad (3.7)$$

The set parameters corresponding to model (3.7) is of the form, $\vartheta_o=(\beta,\sigma_u^2)$.

It is straightforward to see that the difference in the free parameters in models in Equations (3.5) and (3.7) will define the underlying model complexity associated with the models. Thus, Bayesian inference schemes will differ accordingly. Let the defining prior distributions for ϑ_c and ϑ_0 be $g(\vartheta_c)$ and $g(\vartheta_0)$, respectively such that

$g(\vartheta_c)=g(\phi)g(\gamma)g(\sigma_u^2)g(k)$ and $g(\vartheta_o)=g(\beta)g(\sigma_u^2)$.

Posterior inference for ϑ_c and ϑ_0 can be synced using the updated prior distributions.

$$f(\vartheta_c|Z,\omega ¿=f(Z|\omega,\vartheta_c ¿\frac{g(\vartheta_c)}{h(Z|\omega ¿}¿$$

$$=\left\{f(Z_k|\omega_k,\gamma,\sigma_u^2 ¿\frac{g(\gamma)g(\sigma_u^2)g(k)}{h(Z_k|\omega_k ¿}¿\right\}¿$$

$$\alpha\left\{f(Z_k|\omega_k,\gamma,\sigma_u^2 ¿\frac{g(\gamma)g(\sigma_u^2)}{h(Z_k|\omega_k ¿}¿\right\}\left\{f(Z_{k^{¿}}|\omega_{k^{¿}},\phi,\sigma_u^2 ¿\frac{g(\phi)g(\sigma_u^2)}{h(Z_{k^{¿}}|\omega_{k^{¿}} ¿}¿\right\} \tag{3.8}$$

and

$$f(\vartheta_0|Z,\omega ¿=f(Z|\omega,\vartheta_0 ¿\frac{g(\vartheta_0)}{h(Z|\omega ¿}¿$$

$$¿f(Z|\omega,\beta,\sigma_u^2 ¿\frac{g(\beta)g(\sigma_u^2)}{h(Z|\omega ¿}¿ \tag{3.9}$$

where

$$h(Z_k|\omega_k ¿=\iint f(Z_k|\omega_k,\gamma,\sigma_u^2 ¿¿g(\gamma)g(\sigma_u^2)\,d\gamma d\sigma_u^2 \tag{3.10}$$

$$h(Z_{k^{¿}}|\omega_{k^{¿}} ¿=\iint f(Z_{k^{¿}}|\omega_{k^{¿}},\varnothing,\sigma_u^2 ¿¿g(\varnothing)g(\sigma_u^2)\,d\varnothing\, d\sigma_u^2 \tag{3.11}$$

$$h(Z|\omega ¿=\iint f(Z|\omega,\beta,\sigma_u^2 ¿¿g(\beta)g(\sigma_u^2)\,d\beta d\sigma_u^2 \tag{3.12}$$

Variational Bayes Marginal Likelihood for Switching Models

Bayesian switching models involving an appreciable number of parameters often yield Bayesian marginal likelihoods that are somewhat complex and thus complicate posterior inference if not render posterior inference intractable. It can be noticed that the integrals involved both the change point and non-change points models defined in Equations (3.10), (3.11) and (3.12) are multiple integrals. Thus, comes with some inherent complexity and difficulty obtaining the marginal likelihoods associated with the models. Based on Equations (3.8) and (3.9), the following posterior distribution expressions can be deduced for the alternative model (change-point model) and null model (non-change point model) respectively.

$$h'(Z|\omega)=\frac{f(Z|\omega,\vartheta_c)g(\vartheta_c)}{f(\vartheta_c|Z,\omega)}$$

$$=\frac{\{f(Z_k|\omega_k,\gamma,\sigma_u^2)f(Z_{k^{¿}}|\omega_{k^{¿}},\varphi,\sigma_u^2)g(\gamma)g(k)g(\varphi)g(k^{¿})g(\sigma_u^2)\}}{f(\vartheta_c|Z,\omega)} \tag{3.13}$$

$$h'(Z|\omega)=\frac{f(Z|\omega,\vartheta_0)g(\vartheta_0)}{f(\vartheta_0|Z,\omega)}$$

$$=\frac{f(Z|\omega,\beta,\sigma_u^2)g(\beta)g(\sigma_u^2)}{f(\vartheta_0|Z,\omega)} \tag{3.14}$$

The Variational Bayes method Ormerod and Wand, (2010) introduced in Chapter 2 provides a formal probabilistic approach to handle intractable marginal likelihoods by lower bound them with model specific tractable integrals. In what follows, the Variational Bayes techniques for treating complex Bayesian likelihoods are illustrated in brief.

Let $q(\vartheta_c)$ and $q(\vartheta_0)$ be two distributions belonging some distribution families such that

$$\int q(\vartheta_c)d\vartheta_c=1 \quad \text{and} \quad \int q(\vartheta_0)d\vartheta_0=1 \tag{3.15}$$

Then, the marginal likelihoods $h'(Z|\omega)$ and $h(Z|\omega)$ can be written as follows, using Equations (3.13) and (3.14)

$$\int q(\vartheta_c)\log h'(Z|\omega)\,d\vartheta_c=\int q(\vartheta_c)\log\left[\frac{f(Z,\omega,\vartheta_c)}{q(\vartheta_c)}\right]d\vartheta_c+\int q(\vartheta_c)\log\left[\frac{q(\vartheta_c)}{f(\vartheta_c|Z,\omega)}\right]d\vartheta_c$$

(3.16)

$$\int q(\vartheta_0)\log h(Z|\omega)\,d\vartheta_0=\int q(\vartheta_0)\log\left[\frac{f(Z,\omega,\vartheta_0)}{q(\vartheta_0)}\right]d\vartheta_0+\int q(\vartheta_0)\log\left[\frac{q(\vartheta_0)}{f(\vartheta_0|Z,\omega)}\right]d\vartheta_0$$

(3.17)

The decomposition logarithm of the corresponding marginal likelihoods Equations (3.16) and (3.17) into two integrals terms were made possible through the use of properties of the q distributions, $q(\vartheta_c)$ and $q(\vartheta_0)$: The integrals

$$KL_c=\int q(\vartheta_c)\log\left[\frac{q(\vartheta_c)}{f(\vartheta_c|Z,\omega)}\right]d\vartheta_c \qquad (3.18)$$

and

$$KL_0=\int q(\vartheta_0)\log\left[\frac{q(\vartheta_0)}{f(\vartheta_0|Z,\omega)}\right]d\vartheta_0 \qquad (3.19)$$

represents the Kullback-Leiber (KL) divergence (Kullback & Leibler, 1951) for the change-point model and the alternative model, respectively. More importantly, the following statements are true for KL_c and KL_0 (Ormerod & Wand, 2010):

$$KL_c\geq 0\;,\quad KL_0\geq 0 \qquad (3.20)$$

for all $q(\vartheta_c)$ and $q(\vartheta_0)$ densities.

Also if and only if, $q(\vartheta_c)=f(\vartheta_c|Z,\omega)$ and $q(\vartheta_0)=f(\vartheta_0|Z,\omega)$

then, $KL_c = 0$, and $KL_0 = 0$ (3.21)

By Equation (3.20), it is straightforward to deduce from Equations (3.16) and (3.17), the following inequalities,

$$\int q(\vartheta_c) \log h'(Z|\omega) d\vartheta_c \geq \int q(\vartheta_c) \log\left[\frac{f(Z,\omega,\vartheta_c) g(\vartheta_c)}{q(\vartheta_c)}\right] d\vartheta_c \tag{3.22}$$

$$\int q(\vartheta_0) \log h(Z|\omega) d\vartheta_0 \geq \int q(\vartheta_0) \log\left[\frac{f(Z,\omega,\vartheta_0) g(\vartheta_0)}{q(\vartheta_0)}\right] d\vartheta_0 \tag{3.23}$$

Obviously, the roles of the integrals

$$L_c(q) = \int q(\vartheta_c) \log\left[\frac{f(Z,\omega,\vartheta_c) g(\vartheta_c)}{q(\vartheta_c)}\right] d\vartheta_c \tag{3.24}$$

$$L_0(q) = \int q(\vartheta_0) \log\left[\frac{f(Z,\omega,\vartheta_0) g(\vartheta_0)}{q(\vartheta_0)}\right] d\vartheta_0 \tag{3.25}$$

become clearly visible based on Equations (3.22) and (3.23). In particular, $L_c(q)$ and $L_0(q)$ act as lower bounds on the logarithm of $\int q(\vartheta_c) \log h'(Z|\omega) d\vartheta_c$ and $\int q(\vartheta_0) \log h(Z|\omega) d\vartheta_0$ respectively.

Interestingly, the associated lower bounds obtained as a by-product of the Variational method are usually easy to compute with closed-form expressions existing for most models. The choice of the q densities has received much attention for which practical guidance and directions are well postulated in the literature. For further information on the choice of the Variational posterior distributions, readers are referred to (Ormerod & Wand, 2010). The maximization of $L_c(q)$ and $L_0(q)$ results in iterative algorithms termed the Variational Bayes fitting algorithm in the Variational Bayes literature, see, for example, (Ormerod & Wand, 2010; Bishop, 2006; Attias, 2000; Waterhouse et al., 1996).

Application of Variational Inference to Developed Models

Here in the study, we make an application of the Variational inference methods discussed in the first few sections to the models developed. Inference in the Variational Bayes (VB) context is based on the Variational posterior distributions assumed appropriate for approximating the true posterior distribution. Obtaining the optimal Variational posterior distribution is very key in Variational decision making in practice, taking into account the observed data and the assumed prior distributions. In VB techniques, optimum Variational posterior distributions are achieved by either maximizing a Variational lower bound on the logarithm of marginal likelihood or minimizing a Kullback-Leibler divergence for the Variational posteriors. This yields an iterative optimization algorithm termed the Variational Bayes algorithm. For our switching and non-switching models, setting $\vartheta_c=\left(\beta^a,\theta,\sigma^2_{\in},k\right)$ and $\vartheta_0=\left(\beta,\sigma^2_{\in}\right)$, it is possible to consider the VB ideas. Now, applying the Variational approximation technique, to the true change-point and non-change-point posteriors,

$$p\left(\beta^a,\theta,\sigma_\varepsilon^2,k|Y\right)=\frac{p(Y|\beta^a,\theta,\sigma_\varepsilon^2,k)\,p(\beta^a)\,p(\theta)\,p(\sigma_\varepsilon^2)\,p(k)}{p(Y)} \tag{3.26}$$

$$p\left(\beta,\sigma_\varepsilon^2|Y\right)=\frac{p(Y|\beta,\sigma_\varepsilon^2)\,p(\beta)\,p(\sigma_\varepsilon^2)}{p(Y)} \tag{3.27}$$

respectively, the following Variational approximations are adopted

$$q\left(\vartheta_c\right)=q\left(\beta^a\right)q\left(\theta\right)q\left(\sigma_\varepsilon^2\right)q\left(k\right) \tag{3.28}$$

$$q\left(\vartheta_0\right)=q\left(\beta\right)q\left(\sigma_\varepsilon^2\right) \qquad (3.29)$$

where

$$q(\theta)=N\left(\mu_\theta^q,\textstyle\sum_\theta^q\right),\ \ q\left(\beta^a\right)=N\left(\mu_{\beta^a}^q,\textstyle\sum_{\beta^a}^q ¿\right),¿$$

$$q\left(\sigma_\varepsilon^2\right)=IG\left(a_\varepsilon^q,b_\varepsilon^q\right),\ \ q\left(\beta\right)=N¿¿$$

$$p(k)=\begin{cases}\frac{1}{n-2q}, k=q+1,q+2,\ldots,n-q\\ 0, otherwise\end{cases}$$

The corresponding Variational lower bounds based on Equations (3.24) and (3.25) can be expressed as follows.

$$L_c^k(q)=\sum_k \iiint q\left(\beta^a\right)q\left(\theta\right)q\left(\sigma_\varepsilon^2\right)q\left(k\right)\log\left(M^a\right)d\beta^a\, d\theta d\sigma_\varepsilon^2 \qquad (3.30)$$

$$\alpha \iiint q\left(\beta^a\right)q\left(\theta\right)q\left(\sigma_\varepsilon^2\right)\log\left[M^a\right]d\beta^a\, d\theta d\sigma_\varepsilon^2 \qquad (3.31)$$

where

$$M^a=\frac{p\left(\beta^a,\theta,\sigma_\varepsilon^2,k|Y\right)p\left(\beta^a\right)p\left(\theta\right)p\left(\sigma_\varepsilon^2\right)p\left(k\right)}{q\left(\beta^a\right)q\left(\theta\right)q\left(\sigma_\varepsilon^2\right)q\left(k\right)}$$

$$L_0(q)=\int q\left(\beta\right)q\left(\sigma_\varepsilon^2\right)\log\left[\frac{p\left(\beta,\sigma_\varepsilon^2|Y\right)p\left(\beta\right)p\left(\sigma_\varepsilon^2\right)}{q\left(\beta\right)q\left(\sigma_\varepsilon^2\right)}\right]d\beta d\sigma_\varepsilon^2 \qquad (3.32)$$

The lower bounds specified in Equations (3.30) and (3.31) are computable and closed-form expressions are available for the development of optimization algorithms. Details of the derivations and computations are outlined in the Appendix A.2. Optimal parameter values are obtained via iterative Variational algorithms based on the optimization of Equations (3.30) and (3.32). The iterative algorithm for the non-switching model (3.1) is outlined in algorithm 1. Algorithm 2 gives the iterative algorithm corresponding to the switching model (3.2).

Algorithm 1: Variational Algorithm for Null Model (1)

```
$Initialize: b_{\varepsilon}^{q}=0.5, \sum_{\beta}^{q} \text{¿} I, \mu_{\beta}^{q}=\mu_{\beta 0}+\sqrt{\operatorname{var}(y)}$

Set $a_{\varepsilon}^{q} \leftarrow \frac{n}{2}+b_{\varepsilon}^{0}$, $tol<\infty$.

Do until the change in $L_0(q)<tol:$

    • $\sum_{\beta}^{q} \leftarrow \text{¿¿}$
    • $\mu_{\beta}^{q} \leftarrow \sum_{\beta}^{q} \text{¿¿}$
    • $b_{\varepsilon}^{q} \leftarrow \left[\left(y-X\mu_{\beta}^{q}\right)'\left(y-X\mu_{\beta}^{q}\right)+tr\left(X'X\sum_{\beta}^{q}\square\right)\right]$

End
```

Algorithm 2: Variational Algorithm for Switching Model (2)

```
$Initialize: b_{\varepsilon}^{q}=b_{\varepsilon}, \sum_{\beta_a}^{q}=\sum_{\beta_a}\text{¿}, \sum_{\theta}^{q}\text{¿}=\sum_{\theta}\text{¿}, \mu_{\beta_a}^{q}=\mu_{\beta_a}, \mu_{\theta}^{q}=\mu_{\theta}\text{¿¿¿}$

Set $a_{\varepsilon}^{q} \leftarrow \frac{n}{2}+b_{\varepsilon}^{0}$, $tol<\infty$.

For $k$ = 2,…, $n-2$, do until the change in $L_c^k(q)<tol:$

    • $\sum_{\beta_a}^{q} \leftarrow \text{¿¿¿}$
    • $\mu_{\beta_a}^{q} \leftarrow \sum_{\beta_a}^{q} \text{¿¿¿}$
    • $\sum_{\theta}^{q} \leftarrow \text{¿¿}$
    • $\mu_{\theta}^{q} \leftarrow \sum_{\theta}^{q} \text{¿¿}$
    • $b_{\varepsilon}^{q} \leftarrow \text{¿¿}$
    • $D=\left(y_{nk}-X_{nk}\mu_{\theta}^{q}\right)'\left(y_{nk}-X_{nk}\mu_{\theta}^{q}\right)+tr\,\text{¿¿}$

End
```

Variational Bayes Change-Point Detector

This part of the study focuses on the development of appropriate Variational Bayes schemes for detecting switches in linear systems based on the Variational information criteria introduced in Chapter 2. Key to the schemes considered here is the Variational lower bound and information criteria obtained based on the Variational Bayes approximation. The Variational Bayes method usually results in a bye-product termed the Variational lower bound with the capability to serve as a model selection tool. As a result, it has enjoyed extensive usage in myriad scientific problems within the Bayesian modeling community. The computation of Variational information criteria considered here are light-weight, thus does significantly affect the computation complexity of the developed Variational detection algorithms.

Variational Lower Bound-Based Approach

The ability to quantify the information content of a model allows easy development of statistics for assessing some vital properties of the developed model. The Bayesian approach to calibration of information content of a model assumed for a given data via probabilistic modeling provides a principled technique for quantifying the complexity of the data generative model. For a change-point model, the difference in model complexities associated with the non-change-point model and the change-point model provide insights on the development of a formal approach for assessing the presence or otherwise of switching in the underlying structures in the data. This insight on the above difference can be preserved, adopting an appropriate modeling framework that has the potential to properly quantify and keep intact the complexities in both models. Thus, allowing developed fitting algorithms to inherit such features for precise inference on change-point parameters.

The Variational Bayesian formalism allows tractable calibration and quantification of complex model evidence (information) using Variational probability models, simplifying Bayesian decision-making in complex statistical models. With the above and other appealing features of Variational Bayes such as its fast and deterministic nature, it has been widely used in many scientific applications. However, its application of linear change-point problems has received little attention. In particular, a recent application of the Variational Bayesian (VB) technique in analysing speaker change is seen in Signal processing (Valente & Wellekens, 2005a). Valente and Wellekens (2005a) explored the utility of the VB methods to speaker change-point detection with an approximate Variational learning algorithm that improved detection performance of the usual Bayesian Information Criterion (BIC) by 7%; using Hub4 1996 experimental dataset. Their approach focused on Variational lower bound difference between the models generated by the null model $L_0(q)$ and that generated by the alternative model $L_c(q)$ being positive. That is inference on the existence of a change-point was based on the statistic,

$$L_d^k(q) = L_c^k(q) - L_0(q) \tag{3.33}$$

for all possible values of *k* say, $k = k_0, \ldots, n - k_0$ and a decision rule, $L_d > 0$ suggesting existence of change-point in the system. In addition, the set of possible change-point locations are partitioned into subsets termed windows. Detection is done window-wise starting with an initial window and varied according to a pre-specified window length until a change is detected and identified. Although, the above proposal is appealing, it is limited in the following ways. It appears that two or more different conditions may attract the same decision. For example,

1. if $L_c^k(q)=L_0(q)$, $L_d^k(q)=0$, $k=k_0$, …, $n-k_0$
2. if $L_0(q)>L_c^k(q)$, $L_d^k(q)<0$, $k=k_0$, …, $n-k_0$

These two conditions seem to yield the same decision. However, there was no clear exposited in their work. Also, the use of window and the window length variation can be problematic for complex change-point problems, since it can introduce extra computational cost, leading to an increased computational burden associated with change-point analysis. In this regard, we propose the Variational lower bound difference-based as in Equation (3.33) decision rule in a unique way, for detection and estimation of a single change-point in linear change-point data. Let L_m^k denotes the maximum value of $L_c^k(q)$ for each possible value of k as specified in Equation (3.33). Further, let L_m^s denotes the smallest value of $L_c^k(q)$ .Also, let $k^{¿}$ be the unique k that generated L_m^s; so that we can write

$$L_m^s = L(k^{¿}) = \min_{k_0<k<(n-k_0)}\left(L_m^k\right)$$

Then, the VB lower bound change-point detector is based on the statistic

$$L^{¿} = L_m^s - L_0(q) \qquad (3.34)$$

with the decision rule that there exist a change-point if $L^{¿}>0$ if not, then no switch exist in the datasets. The location of the change-point is then estimated by $\hat{k}$ such that $\hat{k}=k^{¿}$ (3.35)

Algorithm 3 outlines the Variational lower bound difference-based change-point detector algorithm.

Algorithm 3: Variational Lower Bound-Based Detector Algorithm

Initialize: $k=k_0,\ b_\varepsilon^q=b_\varepsilon,\ \Sigma_{\beta_a^q}=\Sigma_{\beta_a},\ \Sigma_\theta^q=\Sigma_\theta,\ \mu_{\beta_a}^q=\mu_{\beta_a},\ \mu_\beta^q=\mu_\theta \Sigma_\beta^q=I,$ $\mu_\beta^q=\mu_{\beta_0}+\sqrt{\operatorname{var}(y).}$

- Run algorithm 1 and obtain $L_o(q)$
- For $k=k_0,\ldots,(n-k_0),$ execute algorithm 2.
- Obtain $L_c^\upsilon(q)=\left(L_c^{k_0}(q),L_c^{k_1}(q),\ldots.,L_c^{n-k_0}(q)\right)$.
- Compute $L(k^{¿})=\max\left(L_c^\upsilon(q)\right)$, and set $L_m=L(k^{¿})$.
- Compute $L^{¿}=L_m-L_o(q).$
- If $L^{¿}>0$, change-point exist. Set $k=k^{¿}$ and use variational posterior corresponding to $k^{¿}$ for parameter inference. Otherwise, there is no change-point, use output of algorithm 1 for inference.

End

Variational Information-Based Approach

The Variational information criterion is adopted to develop comparable VB information-based detectors for learning the linear change-point problem. The Variational information criterion is a well-known model selection tool in the Bayesian context. In particular, the Variational Aikake information criterion (VAIC) and Variational Bayes information criterion (VBIC) have been extensively applied for model selection in diverse fields. For detailed information on the above model selection tools in terms of derivation, computation, and application, readers are referred to (You et al., 2014; McGrory & Titterington, 2007; Spiegelhalter et al., 2002). The VAIC computes the Variational approximation to the usual deviance information criterion proposed by (Spiegelhalter et al., 2002). On the other hand, the VBIC can be considered as Variational formulation of the Bayesian information criterion (BIC) of (Schwarz,1978).

The typical information-based model selection inference is centred on the minimum information value. Although the application of the above in model selection problem may seem simple, its use in switching model analysis is non-trivial due to differences in the information under the non-switching and switching models. Nevertheless, it must be pointed out that the Schwarz Information Criterion (SIC) introduced by Schwarz (1978) has been employed in change-point problems (Chen & Gupta, 2012). We propose the use of a Variational information ratio statistic-based schemes for seeking a solution to the linear change-point problems. Let $VAIC_0$ and $VAIC_a$ denote respectively, the Variational Bayesian information for the null and alternative models. Then, we consider the following VAIC and VBIC schemes for change-point detection and estimation.

$$\delta_A^{¿}=\frac{VAIC_s}{VAIC^o} \tag{3.36}$$

$$\delta_B^{¿}=\frac{VBIC_s}{VBIC^o} \tag{3.37}$$

where

$$VAIC_s = VAIC\left(k^{¿}\right) = \min_{k_0<k<(n-k_0)}\left(VAIC^a(k)\right)$$

and

$$VBIC_s = VBIC\left(k^{¿}\right) = \min_{k_0<k<(n-k_0)}\left(VBIC^a(k)\right)$$

The decision rule based on the above information schemes are such that existence of a change-point if $\delta_A^{¿}<1$ and $\delta_B^{¿}<1$ for VAIC and VBIC respectively or otherwise, there is no switching in the data. Further, change-point location is estimated in both information based schemes by k such that $\hat{k}=k^{¿}$.

The computations of the VAIC and VBIC are achieved when the Variational algorithms 1 and 2 have converged. Details of the derivation and computation of the VAIC and VBIA are provided in the Appendix A 3. Algorithm 4 gives the Variational information change-point detection algorithm.

Algorithm 4: VAIC and VBIC Ratio-Based Detection Algorithm

Initialize: $k=k_0,\ b_\varepsilon^q=b_\varepsilon,\ \Sigma_{\beta_a^q}=\Sigma_{\beta_a},\ \Sigma_\theta^q=\Sigma_\theta,\ \mu_{\beta_a}^q=\mu_{\beta_a},\ \mu_\beta^q=\mu_\theta\,\Sigma_\beta^q=I,$ $\mu_\beta^q=\mu_{\beta_0}+\sqrt{\operatorname{var}(y)}.$

- Run algorithm 1 and obtain $VAIC^0, VBIC^0$
- For $k=k_0,$ . . ., $(n-k_0),$ execute algorithm 2.
- Compute $VAIC^a(k), VBIC^a(k),$ and set

$$VAIC^{\upsilon}=\left(VAIC^a(k_0), VAIC^a(k_1), \ldots, VAIC^a(n-k_0)\right)$$

$$VBIC^{\upsilon}=\left(VBIC^a(k_0), VBIC^a(k_1), \ldots, VBIC^a(n-k_0)\right)$$

- Compute

$$VAIC_s = \min\left(VAIC^a(k)\right)$$

$$VBIC_s = \min\left(VBIC^a(k)\right)$$

- Compute

$$-\delta_A^{¿} = \frac{VAIC_s}{VAIC^o}$$

$$-\delta_B^{¿} = \frac{VBIC_s}{VBIC^o}$$

- If $\delta_A^{¿},\ \delta_B^{¿} < 1,$ change-point exist. Set $\hat{k} = k^{¿}$ and use Variational posterior corresponding to $k^{¿}$ for parameter inference. Otherwise, there is no change-point, use output of algorithm 1 for inference.

End ___

Performance Evaluation

In this part of the study, the statistical measures of performance of the proposed methods are considered. In particular, we consider more than one performance measures spanning accuracy of estimation of change-point parameters, model fitting performance, change-point detection performance etc. For the parameter estimation assessment, the focus will be on the accuracy of parameter estimates quantified in statistical measures such as mean absolute error (MAE) and root mean squared error (RMSE). In addition, their relationship with the change-point location *k* will be subjected to analysis to establish dynamics of the errors as the change detection progresses. The MAE and RMSE of change-point parameters were computed based on N = 30 randomly generated datasets. The corresponding MAE and RMSE of the parameters, β^a and θ are defined as follows.

$$MAE(\beta^a)=c_0\sum_{j=1}^{m}|\beta_{j^a}-\beta_{j^a}|, \quad MAE(\theta)=c_0\sum_{j=1}^{m}|\theta_j-\theta_j| \tag{3.38}$$

$$RMSE(\beta^a)=\sqrt{c_0\sum_{j=1}^{m}(\beta_{j^a}-\beta_j^a)^2}, \quad RMSE(\theta)=\sqrt{c_0\sum_{j=1}^{m}(\theta_j-\theta_j)^2} \tag{3.39}$$

Where $c_0=\frac{1}{m}$, β_j^a and θ_j are the posterior means of β^a and θ respectively, for the jth dataset.

The use of the Variational algorithms is dependent on their ability to perform the required task as demanded by the fundamental features of the Variational Bayes theory. Since the VB algorithms are constructed based on the maximization of the Variational objective function termed the lower bound, the nature of the lower bound should provide valuable information for assessing whether the developed algorithm is doing what is expected of it. In this regard, the appropriateness of the developed VB fitting algorithms will be assessed for both the switching model and the non-switching model. This assessment will be fully conducted in the first illustrative example. In the subsequent examples, the focus will be on the detection of change-point, its calibration, and identification. The performance assessment will be based on both simulation studies and real data applications. For the simulations, different scenarios will be considered spanning fixed known change-point locations, random but known change-point locations. Details of the above considerations in terms of simulations settings and assumptions are outlined under each case in this chapter.

Implementation

The implementation of the proposed methods can be done in two major ways. The first way is through the outline of the algorithms provided in pages 67 and 68, and so executing algorithms 1 and 2 separately. This might impact on the computational burden as Variational algorithm 1 and algorithm 2 are implemented twice. An appealing alternative solution with the potential for reducing the computational challenges involves combining the execution of algorithms 3 and 4 in a single run, in which algorithm 1 and algorithm 2 are not run separately. For the programming of the algorithms into executable codes, the R statistical software was utilized. The running of the written codes was done using an Intel (R) Core (TM) i7, 6700 processor Windows PC 3:40 GHz workstation.

Application of Methods to Datasets

The thesis at this point focuses on the application of the developed methods to datasets in order to assess their fitting performance as well as their practical applicability.

Example 1: Simulated Study Based on Fixed *k*

In this example, we consider the first simulation study to illustrate the developed methods for linear change-point analysis. The simulation considered here is centred on the following assumptions. First, we assume that the change-point location in the linear system is known and fixed. Second, we also assume that there exists a linear system in which the response is predicted by four (4) predictors. Based on the above assumptions, our linear change-point dataset is generated as follows. We considered a data size of n = 300; the actual change-point location, $k = 160$; the number of predictors, $p = 4$, and the number of different change-point systems to be N = 30: Next, we generate the true change-point model parameters and such that

$$\beta_t^a \sim^{iid} N(0.55, 0.05^2) \;,\; \theta_t \sim^{iid} N(0.95, 0.05^2) \quad t=1,2,\ldots,5$$

so that an intercept term is considered. Based on the assumed value of k; the linear predictors are generated using the alternative model (4), such that

$$X_k^1 = [1, x_1, x_2, \ldots, x_k]' \;,\; x_j = [x_{j1}, x_{j2}, x_{j3}, x_{j4}] \quad j=1,2,\ldots,k$$

$$X_{nk}^1 = [1, x_{(k+1)}, x_{(k+2)}, \ldots, x_{(n)}] \;,\; x_i = [x_{i1}, x_{i2}, x_{i3}, x_{i4}]' \;,\; i=(k+1),(k+2),\ldots,n$$

where

$$x_{j1} \sim^{iid} N(0.5, 0.01) \;,\; x_{j2} \sim^{iid} N(0,1) \;,\; x_{j3} \sim^{iid} N(0.85, 0.05) \; x_{j4} \sim^{iid} N(0.85, 0.05)$$

$$x_{i1} \sim^{iid} N(0.65, 0.01), \; x_{i2} \sim^{iid} N(0,1), \; x_{i3} \sim^{iid} N(0.85, 0.02^2) \; x_{i4} \sim^{iid} N(0.9, 0.02)$$

Finally, given β^a, θ and the design matrices X_k^1 and X_{nk}^1, the vector of observations $Y = [y_1, \ldots, y_k, y_{(k+1)}, \ldots, y_n]'$ is generated using model (2) in Equation (3.4).

$$y_i \sim N\left((X_k^1 \beta^a)_i, \; 0.002^2 \right) \;,\; i=1,2,\ldots,k$$

$$y_i \sim N\left((X_{nk}^1 \theta)_i, \; 0.002^2 \right) \;,\; i=(k+1),(k+2),\ldots,n.$$

A total of N different linear change-point datasets of each of size n = 300 are simulated based on the above settings for the assessment of the develop methods in this example. On the prior hyper parameter value setting, we consider the following change-point and non-change-point normal prior models, $\beta^0 \sim N_r\left(\mu_{\beta_0}, \Sigma_{\beta_0}\right)$, $\beta^a \sim N_r\left(\mu_\beta, \Sigma_\beta\right)$, $\theta \sim N_r\left(\mu_\theta, \Sigma_\theta\right)$, we set $\mu_{\beta_0} = 0.95$, $\mu_\theta = 0$, $\Sigma_{\beta_0} = \Sigma_\theta = 100 I_r$ so that vague normal priors are utilized.

The remaining inverse gamma prior hyper parameters were set as $a_\varepsilon^0 = a_\varepsilon = 5,$ and $b_\varepsilon^0 = b_\varepsilon = 0.1.$ We run algorithms 1,2, 3 and 4 using the above simulated datasets. In particular, the run of algorithms 1, 2 will be used to assess the appropriateness of developed VB algorithms for the task of linear change-point detection and estimation. The simulation study aided in training the models that were developed before the real dataset was applied.

Example 2: Real Data Application: Manufacturing Process in Refinery

In this example, we consider the application of the developed methods to the analysis of change-point in linear systems generated in the manufacturing processes. In particular, the focus is on a secondary data on manufacturing process in the refinery. The dataset under consideration was obtained as a result of an investigation of a refinery's manufacturing process, in which an octane rating of a specific petroleum product was considered as a function of three raw materials and a variable that characterized the manufacturing conditions. The three raw materials considered are labelled material 1; material 2 and material 3. The resulting dataset is a matrix of dimension 84¿5 on variables characterized as octane rating, amount of material 1, amount of material 2, amount of material 3, and manufacturing conditions. It was of interest to learn the linear relationship existing among the above variables. Most importantly, the question of interest was whether octane rating is predicted by the remaining variables.

Chapter Summary

This chapter aims to presenting the foundation of the statistical methods considered in developing the data generating models namely, the linear switching and non-switching models with Bayesian computation approaches for modeling and inference framework for change-point datasets. The chapter starts by developing non-switching model and switching linear model which contain a single fixed change-point in a linear system which follows Bayesian process. Capturing switching and non-switching information and incorporating such information into statistical models is non-trivial. However, the use of prior probability models and hierarchical modeling concepts within the assumed Bayesian framework allowed an easy encoding of such information and its integration into the assumed models.

The chapter also considered Variational Bayes inference computational approaches which incorporate the switch information for modeling change-point has been introduced and thoroughly discussed. The chapter further considered the Variational Bayes marginal likelihood for switching models and this was used to obtained the Variational lower bounds estimates and also to compute the estimates for the Bayesian information Criteria. The chapter also considered Variational change-point detectors by the two approaches mentioned in the study, namely the Variational lower bound based approach and the Variational information based approach. It also presents the developed Variational Bayes algorithms for the developed models and detectors for the evaluation of the performances using methods of inference developed further in the study. Lastly, the chapter presented the data generative processes for both secondary and simulated data and analysis using the R statistical software.

CHAPTER FOUR

RESULTS AND DISCUSSION

Introduction

The performances of the proposed methods are assessed in comparison with VAIC and VBIC-based methods through simulation studies and real datasets application. In this chapter of the study, we focus on the results obtained from the application of the developed methods to both simulated and real datasets of Bayesian linear switching systems. The real data application involves change-point dataset of linear systems generated in the manufacturing processes. The simulations are based on the statistical models developed in Chapter Three. In addition, the discussions on the results are considered in brief. The presentation begins with the simulation, followed by that of the real data application. All the codes were written in R statistical software and run on an Intel (R) Core (TM) i7 – 6700 processor Windows PC 3.40 GHz workstation.

Application to Simulated Data

We considered the simulation based on the following assumptions.

1. We assume that the change-point location in the linear system is known and fixed.
2. We assume that there exists a linear system in which the response is predicted by four (4) predictors.

Based on the above assumptions, our linear change-point dataset is generated as follows: We set the data such that a total of N = 30 different linear change-point datasets of each of size n = 300; the true change-point location, k = 160; the number of predictors, p = 4, and the number of different change-point systems are simulated based on the above settings for the assessment of the develop methods in this example.

Figure 1 and Figure 2 show the nature of the simulated switching model for twelve randomly selected datasets out of the 30 datasets considered. It can be observed that all the datasets selected exhibit the linear pattern defined by the simulation settings. In particular, the generated regression parameters values assumed fixed and known to be the true values resulted in a positively oriented pattern. Also, the clustering of the points within the switches exhibit some obvious marginal differences with Figure 1 recording a bit of variant pattern for plots from that of Figure 2. In Figure 2, some datasets for example, plots 1, 2 and 5, numbering row-wise and clockwise seem to show unclear location of change-point. However, this can be attributed to the random generation.

Furthermore, although the true location of the switch in each dataset is the same, the varying nature of the observed patterns within each switch existing among the different datasets is clearly evident. In summary, the appropriateness of the simulated datasets for the illustration of the utility of the proposed methods is clearly evident based on the nature of the models.

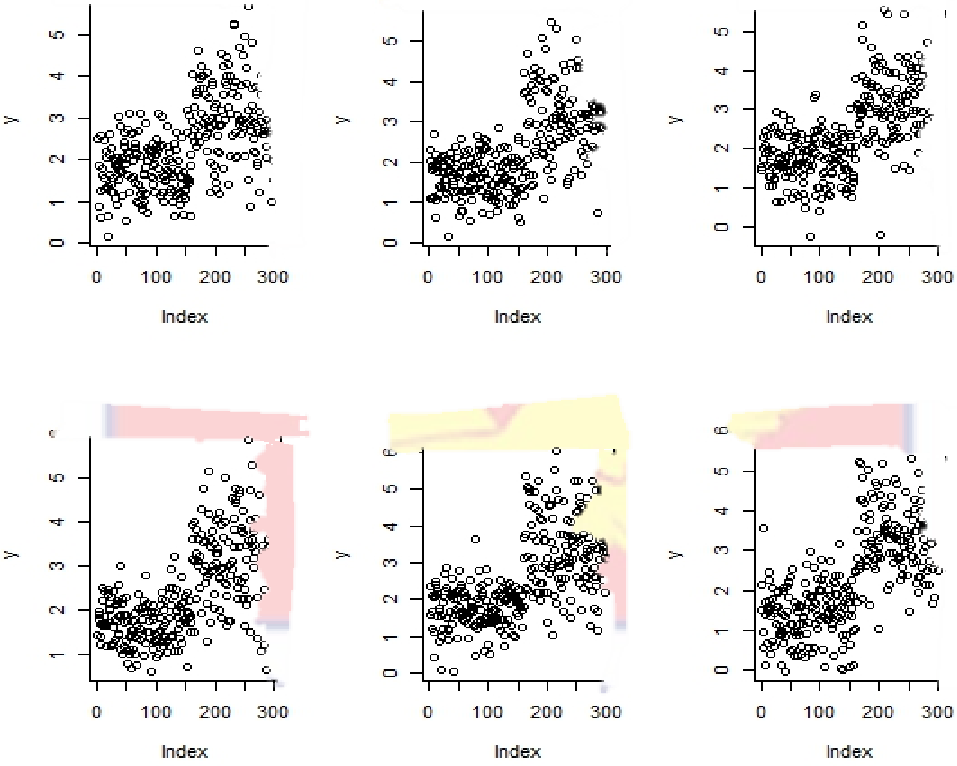

Figure 1: Randomly selected sample simulated datasets for switching linear system.

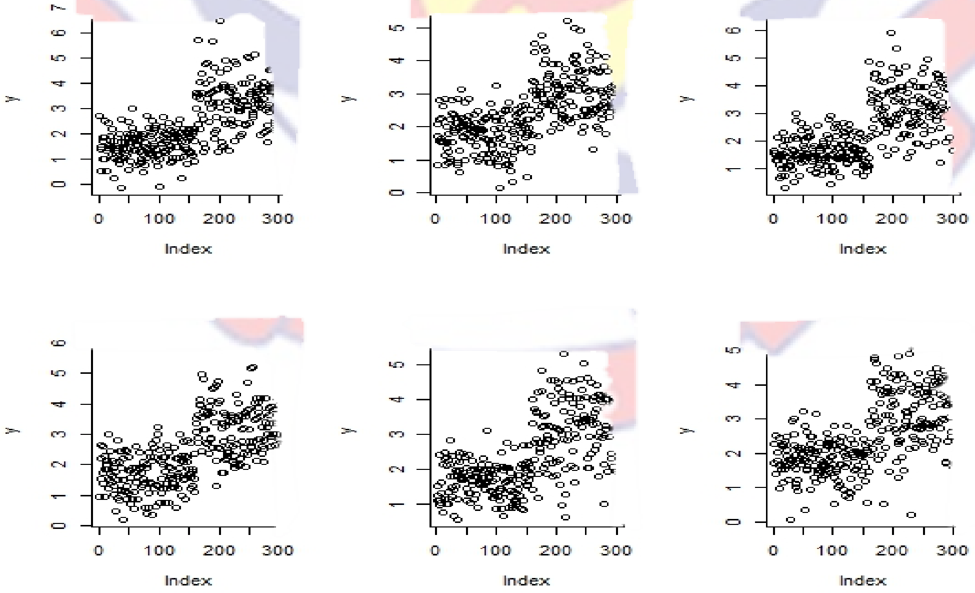

Figure 2: Randomly selected sample simulated linear system.

Example 1: Simulated Based on Fixed *k*

This subsection presents the results for simulated linear switching datasets. First, we access the applicability of the Variational Bayes methods for linear change-point analysis. The Variational Bayes algorithms considered in this study were built based on the maximization of Variational lower bound on the marginal likelihood. Thus, a simple analysis of the pattern exhibited as the iteration progresses can help to check whether the algorithm is appropriate before it is utilized in change-point data modeling. It can be recalled that the simulation studies considered 30 different linear change-point datasets.

Figure 3 presents the lower bound attained at convergence by the Variational algorithm 2 for 6 randomly selected *k* values based a randomly selected dataset out of the 30 datasets considered for the simulation. The corresponding data selected at random is the 11 dataset. From left to right are the plots for the 6 candidates of *k* values, 184; 235; 237; 271; 90 and 81respectively. The dataset was randomly selected.

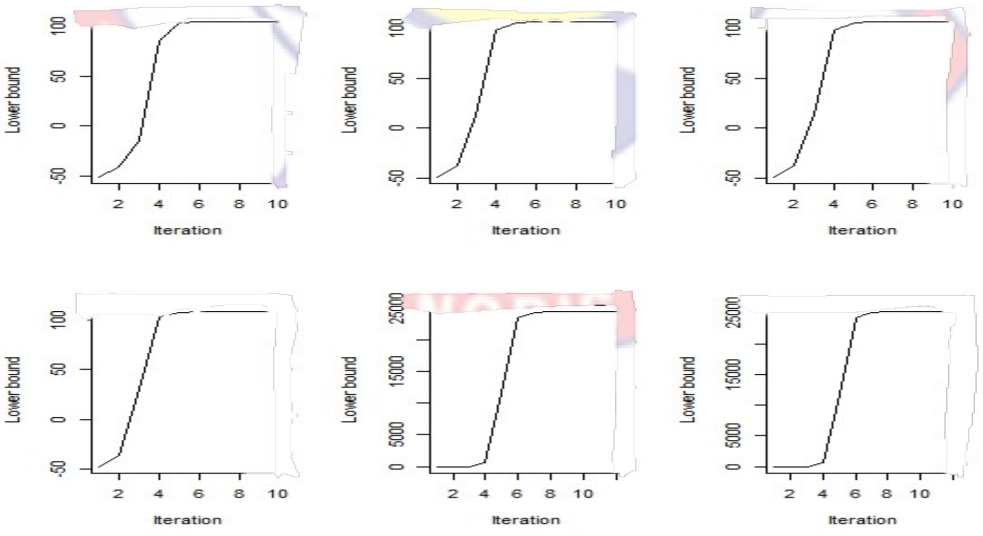

Figure 3: Switching model plot of Variational lower bound attained at convergence for 6 randomly selected *k* candidates for dataset 11.

Figure 4 also presents the nature of the Variational lower bound attained at convergence by algorithm 1 for 6 randomly selected datasets corresponding to datasets 24; 28; 14; 3; 17 and 11: The increasing pattern exhibited by algorithm 1 and algorithm 2 is clearly evident. Thus, the Variational algorithms for both the switching and non-switching models are functioning appropriately as expected of Variational algorithms based on optimizing a lower bound over marginal likelihood. From left to right are the plots for datasets, 24; 28; 14; 3; 17 and 11 respectively.

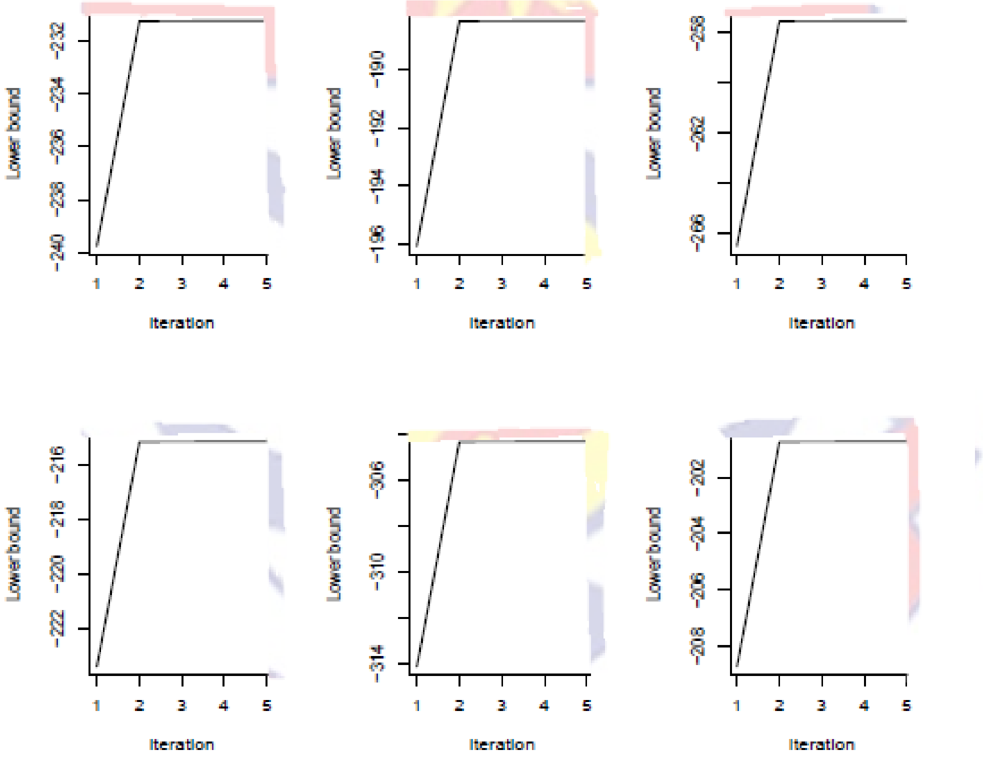

Figure 4: Non-switching model plot of Variational lower bound attained at convergence for 6 randomly selected datasets.

The dynamics of parameter estimation for the randomly selected dataset (dataset 11) as the change-point detection by the developed algorithms progresses are illustrated in Figure 5 and Figure 6. The black curves are the parameter estimates for the first half of the switching model (β s) and the red curves represents the parameter estimates of the second half of the switching model (θs). The blue dash lines are the corresponding true parameter values for the selected simulated dataset. It can be noted that the change-point parameters mimic the switching pattern underlying the simulated linear change-point data as the location *k*, changes.

In particular, the pattern exhibited by the βs clearly shows that they are parameters of the first half of the switch model. In the same way, the θs also depicts a pattern confirming their rightful position with regards to the model. Furthermore, the parameter estimation accuracy is affected by *k* with some saturation or stability for some values of *k*: Considering the first half of the switch model, the β s are poorly estimated at the boundary, particularly for $2 \leq k \leq 4$ or $k \in (2;4)$. However, the estimation accuracy is improved for values of *k* in the range 5 to 160: However, the poor estimation is exhibited for $161 \leq k \leq 298$, evidenced by the vast separation between the blue and black lines. The reverse is seen in the second half of the switching model. The θs are well estimated for values of $k \in$ (161; 289), with the accuracy in estimation dropping for $290 \leq k \leq 298$. Also, a wide departure of the blue line from the black line is depicted for values of *k* in the range 2 -160. Thus, trend exhibited by the parameters based on the estimation progress as a function of *k* clearly confirms the switching nature of the model and the rightful position of β s and θ s.

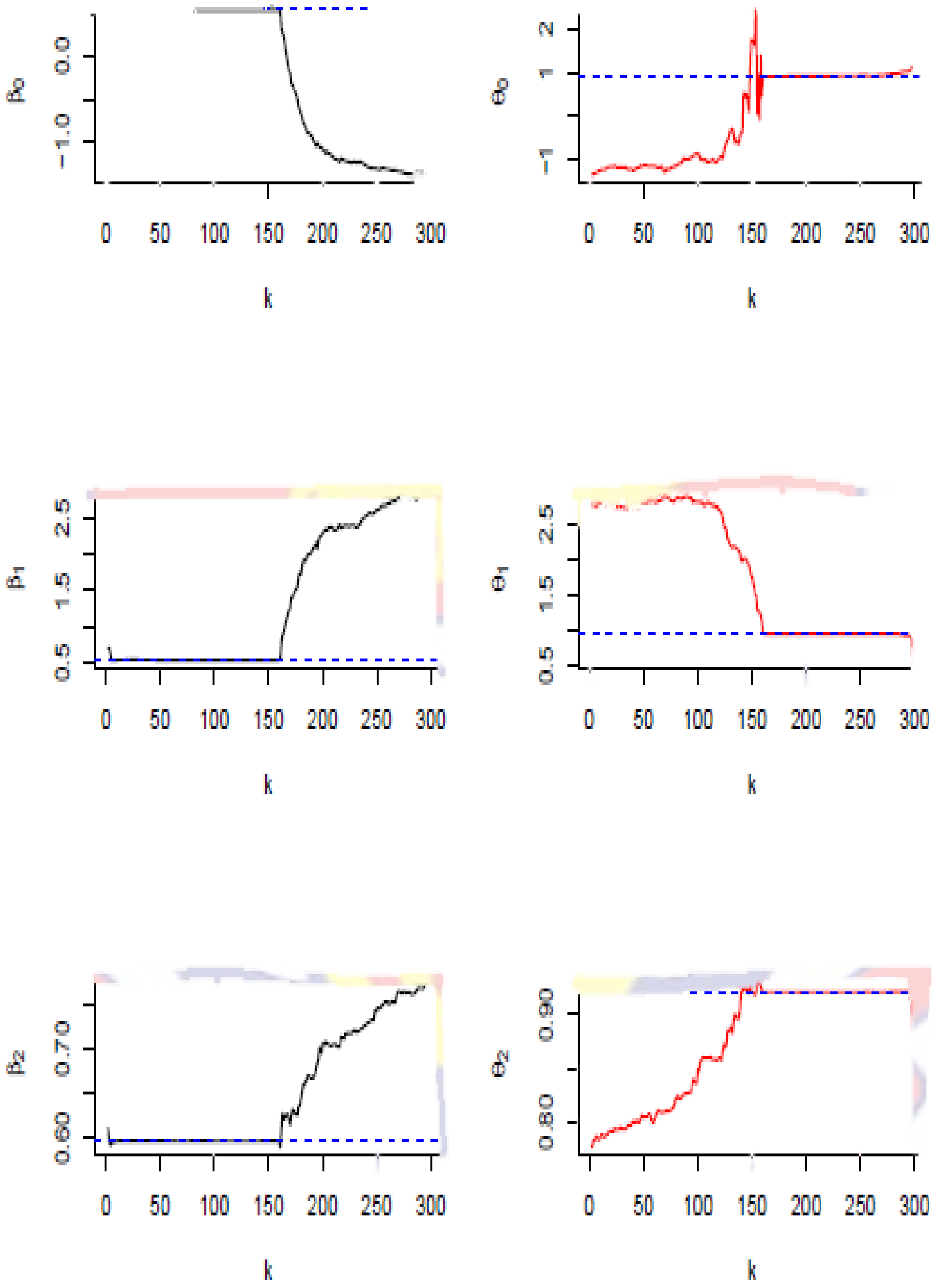

Figure 5: Nature of parameter estimates $\left(\beta_0, \beta_1, \beta_2, \theta_0, \theta_1, \theta_2\right)$ of switching model for a randomly selected dataset over k values.

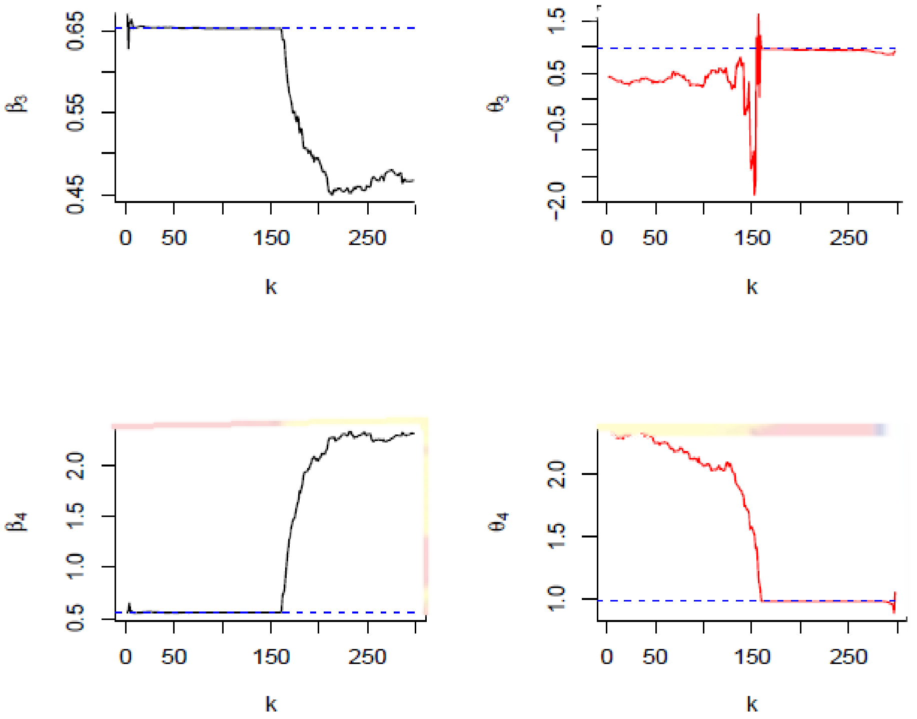

Figure 6: Nature of parameter estimates $(\beta_3, \beta_4, \theta_3, \theta_4)$ of switching model for a randomly selected dataset over *k* values

The overall parameter estimation performance of the developed change-point detectors quantified in statistical measures namely mean absolute error (MAE) and root mean squared error (RMSE) are shown in Figure 7 and Figure 8 respectively. From left to right are the plots MAEs and RMSEs against the possible change-point locations *k* for the β s and θ s respectively. Clearly, the general pattern exhibited by the parameters as, k, increases is shown in the errors (MAE and RMSE). It can be observed that MAE deceases as the *k* increases until *k* = 160 and then it increases from *k* = 161 until *k* = 298: Interestingly, the minimum MAE is attained at *k* = 160:

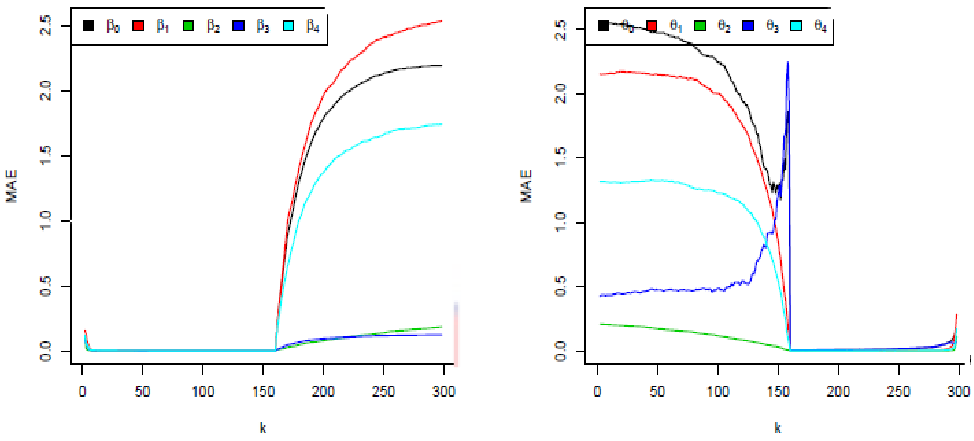

Figure 7: Mean Absolute Error (MAE) pattern of switching model parameters over k values

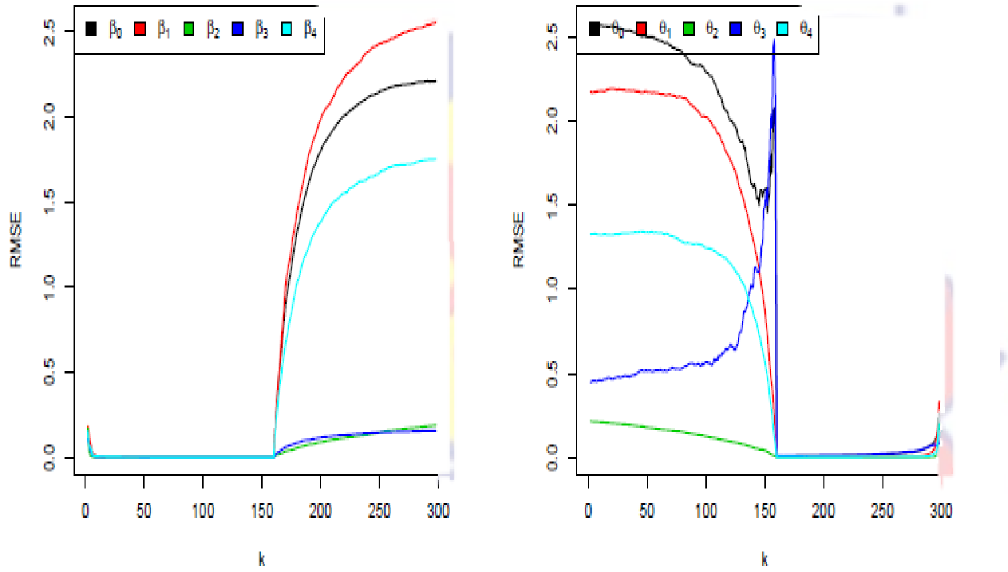

Figure 8: Plot of RMSE switching model parameters against k values.

Figure 9 illustrates the detection dynamics of the Variational lower bound difference VLBD and Variational Bayes information Criteria, VAIC, and VBIC detectors in relation to the possible change-point locations, *k*. The black, red, and green curves are respectively the plots of the VAIC (VAIC^{v}), VBIC (VBIC^{v}) and the maximum Variational lower bound value, obtained for the switching model over the range of *k*. Interestingly, all the three linear change-point detectors are able to detect switches in the simulated linear system and estimate their location correctly. Most importantly, they all report the same estimate of change-point location, this is evident in the common elbow at the same location exhibited by all the detectors. However, the detection path, as well as the dynamics, exhibits visible differences. In particular, the VBLD reaches the true change-point location, *k* = 160 faster than its Variational Bayes information counters. With regards to the VAIC and VBIC, it is clearly evident that the VAIC also gets to the true change-point location, *k* = 160 before VBIC.

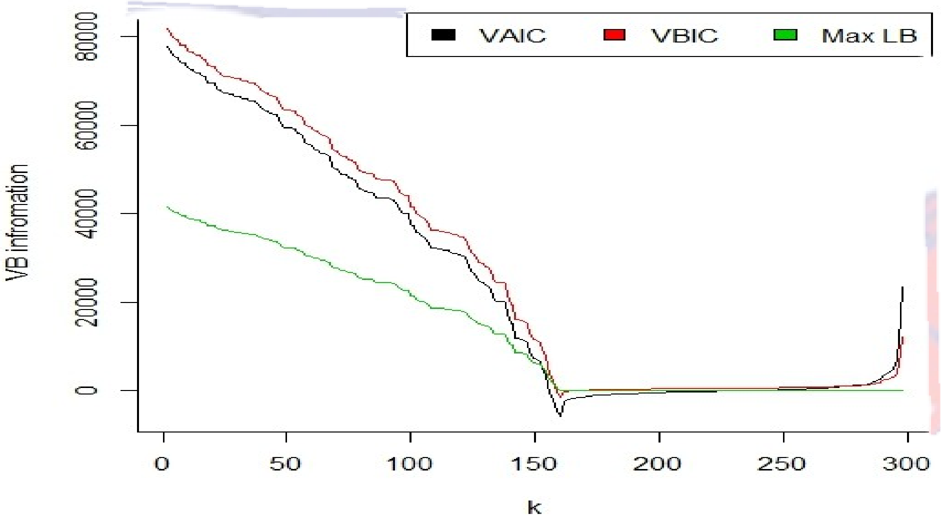

Figure 9: Dynamics of Variational Bayes Information criterion and the Lower Bound over switching positions k.

Considering the common elbow exhibited by all the three detectors in Figure 9, it will be interesting to explore their nature in the neighbourhood of the true change-point location. This will aid in gaining better insight into the typical detection characteristics of the developed linear change-point detectors. In this regard, we explore the detection features of the VBLD, VAIC, and VBIC within $k \in$ (151; 160). Table 4.1 reports some vital detection statistics of the VBLD, VAIC, and VBIC over the range of *k* considered above. Also, the statistics of the null model (H_0) required for change-point analysis are reported. In particular, for the null model (H_0), the lower bound attained at convergence, $L_{(q)}^{0}$, the Variational Akaike information criteria and its Variational Bayes counterpart values computed under H_0 are reported. On the other hand, Table 1 reports the statistics, k, L_m^{s} in columns 1 and 2, VAIC ^{v} and VBIC^{v} in columns 3 and 4, L_m^{k}, $\delta_A^{¿}$, and $\delta_B^{¿}$ in columns 5, 6 and 7 respectively for switching model (H_a). The red-coloured statistics are those that the three detectors yield as linear change-point detection statistics (results) upon application of the algorithms.

Interestingly, it is clearly seen that the maximum lower bound attained at convergence, for the VBLD decreases as k increases till *k* = 160 then an increasing pattern registered again after 160. Thus, the minimum of the values of L_m^{s} is recorded at k = 160 with $L_k^{¿} = -39.1756$

Now comparing with the corresponding H_0 value, $L_{(q)}^{0} = -200.7394$, the value of observed for the statistic, $L_k^{¿}$ is computed as. $L_k^{¿} = -39.1756 - (-200.7394) = 161.5638$. Obviously, VBLD declares that a switch exists in the simulated dataset since $L_k^{¿} > 0$ and in particular, at a location *k* = 160. Therefore, VBLD estimates *k* with $k = 160$.

Also, it can be seen that all the values of $L_k^{¿}$ are positive and the smallest value among all is 161.5638. Furthermore, for VAIC and VBIC detectors, the similar decreasing pattern for increasing change-point location, *k* is exhibited by each one. Interestingly, the minimum VAIC and VBIC values were respectively, -5710.8532 and -1411.1697 occurred at the same position, *k* = 160. These VAIC and VBIC values resulted to the corresponding estimates for the statistics, $\delta_A^{¿}$ and $\delta_B^{¿}$ as $\delta_A^{¿}=-0.117$ and $\delta_B^{¿}=-0.255.$ Since $\delta_A^{¿}<1$ and $\delta_B^{¿}<1$, change-point inference based on both VAIC and VBIC is that there exists a switch at location *k* = 160 and *k* is estimated using k = 160.

Another interesting observation fundamental to VAIC and VBIC based detectors is that the $\delta_A^{¿}$ and $\delta_B^{¿}$ that generated the best estimate of the change-point location *k* are the smallest among all the negative values of $\delta_A^{¿}$ and $\delta_B^{¿}$ respectively. In general, the basic automatic selective feature encoded via the design of the proposals, VBLD, VIAC, and VBIC, allow them to avoid computing all $L_k^{¿}$ values before making a choice. The aspect of the proposals that ensures computational savings.

Table 1: Comparison of Change-Point Statistic Estimates Based on VLBD, VAIC, and VBIC using the Null Model Statistics, $L^{0}_{(q)}$ = - 200. 7394, $VAIC^{0}$= 668.7677, $VBIC^{0}$= 360. 5402.

K	$VLBD$ (L^{s}_{m})	$VAIC$ ($VAIC^{\upsilon}$)	$VBIC$ ($VBIC^{\upsilon}$)	$L^{¿}$	$\delta^{¿}_{A}$	$\delta^{¿}_{B}$
151	6117.693	6787.571	11100.589	6318.433	0.099	0.032
152	6040.421	6637.219	10948.333	6241.161	0.101	0.033
153	5214.143	4982.069	9299.690	5414.882	0.134	0.039
154	4437.701	3419.910	7747.222	4638.441	0.196	0.047
155	3368.350	1337.438	5642.070	3569.089	0.500	0.064
156	2495.223	-411.019	3900.322	2695.963	-1.627	0.092
157	1674.521	-2041.164	2269.554	1875.260	-0.328	0.159
158	1140.980	-3068.361	1225.277	1341.719	-0.218	0.294
159	444.574	-4450.568	-157.846	645.314	-0.150	-2.284
160	-39.176	-5710.853	-1411.170	161.564	-0.117	-0.255
161	20.3530	-4213.280	-859.061	221.092	-0.159	-0.420
162	91.310	2502.365	-286.315	292.050	-0.267	-1.259
163	97.226	-2040.766	136.618	297.966	-0.328	-2.639
164	97.863	-1973.897	-115.028	298.603	-0.339	-3.134
165	99.322	-1798.212	-58.172	300.062	-0.372	-6.198
166	100.336	-1652.887	-11.226	301.075	-0.405	-32.117
167	101.057	-1533.735	27.084	301.797	-0.436	13.312
168	101.320	-1480.797	44.081	302.068	-0.452	8.179
169	101.484	-1447.155	55.131	302.223	-0.462	6.540
170	101.669	-1412.166	66.390	302.408	-0.474	5.431

Source**:** Researcher’s Construct (2021)

Table 2 presents the change-point estimates obtained from the VBLD, VAIC, and VBIC detectors. For the regression parameters, β^{a} and θ, their 95% Bayesian credible intervals are also reported. The values in the round brackets are the true parameter assumed for the simulation. It can be observed that the change-point parameters are well estimated.

Table 2: Comparison of Change-Point Parameter Estimates Based on VLBD, VAIC, and VBIC

H_a K	*VLBD* (k_{VBLD})	*VAIC* (k_{VAIC})	*VBIC* (k_{VBIC})	95% BCI k_{VBLD}	95% BCI k_{VAIC}	95% BCI k_{VBIC}
$\beta_0^a(0.6058)$	0.6053	-	-	[0.353,0.732]	[-]	[-]
$\beta_1^a(0.4743)$	0.4754	-	-	[0.492,0.578]	[-]	[-]
$\beta_2^a(0.5247)$	0.5247	-	-	[0.593,0.602]	[-]	[-]
$\beta_3^a(0.5239)$	0.5244	-	-	[0.427,0.873]	[-]	[-]
$\beta_4^a(0.5186)$	0.5178	-	-	[0.524,0.590]	[-]	[-]
$\theta_0(0.8965)$	0.8686	-	-	[0.698,1.105]	[-]	[-]
$\theta_1(0.8965)$	0.9453	-	-	[0.902,0.985]	[-]	[-]
$\theta_2(0.9214)$	0.9682	-	-	[0.917,0.925]	[-]	[-]
$\theta_3(0.9737)$	0.9530	-	-	[0.749,1.178]	[-]	[-]
$\theta_4(0.9868)$	1.0294	-	-	[0.958,1.016]	[-]	[-]

Source: Researcher’s Construct (2021)

Real Data Application: Manufacturing Process in Refinery

Using the octane rating as a response, y; and the remaining variables as predictors, and with the same prior settings considered in the simulation in Chapter Three, we apply the developed methods to refinery dataset. The nature of the linear switching system underlying the octane rating as well as its relationship with the other variables is shown in Figure 10. It can be seen that the octane rating exhibits multiple fluctuations with the major one occurring at location 75. It is pretty much obvious that the switch at location 75 is the crucial change-point in the data compared to the others. This is because it stands out among the other possible switches. There appears to be small switches at locations 38 and 64, however the most crucial switch occurred at 75.

Again in Figure 10, a careful observations of the behaviour of the various variables against the response variable that is, octane rating indicates some kind of linear relationship between octane rating, the response variable and the predictor variables exists. It is easy to observe from the plot of material 1 that the points on the graph are somewhat clustered around the 65 point. In the second scenario involving the response variable and the material 2, there appears to be a linear relationship oriented in a positive plane. In the third scenario however, it can be seen that there exists a negative linear relationship between the octane rating and material 3(variable x3). The fourth observation however shows a positive linear relationship between the octane rating and the material 4.

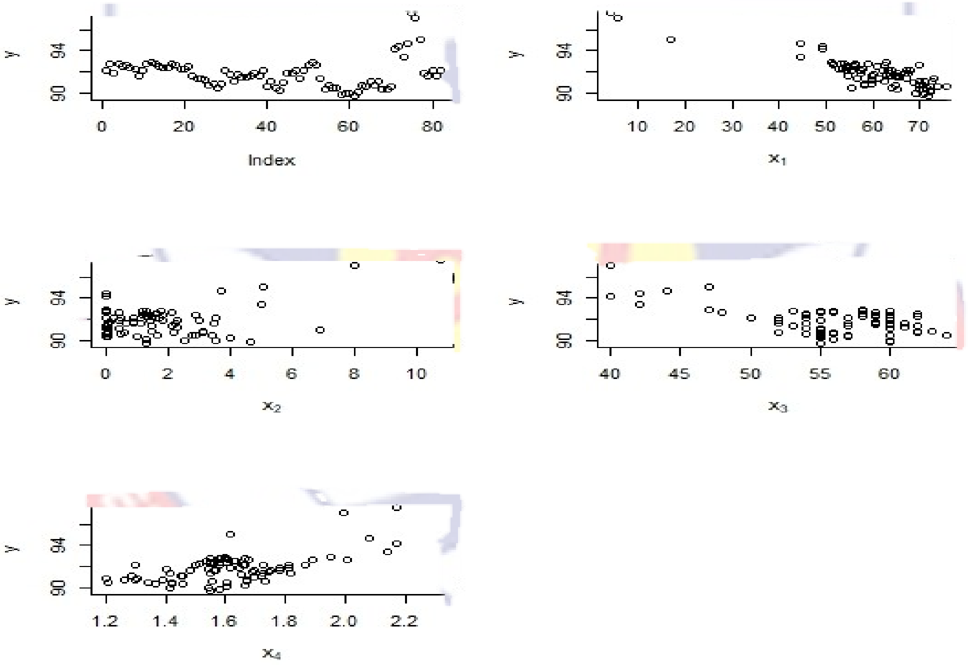

Figure 10: Nature of linear system generated by refinery manufacturing process.

Figure 11 shows the inter-relationships existing among the variables. It can be seen that the correlation coefficient between material 1 and material 2 is a negative value of (-0.59) indicating a strong negative linear relationship between them. It is clearly evident that octane rating has a linear relationship with all the four predictor variables. It can be seen further that material 1 and material 3 shows negative linear relationship with octane rating. However, the predictor variables, material 2 and material 4 indicate positive linear relationship with the octane rating with material 2 recording the least correlation coefficient (0.39).

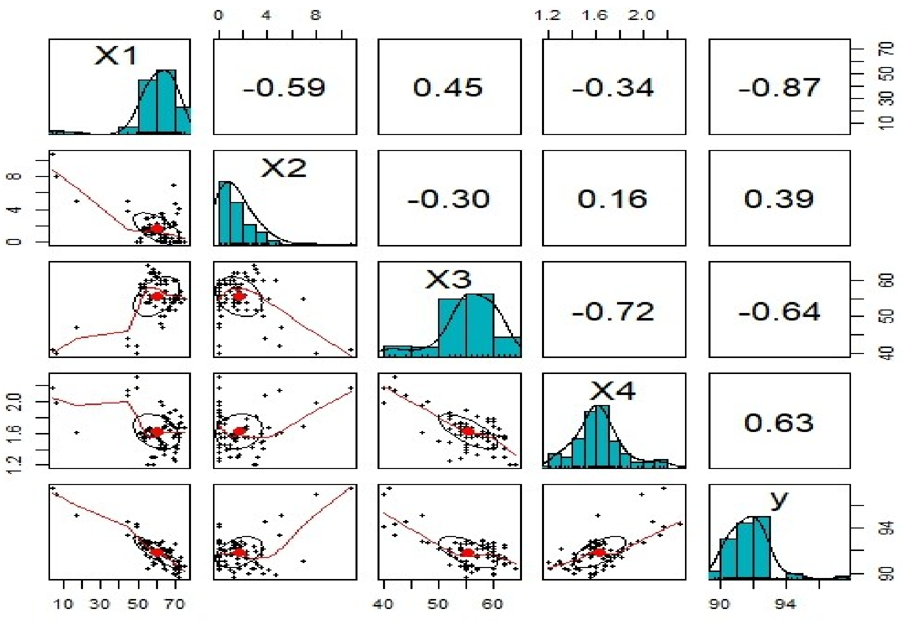

Figure 11: Plot of data obtained from refinery manufacturing process.

Figure 12 presents the nature of the lower bound attained at convergence for 6 randomly selected *k* candidates, namely, 10, 70, 55, 31, 74 and 49, for the real data application. As noted in the stimulated study that the Variational lower bound exhibits an increasing trend, the same observation can be made here too. That is, the lower bound records an upward trend illustrating the appropriateness of the developed Variational algorithms for the purpose of change-point modeling and detection. From left to right are the plots for k values, 10; 70; 55; 31; 74 and 49, respectively**.**

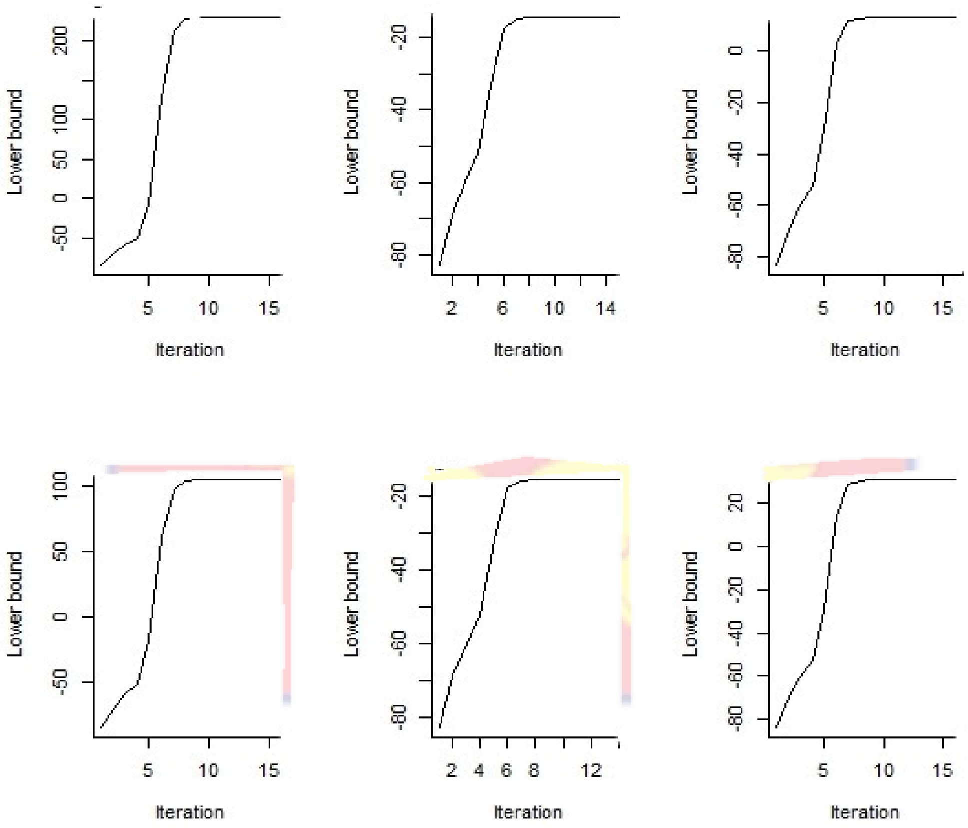

Figure 12: Lower bound attained at convergence for 6 randomly selected *k* candidates for real data application.

The dynamics of parameter estimation for the real dataset as the change-point detection by the developed algorithms progresses are illustrated in Figure 13. The black curves are the parameter estimates for the first phase of the switching model (β s) and the red curves represent the parameter estimates of the second phase of the switching model (θs). It is evident that the pre-change estimates exhibit virtually the same features as that of the corresponding post-change estimates.

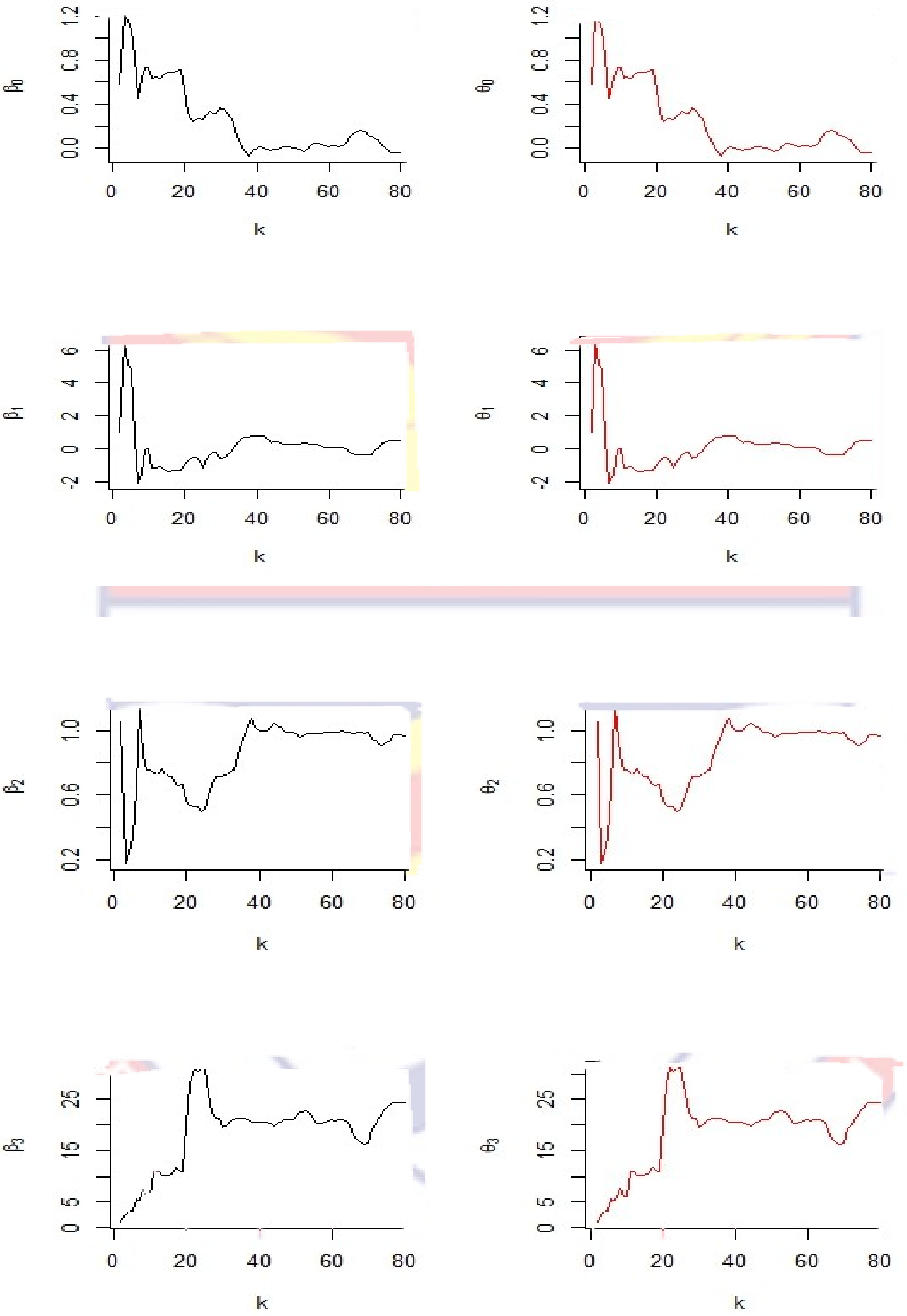

Figure 13: Pattern of switching model parameter estimates over *k*

Table 3 represents summary detection statistics results of the VBLD, VAIC and VBIC over the range of $k \in (61,80)$ for the real dataset under consideration. Also, the null model statistics required for the change-point inference are reported. More importantly, the Variational lower bound attained at convergence, $L_{(q)}^{0}$, the $VAIC^{0}$,and $VBIC^{0}$ values for the null model $\left(H_{0}\right)$ are reported. Further, other statistics, in particular, k, L_{m}^{s}, $VAIC^{v}$, $VBIC^{v}$, $L_{k}^{\text{¿}}$, $\delta_{A}^{\text{¿}}$ and $\delta_{B}^{\text{¿}}$ for the switching model $\left(H_{a}\right)$ are reported in columns 1 to 7 respectively. The coloured estimates are those that the Variational lower bound-based detectors yielded as linear change-point detection results for the real datasets. It can be observed that all the statistics reports some kind of irregular values patterns or fluctuations over the range of $k \in (61,80)$. The maximum lower bound attained at convergence, L_{m}^{s} for the VBLD decreases as *k*, increases till *k* =65, with a sharp increase at *k* = 66.

The estimates are seen to have a decreasing pattern recorded after *k* = 66 till *k* = 75, and then another increasing pattern is evident after *k* = 75. More so, the minimum values of the statistics L_{m}^{s} and $L_{k}^{\text{¿}}$ were recorded as (-15.6571) and (185.2131) respectively and occurred at *k* =75. It can be recalled that, VBLD declares that a switch exists in a dataset, if $L_{k}^{\text{¿}}$ > 0. It is evidently clear from the table that all the values of VBLD detector, $L_{k}^{\text{¿}}$, are positive and that the smallest value among the set of values is recorded as (185.2131) and occurred at the location, *k* = 75. This corresponds to the location of the major change-point existing in the real dataset. Furthermore, for the Variational information-based statistics, corresponding to the VAIC and VBIC, the same decreasing and increasing pattern for increasing change-point location, *k*, is depicted by each one of these schemes.

A careful observation shows that, the minimum estimated values of the VAIC and the VBIC were 850.8256 and 442.1871 respectively, and occurred at the same location, *k* = 64. The estimates for the Variational information ratio statistics, $\delta_A^{¿}$ and $\delta_B^{¿}$ corresponding to the smallest VAIC and VBIC values stated above, were recorded as $\delta_A^{¿}=1.5841$ and $\delta_B^{¿}=1.4197$. It can be recalled that VAIC and VBIC declare the existence of a switch at a certain location, k, if $\delta_A^{¿}<1$ and $\delta_B^{¿}<1$, respectively, for which *k* is estimated using k .

It is clearly evident that, the estimates for both $\delta_A^{¿}$ and $\delta_B^{¿}$ are all positive and greater than one, suggesting that there is no switch existing in the data generative process. This results seems to contradict the fundamental principle of VAIC and VBIC -based detectors proposed that, the detection statistics $\delta_A^{¿}$ and $\delta_B^{¿}$ that generate or compute the best estimates of the location *k*, of the change-point are the smallest among all values of $\delta_A^{¿}$ and $\delta_B^{¿}$ respectively.

Table 3: Real Data: Comparison of Statistic Estimates Based on VLBD, VAIC, and VBIC using the Null Model Statistics

$L^{0}(q)=$ **- 200.8703,** $VAIC^{0}=$ **537.1086,** $VBIC^{0}=$ **311. 448.**

	VLBD	*VAIC*	*VBIC*			
k	(L_m^s)	($VAIC^{\upsilon}$)	($VBIC^{\upsilon}$)	$L^{¿}$	$\delta_A^{¿}$	$\delta_B^{¿}$
61	0.4182	860.4668	453.6293	201.2885	1.6020	1.4565
62	-1.2088	857.2386	450.2683	199.6615	1.5960	1.4457
63	-4.1723	851.6207	444.2210	196.6980	1.5856	1.4263
64	-5.4517	850.8256	442.1671	195.4185	1.5841	1.4197
65	-5.9237	861.7519	444.8441	194.9466	1.6044	1.4283
66	10.4550	896.0747	448.7680	190.4153	1.6683	1.4409
67	-12.1677	915.0724	452.6454	188.7025	1.7037	1.4534
68	-12.3448	926.8837	457.0974	188.5255	1.7257	1.4677
69	-12.3286	941.2118	463.3454	188.5416	1.7524	1.4877
70	-12.9012	952.0461	467.0458	187.9691	1.7725	1.4996
71	-14.5922	980.4615	473.2731	186.2781	1.8254	1.5196
72	-13.8360	1010.5587	485.3799	187.0342	1.8815	1.5585
73	-13.9351	1037.4071	495.3226	186.9351	1.9315	1.5904
74	-14.0258	1088.2290	516.6033	186.8445	2.0261	1.6587
75	-15.6571	1154.8372	543.1124	185.2131	2.1501	1.7438
76	-15.2025	1405.5619	660.8841	185.6677	2.6169	2.1220
77	-12.4214	12872.5603	6377.8585	188.4488	23.9664	20.4781
78	-12.8180	19659.0904	9768.5145	188.0522	36.6017	31.3648
79	-12.3963	31909.6141	15891.9519	188.4740	59.4100	51.0260
80	-11.2465	40306.7856	20084.2658	189.6238	75.0440	64.4867

Source: Researcher’s Construct (2021)

Table 4 presents the change-point estimates obtained from the VBLD, VAIC, and VBIC detectors. For the regression parameters, β^{a} and θ, their 95% Bayesian credible intervals are also reported. The change-point location estimates via the three proposed detectors were reported. It can be observed that, the VBLD-detector, the VAIC-detector and the VBIC-detector estimated k as 75, 38 and 64 respectively. Clearly, the change-point regression parameters estimate for β^{a} and the corresponding, θ, are the same, based on each detector.

Table 4: Real Data. Comparison of Change-Point Parameter Estimates Based on VLBD, VAIC, and VBIC

H_a	*VLBD*	*VAIC*	*VBIC*	95% BCI	95% BCI	95% BCI
k	k_{VBLD}	k_{VAIC}	k_{VBIC}	(k_{VBLD})	(k_{VAIC})	(k_{VBIC})
β_0^a	0.912	0.921	0.912	[0.520,1.2]	[0.330,1.51]	[0.404,1.40]
β_1^a	0.915	0.932	0.915	[0.406,1.4]	[0.334,1.54]	[0.357,1.45]
β_2^a	0.915	0.929	0.927	[0.380,1.4]	[0.357,1.45]	[0.374,1.48]
β_3^a	0.920	0.920	0.927	[0.237,1.4]	[0.346,1.51]	[0.336,1.53]
β_4^a	0.920	0.920	0.927	[0.237,1.4]	[0.346,1.51]	[0.336,1.53]
θ_0	0.912	0.921	0.912	[0.520,1.2]	[0.330,1.51]	[0.404,1.40]
θ_1	0.915	0.922	0.915	[0.406,1.4]	[0.334,1.54]	[0.357,1.45]
θ_2	0.915	0.929	0.927	[0.380,1.4]	[0.357,1.45]	[0.374,1.48]
θ_3	0.920	0.920	0.927	[0.237,1.4]	[0.346,1.51]	[0.336,1.53]
θ_4	1.920	0.920	0.927	[0.237,1.4]	[0.346,1.511]	[0.336,1.54]

Source: Researcher’s Construct (2021)

Figure 14 presents the dynamics of the change-point detectors (VLB difference, VAIC ratio, and VBIC ratio methods) in terms of the statistics of the detectors when applied to the real manufacturing process dataset. It is clearly evident that the two Variational information criteria, VAIC and VBIC exhibit virtually the same features or pattern. It can be seen also that, Max.LB and the LBD detectors exhibit virtually the same decreasing features as the magnitude of location range increases. Interestingly, there exists a clear difference between the patterns exhibited by the Variational information-based detectors (VAIC and VBIC) and VLBD detectors, in comparison with the observations from the simulated data application.

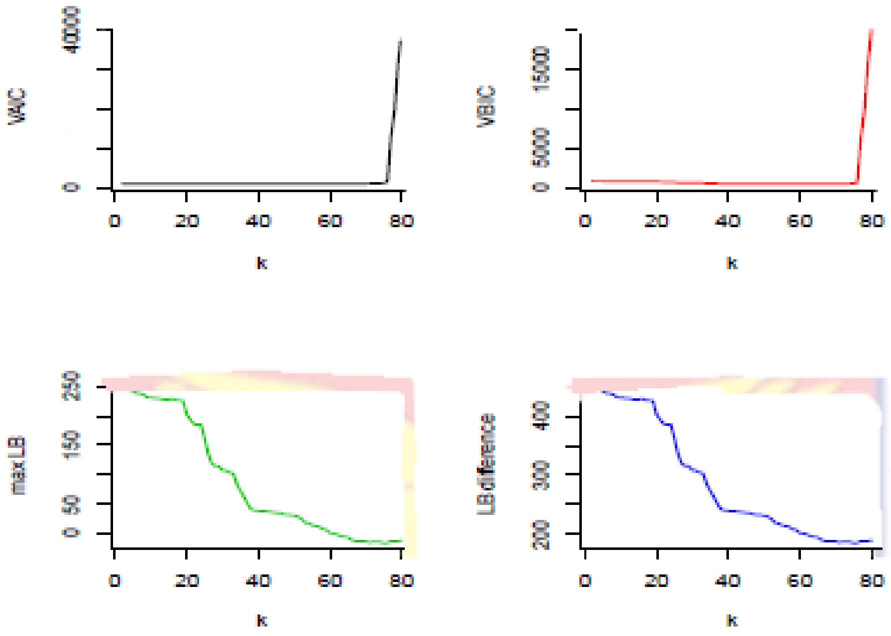

Figure 14: Dynamics of change-point detectors (LB difference, VAIC ratio and VBIC ratio methods)

Figure 15 reports the change-point detection characteristics of the three detectors (VBLD, VAIC, and VBIC), based on the refinery dataset. The coloured lines are the estimated change-point locations detected by each of the proposed detectors. The red, green and blue lines corresponding respectively to the VAIC-ratio detector, VBIC -ratio detector and VLBD detector. Aslo, the calibration in the data in terms of how each detector sees and label a change-point in the data is plotted. It can be observed that the VBLD detector is able to detector the major change-point located at 75, which was of interest. In this regard, we see that the VLBD detector exhibits some robustness than its competitors in the case where there exists more than one change-point with one being dominant.

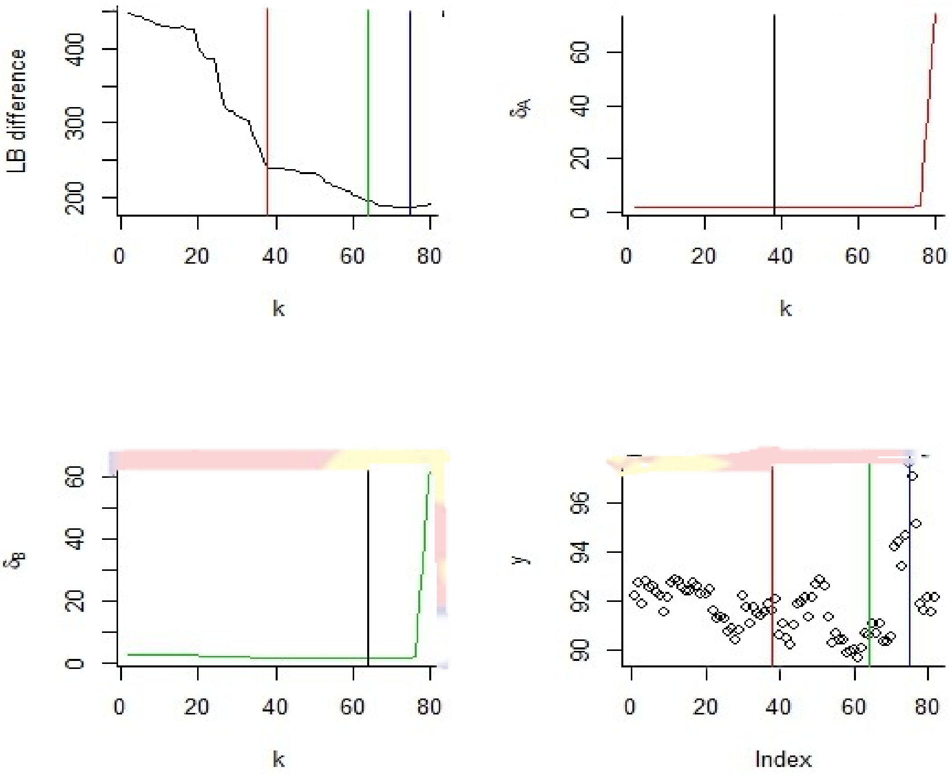

Figure 15: Performance of change-point detectors (LB difference, VAIC ratio and VBIC ratio methods)

Chapter Summary

Chapter four considered the implementation of the developed algorithms using both simulated and real datasets. The simulation was based on the developed variational Bayesian switching and non-switching linear data generative models where specific parameter settings for parameter values as well as prior hyperparameter values were employed in line with existing literature. The real data implementation considered real-time data on the manufacturing process in a refinery. The dataset is a result of the investigation of the manufacturing process in a refinery in which the octane rating of particular petrol product was considered as a function of three raw materials and a variable that characterizes the manufacturing conditions.

The chapter considered the performance of the developed algorithms and the nature of the parameter estimates for randomly selected *k* values for real data and fixed *k* values for simulated data. The results from the simulation indicated that the Variational Lower Bound Difference-based scheme is computational friendly and outperforms all the other Variational information-based schemes in particular VAIC- ratio and the VBIC- ratio in determining the underlying simulated trend based on some statistical measures such as RMSE and MSE.

Furthermore, regarding the ability of the methods to model, detect and locate a single change-point in a linear system, it was established that the VBLBD exhibits better performance and robustness than its information-based counterparts.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

Introduction

This final chapter of the study presents a summary of the findings of the entire study. It includes the conclusions from the study, and some recommendations arrived at from the study.

Summary

This study explored and developed flexible and robust computational methods in the Bayesian framework for modeling and inference for change-point datasets generated by some linear systems in which the linear relationship among the response variable and the predictors is oriented in the positive or negative plane. This was facilitated by the development of various Variational Bayes inference schemes. In Chapter One, the study opened with an introduction to the Variational Bayes concept and looked at issues relating to the background of the study, problem statement raised in the study was how to develop Bayesian solutions for the change-point problems in the linear regression context.

In particular, how the Variational lower bound can be employed for linear change-point analysis. The main objective was to develop an appropriate Bayesian model and inference methods for detecting change-points in linear systems. It was discussed that one importance of the study of this work is to explore the capabilities of the Variational Bayes technique to detect a change in linear systems as this approach is seen to be computationally tractable, fast, and deterministic. One limitation realized in the study was that the framework developed is not an all-encompassing one that can be used to detect all change- points. The chapter ended with a summary of the most salient points identified in the work. Chapter two began with the review of important theorems in the Bayesian framework. Change-point was discussed as an abrupt change in a generative process of a sequence of random variables. Furthermore, a brief review of change-point detection using MCMC was outlined.

Variational Bayesian inference approach was developed in particular a Variational Bayes Lower Bound and Variational Information Criteria for inference about the change-point detection. In addition, an empirical review of some relevant studies on the change-point detection literature was made. The chapter ended with a chapter summary. Bayesian switching and non-switching models were developed for the methodology in chapter three. Other Variational inferences are also developed for the switching and non-switching models. Four algorithms were developed and implemented in chapter three. Simulations were carried out using the algorithms and the datasets generated were used to assess the various schemes that were developed in the course of the study. Also, real data implementation considered real-time data on the manufacturing process in a refinery. The dataset is a result of the investigation of the manufacturing process in a refinery in which the octane rating of petrol was considered as a function of three raw materials and a variable that characterizes the manufacturing conditions. The two application scenarios were reviewed thoroughly. The implementations of the algorithms were evaluated and the results indicated that the algorithms were performing well. The Chapter ended with a chapter summary.

Chapter Four commenced with the results of the application of developed methods to both simulated data and real data of the Bayesian linear switching and non-switching systems that were developed earlier in chapter three. The chapter evaluated the performance of the developed algorithms and characteristics of the parameter estimates for randomly selected datasets for both real and simulated datasets. The results from the real data and simulation applications show that the Variational lower bound difference-based scheme outperforms all the other Variational information-based schemes even though they all performed well. A comparison of the statistical measurement of error parameters also indicated that the Variational lower bound difference scheme performed better than the rest and exhibited robustness. The fourth chapter ended with a chapter summary.

Conclusions

We have proposed and implemented novel Bayesian methods for modeling, fitting, and detecting change-points in linear systems that generate data that exhibits linear patterns. Particularly, Bayesian switching and non-switching models were developed for linear systems. Also, appropriate Variational Bayes inference schemes were developed for parameter inference. Furthermore, the methods adopt the Variational Bayes framework and make use of its bye-product termed Variational lower bound as well as its information criterion to develop appropriate change-point detectors.

The Variational information-based schemes were built based on the properties of the usual Variational Bayesian Information Criteria (VBIC) and Variational Akaike Information Criteria (VAIC). The Variational lower bound-based detector uses the lower bound difference between the switching and non-switching models for all possible candidates of the change-point location, *k*.

The proposed Bayesian non-switching and switching models assume the following general structures respectively.

$$H_0 : Y = X\beta^0 + \varepsilon$$

$$H_1 : Y_k = X_k^1 \beta^a + \varepsilon_k \,, \; Y_{nk}^1 = X_{nk}^1 \theta + \varepsilon_{nk}$$

All unknown parameters were treated as random and modeled with the appropriate probability models.

The change-point detectors depend on the Variational Bayes parameters as a result, switching and non-switching specific Bayesian models are built with Variational inference methods developed for parameter inference. The applicability of the proposed methods for change-point analysis in linear regression is illustrated using both simulation and real data from the refinery manufacturing process dataset. The Variational lower bound–based method exhibits robustness over its Variational information counterparts, especially, in datasets with unclear multiple switches with one outstanding switch. This was the case with the real data application.

Recommendations

This part of the thesis focuses on some vital recommendations based on the proposed methods and its implementation using both simulated and real datasets. Also, some possible directions for extension of the proposals are outlined in brief. Beginning with the recommendations, we consider the following. We recommend

1. the use of Variational lower bound-based statistics such as the difference in lower bounds between switching and non-switching

models for change-point detection and calibration in linear regression models in order to properly model dataset generated by linear systems.

2. the integration of Variational lower bound-based change-point detectors in industrial process control ecosystems since it is fast and light computationally.
3. the proper or accurate calibration of prior models of hyper-parameter in Bayesian change-point models before its use in change-point detection. This is because the wrong choice of prior models yields misleading results. It is thus vital to allow the data to direct the choice of prior settings.

Suggestions for Further Studies

On direction for furthering work of the proposals in this thesis, we outline the following.

1. The linear regression-based change-point detectors considered can be extended to non-linear or functional change-point modeling and detection.
2. In this thesis, change-point detection was restricted to mean change, fixing the variability. Another direction of extension is to consider both mean and variability change-point modeling and detection within the Variational Bayes framework.
3. Another potential extension could be in the direction of modeling the change-point parameter using different probability instead of uniform assumed in this thesis.
4. One can also explore the use of MCMC methods for change-point detection.

REFERENCES

Acitas, S., & Senoglu, B. (2020). Robust change point estimation in two-phase linear regression models: An application to metabolic pathway data. *Journal of Computational and Applied Mathematics*, *363*, 337–349.

Adams, R. P., & MacKay, D. J. (2007). Bayesian online changepoint detection. *arXiv preprint arXiv:0710.3742*.

Akaike, H. (1973). Theory and an extension of the maximum likelihood principal. In *International symposium on information theory. Budapest, Hungary: Akademiai Kaiado*.

Andrade, J. A. A., & O'Hagan, A. (2006). Bayesian robustness modeling using regularly varying distributions. *Bayesian Analysis*, *1*(1), 169-188.

Attias, H. (2000). A Variational Bayesian framework for graphical models, in Advances in Neural Information Processing Systems 13. MIT Press.

Bacon, D. W., & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika*, *58*(3), 525–534.

Barry, D., & Hartigan, J. A. (1992). Product partition models for change point problems. *The Annals of Statistics*, 260–279.

Barry, D., & Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, *88*(421), 309–319.

Basalamah, D., Said, K. K., Ning, W., & Tian, Y. (2021). Modified information criterion for linear regression change-point model with its applications. *Communications in Statistics-Simulation and Computation*, *50*(1), 180–197.

Beck, J. L. (2010). Bayesian system identification based on probability logic. *Structural Control and Health Monitoring*, *17*(7), 825–847.

Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian analysis*, *1*(3), 385-402.

Bernardo, J. M., & Smith, A. F. M. (2009). *Bayesian theory* (Vol. 405). John Wiley & Sons.

Bishop, C. M. (2006). Pattern recognition. *Machine Learning*, *128*(9).

Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, *112*(518), 859–877.

Box, G. E. P., & Tiao, G. C. (2011). *Bayesian inference in statistical analysis* (Vol. 40). John Wiley & Sons.

Brodsky, E., & Darkhovsky, B. S. (1993). *Nonparametric methods in change point problems* (Vol. 243). Springer Science & Business Media.

Broemeling, I. D. (1972). Bayesian procedures for detecting a change in a sequence of random variables.

Cai, X., Said, K. K., & Ning, W. (2016). Change-point analysis with bathtub shape for the exponential distribution. *Journal of Applied Statistics*, *43*(15), 2740–2750.

Carlin, B. P., Gelfand, A. E., & Adrian F. M. Smith. (1992). Hierarchical Bayesian Analysis of Changepoint Problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *41*(2), 389–405. https://doi.org/10.2307/2347570.

Chaturvedi, A., & Shrivastava, A. (2016). Bayesian analysis of a linear model involving structural changes in either regression parameters or disturbances precision. *Communications in Statistics - Theory and Methods*, *45*(2), 307–320. https://doi.org/10.1080/03610926.2013.806666.

Chen, C. W., Chan, J. S., Gerlach, R., & Hsieh, W. Y. (2011). A comparison of estimators for regression models with change points. *Statistics and Computing*, *21*(3), 395-414.

Chen, J., Gupta, A. K., & Pan, J. (2006a). Information criterion and change point problem for regular models. *Sankhy{\=a}: The Indian Journal of Statistics*, 252–282.

Chen, J., Gupta, A. K., & Pan, J. (2006b). Information criterion and change point problem for regular models. *Sankhya: The Indian Journal of Statistics*, *68*(2), 252–282.

Chen, J. (1998). Testing for a change point in linear regression models. *Communications in Statistics - Theory and Methods*, *27*(10), 2481–2493. https://doi.org/10.1080/03610929808832238.

Chen, J, & Gupta, A. K. (2001). ON CHANGE POINT DETECTION AND ESTIMATION. *Communications in Statistics - Simulation and Computation*, *30*(3), 665–697. https://doi.org/10.1081/SAC-100105085.

Chen, J., & Gupta, A. K. (2012). *Parametric Statistical Change Point Analysis*. Boston: Birkhäuser Boston. https://doi.org/10.1007/978-0-8176-4801-5.

Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *The Annals of Mathematical Statistics*, *35*(3), 999–1018.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of Econometrics*, *86*(2), 221–241.

Chick, S. E. (2006). Subjective probability and Bayesian methodology. *Handbooks in Operations Research and Management Science*, *13*, 225–257.

Choy, J. H.C, & Broemeling, L. D. (1980). Some Bayesian inferences for a changing linear model. *Technometrics*, *22*(1), 71–78.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591–605.

Corduneanu, A., & Bishop, C. M. (2001). Hyperparameters for Soft Bayesian Model Selection. In T. S. Richardson & T. S. Jaakkola (Eds.), *Proceedings of the Eighth International Workshop on Artificial Intelligence and Statistics* (Vol. R3, pp. 63–70). PMLR. Retrieved from https://proceedings.mlr.press/r3/corduneanu01a.html.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American Journal of Physics*, *14*(1), 1–13.

Cox. R.T (1961). Probability: The Algebra of Probable Inference., *134*(3478), 551.

Dawid, A. P., Stone, M., & Zidek, J. V. (1973). Marginalization paradoxes in Bayesian and structural inference. *Journal of the Royal Statistical Society: Series B (Methodological)*, *35*(2), 189–213.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1–22.

Du, H., Edwards, M. C., & Zhang, Z. (2019). Bayes factor in one-sample tests of means with a sensitivity analysis: A discussion of separate prior distributions. *Behavior Research Methods*, *51*(5), 1998–2021. https://doi.org/10.3758/s13428-019-01262-w.

Elsner, J. B., Niu, X., & Jagger, T. H. (2004). Detecting shifts in hurricane rates using a Markov chain Monte Carlo approach. *Journal of Climate*, *17*(13), 2652–2666.

Farley, J. U., & Hinich, M. J. (1970). A test for a shifting slope coefficient in a linear model. *Journal of the American Statistical Association*, *65*(331), 1320–1329.

Ferreira, P. E. (1975). A Bayesian analysis of a switching regression model: known number of regimes. *Journal of the American Statistical Association*, *70*(350), 370–374.

Fryzlewicz, P., & Rao, S. S. (2011). BaSTA: consistent multiscale multiple change-point detection for ARCH processes. *Preprint.*

Gelfand, A. E., Hills, S. E., Racine-Poon, A., & Smith, A. F. M. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, *85*(412), 972–985.

Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, *85*(410), 398–409. https://doi.org/10.1080/01621459.

1990. 10476213.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *PAMI-6*(6), 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

Geng, J., Zhang, B., Huie, L. M., & Lai, L. (2019). Online change-point detection of linear regression models. *IEEE Transactions on Signal Processing*, *67*(12), 3316–3329.

Ghaderinezhad, F., & Ley, C. (2020). On the Impact of the Choice of the Prior in Bayesian Statistics. In *Bayesian Inference on Complicated Data*. IntechOpen. https://doi.org/10.5772/intechopen.88994.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*(4), 711–732.

Hahn, G., Banerjee, M., & Sen, B. (2017). Parameter estimation and inference in a continuous piecewise linear regression model. *Manuscript, Department of Statistics, Columbia University (December 2016)*, *32*(2), 407–451.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications.

Henderson, R., & Matthews, J. N. S. (1993). An investigation of changepoints in the annual number of cases of haemolytic uraemic syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *42*(3), 461–471.

Hinkley, D., Chapman, P., & Runger, G. (1980). *Change-point problems*. University of Minnesota.

Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika*, *56*(3), 495–504.

Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables.

Hinkley, D. V. (1971). Inference in two-phase regression. *Journal of the American Statistical Association*, *66*(336), 736–743.

Holbert, D. (1982). A Bayesian analysis of a switching linear model. *Journal of Econometrics*, *19*(1), 77–87. https://doi.org/10.1016/0304-4076(82) 90051 -3.

Jaakkola, T. S., & Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, *10*(1), 25–37.

James, B., James, K. L., & Siegmund, D. (1987). Tests for a change-point. *Biometrika*, *74*(1), 71–83.

Jang. E. (2016). A Beginner's Guide to Variational Methods: Mean-Field Approximation. Retrieved October 1, 2021, from http://blog.evjang.com/2016/08/variational-bayes.html, 2016.

Jaynes, E. T. (2003). *Probability theory: The logic of science*. Cambridge university press.

Jiang, T. (2015). *Information Approach for Change Point Detection of Weibull Models with Applications*. Bowling Green State University.

Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*(1), 140–155.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., & Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine Learning*, *37*(2), 183–233.

Kang, S. (2015). *Bayesian change-point analysis in linear regression model with scale mixtures of normal distributions*. Michigan Technological University, Houghton, Michigan. https://doi.org/10.37099/mtu.dc.etds /920.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795.

Khodadadi, A., & Asgharian, M. (2008). Change-point problem and regression: an annotated bibliography. *COBRA Preprint Series*, 44.

Konishi, S., Ando, T., & Imoto, S. (2004). Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika*, *91*(1), 27-43.

Kullback, S. (1997). *Information theory and statistics*. Courier Corporation.

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The annals of mathematical statistics*, *22*(1), 79-86.

Lavielle, M., & Lebarbier, E. (2001). An application of MCMC methods for the multiple change-points problem. *Signal Processing*, *81*(1), 39–53.

Lee, K. H. (2004). Current Developments in the Discovery and Design of New Drug Candidates from Plant Natural Product Leads. In *Journal of Natural Products*. https://doi.org/10.1021/np030373o.

Liu, Y., Zou, C., & Zhang, R. (2008). Empirical likelihood ratio test for a change-point in linear regression model. *Communications in Statistics-Theory and Methods*, *37*(16), 2551–2563.

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

MacNeil, I. B., & Mao, Y. (1993). Change-point analysis for mortality and morbidity rate. *Journal of Applied Statistical Science 1*, *43*, 359–377.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., & Hawkins, D. M. (2007). A change point method for linear profile data. *Quality and Reliability Engineering International*, *23*(2), 247–268.

McGrory, C. A., & Titterington, D. (2007). Variational approximations in Bayesian model selection for finite mixture distributions. *Computational Statistics & Data Analysis*, *51*(11), 5352–5367.

McGrory, C. A., Titterington, D. M., Reeves, R., & Pettitt, A. N. (2009). Variational Bayes for estimating the parameters of a hidden Potts model. *Statistics and Computing*, *19*(3), 329.

Mensah, D. K. (2010). *Bayesian Methods for Measuring Acute Malnourishment*. Plymouth University, United Kingdom.

Ngunkeng, G., & Ning, W. (2014). Information approach for the change-point detection in the skew normal distribution and its applications. *Sequential Analysis*, *33*(4), 475–490.

Ninomiya, Y. (2015). Change-point model selection via AIC. *Annals of the Institute of Statistical Mathematics*, *67*(5), 943–961.

O'Hagan, & Jonathan Forster. (2005). Kendall's Advanced Theory of Statistics, Vol. 2B: Bayesian Inference (2nd ed.). *Journal of the American Statistical Association*, *2B*, 1465–1466. Retrieved from https://econpapers.repec.org/RePEc:bes:jnlasa:v:100:y:2005:p:1465-1466.

Ormerod, John T, & Wand, M. P. (2010). Explaining variational approximations. *The American Statistician*, *64*(2), 140–153.

Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, *41*, 100–115.

Page, E. S. (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, *42*, 523–527.

Page, E. S. (1957). On problems in which a change in a parameter occurs at unknown point. *Biometrika*, *44*, 248–252.

Pandya, M. M., & Sheth, P. K. (2016). Study on Two Phase Linear Regression Model: Bayesian Approach. *International Journal of Engineering and Management Research (IJEMR)*, *6*(2), 528–534.

Pettitt, A. N. (1979). A non-parametric approach to the change-point problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(2), 126–135..

Pool, J. G., & Borchgrevink, C. F. (1964). Comparison of rat liver response to coumarin administered in vivo versus in vitro. *American Journal of Physiology-Legacy Content*, *206*(1), 229–238.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association*, *53*(284), 873–880.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *Journal of the American Statistical Association*, *55*(290), 324–330.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 111–163.

Rahman, A., Gao, J., D’Este, C., & Ahmed, S. E. (2016). An Assessment of the Effects of Prior Distributions on the Bayesian Predictive Inference. *International Journal of Statistics and Probability*, *5*(5), 31.

https://doi.org/10.5539/ijsp.v5n5p31.

Robert, C. P., Ryden, T., & Titterington, D. M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *62*(1), 57–75.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.

Seber, G. A. F., & Lee, A. J. (2003). *Linear Regression Analysis*. *Linear Regression Analysis*. Wiley. https://doi.org/10.1002/9780471722199.

Seidou, O., Asselin, J. J., & Ouarda, T. B. M. J. (2007). Bayesian multivariate linear regression with application to change point models in hydrometeorological variables. *Water Resources Research*, *43*(8).

Shaban, S. A. (1980). Change Point Problem and Two-Phase Regression: An Annotated Bibliography. *International Statistical Review / Revue Internationale de Statistique*, *48*(1), 83–93. Retrieved from http://www.jstor.org/stable/1402408.

Sharma, S., Swayne, D. A., & Obimbo, C. (2016). Trend analysis and change point techniques: a survey. *Energy, Ecology and Environment*, *1*(3), 123–130.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike’s information criterion. *Biometrika*, *63*(1), 117–126.

Smith, A. F. M. (1975). A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika*, *62*(2), 407–416.

Son, Y. S., & Kim, S. W. (2005). Bayesian single change point detection in a sequence of multivariate normal observations. *Statistics*, *39*(5), 373–

387.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *64*(4), 583–639.

Sprent, P. (1961). Some hypotheses concerning two phase regression lines. *Biometrics*, *17*(4), 634–645.

Stephens, M. (2000). Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of Statistics*, 40–74.

Titterington, D. M. (2004). Bayesian methods for neural networks and related models. *Statistical Science*, 128–139.

Titterington, D. M., & Wang, B. (2006). Convergence properties of a general algorithm for calculating variational Bayesian estimates for a normal mixture model. *Bayesian Analysis*, *1*(3), 625–650.

Truong, C., Oudre, L., & Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing*, *167*. https://doi.org/10.1016/j.sigpro.2019.107299.

Tsurumi, H. (1980). A Bayesian estimation of structural shifts by gradual switching regressions with an application to the US gasoline market. *Bayesian Analysis in Econometrics and Statistics*, 213–240.

Ueda, N., & Ghahramani, Z. (2002). Bayesian model search for mixture models based on optimizing variational bounds. *Neural Networks*, *15*(10), 1223–1241.

Valente, F., & Wellekens, C. (2005a). Variational Bayesian speaker change detection. In *INTERSPEECH* (pp. 693–696).

Valente, F., & Wellekens, C. (2005b). Variational bayesian speaker change detection. *9th European Conference on Speech Communication and Technology*, 693–696. https://doi.org/10.21437/interspeech.2005-199.

Vatsa, R. (2011). *Variational Bayes Approximation for Inverse Regression Problems*. Trinity College Dublin.

Vostrikova, L. Y. (1981). Detecting disorder in multidimensional random processes. In *Doklady akademii nauk* (Vol. 259, pp. 270–274).

Waterhouse, S., MacKay, D., Robinson, T., & others. (1996). Bayesian methods for mixtures of experts. *Advances in Neural Information Processing Systems*, 351–357.

Worsley, K J. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, *74*(366a), 365–367.

Worsley, Keith J. (1986). Confidence regions and tests for a change-point in a sequence of exponential family random variables. *Biometrika*, *73*(1), 91–104.

Yang, Y. (2005). Can the strengths of AIC and BIC be shared? A conflict between model indentification and regression estimation. *Biometrika*, *92*(4), 937–950.

You, C., Ormerod, J. T., & Mueller, S. (2014). On variational Bayes estimation and variational information criteria for linear regression models. *Australian & New Zealand Journal of Statistics*, *56*(1), 73–87.

Zhang, N. R., & Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, *63*(1), 22–32.

Zhou, H., & Liang, K.-Y. (2008). On estimating the change point in generalized linear models. *Beyond Parametrics in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen*, *1*, 305–320. https://doi.org/10.1214/193940307000000239.

APPENDICES

APPENDIX A

DERIVATIONS FOR THE NULL MODEL

This section of the appendix is tailored towards the derivations of the required Variational updating equations for the development of the non-switching model, the Variational optimization function, and the derivation of the Variational Akaike Information Criteria (VAIC) and the Variational Bayes Information Criteria (VBIC) and their derivatives for change-point detection'

Appendix A. 1: Derivation of Variational Updates for Null Model

In this section, the Variational updates for the null models are derived. The derivation begins with the identification of the joint posterior distribution and its Variational counterpart. The joint posterior based on the assumed priors and defining data likelihood function for the null can be express in the form;

$$\begin{aligned} p(\vartheta_0|y) &\propto p(y|\vartheta_0)p(\vartheta_0) \\ &\propto p(y|\beta^0,\sigma^2_\in)p(\beta^0)p(\sigma^2_\in) \end{aligned} \qquad (35)$$

where $\vartheta_0=(\beta^0,\sigma^2_\in)$ denotes the set of parameters in the model. The corresponding Variational distributions assumed for (47) is separable of the form

$$q(\vartheta_0)=q(\beta^0)q(\sigma^2_\in) \qquad (36)$$

where $q(\beta^0)\sim N(\mu^q_\beta,\Sigma^q_\beta)$, , $q(\sigma^2_\in)\sim IG(a^q_\in,b^q_\in)$. Using equation (47), the required updates for the parameters are obtained based on the following equations.

$$q(\beta^0) \propto \exp\left[E_{-q(\beta^0)}\log\{p(y|\beta^0,\sigma^2_\in)p(\beta^0)\}\right] \qquad (37)$$

and

$$q\left(\sigma_\in^2\right) \propto \exp\left[E_{-q\left(\sigma_\in^2\right)}\log\left\{p\left(y|\beta^0,\sigma_\in^2\right)p\left(\sigma_\in^2\right)\right\}\right] \qquad (38)$$

Based on equations (48) and (51), the following are true.

$$q\left(\beta^0\right) \propto \exp\left[E_{-q\left(\beta^0\right)}\log\left\{p\left(y|\beta^0,\sigma_\in^2\right)p\left(\beta^0\right)\right\}\right] \qquad (39)$$

$$\propto \exp\left[E_{-q\left(\beta^0\right)}\log\left\{N\left(y;X\beta,\sigma_\in^2 I_n\right)N\left(\beta^0,\mu_\beta^0,\Sigma_\beta^0\right)\right\}\right] \qquad (40)$$

It can be deduced that

$$N\left(y;X\beta^0,\sigma_\in^2 I_n\right)N\left(\beta^0;\mu_\beta^0,\Sigma_\beta^0\right) \propto$$

$$\exp\left\{-\frac{1}{2\sigma_\in^2\Sigma_\beta^0}\left[\Sigma_\beta^0 D\left(y,X,\beta^0\right)+\sigma_\in^2 D\left(\beta^0,\mu_\beta^0\right)\right]\right\} \qquad (41)$$

Where $D\left(y;X,\beta^0\right)=\left(y-X\beta\right)'\left(y-X\beta\right)$, $D\left(\beta^0,\mu_\beta^0\right)=\left(\beta^0-\mu_\beta\right)'\left(\beta^0-\mu_\beta\right)$.

Completing squares in β of the exponent gives

$$N\left(y;X\beta^0,\sigma_\in^2 I_n\right)N\left(\beta^0;\mu_\beta^0,\Sigma_\beta^0\right) \propto \exp\left\{-\frac{1}{2\Sigma_\beta^{\text{¿}}}\left[\left(\beta-\mu_\beta^{\text{¿}}\right)'\left(\beta-\mu_\beta^{\text{¿}}\right)\right]\right\},$$

where

$$\Sigma_\beta^{\text{¿}}=\frac{\sigma_\in^2\Sigma_\beta^0}{\Sigma_\beta^0 X'X+\sigma_\in^2}, \quad \mu_\beta=\frac{\Sigma_{\beta_0}y'X+\sigma_\in^2\mu_\beta^0}{\Sigma_\beta^0 X'X+\sigma_\in^2 I_n}$$

We now take log, and compute the expectation with respect to variational distributions $q\left(\sigma_\in^2\right)$. After this, $q\left(\beta^0\right)$ can be expressed as

$$q\left(\beta^0\right)\sim N\,\text{¿¿} \qquad (42)$$

$$\Sigma_\beta^{q*\text{¿}=\left[\frac{a_\in^q}{b_\in^q}X'X+\Sigma_\beta^{0^{-1}}\right]^{-1}}\text{¿}, \quad \mu_\beta^{q*\text{¿}=\Sigma_\beta^{q*\text{¿}\left[\frac{a_\in^q}{b_\in^q}y'X+\mu_\beta^0\Sigma_\beta^{0^{-1}}\right]\text{¿}}}\text{¿}$$

Now, the updating equations for $q\left(\beta^0\right)$ are obtained by comparing $N\left(\mu_\beta^q,\Sigma_\beta^q\right)$ with $N\,\text{¿¿}$ in (42). This leads to the following expressions for estimating μ_β^q and Σ_β^q.

$$\Sigma_\beta^q \leftarrow \Sigma_\beta^{q*\text{¿}}\text{¿}, \quad \mu_\beta^q \leftarrow \mu_\beta^{q*\text{¿}}\text{¿}$$

The updating equations for $q\left(\sigma_{\in}^{2}\right)$ can be obtained using (51) as follows.

$$q\left(\sigma_{\in}^{2}\right) \quad \alpha \quad \exp\left[E_{-q(\sigma_{\in}^{2})}\log\left(p\left(y|\beta,\sigma_{\in}^{2}\right)\right)p\left(\sigma_{\in}^{2}\right)\right]$$

$$\alpha \quad \exp\left[E_{-q(\sigma_{\in}^{2})}\left[\log\left\{N\left(y;X\beta^{0},\sigma_{\in}^{2}I_{n}\right)\right\}+\log\left\{IG\left(\sigma_{\in}^{2},a_{\in}^{0},b_{\in}^{0}\right)\right\}\right]\right]$$

$$\alpha \quad \exp\left[\log\left(\sigma_{\in}^{2}\right)^{-\left(\frac{n}{2}+a_{\in}^{0}\right)-1}-\frac{1}{\sigma_{\in}^{2}}\left[\frac{1}{2}\left\{\mu_{y}\right\}+b_{\in}^{0}\right]\right]$$

$$\sim \quad IG\left(\left(\frac{n}{2}+a_{\in}^{0}\right),\frac{1}{2}\mu_{y}+b_{\in}^{0}\right) \qquad (43)$$

where $\mu_{y}=\left(y-X'\mu_{\beta}^{q}\right)'\left(y-X'\mu_{\beta}^{q}\right)+tr\left(X'X\Sigma_{\beta}^{q}\right)$. By comparing Now, $IG\left(a_{\in}^{q},b_{\in}^{q}\right)$ and (43), it is easy to see the resulting updating equations for estimating the scale and shape parameters of $q\left(\sigma_{\in}^{2}\right)$ are

$$a_{\in}^{q}\leftarrow\frac{n+2a_{\in}^{0}}{2}, \quad b_{\in}^{q}\leftarrow\frac{1}{2}\left[\left(y-X'\mu_{\beta}^{q}\right)'\left(y-X'\mu_{\beta}^{q}\right)+tr\left(X'X\Sigma_{\beta}^{q}\right)\right]+b_{\in}^{0}$$

Appendix A. 2: Computation of Variational Optimization Function.

The Variational lower bound defined by the null model based on the assumed Variational distributions in (26) can be written as;

$$L_{0}(q)=E_{q}\left[\log p\left(y|\beta^{0},\sigma_{\in}^{2}\right)\right]+E_{q}\left[\log p\left(\beta^{0}\right)\right]E_{q}\left[\log p\left(\sigma_{\in}^{2}\right)\right]-E_{q}\left[\log q\left(\beta^{0}\right)\right]-E_{q}\left[\log q\left(\sigma_{\in}^{2}\right)\right] \qquad (44)$$

The component of (44) are computed as follows.

$$E_{q}\left[\log p\left(y|\beta^{0},\sigma_{\in}^{2}\right)\right]=E_{q}\left[\log N\left(y;X'\beta^{0}\right)\right]$$

$$=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\left[\log\left(b_{\in}^{q}\right)\right]-\psi\left(a_{\in}^{q}\right)-\frac{a_{\in}^{q}}{2b_{\in}^{q}}\left[\left(y-X'\mu_{\beta}^{q}\right)'\left(y-X'\mu_{\beta}^{q}\right)+tr\left(X'X\Sigma_{\beta}^{q}\right)\right]$$

$$E_{q}\left[\log p\left(\beta^{0}\right)\right]=E_{q}\left[\log N\left(\beta^{0};\mu_{\beta}^{0},\Sigma_{\beta}^{0}\right)\right]$$

$$=-\frac{r}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma_{\beta}^{0}|-\frac{1}{2}\left[\left(\mu_{\beta}^{q}-\mu_{\beta}^{0}\right)'\left(\mu_{\beta}^{q}-\mu_{\beta}^{0}\right)+tr\left(\Sigma_{\beta}^{0^{-1}}\Sigma_{\beta}^{q}\right)\right]$$

$$E_q\left[\log p\left(\sigma_{\in}^2\right)\right]=E_q\left[\log IG\left(\sigma_{\in}^2;a_{\in}^0,b_{\in}^0\right)\right]$$

$$=a_{\in}^0\log(b_{\in}^0)-\log\Gamma(a_{\in}^0)-(a_{\in}^0+1)\left[\log(b_{\in}^q)-\Psi(a_{\in}^q)\right]-\frac{b_{\in}^0 a_{\in}^q}{b_{\in}^q}$$

$$E_q\left[\log p(\beta)\right]=E_q\left[\log N\left(\beta;\mu_\beta^q,\Sigma_\beta^q\right)\right]$$

$$=-\frac{r}{2}\log(2\pi)-\frac{1}{2}\log|\Sigma_\beta^q|-\frac{r}{2}$$

$$E_q\left[\log p\left(\sigma_{\in}^2\right)\right]=a_{\in}^q\log(b_{\in}^q)-\log\Gamma(a_{\in}^q)-(a_{\in}^q+1)\left[\log(b_{\in}^q)-\Psi(a_{\in}^q)\right]-a_{\in}^q$$

Appendix A.3: Computation of VAIC and VBIC for Null Model

In this subsection of the thesis, the VAIC and VBIC computational expressions are derived for the non-switching model. First we consider the general structures of the above information criterions. The VAIC and VBIC can be defined for the null model as;

$$VAIC=-2\log p\left(y|\vartheta_0^{¿}\right)+2p_D^{¿} \tag{45}$$

$$p_D^{¿}=2\log p\left(y|\vartheta_0^{¿}\right)-2E_q\left[\log p\left(y|\vartheta_0\right)\right]$$

$$VBIC=-2L_0(q)+2E_q\left[\log p\left(\vartheta_0\right)\right] \tag{46}$$

The component can be obtained from the Appendix A .1. and Appendix A. 2. In particular, we have

$$E_q\left[\log p\left(\vartheta_0\right)\right]=E_q\left[\log p\left(\beta^0\right)\right]+E_q\left[\log p\left(\sigma_{\in}^2\right)\right]$$

$$E_q\left[\log p\left(y|\vartheta_0\right)\right]=E_q\left[\log p\left(y|\beta^0,\sigma_{\in}^2\right)\right]$$

$$\log p\left(y|\vartheta_0^{¿}\right)=\log p\left(y|\beta^{0*¿,\sigma_{\in}^{2*¿)-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\sigma_{\in}^{2*¿)-\frac{1}{2\sigma_{\in}^{2*¿}}\left[(y-X\mu_\beta^q)'(y-X\mu_\beta^q)\right]¿}¿}¿}\right.$$

where $\sigma_{\in}^{2*¿=E_q\left[q(\sigma_{\in}^2)\right]=\frac{b_{\in}^q}{a_{\in}^q-1},a_{\in}^q>1.}$ ¿

APPENDIX B

DERIVATIONS FOR THE SWITCHING MODEL

This section of the Appendix focuses on the derivations for the change-point model. The computation of the Variational updates for the Variational algorithm, derivation of computational formulas for the Variational information namely, the VAIC, VBIC and the Variational optimization function termed the lower bound.

Appendix B.1: Derivation of Variational Updates for Switching Model

We consider the Variational parameter updating equations required for fitting algorithm. The defining structure for the

$$\begin{aligned} p(\vartheta_c / y) &\propto p(y / \vartheta_c) p(\vartheta_c) \\ &\propto p(y / \beta^a, \theta, k, \sigma^2_\in) p(\beta^a) p(\theta) p(\sigma^2_\in) p(k) \end{aligned} \qquad (47)$$

where $\vartheta_c = (\beta^a, \theta, k, \sigma^2_\in)$ denotes the set of parameters in the switching model. The Variational updating equations are derived as follows.

$$\begin{aligned} q(\beta^a) &\propto \exp\left[E_{-q(\beta^a)} \log\{p(y_k ¿ \beta^a, \theta, k, \sigma^2_\in) p(\beta^a)\}\right] \qquad (48) \\ &\propto \exp\left[E_{-q(\beta^a)} \log\{N(y_k; X_k \beta^a, \sigma^2_\in I_k) N(\beta^a; \mu_\beta, \Sigma_\beta)\}\right] \end{aligned}$$

It can be deduced that

$$N(y_k; X_k \beta^a, \sigma^2_\in I_k) N(\beta^a; \mu_\beta, \Sigma_\beta) \propto$$

$$\exp\left\{-\frac{1}{2\sigma^2_\in \Sigma_\beta}\left[\Sigma_\beta D(y_k, X_k, \beta^a) + \sigma^2_\in D(\beta^a, \mu_\beta)\right]\right\}, \qquad (49)$$

where $D(y_k, X_k, \beta^a) = (y_k - X_k \beta^a)'(y_k - X_k \beta^a)$, $D(\beta^a, \mu_\beta) = (\beta^a - \mu_\beta)'(\beta^a - \mu_\beta)$. Completing squares in β yields S

$$N(y_k; X_k\beta^a, \sigma_\in^2 I_k) N(\beta^a; \mu_\beta, \Sigma_\beta) \;\alpha\; \exp\left\{-\frac{1}{2\sigma_\in^2 \Sigma_\beta}\left[(\beta-\mu_{\beta^a}^{¿})'(\beta-\mu_{\beta^a}^{¿})\right]\right\},$$

where

$$\Sigma_{\beta^a}^{¿} = \frac{\sigma_\in^2 \Sigma_\beta}{\Sigma_\beta X_k' X_k + \sigma_\in^2 I_k}, \quad \mu_\beta = \frac{\Sigma_\beta y_k' X_k + \sigma_\in^2 \mu_\beta}{\Sigma_\beta X_k' X_k + \sigma_\in^2 I_k}$$

Taking log, followed by expectation with respect to $q(\sigma_\in^2)$, $q(\beta^a)$ yields

$$q(\beta^a) \sim N(\mu_{\beta^a}^{q*¿}, \Sigma_{\beta^a}^{q*¿}), ¿ \qquad (50)$$

where

$$\Sigma_{\beta^a}^{q*¿} = \left[\frac{a_\in^q}{b_\in^q} X_k' X_k + \Sigma_\beta^{-1}\right]^{-1} ¿, \quad \mu_{\beta^a}^{q*¿} = \Sigma_{\beta^a}^{q*¿}\left[\frac{a_\in^q}{b_\in^q} y_k' X_k + \mu_\beta \Sigma_\beta^{-1}\right]. ¿ ¿$$

The updates for $q(\beta^0)$ are obtained by comparing $N(\mu_\beta^q, \Sigma_\beta^q)$ with $N(\mu_{\beta^a}^{q*¿}, \Sigma_{\beta^a}^{q*¿}) ¿$ in (53). Thus, we have the following expressions for estimating $\mu_{\beta^a}^q$ and $\Sigma_{\beta^a}^q$.

$$\Sigma_{\beta^a}^q \leftarrow \Sigma_{\beta^a}^{q*¿} ¿, \quad \mu_{\beta^a}^q \leftarrow \mu_{\beta^a}^{q*¿} ¿$$

Next, we consider the updates for $q(\theta)$.

$$q(\theta) \;\alpha\; \exp\left[E_{-q(\beta^a)} \log\left\{N(y_{nk}; X_{nk}\theta, \sigma_\in^2 I_{nk}) N(\theta; \mu_\theta, \Sigma_\theta)\right\}\right] \qquad (51)$$

It can be deduced that,

$$N(y_{nk}; X_{nk}\theta, \sigma_\in^2 I_{nk}) N(\theta; \mu_\theta, \Sigma_\theta) \;\alpha$$

$$\exp\left\{-\frac{1}{2\sigma_\in^2 \Sigma_\theta}\left[\Sigma_\theta D^1(y_{nk}, X_{nk}, \theta) + \sigma_\in^2 D^1(\theta, \mu_\theta)\right]\right\} \qquad (52)$$

where $D^1(y_{nk}; X_{nk}, \theta) = (y_{nk} - X_{nk}\theta)'(y_{nk} - X_{nk}\theta)$, $D^1(\theta, \mu_\theta) = (\theta - \mu_\theta)'(\theta - \mu_\theta)$. Completing squares in β yields.

$$N(y_{nk};X_{nk}\theta,\sigma_{\in}^{2}I_{nk})N(\theta;\mu_{\theta},\Sigma_{\theta}) \quad \alpha \quad \exp\left\{-\frac{1}{2\sigma_{\in}^{2}\Sigma_{\theta}^{\text{¿}}}\left[(\theta-\mu_{\theta}^{\text{¿}})'(\theta-\mu_{\theta}^{\text{¿}})\right]\right\},$$

where

$$\Sigma_{\theta}^{\text{¿}}=\frac{\sigma_{\in}^{2}\Sigma_{\theta}}{\Sigma_{\theta}X_{nk}^{'}X_{nk}+\sigma_{\in}^{2}I_{nk}}, \quad \mu_{\theta}=\frac{\Sigma_{\theta}y_{nk}^{'}X_{nk}+\sigma_{\in}^{2}\mu_{\theta}}{\Sigma_{\theta}X_{nk}^{'}X_{nk}+\sigma_{\in}^{2}I_{nk}}$$

Taking log, followed by expectation with respect to $q(\sigma_{\in}^{2})$, $q(\theta)$ yields

$$q(\theta)\sim \; N(\mu_{\theta}^{q*\text{¿},\Sigma_{\theta}^{q*\text{¿}),\text{¿}}}\text{¿}$$

where $\Sigma_{\theta}^{q*\text{¿}=\left[\frac{a_{\in}^{q}}{b_{\in}^{q}}X_{nk}^{'}X_{nk}+\Sigma_{\theta}^{-1}\right]^{-1}}\text{¿}$, $\mu_{\theta}^{q*\text{¿}=\Sigma_{\theta}^{q*\text{¿}\left[\frac{a_{\in}^{q}}{b_{\in}^{q}}y_{nk}^{'}X_{nk}+\mu_{\theta}\Sigma_{\theta}^{-1}\right].}}\text{¿}\text{¿}$

The updates for $q(\theta)$ are obtained by comparing $N(\mu_{\theta}^{q},\Sigma_{\theta}^{q})$ with $N(\mu_{\theta}^{q*\text{¿},\Sigma_{\theta}^{q*\text{¿}}\text{¿}}}\text{¿}$ in (53). Thus, we have the following expressions for estimating μ_{θ}^{q} and Σ_{θ}^{q}.

$$\Sigma_{\theta}^{q}\leftarrow\Sigma_{\theta}^{q*\text{¿}}\text{¿}, \quad \mu_{\theta}^{q}\leftarrow\mu_{\theta}^{q*\text{¿}}\text{¿}.$$

$$q(\sigma_{\in}^{2}) \quad \alpha \quad \exp\left[E_{-q(\sigma_{\in}^{2})}\log(p(y/\beta,\theta,k,\sigma_{\in}^{2})p(\sigma_{\in}^{2})\right]$$

$$\alpha \quad \exp[E_{q(\sigma_{\in}^{2})}[\log\{N(y_{k};X_{k}\beta^{a},\sigma_{\in}^{2}I_{k})N(y_{nk};X_{nk}\theta,\sigma_{\in}^{2}I_{nk})\} +$$

$$+\log\{IG(\sigma_{\in}^{2},a_{\in}^{0},b_{\in}^{0}\}]]$$

$$\alpha \quad [\log(\sigma_{\in}^{2})^{-\left(\frac{n}{2}+a_{\in}\right)-1}-\frac{1}{\sigma_{\in}^{2}}\left[\frac{1}{2}\{B_{1}(y_{k},X_{k},\beta^{a})+B_{2}(y_{nk},X_{nk},\theta\}+b_{\in}\right]]$$

$$\sim \mathrm{IG}\left(\left(\frac{n}{2}+a_{\in}\right),\frac{1}{2}[B_{1}+B_{2}]+b_{\in}\right) \qquad (54)$$

Where

$$B_{1}=(y_{k}-X_{k}^{'}\mu_{\beta^{a}}^{q})'(y_{k}-X_{k}^{'}\mu_{\beta^{a}}^{q})+tr(X_{k}^{'}X_{k}\Sigma_{\beta^{a}}^{q}),$$

$$B_{2}=(y_{nk}-X_{nk}^{'}\mu_{\theta}^{q})'(y_{nk}-X_{nk}^{'}\mu_{\theta}^{q})+tr(X_{nk}^{'}X_{nk}\Sigma_{\theta}^{q}).$$

By comparing now, IG $\left(a^q_{\in}, b^q_{\in}\right)$ and (54), we have the updating equations for $q\left(\sigma^2_{\in}\right)$.

$$a^q_{\in} \leftarrow \tfrac{n}{2} + b_{\in} \;, \quad b^q_{\in} \leftarrow \tfrac{1}{2}\left(B_1 + B_2\right) + b_{\in} \;.$$

Appendix B. 2: Computation of Variational Optimization Functions for Switching Model.

Based on the Variational approximation defined in (25), the structure of the variational lower bound defined by the switching model is of the form,

$$L_c(q) = E_q\left[\log p\left(y/\beta^a, \theta, \sigma^2_{\in}\right] + E_q\left[\log p\left(\beta^a\right)\right] + E_q\left[\log p(\theta)\right] + E_q\left[\log p\left(\sigma^2_{\in}\right)\right]\right.$$

$$+ E_q\left[\log p(k)\right] - E_q\left[\log q\left(\beta^a\right)\right] - E_q\left[\log q(\theta)\right] - E_q\left[\log q\left(\sigma^2_{\in}\right)\right] - E_q\left[\log q(k)\right]. \quad (55$$

)

The computational expressions for the components of (55) are as follows.

$$E_q\left[\log p\left(y/\beta^a, \theta, \sigma^2_{\in}\right)\right] = \; E_q\left[\log N_k\left(y_k; X_k \beta^a, \sigma^2_{\in} I_k\right)\right]$$

$$+ E_q\left[\log N_{(n-k)}\left(y_{(n-k)}; X_{(n-k)}\theta, \sigma^2_{\in}\right)\right] - \tfrac{n}{2}\log(2\pi) - \tfrac{n}{2}\left[\log\left(b^q_{\in}\right) - \psi\left(a^q_{\in}\right)\right]$$

$$-\frac{a^q_{\in}}{2b^q_{\in}}\left[\left(y'_k y_k + y'_{nk} y_{nk} - 2y'_k X_k \mu^q_{\beta} - 2y'_{nk} X_{nk} \mu^q_{\theta} + D_1 + D_2\right]\right.$$

$$E_q\left[\log p\left(\beta^a\right)\right] = E_q\left[\log N\left(\beta^a; \mu_{\beta}, \Sigma_{\beta}\right)\right]$$

$$= -\tfrac{r}{2}\log(2\pi) - \tfrac{1}{2}\log\left|\Sigma_{\beta}\right| - \tfrac{1}{2}\left[\left(\mu^q_{\beta^a} - \mu_{\beta}\right)'\left(\mu^q_{\beta^a} - \mu_{\beta}\right) + tr\left(\Sigma^{-1}_{\beta}\Sigma^q_{\beta^a}\right)\right]$$

$$E_q\left[\log p(\theta)\right] = E_q\left[\log N\left(\theta; \mu_{\theta}, \Sigma_{\theta}\right)\right]$$

$$= -\tfrac{r}{2}\log(2\pi) - \tfrac{1}{2}\log\left|\Sigma_{\theta}\right| - \tfrac{1}{2}\left[\left(\mu^q_{\theta} - \mu_{\theta}\right)'\left(\mu^q_{\theta} - \mu_{\theta}\right) + tr\left(\Sigma^{-1}_{\theta}\Sigma^q_{\theta}\right)\right]$$

$$E_q\left[\log p\left(\sigma^2_{\in}\right)\right] = E_q\left[\log IG\left(\sigma^2_{\in}; a_{\in}, b_{\in}\right)\right]$$

$$= a_{\in}\log\left(b_{\in}\right) - \log\Gamma\left(a_{\in}\right) - \left(a_{\in} + 1\right)\left[\log\left(b^q_{\in}\right) - \psi\left(a^q_{\in}\right)\right] - \frac{b_{\in} a^q_{\in}}{b^q_{\in}} \;.$$

$$E_q[\log p(k)] = E_q\left[\log \frac{1}{(n-2p)}\right] = -\log(n-2p)$$

$$E_q[\log q(\beta)] = -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma^q_{\beta^a}| - \frac{r}{2}$$

$$E_q[\log q(\theta)] = -\frac{r}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma^q_\theta| - \frac{r}{2}$$

$$E_q\left[\log q(\sigma^2_\in)\right] = a^q_\in \log(b^q_\in) - \log\Gamma(a_\in) - (a_\in + 1)\left[\log(b^q_\in) - \psi(a^q_\in)\right] - a^q_\in$$

$$E_q[\log q(k)] = -\log(n-2p)$$

$D_1 = tr\left(\left(\mu^q_\beta \mu^{q'}_\beta + \Sigma^q_\beta\right) X'_k X_k\right)$, $D_1 = tr\left(\left(\mu^q_\theta \mu^{q'}_\theta + \Sigma^q_\theta\right) X'_{nk} X_{nk}\right)$, $X_{nk} = X_{(n-k)}$,

$y_{nk} = y_{(n-k)}$.

Appendix B. 3: Computation of VAIC and VBIC for the Switching Model.

The computation of the VAIC and VBIC for the switching model follows from that of the non-switching model in Appendix A. 3. We can write

$$VAIC = -2\log p(y/\vartheta^{¿}_c) + 2p^{¿}_D \tag{56}$$

$$p^{¿}_D = 2\log p(y/\vartheta^{¿}_c) - 2E_q\left[\log p(y/\vartheta_c)\right]$$

and

$$VBIC = -2L^k_c(q) + 2E_q\left[\log p(\vartheta_c)\right] \tag{57}$$

$$E_q\left[\log p(\vartheta_c)\right] = E_q\left[\log p(\beta^a)\right] + E_q\left[\log p(\theta)\right] + E_q\left[\log p(\sigma^2_\in)\right] + E_q\left[\log p(k)\right]$$

$$E_q\left[\log p(y/\vartheta^{¿}_c)\right] = E_q\left[\log p(y/\beta^a, \theta, \sigma^2_\in)\right]$$

$$\log p(y/\vartheta^{¿}_c) = \log p(y/\beta^{a*¿,\theta^{¿},\sigma_\in^{2*¿) - \frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_\in^{2*¿)}¿}¿}¿$$

$$-\frac{1}{2(\sigma^{2*¿)}_\in}\left[(y_k - X_k\mu^q_{\beta^a})'(y_k - X_k\mu^q_{\beta^a}) + (y_{nk} - X_{nk}\mu^q_\theta)'(y_{nk} - X_{nk}\mu^q_\theta)\right], ¿$$

where $\sigma_\in^{2*¿ = E_q\left[q(\sigma^2_\in)\right] = \frac{b^q_\in}{a^q_\in - 1}¿}$,