A BAYESIAN MODEL FOR FORECASTING THE CHOICE OF CANDIDATE IN A PRESIDENTIAL ELECTION

JEREMIAH KIMANI KIINGATI

DOCTOR OF PHILOSOPHY
(Applied Statistics)

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

2021

A Bayesian Model for Forecasting the Choice of Candidate in a Presidential Election

Jeremiah Kimani Kiingati

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Applied Statistics of the Jomo Kenyatta University of Agriculture and Technology

2021

DECLARATION

This thesis is my original work and has not been presented for a degree in any other University.

Signature: ______________________ Date: ______________________

Jeremiah Kimani Kiingati

This thesis has been submitted for examination with our approval as University Supervisors.

Signature: ______________________ Date: ______________________

Prof. Samuel M. Mwalili

JKUAT, Kenya

Signature: ______________________ Date: ______________________

Prof. Anthony Waititu

JKUAT, Kenya

DEDICATION

To my maternal grandmother Njeri wa Githu, my loving mother Jane Wambui, the memory of my son Juan Bayes, my wife Vero and my children Asuntah, Arleen and Jaden.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep appreciation to the most high and sovereign God who has been my strong pillar and my source of inspiration, wisdom, knowledge and understanding. He has been the source of my strength throughout this program and it's such an honour and a privilege. Secondly, I would like to thank the following people, without whom I would not have been able to complete this research.

Special thanks to my patient and supportive supervisors, professors Samuel Mwalili and Anthony Waititu, for their consistent support, encouragement, insightful guidance and feedback which always made me feel confident in my abilities throughout this project. You have such a remarkable hope and patience knowing when to push and when to let up. I really appreciate.

The Kenyan media houses and opinion pollsters, member of the fourth and fifth estates, need to be applaud for they are the making sources of these Kenyan opinion polls data and without them I would have no content for my thesis.

I must acknowledge as well the many friends, colleagues, and church mates who shared their memories and experiences. Thank you for our friendly chats, prayers and personal support in my academic endeavours. The meetings and conversations were vital in inspiring me to think. I am extremely grateful to Robert Kasisi and my co-mentors Henry and Caroline for their invaluable advice and consistent encouragement, as well as insightful editing assistance in the final stretch.

Finally, I cannot forget to offer my gratitude and deep appreciation to my family who were always willing and enthusiastic to assist in any way they could throughout the research. For my kids, sorry for being even grumpier than normal whilst I wrote this thesis and for my wife Veronicah, thanks for all your support, without which I would have stopped these studies a long time ago. You have been amazing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS AND ACRONYMS

ARMA	Autoregressive-Moving Average
Be	Beta
GDP	Gross Domestic Product
CORD	Coalition for Reforms and Democracy
KANU	Kenya National African Unity
KNBS	Kenya National Bureau of Statistics
MCMC	Markov chain Monte Carlo
M-H	Metropolis-Hastings
SES	Socio-Economic Status

ABSTRACT

Opinion poll plays a prominent role as a source of information in different societies of the world and are permanent feature of contemporary politics. Such election surveys have several purposes, including forecasting elections outcomes and studying the distribution of votes as they vary over geographic, demographic and political variables. Most of the opinion polls data are analyzed superficially using discrete or summary statistics. This work provides an in-depth statistical inference to these data. The goal of this thesis is to formulate and apply Bayesian model for comparing the two leading candidates in the presidential voting process. Although much of the media attentions during presidential election years focuses on polls tracking popular support for the major candidate, the vital role played by Bayesian statistical analysts in predicting elections incorporated in order to address forecasting election outcomes. We considered a Bayesian hierarchical model in predicting Kenya's presidential elections outcomes based on pre-election polls collected at most four months prior to the 2007 and 2013 general elections taking into account the evolution of opinions during campaigns Kenya's presidential elections are predictable and we were able to come up with a powerful methodological option for predicting the outcome of Kenya's presidential elections that uses Bayesian estimation approach and incorporates polling data to account for the evolution of opinions during campaigns. The results show that the leading candidate in the polls will win the election if the observed pattern does not portray misclassification; otherwise, the race is too close to call if there is underlying uncertainty. In conclusion, the research obtained predictions which suffices to prove that the main novel points of the analysis - namely the use of representative areas, the Bayesian analysis of an appropriately chosen hierarchical model and the probabilistic classification of undecided vote in opinion polls – are certainly important points in the right direction.

CHAPTER ONE

INTRODUCTION

1.1 Background

Understanding voter preferences and forecasting the final outcome of elections is of critical importance to politicians, as they can use the insights gained from the exercise to fine-tune their campaign strategies. Consequently a substantial literature on the prediction of election is available Some authors like Campbell (2012, 2008) and Kennedy *et al.* (2017) have used regression models for nationwide polling data to forecast the outcome of the popular vote in the US elections while others use state-level polls to make quantitative estimate of the proportion of votes for the two major party candidates in each state (Abrams, 1970; Cohen, 1998; Holbrook and DeSart, 1999; Cohen and Hamman, 2003; Cohen, 2015; Wang *et al.*, 2015; Campbell *et al.*, 2017; Jaidka *et al.*, 2019).

However, prediction of who will actually win the presidency, the issue of principal interest, is not addressed by the analyses of popular opinion trends Gelman *et al.* (2004, 2016) employed a multilevel logistic regression model to generate estimates of state-level vote shares in US. Their model employs national opinion data and state level demographic covariates to obtain estimates in a manner that is related to the small-area estimation problem.

Predicting election day voting outcomes based on early pre-election polling is a very complicated problem because such a prediction would require a consideration of opinion trends, future campaign spending, and historical voter behaviour. Further, the actual election day results will be affected by many unpredictable factors arising in the final days of the campaign, including world events and candidate mistakes. Effectively, the majority will, is difficult to measure and fundamentally ambiguous (Gelman *et al.*, 2002; Feld and Grofman, 2010; Gayo-Avello, 2013). Consequently, we focus on the simpler problem of estimating the probability that the incumbent president would win the election if it were held on the day of the recent poll

Since the advent of scientific polling in the 1930s, most notable the successful prediction of the re-election of Franklin Roosevelt by Gallup in the 1936 US presidential elections, there has been a growing interest by political pundits and scholars alike to predict the winner of the presidential elections., This unquenchable urge of beeping into the future of an election has become a vibrant discipline in political research. Recently, extensive and statistically based polling procedures draws a lot from the work of the Gallup Organization (Cameron and Crosby, 2000; Jackman, 2005; Wolfers and Leigh, 2002; Wattenberg, 2003; Cuzán *et al.*, 2005).

Technological advancement and availability of enormous credible data has made the task of predicting election relatively easier in advanced democracies. Since 1970s onward, a wide range of forecasting techniques have been developed. Several authors have used structural models for countrywide polling data to forecast the outcome of the presidential race. Others have employed the space state models while others used the time series models. Although forecasting has now quite a history especially in developed countries, analyses of popular opinion trends in sub–Saharan Africa has numerous challenges the most predominant being limited and flawed polling data that depicts marginal fluctuations among the pollsters.

Nevertheless, predicting the effectual winner of an election has not been a rosy affair for political scientist even in developed economies. According to Biemer (2010) and Groves and Lyberg (2010), polls are subject to various types of error and they are far from being perfect predictors of election outcomes themselves. Buchanan (1986) and Shirani-Mehr *et al.* (2018) found that the empirical error of polls is about twice as large as the estimated sampling error. This has given critics room for discussion.

The dynamics of opinion poll data due to factors like change of campaign strategies (such as the use of propagandas), demographic covariates, economic variables (such as inflation and unemployment), abstract ideas (such as change of status quo) and the effect of news is quite a puzzle to political science researchers. As such the question of how to treat the previous and current pre-election polls data is inevitable. Some researchers consider only the most recent poll others Combine all previous polls up the present time and treat it as a single sample, weighting only by sample size, while others Combine all previous polls but adjust the sample size according to a weight function depending on the day the poll is taken. In predicting Kenya's presidential elections outcomes based on pre-election polls we will apply a sequential Bayesian model.

In growing economies, ethnicity and the theory of mistrust are key drivers of political activities. Nearly all trilling candidates doubt the poll and/or pops holes on the credibility of the pollster/ institution. Luckily, the contemporary view of competitive elections as the hallmark of modern democracy, evidenced by the media freedom and freedom of expression, has recently helped to spur growth in forecasting research due to the increase in the number of polls.

Historically, one of the main challenges associated with forecasting election outcomes in sub Saharan Africa, and in particular Kenya, has been the lack of credible pre-election poll data (Cohen, 1998). However, as the growing economies transit from long historical experience of authoritarianism and dictatorship including a period of "cultures of silence," to freedom of expression, the growth of the opinion polls industry is phenomenal and their published results are now becoming easily accessible through various platforms such as social media and Internet. For instance, the 2007 presidential election recorded the highest number of opinion polls ever conducted and published in Kenya.

Although pre-election poll data is inevitably flawed, they can still provide much insight about national and regional trends and the nontrivial biases inherent in such data provides a vehicle for discussion about data validity and the associated validity of statistical inferences.

Scholars have argued that presidential pre-election polling data may not be useful until one year to the election to predict the winning president between any two strong aspirants. Thus, this research describes a method used to predict the outcome of presidential election by monitoring how people intend to vote throughout the electoral campaign. The argument lies entirely within the Bayesian framework. We argue that the closeness of recent Kenyan presidential opinion polls coupled with the wide accessibility of data should change how presidential election forecasting is conducted. We present a Bayesian forecasting model that concentrates on the national wide pre-election polls prior to 2007 general elections and consider finer details such as third-party candidates and self-proclaimed undecided voters. We incorporate our estimators into WinBUGS to determine the probability that a candidate will win an election. The model predicted the outright winner for the 2007 Kenyan election.

1.2 The 2007 and 2013 Elections in Kenya

Kenya become a multi-party democracy in 1992. There has been general multi-party elections in 1992, 1997, 2002, 2007, 2013 and 2017. In this Section we show the results for the 2007 and 2013 elections only. Figure 1.1 in the left and right panel respectively show the presidential results of the leading candidates per 8 former provinces and 47 counties. This election was won by Mr Kibaki at 46.4%

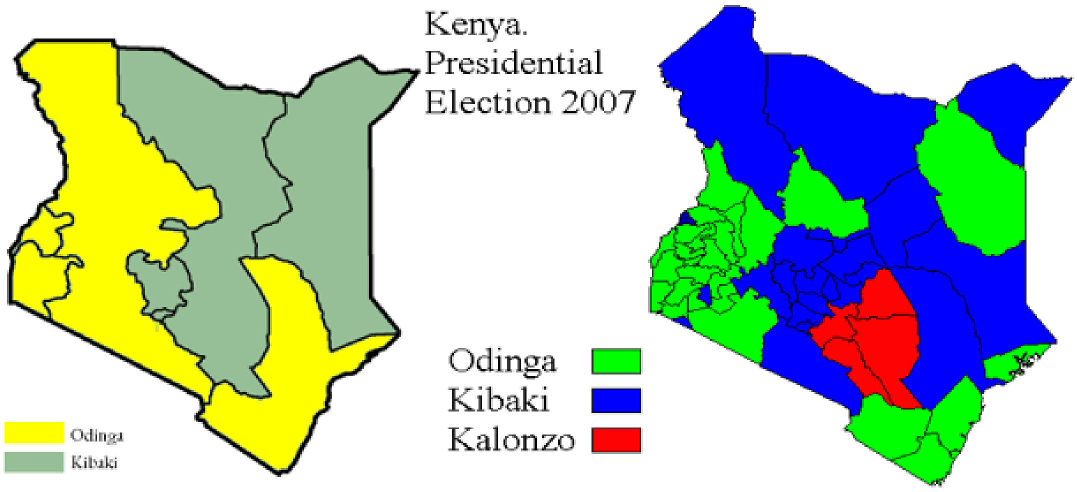

Figure 1.1: *Distribution of the Kenya 2007 presidential election votes.*

Figure 1.2 shows the 2013 election per county. The two-colour scheme shows the intensity of the support for the CORD (Coalition for Reforms and Democracy) the opposition party in shade of RED and JUBILEE in shade of BLUE. There was no incumbent in this election. The election appears to have been evenly contested. In this election the JUBILEE party won by 50.1%

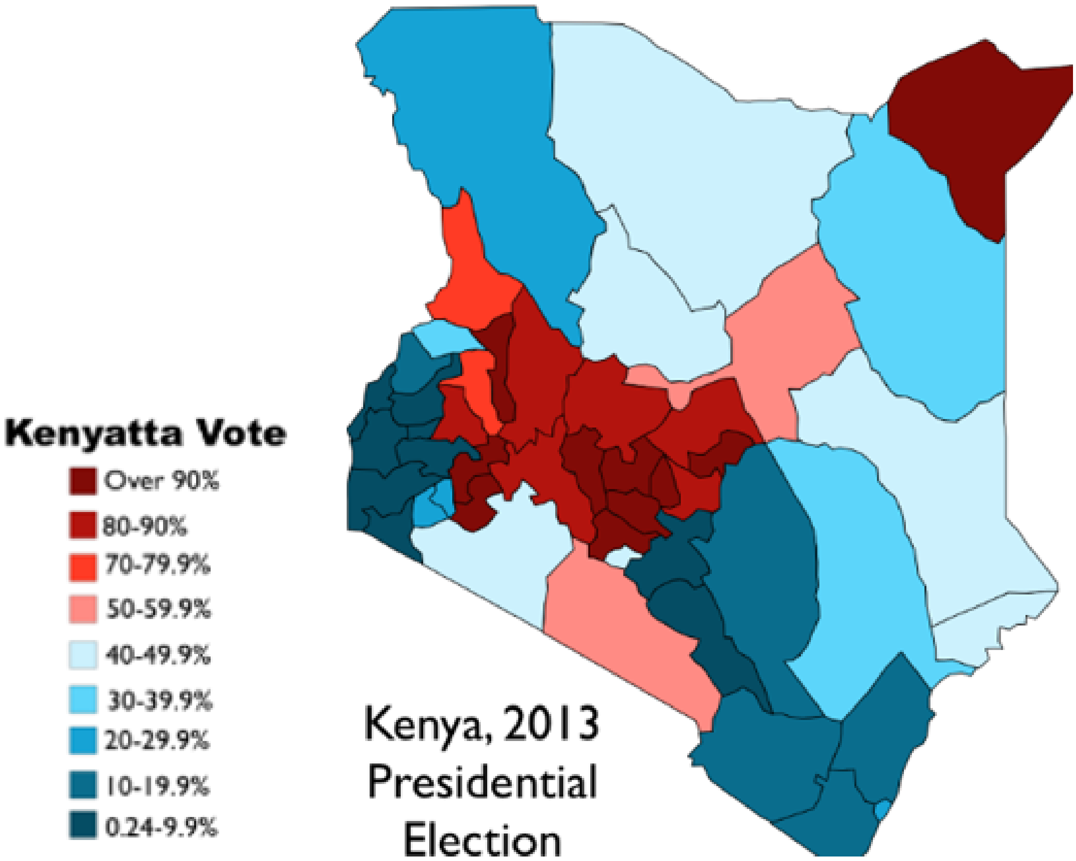

Figure 1.2: *Distribution of the Kenya 2013 presidential election votes across counties.*

1.3 Development of Opinion Polling in Kenya

Some democracies have discussed and/or regulated the publication of political opinion polls fearing the undue influence on voters. In Kenya, the Statistics Act, 2006, that became law in September 2006, mandated the Kenya National Bureau of Statistics to regulate opinion polling but since 2006 was the eve of an election year, this move was perceived as an effort to muzzle pollsters in order to keep politically-based opinion polls which are not in the government's favour away from the public. As such this Act has never been invoked.

Opinion poll plays a prominent role as a source of information in different societies of the world. Basically, they have become a permanent feature of contemporary politics, in fact, in the last few decades, the dependence on opinion polling in politics has increased in some liberal democracies that where the media is referred to as the “Fourth Estate,” opinion polling has been reckoned as the “Fifth Estate” in some recent studies. Unfortunately, in sub-Saharan Africa little of such studies have been happening.

Kenya particularly has never been adept at predicting election outcomes before 2002 few opinion polls were being done. Censorship during the KANU era caused fear and many firms could not conduct opinion polls and media organizations could not even publish the results. In 1997 and 2002, a few opinion polls were conducted, which predicted that the then President Daniel Moi would win the election but none were made public. After 2002 with the increased democratic space, opinion polls have been regularly conducted with two target groups: the general public and business leaders. General public opinion polls seek to provide systematic and representative public perceptions on social, economic and cultural issues. Knowing how to evaluate and place this type of information into context is very essential to any user.

In Kenya, pollsters began to openly conduct and release opinion polls to the media after the 2002 elections. The results of these opinion poll have elicited mixed reactions. As expected, opposing political parties make charges and counter charges on the efficacy and genuineness of various polls. This depicts the high level of competition and dependence on the polls by various political parties, in using them as campaign inputs. It is common for politicians to use favourable findings in their election speeches.

Since each election is fought on certain themes and issues which are propped up by either the electorate, political parties, interest groups or media, opinion polls reflecting the public opinions and expectations on various issues can bring about a meaningful political debate, which undoubtedly will strengthen the political process.

Some people have argued that publishing opinion polls would impact the political reality by causing a bandwagon effect, or provide an “underdog” effect, possibly revealing sympathy and support for the apparent loser. These two possible effects have received a fair amount of attention in literature. Although the polls themselves attract attention over sample size and methodology, there has been no serious research conducted in Kenya to gauge their impact on the election outcome. Such processes are of theoretical interest as they affect stable prediction leading to compromised models of the social sciences.

Interestingly, a large majority of the Kenyan electorate have faith in the opinion polls and their potential use in the process of Kenya's emerging democracy. Sunday Nation, December 9, 2007 pp 2. The highest number of opinion polls ever conducted in Kenya, which elicited mixed reactions from the electorate and politicians alike, was registered during the run up to the 2007 general election. This controversy raises two scenarios: the manipulative power of the opinion polls on the psyche of the electorate, and secondly, the genuineness and scientific basis of various polls.

1.4 Statement of the Problem

Several researchers in the field of political science focus their attention on the historic problem of understanding and predicting election outcomes. Understanding voter preferences and forecasting the final outcome of elections is of critical importance to politicians, as they can use the insights gained from the exercise to fine-tune their campaign strategies. Most studies have used regression models for nationwide polling data to forecast the outcome of the popular vote. The prediction of who will actually win the presidency, the issue of principal interest in many polls, is not addressed by the analyses of popular opinion trends. Thus, showing a need for robust statistical method for analysing opinion poll data.

The actual Election Day results will be affected by many unpredictable factors arising in the final days of the campaign, including world events and candidate mistakes. Effectively the will of majority is difficult to measure and fundamentally ambiguous. Consequently, we focus on the simpler problem of estimating the probability that the incumbent president would win the election if it were held on the day of the recent poll. A Bayesian approach present a powerful tool for overcoming this problem by combining prior information from various and repeated opinion polls as prior distribution.

1.5 Objectives of the Study

1.5.1 Main objective

The main aim of this research is to develop a Bayesian model for forecasting the choice of candidates in a presidential election.

1.5.2 Specific objectives

The Specific objectives of this study are:

1. To formulate a Bayesian model for comparing two candidates in presidential voting process
2. To formulate a multinomial Bayesian model for comparing three or more candidates in presidential voting process
3. To forecast presidential election performance using a simulation study
4. To apply the formulated Bayesian mode to real election data in Kenya

1.6 Significance of the Study

Currently in Kenya, very few people do understand the political voting process. In addition, results from pollsters have been highly doubted due to lack of statistical based approach and backing. This research therefore seeks to explore statistical approaches that will enable Kenyans to appreciate and understand the voting process, determinants of a vote as well as prediction of the future outcomes. Moreover this work will lay a statistical bases for the contemporary pollsters in application of statistical tools for analysing their results. This research is also geared towards helping policy makers to change their views and approaches as they will better understand the political voting system in Kenya.

1.7 Structure of the Thesis

This thesis has five chapters. Subsequent to this introductory chapter is the Chapter 2, which present literature review relevant to our study. Here a review is made of the methods in use as well as some earlier studies. Chapter 3 consists of methodology used in carrying out this study. In this chapter, Bayesian methods for analysing opinion poll data are formulated. Chapter 4 presents simulated results as well as the results and discussion of applying the methods to Kenyan situation. Finally, Chapter 5 gives conclusion and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Opinion poll plays a prominent role as a source of information in different societies of the world. Basically they have become a permanent feature of contemporary politics, in fact, so much has the dependence on opinion polling in politics increased in the last few decades in some liberal democracies that where the media is referred to as the "Fourth Estate," then some recent studies reckon opinion polling to be the "Fifth Estate." Consequently, knowing how to evaluate and place this type of information into context is very essential to any user.

Since the successful prediction of the re-election of Franklin Roosevelt by Gallup in the 1936 US presidential elections, the unquenchable urge of beeping into the future of an election has become a vibrant discipline in political research. Recently, extensive and statistically based polling procedures draws a lot from the work of the Gallup Organization (Cameron and Crosby, 2000; Wolfers and Leigh, 2002; Jackman, 2005; Cuzán *et al.*, 2005; Armstrong and Graefe, 2011).

Adoption of elections to recruit political leaders began in Western Europe in the fourteenth and fifteenth centuries (Hogan and Hogan, 1987). Since then liberal democracies has institutionalized the method to legitimize leaders and governments (Lakeman, 1974; Moyo, 1992). Wanyande (2006) argued that elections give the electorate opportunities to indirectly participate in governance or influence the way they are governed. In growing democracies, state power confers many advantages orchestrating hotly contested elections which are usually characterized by accusations of rigging and unfairness in the electoral process.

Several theoretical and empirical researches which identifies functions performed by elections in liberal democracies exist globally (Hogan and Hogan, 1987; Lakeman, 1974; Moyo, 1992; Ashworth, 2012; Hendrix, 2010; Mahler, 2017). However, studies on elections in Africa pays scanty attention to the factors that influence voter behaviour and turnout which is a core concept for understanding any election. Understanding voter preferences and forecasting the final outcome of elections is of critical importance to politicians as they can use the insights gained from the exercise to fine-tune their campaign strategies.

Forecasting election is a relatively recent and increasingly popular discipline in political science. The local straw poll conducted by The Harrisburg Pennsylvanian in 1824 (Tankard Jr, 1972; Smith, 1990), showing Andrew Jackson leading John Quincy Adams by 335 votes to 169 in the contest for the United States Presidency being the earliest documented form of election predicting. The straw poll correctly predicted Jackson popular vote win in that state. Consequently, such straw votes became gradually more popular, though they remained local. The 1916 Literary Digest national survey correctly predicted Woodrow Wilson's election as president. They also correctly predicted the victories of Warren Harding n 1920, Calvin Coolidge in 1924, Herbert Hoover in 1928, and Franklin Roosevelt in 1932 (Campbell, 2020; Walton *et al.*, 2017; Lohr and Brick, 2017).

The Gallup successful prediction of Franklin Roosevelt overwhelming victory over Alf Landon in the 1936 United States presidential race drew a surge of interest in statistically based polling procedures (Venkataramani, 1960; Grant Jr, 1992). Since that election over 80 years ago, most election surveys have invoked statistical sampling methods and statistical data analysis techniques. Moreover, a wide range of statistical forecasting techniques have been developed from 1970s onwards.

Election forecasting is becoming a more diverse and relevant as evidenced by the growing interest by political pundits and scholars alike to predict the winner of the presidential elections. This has been catalyzed by the ideology that competitive elections are the hallmark of modern democracy. The ability to foreshadow winner is a tantalizing skill that has garnered significant scientific attention (Butler and Kavanagh, 1997; Walther, 2015; Gelman *et al.*, 2004; Lewis-Beck and Dassonneville, 2015; Jackman, 2005). The goal of this delimited task to most researchers is not to explain election outcomes but to describe and predict them.

Technological advancement and availability of enormous credible data has made the task of predicting election relatively easier in advanced democracies (Mavragani and Tsagarakis, 2016; Newman, 2017). Several authors have used regression models for countrywide polling data to forecast the outcome of the presidential race and state-level polls to predict the outcome of the election in each state while others used the time series models. Such election surveys, particularly those released in the election year, appeals to a basic human urge to peek into the future and they stimulate considerable debate and speculation amongst the media, the public and politicians.

Basically, predicting the effectual winner of an election has not been a rosy affair for political scientist even in developed economies. Despite the many successful forecasts and numerous scholarly works, historic election surveys records depict mixed results. In the 1936 US presidential election for example, literacy digest predicting a comfortable win for Landon (57%) against incumbent Roosevelt's (43%). The actual elections results were 62% for Roosevelt against 38% for Landon. Their 19% sampling error was majorly attributed to sample bias (Katz, 1941; Venkataramani, 1960; Grant Jr, 1992). On contrary using Quota sampling George Gallup correctly predict a victory for Roosevelt using a sample of about 50,000 people. He also predicted the winner of the 1940 and 1944 elections. To minimize nonresponse bias for the 1948 election between Thomas Dewey and Harry Truman, Gallup employed professionals to conduct the interviews on each of the individuals in a fairly small sample of about 3250 respondents (Hogan, 1997; Sitkoff, 1971; McDonald *et al.*, 2001; Topping, 2004) . This notwithstanding election forecasting has garnered increasing attention in the recent years. Biemer (2010); Groves and Lyberg (2010) argued that polls are subject to various types of error implying they are far from being perfect predictors of election outcomes. Buchanan (1986); Shirani-Mehr *et al.* (2018) found that the empirical error of polls is about twice as large as the estimated sampling error. Even though the use of expert judgment in forecasting elections dates back long before the emergence of scientific polling (Kernell, 2000), we know surprisingly little about the relative accuracy of experts and polls. Rampant forecast from betting markets is expected during the last few days prior to the elections. Ideally experts are expected to make useful predictions for problems for which they get good feedback about their prediction accuracy as they know the situation well (Greene, 1993).

However, the outcomes of elections, the basis upon which the accuracy and bias of forecasters is judged, can be used by political experts to draw on a vast amount of theory and empirical evidence about electoral behavior which should help them to progressively develop techniques that can also be used in many other related disciplines as well. For instance, Erikson and Wlezien (2013) argued that polls tend to tighten. Researchers have also shown that certain campaign events like party conventions (Campbell *et al.*, 1992) and candidate debates (Benoit *et al.*, 2003) can yield predictable shifts in the candidates' polling numbers, not necessarily by affecting people's vote preference but rather their willingness to participate in a poll (Gelman *et al.*, 2016)

Although forecasting has now quite a history especially in developed countries, analyses of popular opinion trends has been highly doubted by majority of the contestants in sub Saharan Africa probably due to lack of credible pre-election data or due to the tyrannical nature of the governing elites who have had little interest in such externally-determined statements about their ability to deliver what the voters want (Cheeseman, 2008; Muhati, 2014). In Kenya for instance, most books on politics and even elections simply do not mention opinion polls. They were never carried out under the single party state, either because they were logistically impossible, technically difficult (prior to personal compuerts (PCs) and statistical software packages), considered pointless (rightly or wrongly) in a one-party state or inhibited or prevented by the government. There is therefore very little published history on this issue until the 1990s, when the introduction of multi-party democracy was to result in a rapid expansion of interest in and execution of opinion polls (Grignon *et al.*, 2001).

Throughout the single-party era, the Kenyan government had little or no interest in the public opinion on any policy issue, bar those considered 'social' rather than political or economic (such as divorce law). Historically Kenya's political culture has been an undemocratic one, characterized by low political awareness and socialization, intense ethnic antagonism, low political morality, routine electoral fraud, lack of accountability, physical insecurity, corruption and apathy (Buglass, 1997; Ajulu, 1992). This left little place for solicitation of public opinion in any form.

Election forecasting is highly data-driven, focused on a very concrete and delimited task, and in most studies the goal is not to explain election outcomes but rather to describe and predict them. Pre- election opinion polls that are conducted scientifically and impartially, are essential in projecting voting intentions of the electorate in a democratic polity (Brouwer, 1955; Webb, 2010). They measure not only support for political parties and candidates, but also public opinion on a wide range of social, economic and political issues. Public opinion is a critical force in shaping and transforming the society.

Much of the extensive literature devoted to forecasting elections outcomes focuses on using pre-election polls or prediction markets to predict election results. Scholars like Rothschild (2009), Holbrook-Provow (1987), Kaplan and Barnett (2003) and DeSart and Holbrook (2003) showed that fairly accurate forecasts can be made using polls taken just before an election. Although reasonably accurate forecasts of election results can be made just before an election using both polls and prediction markets, reliability of these methods drops drasticaaly when used months before an election (Arrow, 1971).

Nevertheless, some scholars have developed unique methods of forecasting election results that may hold more promise than using polls or prediction markets months before the election. Several techniques have been explored for forecasting the results of elections For instance Armstrong and Graefe (2011) approach that used biographical information, Graefe and Armstrong (2012) model that used measures of how well candidates would be expected to handle particular issues while Jones (2002) and Lewis-Beck and Dassonneville (2015) surveyed experts and/or voters for their predictions.

A common and more viable election forecasting approach which is independent of polls or prediction markets involves using econometric models. These methods use a wide range of economic and political indicators such as economic growth rates, results of previous elections, incumbency, and a variety of other possible considerations to predict the likely outcome of an election. However, while there is an extensive literature on forecasting elections using econometric models, so far, the vast majority of this literature has focused on forecasting nationwide results.

2.2 Fundamentals of Bayesian Inference

Bayesian inference is the process of fitting a probability model to a set of data and summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations (Gelman *et al.*, 1992; Geweke *et al.*, 1991; Gelman *et al.*, 2004, 2013). It uses the prior information or expert opinion to augment the existing data of interest. As quantitative political research becomes increasingly sophisticated, the more complex, but more capable, Bayesian approach is likely to grow in popularity (Casella and George, 1992). The Bayesian inferential engine is a coherent set of axioms that converts prior information to posterior evidence by conditioning on observed data Thus, stipulating prior distributions for unknown quantities is a requirement, and this requirement has been a long-standing source of controversy.

Bayesian statistics provides a number of ways to define prior information, and the strength of these prior assertions can vary considerably within the same inferential framework. Recent Bayesian work in fields other than political science has exploited the elicited prior as a means of drawing information from subject-area experts with the goal of constructing a probability structure that reflects their specific qualitative knowledge, and perhaps experiential intuition, about the studied effects (Gelman *et al.*, 2004, 2013). These informed priors derive their name from the way in which the information is elicited from non-statisticians who have a great deal of information about the substantive question but are not involved in the model construction process. Such experts can be physicians, policy-makers, theoretical economists, historians, previous study participants, outside experts, politicians, community leaders, and others.

The promise of this approach that it has the potential to tie together the seemingly antithetical research approaches of qualitative area studies with data-oriented work based on statistical methods, Early critics of the Bayesian paradigm (Pearson, 1920) focused on the almost exclusive use of uniform (flat) priors at the time as a method for expressing prior ignorance or uncertainty. Their concern was the effect that this prior has with small samples (since large enough samples produce standard likelihood analysis results), and the fact that uniformness does not represent a genuine lack of information about a parameter. A great deal of Bayesian work in the middle of the twentieth century dwelt on the quixotic goal of finding an "objective" alternative (Lindley, 1961) to mitigate concern about arbitrarily interjecting subjective information through the prior. This effort proved to be misguided since all statistical models are subjective and a substantial advantage to the Bayesian choice is that previously known information can be directly and transparently included in the model specification. The posterior distribution of voter preference θ given some data y is (Gelman *et al.*, 2013):

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int_\Theta f(y|\theta)\pi(\theta)\,d\theta}, \tag{2.1}$$

$$\propto f(y|\theta)\pi(\theta) \tag{2.2}$$

where $\pi(\theta)$ is the prior distribution and $f(y|\theta)$ is the likelihood function.

Recent efforts, which includes this article, focuses on applying simulation tools from Bayesian statistics (i.e., Markov chain Monte Carlo) to solve previously intractable problems. This computational perspective mostly avoids the specification of deeply informed priors in favour of diffuse (very spread-out) forms (Jac). The useful purposes of such priors, particularly in dealing with so-called nuisance parameters, greatly supersedes their limitation of fully exploit Bayesian capabilities.

Some mathematics arises in the analytical manipulation of the probability distributions, notably in transformation and integration in multi-parameter problems (Gelman *et al.*, 1992). In this research emphasize is put on stochastic simulation (Geweke *et al.*, 1991), and the combination of mathematical analysis and simulation, as general methods for summarizing distributions.

2.2.1 Sequential Bayesian model for Bernoulli opinion polls

The problem of understanding and predicting election outcomes has long been part of political science research. However, lack of pre-election poll data especially in developing countries is one of the main challenges associated with forecasting election outcomes (Cohen, 1998). Nevertheless, in developed countries opinion polls are now easily accessible through Online polling. The online polling is overhauling traditional phone polls, for instance, the analysis of the 2016 US presidential election campaign (Carr *et al.*, 2018).

The seminal work of Goodhart and Bhansali (1970), spurred numerous studies which examined the evolution of voting intentions, as measured by opinion polls, and in particular the relationship between political popularity, ethnicity, youth factor and economic variables such as inflation, gross domestic product, personal producer index and unemployment. See for example (Miller and Mackie, 1973; Powell Jr and Whitten, 1993; Rattinger, 1991). An empirical issue of particular relevance to the present study is the degree of persistence in political popularity. Building on the rational expectations' version of the permanent income hypothesis Sheffrin *et al.* (1996) argued that the effect of news about the economy on voting intentions would be permanent (see also, Byers *et al.* (1997); Sargent (2010)). The practical implication of their model is that the time series of opinion data should behave like a random walk, with the Autoregressive-Moving Average representation of the time series containing an autoregressive root of unity. Such processes are nonstationary, and exhibit no mean-reversion tendencies.

Further analysis of the UK data in Byers *et al.* (1997) rejected the unit root hypothesis in favour of stationary Autoregressive-Moving Average models, although with autoregressive coefficients close to unity. Similar results are reported by Scott *et al.* (1977). Such models would imply that the effect of news on voting intentions, although it could be quite persistent in practice, is in principle transitory. As a consequence of aggregating heterogeneous poll responses under certain assumptions about the evolution of individual opinion, Byers *et al.* (1997) concluded that the time series of poll data should exhibit long memory characteristics. In an analysis of the monthly Gallup data on party support in the UK, Byers *et al.* (1997) confirmed that the series are long memory, and virtually pure 'fractional noise' processes. However, even though this time series approach is appealing it requires data observed over a long period of time which is a limitation to us. The alternative approach is the frequentist regression modelling which does argument/ update opinion polls as series of observations. However, holds our model to be updated once the data set is updated sequentially from time to time. In other words, our expression must include the past information which serves as a prior information. It follows therefore that a Sequential Bayesian Analysis is the best candidate for this type of model. Basically, a simple model of political popularity, as recorded by opinion polls of voting intentions, is proposed; in particular, the Sequential Bayesian Analysis.

If the response y is a binary success (1) vversus failure (0) indicator, the canonical family is the *Bernoulli* distribution with density

$$y \sim \text{Bernoulli}(\theta) = \theta^y (1-\theta)^{1-y},$$

where $\theta \in [0, 1]$ can be interpreted as the success probability. The sequential Bayesian approach combines all previous polls up to the present, where the posterior distribution for sample y_t serves as the prior for the sample y_{t+1}.

Bayesian estimation and inference has a number of advantages in statistical modelling and data analysis. These includes:- (a) Provision of confidence intervals on parameters and probability values on hypotheses that are more in line with common sense interpretations; (b) provision of a way of formalizing the process of learning from data to update beliefs in accordance with recent notions of knowledge synthesis; (c) assessing the probabilities on both nested and non-nested models (unlike classical approaches) and; (d) using modern sampling methods, is readily adapted to complex random effects models that are more difficult to fit using classical methods (Gelman *et al.*, 1995).

Unlike in the past when statistical analysis based on the Bayes theorem was often daunting due to the numerical integrations needed. Recently developed computer-intensive sampling methods of estimation have revolutionised the application of Bayesian methods, and such methods now offer a comprehensive approach to complex model estimation, for instance in hierarchical models with nested random effects (Raftery *et al.*, 1996). They provide a way of improving estimation in sparse datasets by borrowing strength (Stroud, 1994; Richardson and Best, 2003), and allow finite sample inferences without appeal to large sample arguments as in maximum likelihood and other classical methods. Sampling-based methods of Bayesian estimation provide a full density profile of a parameter so that any clear non-normality is apparent, and allow a range of hypotheses about the parameters to be simply assessed using the collection of parameter samples from the posterior.

Bayesian methods may also improve on classical estimators in terms of the precision of estimates. This happens because specifying the prior brings extra information or data based on accumulated knowledge, and the posterior estimate in being based on the combined sources of information (prior and likelihood) (Gelman *et al.*, 2004, 1992, 2013). Indeed, a prior can often be expressed in terms of an equivalent 'sample size'. The relative influence of the prior and data on updated beliefs depends on how much weight is given to the prior (how 'informative' the prior is) and the strength of the data. For example, a large data sample would tend to have a predominant influence on updated beliefs unless the prior was informative. If the sample was small and combined with a prior that was informative, then the prior distribution would have a relatively greater influence on the updated belief:

How to choose the prior density or information is an important issue in Bayesian inference, together with the sensitivity or robustness of the inferences to the choice of prior, and the possibility of conflict between prior and data (Andrade and O'Hagan, 2006; Berger *et al.*, 1994). In some situations, it may be possible to base the prior density for θ on cumulative evidence using a formal or informal meta-analysis of existing studies. A range of other methods exist to determine or elicit subjective priors (Garthwaite *et al.*, 2005; Moala and O'Hagan, 2010; Elfadaly and Garthwaite, 2013, 2017). A simple technique known as the histogram method divides the range of θ into a set of intervals (or 'bins') and elicits prior probabilities that θ is located in each interval; from this set of probabilities, may be represented as a discrete prior or converted to a smooth density. Another technique uses prior estimates of moments along with symmetry assumptions to derive a normal prior density including estimates and of the mean and variance.

Other forms of prior can be re-parameterised in the form of a mean and variance (or precision); for example beta priors for probabilities can be expressed as $B(m\tau, (1 - m)\tau)$ where m is an estimate of the mean probability and τ is the estimated precision (degree of confidence in) that prior mean.

2.2.2 Bayesian multinomial polls

There has been a growing interest by political pundits and scholars alike to predict the winner of the presidential elections. Although forecasting has now quite a history, we argue that the closeness of recent Kenyan presidential opinion polls and the wide accessibility of data should change how presidential election forecasting is conducted.

We present a Bayesian forecasting model that concentrates on the national wide pre-election polls prior to 2013 general elections and considers finer details such as third-party candidates and self-proclaimed undecided voters. We incorporate our estimators into WinBUGS to determine the probability that a candidate will win an election. The model predicted the outright winner for the 2013 Kenyan election.

2.3 Forecasting from Opinion Polls

In most democratic countries, forecasting the outcome of voting events has long been of great interest. Quite a good number of statistical models have been built, based on pre-election opinion polls, to describe methods for predicting political parties and presidential candidates share of votes in national elections (Campbell *et al.*, 1992; Jackman, 2005; Linzer, 2013). Some scholars have used exit polls in their models for forecasting the outcome of an election (Brown and Payne, 1975; Curtice and Firth, 2008). Most of these models have been successful in predicting the outcome before the final result is declared. Polls fluctuates with time thus they lose precision if they are taken too early before the elections (Campbell, 2008, 2012). Nevertheless, they can be useful in identifying trends in preferences of the electorate.

Several models used to predict the united states presidential election outcomes, have employed fundamentals (or the covariate information such as economic growth, unemployment rate and geographical variation) to create initial forecasts which are in turn used as prior input into a binomial model (Campbell *et al.*, 1992; Lock and Gelman, 2010; Linzer, 2013). According to theories of retrospective voting, which are the basis for fundamentals, voters tend to punish the incumbent or the incumbent party for social or economic crises (Duch and Stevenson, 2008; Linzer, 2013; Lewis-Beck and Dassonneville, 2015). In a Bayesian sense the data, here the opinion poll, can be combined with prior distributions, which can be represented as follows (Lock and Gelman, 2010)

Opinion Poll :

$$\hat{d}_{s,t}|d_{s,0}\ N\left(d_{s,0}, \frac{p_{s,0}(1-p_{s,0})}{n_{s,t}} + var(d_{s,t}|d_{s,0})\right)$$

Prior:

$$\hat{d}_{s,0}|d_{s,2004}\ N\left(d_{s,2004}, var(d_{s,0}|d_{s,2004})\right)$$

Here, $d_{s,0}$ is equivalent to the notation $d_{s,2008}$ referring to the relative position of state s at the time of the 2008 election; $d_{s,t}$ is the polls data at time t and $p_{s,0}$ is the probability of winning the polls. The posterior will be the product of the two equations above.

Using a linear regression with 16 covariates measured at national, regional and state levels, Campbell *et al.* (1992) predicted the outcome of a presidential election for each state in the US. His covariates included, amongst others, macroeconomic variables, opinion polls from early campaign, state's voting record in the previous two elections, incumbency of a candidate and measures of partisan shifts over time.

To analyse the deviations of opinion on the state level from the nationwide average (Lock and Gelman, 2010), used a Bayesian forecasting model which was based on a normal approximation to a binomial outcome. They integrated past election data to form prior distributions for each state outcome, and combines them with information from the state-level opinion polls to create posterior distributions of the shares of vote before the election. This is to account for overdispersion due to survey issues, such as weighting or clustering as well as uncertainty about opinion.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This study seeks to obtain a good prediction procedure, which should identify the key factors influencing the voting process in Kenya and control for the possible account for the cost of misclassification. In this chapter, we propose Bayesian approach to analysing opinion polls data. We formulated models for both univariate, bivariate as well as multivariate analysis where more than two candidate are compared.

3.2 Data Selection

Each province is divided into a large number of wards. The interest lies in identifying which wards will return more representative results, in the sense of yielding estimates of the percentage of votes for each candidate that would be similar to those that will be obtained for the whole area. This would eliminate possibly atypical areas and would reinforce the plausibility of the assumptions of a simple model. Thus, a need for an appropriate distance among the corresponding probability distributions. A Bayesian estimate of the probability $\theta_{i,j}$ that an elector with characteristic similar to those living in the area covered by ward I, will vote for candidate j is

$$\theta_{i,j} = \frac{(n_{i,j} + \frac{1}{2})}{n_i + \frac{m}{2}} \tag{3.1}$$

$$\text{where} \quad n_i = \sum_{i=1}^{m} n_{i,j}$$

For each of the wards, the expected loss of will be computed using an approximate distribution rather than the true distribution to predict the proportion θ_i of the vote that the j^{th} candidate. $l_i = \sum_{i=1}^{m} \log \frac{\theta_j}{\theta_{i,j}}$ (which is non negative). Where θ_j is the proportion of votes obtained in the province by candidate j in the previous elections (Bernardo, 1984). According to this criterion, the smaller l_i is the more representative ward i is. It will be assumed that this representability remains essentially unchanged, and accordingly, limited sampling to the area covered by the more representative ward. Empirical evidence suggests that representative area remains representative thus allowing accurate prediction based solely on them.

3.3 Overview of Bayes' Theorem

3.3.1 Introduction to Bayes' theorem

Thomas Bayes $(1702 - 1761)$ studied conditional probability of the sort we are confronted with daily, for example, when we read that youth drivers are more likely to violate traffic rules than older people or religions tend to be politically conservative or that better educated people live longer (we ignore whether the claim are true) The original Bayes' theorem applies to point probabilities. The basic theorem states:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)} = \frac{p(B|A)p(A)}{\sum_A p(B|A)p(A)} \tag{3.2}$$

The basic process of concern in Bayesian statistics is re-computation of the probability of interest using the updated prior. In other words, a Bayesian perspective, begins with some prior probability for some event, and this prior is updated probability with new information to obtain a posterior probability. The posterior probability can then be used as a prior in subsequent analysis. This is an appropriate strategy for conducting scientific research. In a layman language, Bayesian analysis is applicable in situations where data d and hypothesis h are combined implying

$$p(h|d) = \frac{p(d|h)p(h)}{p(d)} \tag{3.3}$$

Reasoning from data to hypothesis is the heart of scientific research and bayes insight helps us operationalize some aspects of this in a useful fashion.

3.3.2 Bayes' theorem and probability distribution

Bayesian statistics typically involves using probability distributions rather than point probabilities for the quantities in the theorem. The inclusion of a prior probability distribution ultimately produces a posterior probability that is also a probability distribution as well. Generally, the goal of Bayesian statistics is to represent prior uncertainty about model parameters with a probability distribution and to update this prior uncertainty with current data to produce a posterior probability distribution for the parameters that contains less uncertainty. This perspective implies a subjective view of probability - probability represents uncertainty- and it contrasts with classical perspective.

From a Bayesian approach any quantity for which the true value is uncertain, including model parameters, can be represented by a probability distribution. However, from a classical point of view, it is unacceptable to place a probability distribution on parameters because parameters are assumed to be fixed quantities; only the data are random and thus probability distribution can only be used to represent data. In terms of probability distribution, Bayes' theorem can be written as

$$p(\theta|data) = \frac{p(data|\theta)p(\theta)}{p(data)} \tag{3.4}$$

Where $p(\theta|data)$ is the posterior distribution for the parameter θ and $p(data|\theta)$ is the sampling density for the data. $p(\theta)$ is the prior distribution for the parameter and $p(data)$ is the marginal probability of the data. If the sample space is continuous this marginal probability can be computed as:

$$p(data) = \int p(data|\theta)p(\theta)d\theta \tag{3.5}$$

This quantity is sometimes called the marginal likelihood for the data and serves as a normalizing constant to make the posterior density proper. Since the sampling density is proportional to the likelihood function then Bayes theorem for probability

$$\text{posterior} \propto \text{Likelihood} \times \text{prior} \tag{3.6}$$

Specification of an appropriate prior distribution for the parameters is the most substantial aspect of a Bayesian analysis that differentiates it from a classical analysis. The prior distribution should represent the plausible values of the prior probabilities and their relative merit (eg giving prior weight). An appropriate prior distribution for unknown proportion is a beta distribution. The posterior distribution (3.4) is obtained via Markov chain Monte Carlo (MCMC) sampling. This is explained at the end of this Chapter.

3.4 Sequential Bayesian Analysis of Bernoulli Opinion Polls

3.4.1 Binomial Data

Consider a binary outcome variable T defined as;

$$Z_i = \begin{cases} 1 \quad \text{if the} \quad i^{th} \quad \text{respondent voted for the incumbent} \\ 0 \quad \text{if the} \quad i^{th} \quad \text{respondent voted for the Challenger} \end{cases} \tag{3.7}$$

Therefore in an opinion poll of size n where x respondents voted for the incumbent and $n - x$ for the challenge, the random variable $X = \sum_{i=1}^{n} Z_i$ has a binomial distribution with parameter (i.e. the probability that respondent i will vote for the incumbent) θ . The probability density function of X given θ is as below

$$p(X|\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0, 1, 2 \ldots, n \tag{3.8}$$

3.4.2 The sequential binomial model

We can use a distribution to represent our prior knowledge and uncertainty regarding unknown parameter θ. An appropriate and a conjugate prior distribution for our unknown parameter θ is a beta distribution denoted by $Be(\alpha, \beta)$. The probability density function of a beta distribution is:

$$\pi(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \tag{3.9}$$

where $\Gamma(\alpha)$ is the gamma function applied to α and $0 < \theta < 1$. The parameters α and β can be thought of as prior "successes" and "failures," respectively. This prior density can also be expressed using the proportionality sign as;

$$\pi(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

The prior mean $\mathbb{E}(\theta) = \mu = \frac{\alpha}{\alpha+\beta}$ and variance $var(\theta) = \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ The posterior density of θ is a product of a prior and the likelihood. Implying a beta $\alpha - x$ and $\beta + n - x$ distribution with parameters , specifically:

$$\pi(\theta|X) = p(X|\theta)\pi(\theta) = Be(\alpha + x, \beta + n - x)$$
$$\pi(\theta|X) \propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1} \tag{3.10}$$

Where $Be(\cdot,\cdot)$ denotes $Beta(\cdot,\cdot)$ density function. We shall denote this posterior by $p(\theta|X_{(1)})$, Now, if we observe another sample $X_{(2)}$ then the posterior becomes

$$\pi(\theta|X_{(2)}) = p(X = x_{(2)|\theta})p(\theta|X_{(1)}) = Be(\alpha + x_1 + x_2, \beta + n_1 + n_2 - x_1 - x_2) \tag{3.11}$$

Recursively, for the k^{th}sample we have the posterior for $X_{(k)}$ as

$$\pi(\theta|X_{(k)}) = p(X = x_{(k)}|\theta)p(\theta|X_{(k-1)}) = Be(\alpha + \sum_{i=1}^{k} x_i, \beta + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} x_i) \tag{3.12}$$

This gives a better estimate than the one obtained by just aggregating all the previous pre-election polls in a single prior. The choice of α and β for our prior distribution depends on at least two factors: (1) the amount of information about the parameter available prior to this poll; (2) the amount of stock we want to put into this prior information. Contrary to the view that this is a limitation of Bayesian statistics, the incorporation of prior information can actually be an advantage and provides us considerable flexibility. If we have little or no prior information, or we want to put very little stock in the information we have, we can choose values for α and β that reduce the distribution to a uniform distribution. For instance, choosing $\alpha = 1$ and $\beta = 1$, we get

$$\pi(\theta|\alpha,\beta) \propto \theta^0(1-\theta)^0 = 1$$

which is proportional to a uniform distribution on the allowable interval for θ . That is, the prior distribution is flat, not producing greater a priori weight for any value of θ over another. Thus, the prior distribution will have little effect on the posterior distribution. For this reason, this type of prior is called "noninformative." On the other hand, if we have considerable prior information that we wish to weigh heavily relative to the current data, large values of α and β are used. A little massage of the formula for the variance reveals that, as α and β increase, the variance decreases, which makes sense, because adding additional prior information ought to reduce our uncertainty about the parameter. Thus, adding more prior successes and failures (increasing both parameters) reduces prior uncertainty about the parameter of interest θ. Lastly, if we have considerable prior information that we do not wish to weigh heavily in the posterior distribution, moderate values of α and β are chosen that yield a mean that is consistent with the previous research but that also produce a variance around that mean that is broad.

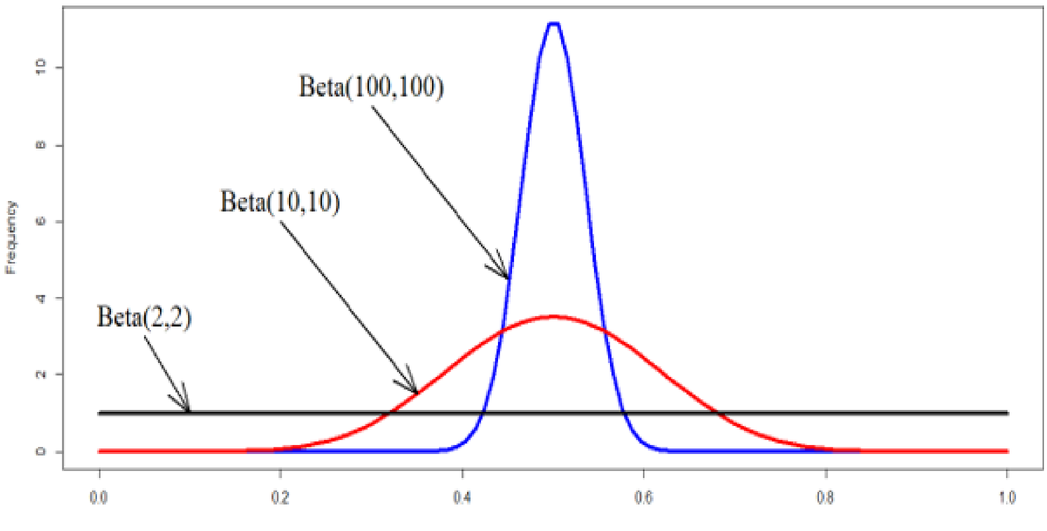

Figure 3.1: *Various beta distributions with mean =0.5 for various choice of parameters*

In order to clarify these ideas, we illustrate using beta distributions plots with different values of α and β . All the three beta distributions, displayed in Figure 1, have a mean of 0.5; but different variances as a result of having α and β parameters of different magnitude.

The most-peaked beta distribution has parameters $\alpha = \beta = 100$. The leastpeaked distribution is almost flat—uniform—with parameters $\alpha = \beta = 2$. As with the binomial distribution, the beta distribution becomes skewed if α and β are unequal, but the basic idea is the same: the larger the parameters, the more prior information and the narrower the density

Throughout the fall of every general election year in Kenya, many pollsters conduct a number of polls attempting to predict whether candidate A or candidate B would win the presidential election. One of the hotly contested general election was the 2007 elections the battleground predominantly between the incumbent (here demoted as K) and the challenger (here denoted as R). The polls leading up to the election showed the two candidates claiming proportions of the votes that were statistically indistinguishable in the nation.

Figure 2 shows the prior, likelihood, and posterior densities. The likelihood function has been normalized as a proper density for θ , rather than X. Clearly the posterior density is a compromise between the prior distribution and the likelihood (current data). The posterior is between the prior distribution and the likelihood, but closer to the prior. The reason the posterior is closer to the prior is that the prior contained more information than the likelihood: There were 1,950 previously sampled persons and only 1,067 in the current sample.

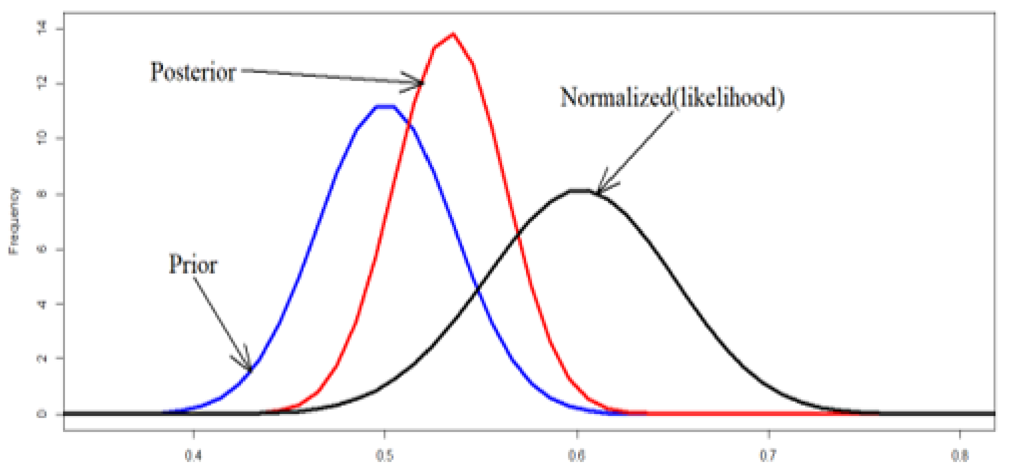

Figure 3.2: *Prior, likelihood, and posterior for 2007 polling data for Kenya*

With the posterior density determined, we now can summarize our updated knowledge about the proportion of voters who will vote for incumbent, and answer our question of interest: What is the probability that the incumbent would win? A number of summaries are possible, given that we have a posterior distribution with a known form (a beta density). First, the mean of incumbent K is 1498/(1498 + 1519) = 0.497, and the median is also 0.497. The variance of this beta distribution is .00008283 (standard deviation=.0091). If we assume that this beta distribution is approximately normal, then the approximate a 95%confidence interval of K is [0.479 – 0.515].

3.4.3 A Bayesian multinomial model

Consider a particular province and let i, j denote the unknown probability that an elector from ward i will vote for candidate j and assume that a random sample of size n_i is taken from electors of ward i. Let n_{ij} denote those who will vote for candidate j. So $n_i = \sum_{i=1}^{m} n_{ij}$ It will be assumed that n_{ij} is a random sample from a multinomial distribution with parameters θ_{ij} and $\sum_{i=1}^{m} \theta_{ij}$. From a Bayesian point of view, the information provided by n_{ij} about $\theta_{ij} = 1, 2, ...m$ is encapsulated in the corresponding posterior distribution of the θ_{ij}. A Dirichlet prior is mathematically convenient and not difficult to assess. The probability density function of a Dirichlet random vector is given by

$$p(\theta_1, \theta_2, \ldots, \theta_m) = \Gamma\left(\sum_{i=1}^{m} \alpha_j\right) \prod_{i=1}^{m} \left[\frac{\theta_j^{\alpha_j^{-1}}}{\alpha_j}\right] \tag{3.13}$$

and it's a conjugate prior for the multinomial distribution. Historically conjugacy has been very important to Bayesian statistician in that using conjugate prior/likelihood with known form ensured that the posterior would be a known distribution that could be easily evaluated to answer the scientific question of interest. Combining a Dirichlet distribution as a prior with multinomial distribution likelihood, the resulting posterior distribution is :

$$p(\theta_1, \theta_2, \ldots, \theta_m | x_1, x_2, \ldots, x_m) \propto f(x_1, x_2, \ldots, x_m | \theta_1, \theta_2, \ldots, \theta_m) \pi(\theta_1, \theta_2, \ldots, \theta_m)$$

$$\propto \text{multinomial}(\boldsymbol{x}|\boldsymbol{\theta})\text{dirichlet}(\boldsymbol{\theta}, \boldsymbol{\alpha})$$

$$\propto \text{dirichlet}(\boldsymbol{\theta}, \boldsymbol{\alpha} + \boldsymbol{x})$$

$$\propto \prod_{i=1}^{m} \theta_{ij}^{x_i + \alpha_j} \tag{3.14}$$

In this research it is of interest to consider:

1. The n opinion poll data sets as separate samples from the same population each one providing conditionally independent information regarding the prior parameters
2. Each poll results as the results of a poll specific parameter θ_{ij} with θ_{ij} being a random realization from a Dirichlet distribution with hyper parameters $\alpha_j'^S$

His approach yields a hierarchical model with the following structure

$$pr(\boldsymbol{\alpha}, \boldsymbol{\theta}|\boldsymbol{x}) \propto pr(\boldsymbol{x}|\boldsymbol{\theta}) pr(\boldsymbol{\theta}|\boldsymbol{\alpha}) pr(\boldsymbol{\alpha}) \tag{3.15}$$

The likelihood part of the model is the product of the sampling density of the n Polls

$$pr(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{i=1}^{n} \prod_{j=1}^{m} \theta_{ij}^{x_i + \alpha_j} \tag{3.16}$$

The prior density for θ_{ij} is dirichlet, their product is the full prior density

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) \propto \prod_{i=1}^{n} \Gamma\left(\sum_{i=1}^{m} \alpha_j\right) \prod_{i=1}^{m} \left[\frac{\theta_j^{\alpha_j^{-1}}}{\alpha_j}\right] \tag{3.17}$$

Effectively, this article considers previous polling data as a random process generated by the hyper parameters α_j^{-1S}.

Our key interest centers particularly on α_j^{-1S} which are thought to be the population parameters governing the proportion of voters who would vote for candidate j and which drives each individual poll results. Lastly the hyper prior for the hyper parameters j need to be determined. Ideally we will prefer a t hyper prior that is relatively noninformative to reduce the degree of subjectivity on the hyper prior. As such standard reference uninformative hyper prior

$$pr(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{i=1}^{n}\prod_{j=1}^{m}\theta_{ij}^{\frac{i}{2}} \tag{3.18}$$

which ensures nonnegativity of the prior parameters α_j (Tierney, 1994; Gelman et al., 1995) are used. The corresponding posterior distribution

$$p(\boldsymbol{\alpha},\boldsymbol{\theta}|\boldsymbol{x}) \propto \prod_{i=1}^{n}\prod_{j=1}^{m}\theta_{ij}^{x_i+\alpha_j}\prod_{i=1}^{n}\Gamma\left(\sum_{i=1}^{m}\alpha_j\right)\prod_{i=1}^{m}\left[\frac{\theta_j^{\alpha_j^{-1}}}{\alpha_j}\right]\prod_{i=1}^{n}\prod_{j=1}^{m}\theta_{ij}^{\frac{i}{2}} \tag{3.19}$$

Using Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; Gilks,1996), the conditional posterior distribution for after eliminating terms that do not involve them are easily seen to be Dirichlet with parameters $x_i + \alpha_j - \frac{1}{2}$, i.e.:

$$p(\boldsymbol{\alpha},\boldsymbol{\theta}|\boldsymbol{x}) \propto \prod_{j=1}^{m}\theta_{ij}^{x_i-\frac{3}{2}} \tag{3.20}$$

The conditional posterior distribution for $\alpha_j b$ is not simple. Consider a posterior for a general α_j by eliminating terms not involving α_j the posterior for α_j is

$$\left[\frac{\Gamma(\sum_{i=1}^{m}\alpha_j)}{\prod_{i=1}^{m}\Gamma(\alpha_j)}\right]^n\prod_{i=1}^{n}\theta_{ij}^{x_i-\frac{3}{2}} \tag{3.21}$$

Equation (3.21) simplify to

$$\left[\frac{\Gamma(\sum_{i=1}^{m}\alpha_j)}{\prod_{i=1}^{m}\Gamma(\alpha_j)}\right]^n \exp\left[\ln\left(\prod_{i=1}^{n}\theta_{ij}^{x_i-\frac{3}{2}}\right)\right] \tag{3.22}$$

$$= \left[\frac{\Gamma(\sum_{i=1}^{m}\alpha_j)}{\prod_{i=1}^{m}\Gamma(\alpha_j)}\right]^n \exp\left[\sum_{i=1} n(x_i-\frac{3}{2})\ln\theta_{ij}\right] \tag{3.23}$$

$$\propto \left[\frac{\Gamma(\sum_{i=1}^{m}\alpha_j)}{\prod_{i=1}^{m}\Gamma(\alpha_j)}\right]^n \exp\left(\sum_{i=1} n\alpha_j\ln\theta_{ij}\right) \tag{3.24}$$

Although the θ_j parameters conditioned on the data and values for $\alpha_j'^S$ can be drawn directly from a Dirichlet distribution, the distribution of $\alpha_j'^S$ are unknown forms and must therefore be simulated using MH step.

3.4.4 Bayesian trivariate model description

More formally, define θ_i to be the true proportion of voters in a county who intend to vote for candidate i in the election (for simplicity, let $i = 1$ correspond to the incumbent candidate, $i = 2$ correspond to the main opposition candidate, $i = 3$ collectively correspond to all the other minor candidates (also called third force), and $i = 4$ correspond to no candidate or voters who have declared that they are still undecided). These proportions are assumed to be continuous (between 0 and 1) and sum to 1.

The joint prior distribution for $\boldsymbol{\theta} = \theta_1, \theta_2, \theta_3$ is assumed to be a conjugate prior distribution (i.e., the resulting posterior belongs to the same distributional family as the prior distribution). To satisfy this requirement, assume that θ follows a Dirichlet distribution, $\boldsymbol{\theta} \sim$ Dirichlet($\alpha_1, \alpha_2, \alpha_3$), which is a multivariate generalization of the beta distribution and is often used as a prior for the probability of a success in Bernoulli trials. Therefore, the joint probability density function of θ can be written as

$$\pi(\theta_1,\theta_2,\theta_3) \propto \theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\theta_3^{\alpha_3-1}, \quad \theta_i > 0, i = 1,2,3 \quad \text{and} \quad \sum\theta_i \leq 1 \tag{3.25}$$

The probability that a candidate wins a given county can be computed using the marginal probability densities. To obtain these marginals, we sequentially integrated the remaining variables out of the joint Dirichlet probability density function. We now illustrate this process by first rewriting the joint Dirichlet probability density function as

$$\pi(\theta_1, \theta_2) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} (1 - \theta_1 - \theta_2) \tag{3.26}$$

$$\theta_1, \theta_2 \geq 0, i = 1, 2 \quad \text{and} \quad \sum \theta_i \leq 1$$

Integrating overθ_3 leads to an expression for the joint probability density of θ_1 and θ_2

$$\pi(\theta_1, \theta_2) = \int_0^{1-\theta_1-\theta_2} c\theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} (1 - \theta_1 - \theta_2)^{\theta_3 - 1} d\theta_3 \tag{3.27}$$

$$\pi(\theta_1, \theta_2) = c\theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \int_0^{1-\theta_1-\theta_2} (1 - \theta_1 - \theta_2)^{\theta_3 - 1} d\theta_3 \tag{3.28}$$

These results leads to the expression

$$\pi(\theta_1, \theta_2) \propto \theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} (1 - \theta_1 - \theta_2)^{\theta_3 - 1}, \theta_1, \theta_2 \geq 0, \quad \text{and} \quad \theta_1 + \theta_2 \leq 1 \tag{3.29}$$

Integrating over all possible values of θ_2 gives the marginal density of θ_1,

$$\pi(\theta_1) = \int_0^{1-\theta_1-\theta_2} \mathbb{K}\theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} (1 - \theta_1 - \theta_2)^{\theta_3 - 1} d\theta_2 \tag{3.30}$$

$$\pi(\theta_1) = \mathbb{K}\theta_1^{\alpha_1 - 1} \theta_2^{\alpha_2 - 1} \int_0^{1-\theta_1-\theta_2} (1 - \theta_1 - \theta_2)^{\theta_3 - 1} d\theta_2 \tag{3.31}$$

The results leads to the expression

$$\pi(\theta_1) \propto \theta_1^{\alpha_1 - 1} (1 - \theta_1)^{\theta_2 + \theta_3 - 1}, \quad 0 \leq \theta_1 \leq \tag{3.32}$$

Therefore, by the form of $f(\theta_1), \theta_1$ is distributed as a beta random variable with parameters $\alpha_2 + \alpha_3$. Using the identical argument, θ_2 and θ_3 are also distributed as beta random variables; hence, $\theta_i \sim Beta(\alpha_i, \sum \alpha_i - \alpha_i), i = 1, 2, 3$

3.4.5 Trivariate choice of prior parameters

The beta prior distribution is characterized by 2 shape parameters, here, α_1 and $\alpha_2 + \alpha_3$ which must be chosen. Different choices can be incorporated, which would result in quite varied substantive implications. For example, one possibility is to set these values so that the expected value and the variance, or the first and second moments, for $\theta_i, i = 1, 2, 3$, closely match observed elections. However, there are an infinite number of combinations of α_1, α_2 and α_3 that result in the same values for the expectations and the variances for θ_1, θ_2 and θ_3. That is, if the expectation and variance of a beta random variable are given by

$$\mathbb{E}(\theta_i) = \frac{\alpha_i}{\sum_{i=1}^{3}} \quad \text{and} \tag{3.33}$$

$$var(\theta_i) = \frac{\alpha_i(\sum_{i=1}^{3} \alpha_i - \alpha_i)}{(\sum_{i=1}^{3})^2(\sum_{i=1}^{3} \alpha_i + 1)} \tag{3.34}$$

The choice of the shape parameters is essentially arbitrary if we do not issue any constraints or do not use any substantive guidance. Fortunately, in presidential forecasting, we have a great deal of substantive knowledge that can be integrated. One way to constrain our choices is to choose the $\alpha_i^{\prime s}$ so that

1. $\dfrac{\alpha_i}{\sum \alpha_i}$ equals what θ_i is expected to be, prior to observing the polling data, and
2. The spread of the prior distribution for θ_i (determined by $\sum \alpha_i$ with larger values indicating less uncertainty) reflects the perceived uncertainty in θ_i.

3.5 Markov Chain Monte Carlo Sampling

Markov chain Monte Carlo (MCMC) methods are a class of algorithms for sampling from a probability distribution. By constructing a Markov chain one can obtain a sample of the desired distribution as its limiting (stationary) distribution. The idea of MCMC sampling is to simulate a random walk in the space of parameters of interest,

$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_d)'$, which converges to the joint posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$. The samples are drawn sequentially, with the distribution of the sampled draws depending on the last value drawn; hence, the draws form a Markov chain. The states of the chain after a large number of iterations is then used as a sample from the desired posterior distribution.

3.5.1 The Metropolis algorithm

The Metropolis algorithm is a widely used procedure for sampling from a specified distribution on a large finite set. Given a target posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y})$, known up to a normalizing constant, the Metropolis algorithm creates a sequence of random vectors $(\theta^{(1)}, \theta^{(2)}, \cdots)$ whose distribution converges to the target distribution. Each sequence can be considered a random walk whose stationary distribution is $p(\boldsymbol{\theta}|\boldsymbol{y})$. The algorithm proceeds as follows (see, e.g.,Tierney, 1994; Gelman *et al.*, 2004). Start with some initial value θ^0. For $t = 1, 2, \cdots$, obtain $\theta^{(t)}$ from $\theta^{(t-1)}$ using the following steps:

1. Sample a *candidate point* θ^* from a *proposal distribution* at time t, $q(\theta^*|\theta^{(t-1)})$. The proposal distribution distribution must be *symmetric*; that is, $q(\theta_a|\theta_b) = q(\theta_b|\theta_a)$ for all θ_a and θ_b.

2. Calculate the ratio of the densities,
$$r = \frac{p(\theta^*|y)}{p(\theta^{(t-1)}|y)}.$$

3. Set
$$\theta^{(t)} = \begin{cases} \theta^* & \text{with probablity } \min(r, 1), \\ \theta^{(t-1)} & \text{otherwise.} \end{cases}$$

The algorithm requires the ability to draw θ^* from the proposal (jumping) distribution $q(\theta^*|\theta)$ for all θ.

3.5.2 The Metropolis-Hastings algorithm

Metropolis–Hastings (M-H) algorithm is a Markov chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution from which direct sampling is difficult. M-H algorithm generalizes the basic Metropolis algorithm, described above, in two ways. First, the proposal distribution q needs no longer to be symmetric. That is, there is no requirement that $q(\theta_a|\theta_b) = q(\theta_b|\theta_a)$. Secondly, to correct for the asymmetry in the proposal density the acceptance ratio is now (see, e.g., Tierney, 1994; Gelfand and Smith, 1990; Gelman *et al.*, 2004)

$$r = \frac{p(\theta^*|y)q(\theta^{(t-1)}|\theta^*)}{p(\theta^{(t-1)}|y)q(\theta^*|\theta^{(t-1)})}.$$

Allowing an asymmetric proposal distribution can be useful in increasing the speed of the random walk.

3.5.3 Gibbs sampler

Gibbs sampling or a Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate probability distribution, when direct sampling is difficult. The Gibbs sampler (Gelfand and Smith, 1990; Gelman *et al.*, 2004) is a MCMC algorithm that has been found very useful in multidimensional problems. Let $\mathbf{Z_i} = (X_i, Y_i)'$ be a Markov chain. The Gibbs sampler can be used to generate specific multivariate distributions. Let $f(x, y)$ be a given joint density and $f(x|y)$ and $f(y|x)$ to be conditional densities. The Gibbs sampling algorithm is given by

1. Generate $\mathbf{Z}_0 = (X_0, Y_0)'$. Set $i = 1$.
2. Generate $X_i \sim f(x_i|Y_{i-1} = y_{i-1})$
3. Generate $Y_i \sim f(y_i|X_i = x_i)$
4. Set $i = i + 1$ and goto step 2.

In general, Gibbs sampling algorithm defined in terms of subvectors of θ. At each iteration t, the Gibbs sampler cycles through the subvectors of θ, drawing θ_j from the conditional distribution given all the remaining components of θ:

$$p_j(\theta_j|\theta_{(-j)}^{(t-1)}, y),$$

where $\theta_{(-j)}$ represents all the components of θ, except for θ_j, i.e. $\theta_{(-j)} = (\theta_1, \cdots, \theta_{j-1}, \theta_{j+1}, \cdots, \theta_d)'$. This suggests the following MCMC scheme.

1. Generate $\theta_1^{(t)}$ from $p_1(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \cdots, \theta_d^{(t-1)}, y)$

2. Generate $\theta_2^{(t)}$ from $p_2(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \cdots, \theta_d^{(t-1)}, y)$

   $\vdots$

d. Generate $\theta_d^{(t)}$ from $p_d(\theta_d|\theta_1^{(t)}, \theta_2^{(t)}, \cdots, \theta_{d-1}^{(t)}, y)$

At the completion of these steps, the vector $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \cdots, \theta_d^{(t)})'$ provides the simulated value of θ at the tth iteration of sampling. The d steps of this Gibbs sampling scheme completes one iteration of the simulation method.

After a large number, T, of iterations, we obtain $\boldsymbol{\theta}^{(T)}$. Gelman *et al.* (1995) have shown that under mild conditions, the joint distribution $\boldsymbol{\theta}^{(T)}$ converges at an exponential rate to $p(\boldsymbol{\theta}|\boldsymbol{y})$ as $T \to \infty$. The desired joint posterior distribution, $p(\boldsymbol{\theta}|\boldsymbol{y})$, can be approximated by the empirical distribution of M values $\boldsymbol{\theta}^{(t)}$ for $t = T+1, T+2, \cdots, T+M$, where T is large enough so that the Gibbs sampler has converged and M is chosen to give sufficient precision to the empirical distribution of interest.

The WinBUGS software, described in the next section, uses Gibbs sampling and a Metropolis-within-Gibbs routine to draw MCMC samples from complex statistical models.

3.5.4 WinBUGS for Bayesian inference

`WinBUGS` (MS Windows operating system version of the `BUGS`: **B**ayes-ian analysis **U**sing **G**ibbs **S**ampling) is flexible software for the Bayes-ian analysis of complex statistical models using MCMC methods. The software is currently distributed electronically from the BUGS project web site.

`http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml`

More details can be obtained from `WinBUGS`' extensive user manual (Spiegelhalter *et al.*, 2003).

The versatility of the `WinBUGS` package allows for a wide variety of posterior models. Firstly, it is possible to sample from a large number of statistical models including, for example, the Bernoulli, Poisson, normal, multinomial and gamma distributions. Secondly, with recent developments in the software, it is also possible to draw samples from non-standard distributions using the "ones" trick (see `WinBUGS`'s manual). Thirdly, various sampling algorithms are implemented including Gibbs sampler, Metropolis, slice sampler etc.

A Bayesian p-value can be obtained from the WinBUGS program by using the `step( )` function. The function creates a Boolean variable that counts the number of simulations in which, for example, $\beta_x > 0$ is true. As a result, the $0/1$ values from the `step( )` function can be used to compute left- or right-tail areas. For example, if we have M samples of $p_{\beta_x} = \texttt{step}(\beta_x)$ after T burn-in samples, an equal-tail two-sided Bayesian p-value of β_x is given by

$$2\min\left(1 - \frac{1}{M}\sum_{t=T+1}^{T+M} p_{\beta_x}^{(t)}, \frac{1}{M}\sum_{t=T+1}^{T+M} p_{\beta_x}^{(t)}\right).$$

While the MCMC algorithms, e.g. using `WinBUGS`, have the potential to be quicker than the numerical approximations, the convergence rate of any algorithm cannot be guaranteed. In the next section we describe methods for assessing the rate of convergence of MCMC samples.

3.5.5 MCMC convergence

Convergence is diagnosed when the chains have 'forgotten' their initial values, and the output from all chains is indistinguishable. Geweke *et al.* (1991) proposed a convergence diagnostic for MCMC samples based on a test for equality of the means of the first and last part of a *single* chain (by default the first 10% and the last 50%). Geweke's approach involves calculation of the sample mean and asymptotic variance in each window, the latter being determined by spectral density estimation. His convergence diagnostic Z is the difference between these two means divided by the asymptotic standard error of their difference. As the chain length $\rightarrow \infty$, the sampling distribution of the chain has converged. Hence values of $Z \rightarrow N(0, 1)$ which fall in the extreme tails of a standard normal distribution, ± 2, suggest that the chain has not fully converged.

Gelman *et al.* (1992) and Gelman *et al.* (2004) proposed a general approach to monitoring convergence of MCMC output in which two or more parallel chains are run with starting values that are overdispersed relative to the posterior distribution. Convergence for multiple chains may be assessed using Gelman-Rubin scale factor reduction factors that compare variation of the samples parameter values within and between chains. It is based on a comparison of within-chain and between-chain variances, and is similar to a classical analysis of variance. To measure the variability of sample $\theta_j^{(t)}$ within the chain $(j = 1, \cdots, J)$ define

$$V_j = \sum_{t=T+1}^{T+M} \left(\theta_j^{(t)} - \bar{\theta}_j\right)^2 /(M-1)$$

over M iterations after an initial burn-in of T iterations, where $\bar{\theta}_j$ is the average of $\theta_j^{(t)}$ $(t = T+1, \cdots, T+M)$. Ideally, the burn-in period is the initial set of samples where the effect of initial parameter values tails off. Convergence is therefore assessed from $T+1$ to $T+M$. Variability within chains V_W is the average of V_js. Between chain variance is measured by

$$V_B = \frac{M}{J-1} \sum_{j=1}^{J} \left(\bar{\theta}_j - \bar{\theta}\right)^2$$

where $\bar{\theta}$ is the average of $\bar{\theta}_j$s. The scale factor reduction (SRF) compares a pooled estimator of $var(\theta)$, given by $V_P = V_B/M + MV_W/(M-1)$, to V_W. More specifically, SRF $= \sqrt{V_P/V_W}$ with values under 1.2 (Congdon, 2010, p. 19) indicating convergence.

More recently Brooks and Roberts (1998) proposed a convergence statistic known as Brooks-Gelman-Rubin (BGR). This is a ratio of parameter interval lengths, where for chain j the length of the $100(1-\alpha)\%$ interval for parameter θ is obtained, i.e. the gap between 0.5α and $(1-0.5\alpha)$ points from M simulated values. This provides J within-chain interval lengths, with mean I_w. For the pooled output of MJ samples, the same $100(1-\alpha)\%$ interval I_B is obtained. The ratio I_B/I_W should converge to one if there is convergent mixing over different chains.

The above MCMC convergence diagnostics are implemented in two `R`-packages, namely CODA (**C**onvergence **D**iagnosis and **O**utput **A**nalysis) and BOA (**B**ayesian **O**utput **A**nalysis). These packages are downloadable from `http://cran.r-project.org/`. The packages compute convergence diagnostics and statistical and graphical summaries for the MCMC samples. Even though BOA is designed to be faster and more efficient than CODA, it is not flexible in terms of data manipulation than CODA. That is, CODA offers more analysis options and better graphical tools than BOA.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Introduction

This study seeks to model the vote share for the incumbent and the closest challenger(s) in Kenya's general election. Previous elections in Kenya have shown that presidential election is a two horse-race. There is evidence that presidential strongholds are based largely on ethnicity, implying that a candidate is likely to garner more votes in region where he/she comes from even without campaigning. The study also looked at DDP, inflation, PPI, time to election and unemployment as import predictors to candidates vote share. However, PPI, GDP and unemployment were insignificant and therefore were removed from the model.

4.2 Simulation Results for Comparison of Two Candidates

In order to understand the concept of sequential Bayesian analysis, we will consider a case of two candidates (incumbent denoted by K and challenger denoted by R) with four different scenarios forming our simulation set ups.

4.2.1 Simulations – scenario one

To begin with, let us consider the case where the two candidates have roughly equal popularity proportions but with some observable fluctuations. The first 100 iterations yielded results shown in Figure (4.1). Even after 20 iterations, the chain tends to the true value.

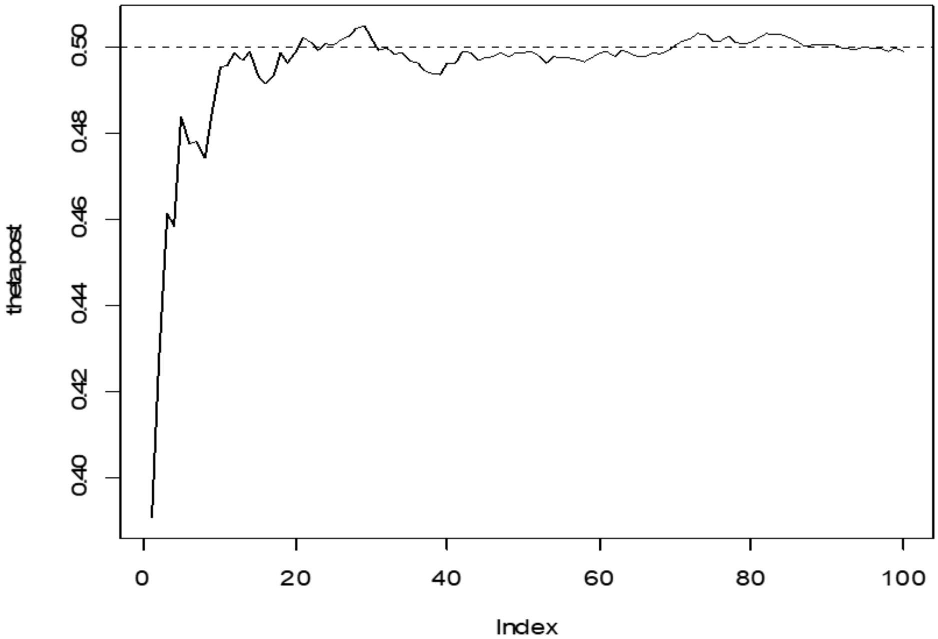

Figure 4.1: *History plot for the first 100 iterations to demonstrate initiation of the chain*

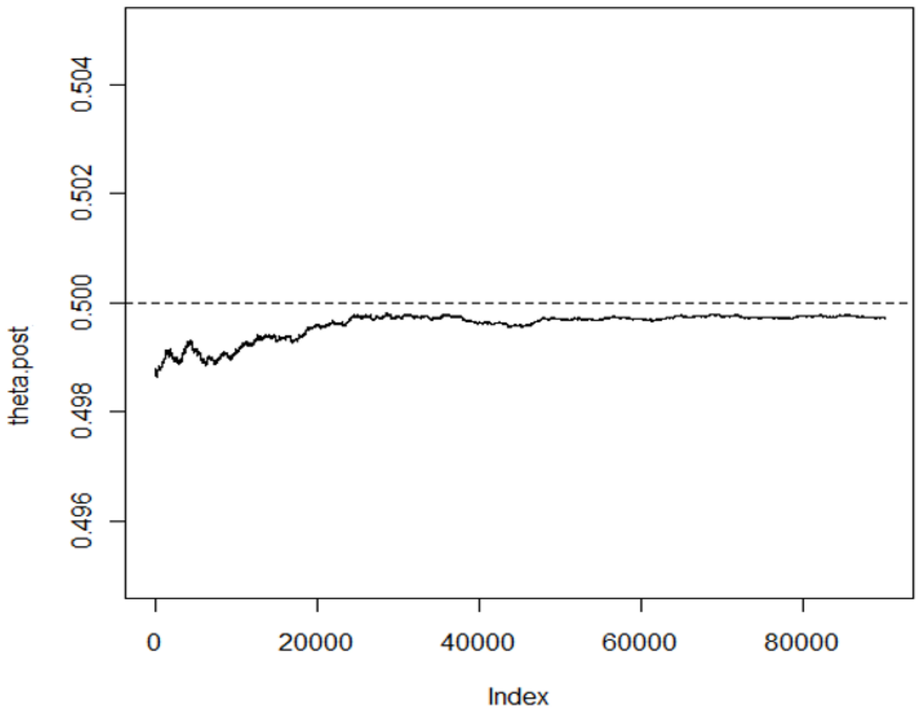

Figure 4.2: *Trace plot after burn-in of 10,000 iterations predicting the success rate of the incumbent*

4.2.2 Simulations – scenario two

We now consider the case where the popularity of one candidate (say the challenger) is increasing implying that the popularity of the other is decreasing over time. As shown in Figure certainly the incumbent will win the presidential race as the posterior probability is well above 0.5.

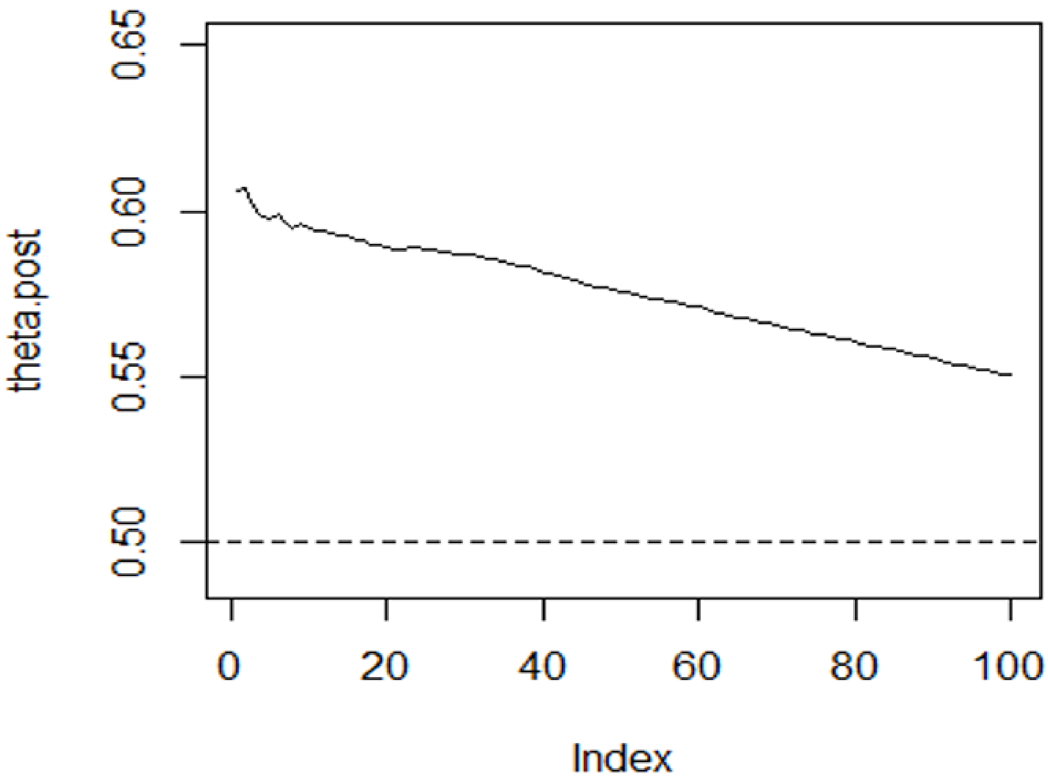

Figure 4.3: *Trace plot after 10,000 burn-in where the probability of the incumbent is assumed over 0.5*

4.2.3 Simulations – scenario three

Thirdly, we consider the case where the popularity of one candidate (say the challenger) being constantly slightly higher but with misclassification in favour of the other (say the Incumbent). The misclassification for this scenario is as follows:

- No misclassification
- Low misclassification: p01=.05 p10=.10
- Misclassification: p01=.05 p10=.15
- Misclassification: p01=.10 p10=.10

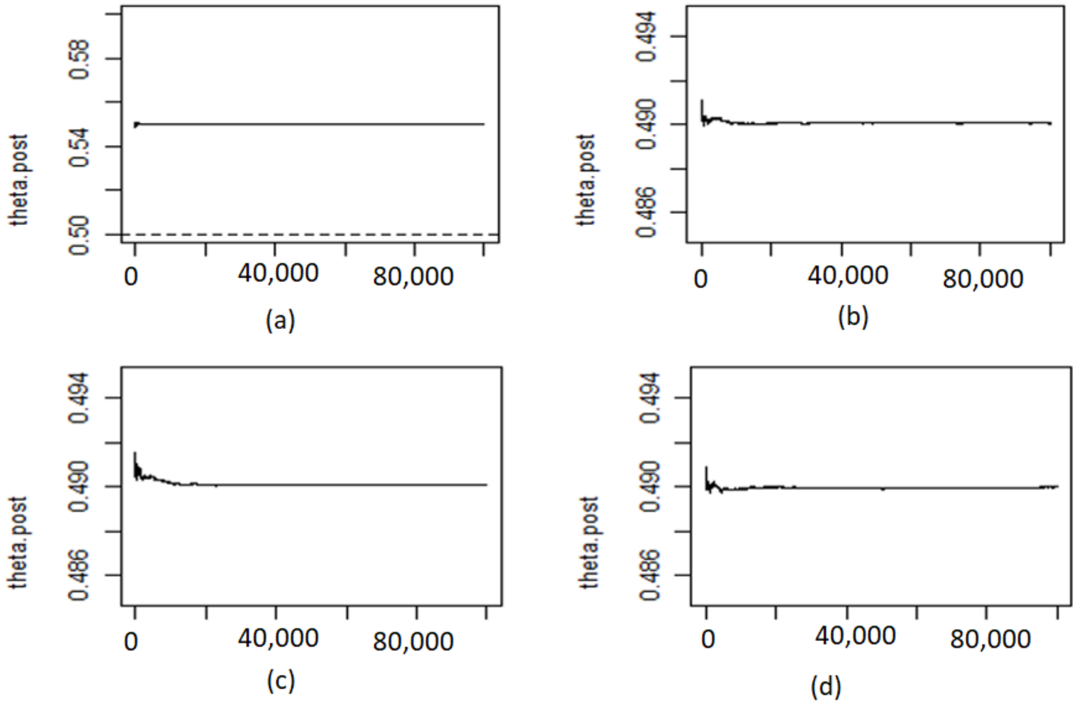

Figure 4.4: *History plot showing the effect of misclassification (panels a-d)*

The simulations results show that without the misclassification, panel (a), the challenger will win the election with a good margin. However, with misclassification, panels (b) to (d), the challenger will narrowly lose the election as his popularity eventually stabilizes around 0.491.

4.3 Results from the Bayesian Model Fit

A Bayesian was fitted to the Kenyan 2013 opinion data comparing three candidates, the two leading candidates and the other, which comprises of all the other remaining candidates. The parameters of interest here are $\alpha_1, \alpha_2, \alpha_3$ and $\theta_1, \theta_2, \theta_3$ which predict the probability of each of the three candidates. The model was fitted in WinBUGS with post-analysis in R.

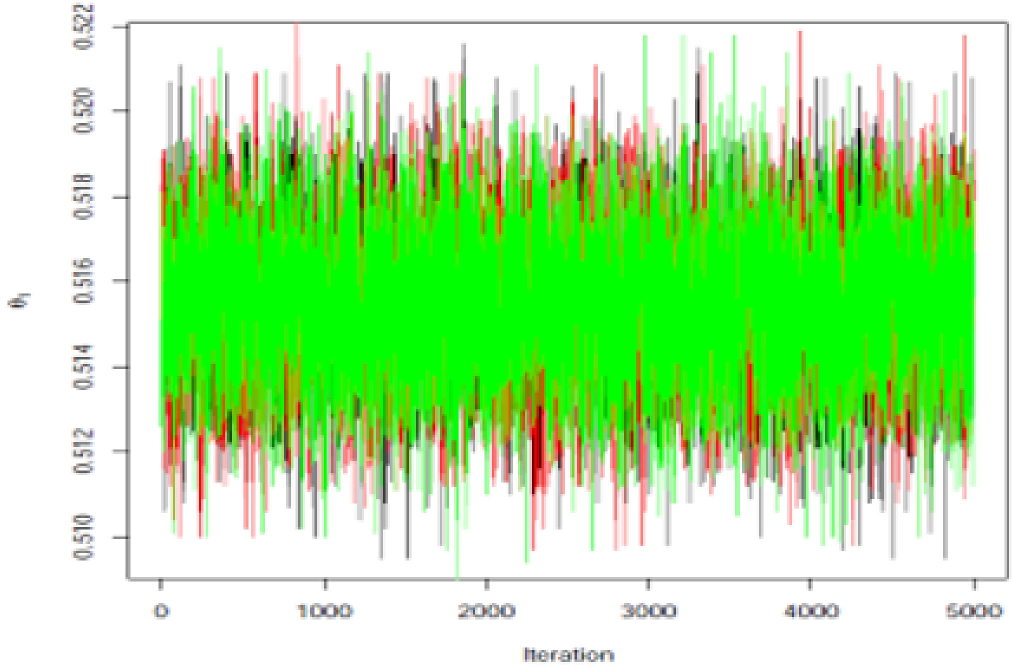

Figure 4.5: *Trace plot of the θ_1, the parameter for the leading candidate, from the trivariate model of the* 2013 *Kenyan poll*

From plots of the three chains, for each parameter, as can be seen in Figures (4.5) - (4.10) , it is immediately clear that the posterior estimates of the parameters are converging because the chain mix like spaghetti.

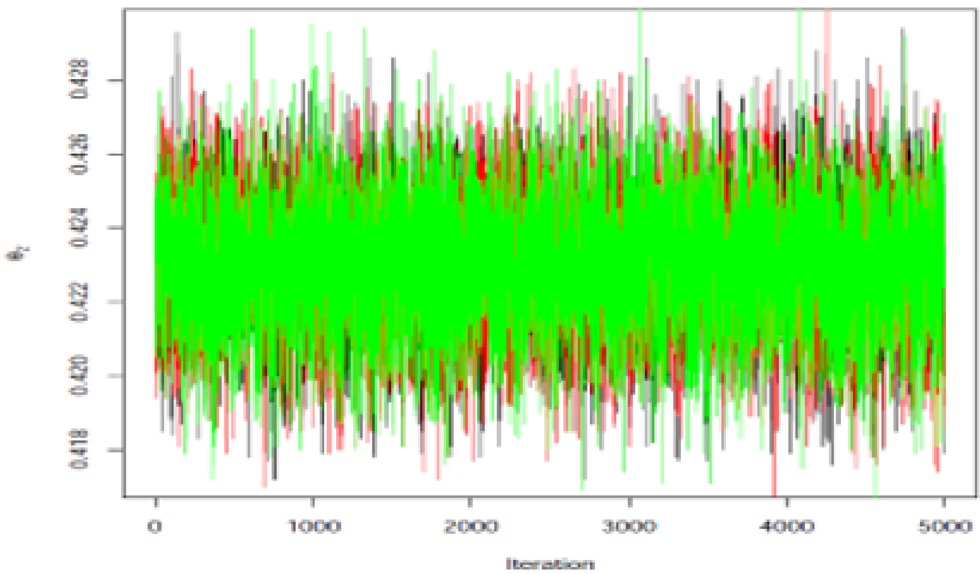

Figure 4.6: *Trace plot of the θ_2, the parameter for the second candidate, from the trivariate model of the 2013 Kenyan poll*

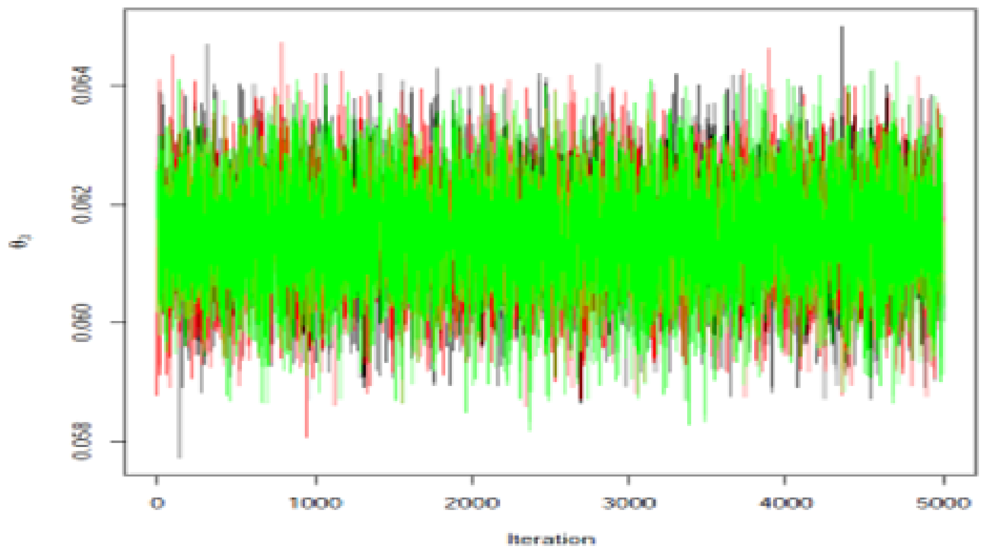

Figure 4.7: *Trace plot of the θ_3, the parameter for the other candidates, from the trivariate model of the 2013 Kenyan poll*

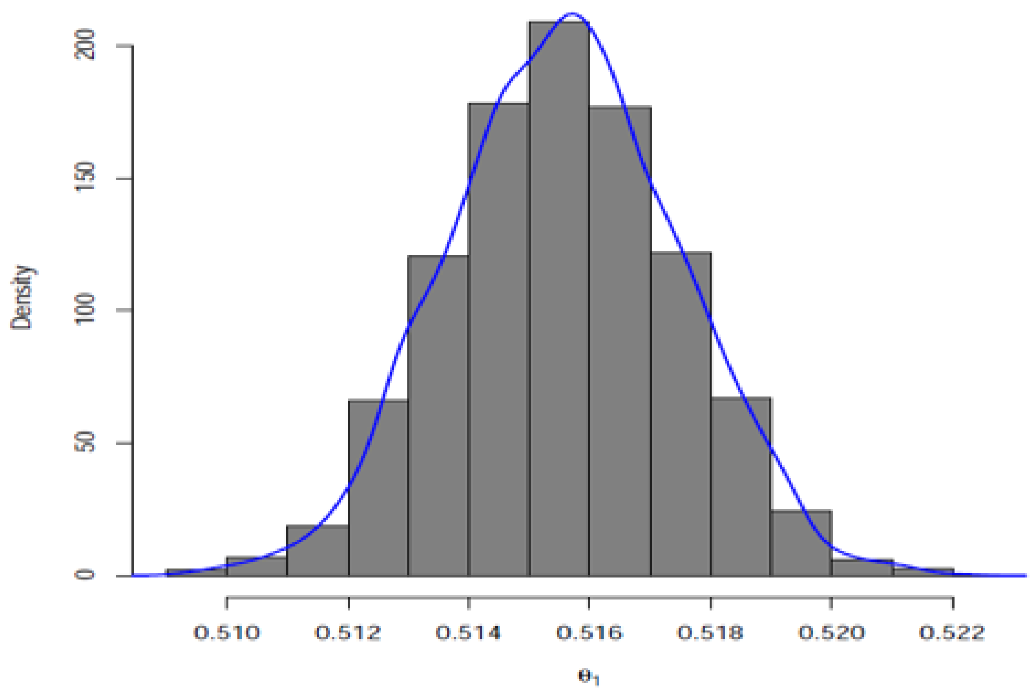

Figure 4.8: *Density plot of the θ_1, the parameter for the leading candidate, from the trivariate model of the 2013 Kenyan poll*

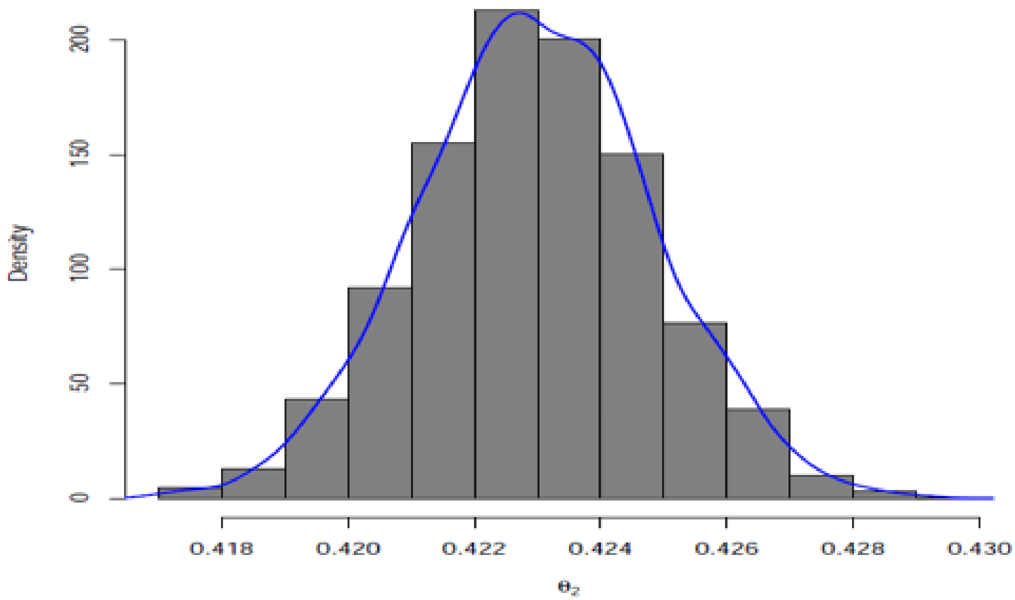

Figure 4.9: *Density plot of the θ_2, the parameter for the leading candidate, from the trivariate model of the 2013 Kenyan poll*

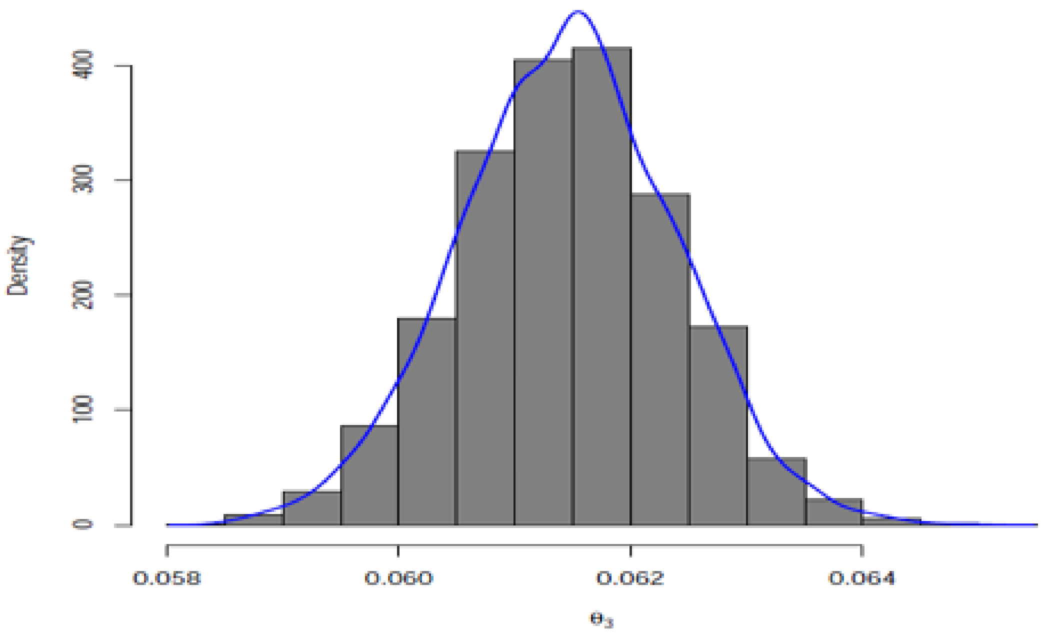

Figure 4.10: *Density plot of the θ_3, the parameter for the leading candidate, from the trivariate model of the 2013 Kenyan poll*

The density plots of the parameters of interested, shown in Figure (4.8) - (4.10), indicate that the posterior distributions of $(\theta_1, \theta_2, \theta_3)$ are symmetrical and take nearly a normal distribution shame. This implies that the posterior point estimates can estimated by posterior mean or median as shown in Table 4.1.

Table 4.1: *The posterior point estimates and 95% credible interval from the trivariate model of the 2013 Kenyan poll*

Parameters	Estimates	S.E	25% Credible Interval	95% Credible Interval
α_1^*	0.000	-	-	-
α_2	-0.198	0.0078	-0.213	-0.183
α_3	-2.127	0.0166	-2.160	-2.095
θ_1	0.515	0.0019	0.512	0.519
θ_2	0.423	0.0019	0.419	0.427
θ_3	0.0615	9.3E-04	6.0E-02	0.063

α_1^* is set to 0 as a baseline parameter. From these results the leading Candidate is projected to win with 51.5% (95% CI 51.2-51.9) with the second candidate with 42.3% (95% CI 41.9-42.7) and the other remaining candidate at 6.2% (95% CI 6.0-6.3). Since the 95% credibility interval do not overlap, the model seems to predict incumbent as the outright winner.

4.4 Assessing Bayesian Convergence Diagnostics

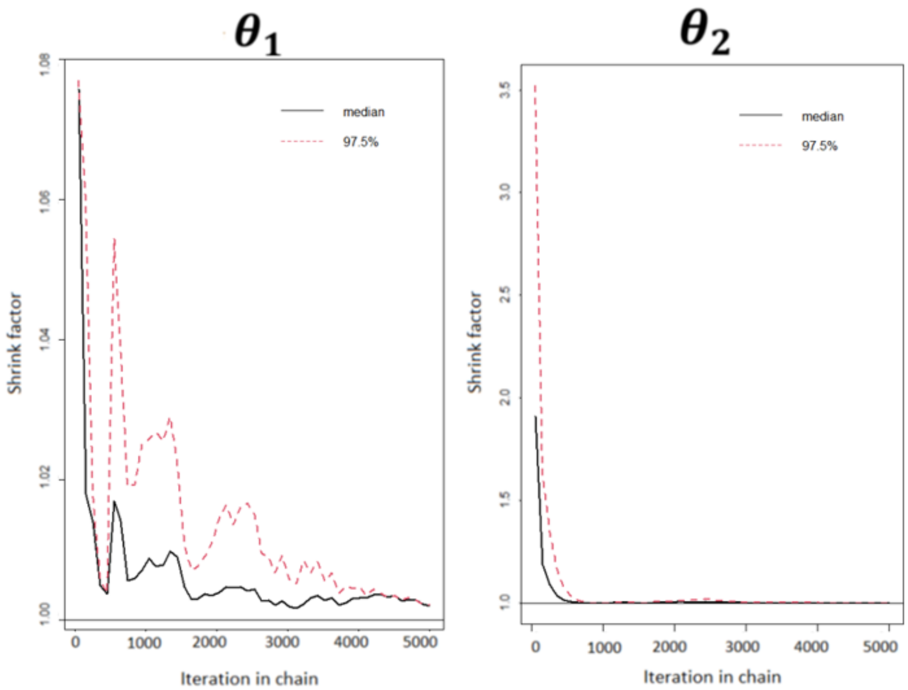

Figure 4.11: *Gelman convergence dignostic for θ_1 and θ_2*

As can be seen in Figure 4.11, both with Gelman potential scale reduction equal 1.0. This is a good indicator of convergence of the posterior draws. Furthermore, Figure 4.12, after lag of about 15 all parameters decayed well.

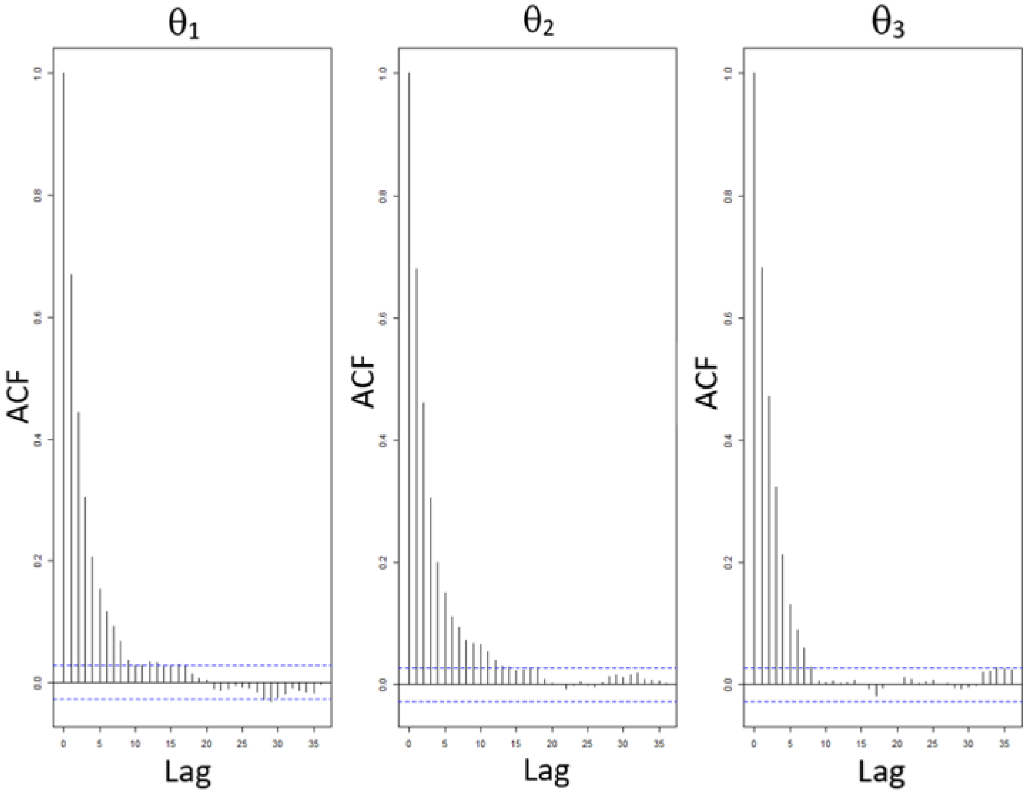

Figure 4.12: *Auto-correlation function for θ_1, θ_2 and θ_3*

CHAPTER FIVE

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusions

In this research, we have developed the basis of the Bayesian approach to statistical inference of opinion poll data. Bayesian approach handles various scenario in the full projection including the aspect of misclassification error. Further, even where polling data are scanty, it incorporates prior distribution to express the model uncertainty.

Bayesian analysis provided a very powerful approach to modelling opinion poll data. The posterior mode gives unbiased estimate of the parameter of interest. Furthermore, it averaged over many data points, thus taking care of possible inconsistencies in some data points. Even though in this case we considered three candidate, the model can be easily extend to more than three candidates.

We have provided a flexible way of comparing the two leading candidates since in most election there is always two candidates who lead the pack. Our approach, though applied to Kenyan opinion polls, can be applied anyway in the world. The developed model can be extended easily to multinomial or count outcomes since it is only the parent distribution (the likelihood) that changes. Otherwise, the philosophy and the ideology remains the same.

5.2 Recommendations and Further Research

This work has demonstrated the importance of the opinion polls and the critical role it plays in the presidential electioneering process. It is clear that in most election the incumbent will have an upper hand primarily because of the prior experience and unequivocal support from the current government. Opinion polls, however, may suffer from self-reporting bias. As part of the further research, we recommend a model that adjust for the resulting bias. In addition, future research work should look at the house effects, underdog and bandwagon effects as well as the effect of news on voters choice. Still the assumption that the pollsters were independent and that the publication of such polls had no impact on the voters choice of candidate is something that need to be relaxed.

REFERENCES

Abrams, M. (1970). The opinion polls and the 1970 british general election. *Public Opinion Quarterly*, pages 317–324.

Ajulu, R. (1992). Kenya: The road to democracy.

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

Armstrong, J. S. and Graefe, A. (2011). Predicting elections from biographical information about candidates: A test of the index method. *Journal of Business Research*, 64(7), 699–706.

Arrow, K. J. (1971). The theory of risk aversion. *Essays in the theory of risk-bearing*, pages 90–120.

Ashworth, S. (2012). Electoral accountability: recent theoretical and empirical work. *Annual Review of Political Science*, 15, 183–201.

Benoit, W. L., Hansen, G. J. and Verser, R. M. (2003). A meta-analysis of the effects of viewing us presidential debates. *Communication Monographs*, 70(4), 335–350.

Berger, J. O., Moreno, E., Pericchi, L. R., Bayarri, M. J., Bernardo, J. M., Cano, J. A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B. *et al.* (1994). An overview of robust bayesian analysis. *Test*, 3(1), 5–124.

Biemer, P. P. (2010). Total survey error: Design, implementation, and evaluation. *Public Opinion Quarterly*, 74(5), 817–848.

Brooks, S. P. and Roberts, G. O. (1998). Convergence assessment techniques for markov chain monte carlo. *Statistics and Computing*, 8(4), 319–335.

Brouwer, M. (1955). Twenty questions about polls. *Gazette (Leiden, Netherlands)*, 1(3), 175–180.

Brown, P. and Payne, C. (1975). Election night forecasting. *Journal of the Royal Statistical Society: Series A (General)*, 138(4), 463–483.

Buchanan, W. (1986). Election predictions: An empirical assessment. *Public Opinion Quarterly*, 50(2), 222–227.

Buglass, I. (1997). The democratization process in sub-saharan africa: Rhetoric or reality (ethiopia, kenya, cameroon, sierra leone).

Butler, D. and Kavanagh, D. (1997). *The British general election of 1997*. Springer.

Byers, D., Davidson, J. and Peel, D. (1997). Modelling political popularity: an analysis of long-range dependence in opinion poll series. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 471–490.

Cameron, L. and Crosby, M. (2000). It's the economy stupid: Macroeconomics and federal elections in australia. *Economic Record*, 76(235), 354–364.

Campbell, J. E. (2008). Editor's introduction: Forecasting the 2008 national elections. *PS: Political Science & Politics*, 41(4), 679–682.

Campbell, J. E. (2012). Forecasting the 2012 american national elections. *PS, Political Science & Politics*, 45(4), 610.

Campbell, J. E., Cherry, L. L. and Wink, K. A. (1992). The convention bump. *American Politics Quarterly*, 20(3), 287–307.

Campbell, J. E., Norpoth, H., Abramowitz, A. I., Lewis-Beck, M. S., Tien, C., Erikson, R. S., Wlezien, C., Lockerbie, B., Holbrook, T. M., Jerôme, B. *et al.* (2017). A recap of the 2016 election forecasts. *PS: Political Science & Politics*, 50(2), 331–338.

Campbell, W. J. (2020). 2."a time of polls gone mad": The literary digest debacle of 1936. In *Lost in a Gallup*, pages 40–61. University of California Press.

Carr, L., Newtson, A. J. and Joshi, J. (2018). Towards modernizing the future of american voting. In *2018 IEEE 4th International Conference on Collaboration and Internet Computing (CIC)*, pages 130–135. IEEE.

Casella, G. and George, E. I. (1992). Explaining the gibbs sampler. *The American Statistician*, 46(3), 167–174.

Cheeseman, N. (2008). The kenyan elections of 2007: an introduction. *Journal of Eastern African Studies*, 2(2), 166–184.

Cohen, J. E. (1998). State-level public opinion polls as predictors of presidential election results: The 1996 race. *American Politics Quarterly*, 26(2), 139–159.

Cohen, J. E. (2015). *Presidential leadership in public opinion*. Cambridge University Press.

Cohen, J. E. and Hamman, J. A. (2003). The polls: Can presidential rhetoric affect the public's economic perceptions? *Presidential Studies Quarterly*, 33(2), 408–422.

Congdon, P. D. (2010). *Applied Bayesian hierarchical methods*. CRC Press.

Curtice, J. and Firth, D. (2008). Exit polling in a cold climate: the bbc–itv experience in britain in 2005. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(3), 509–539.

Cuzán, A. G., Armstrong, J. S. and Jones, R. (2005). Combining methods to forecast the 2004 presidential election: The pollyvote. In *annual meeting of the Southern Political Science Association. New Orleans. January*. Citeseer.

DeSart, J. A. and Holbrook, T. M. (2003). Statewide trial-heat polls and the 2000 presidential election: A forecast model. *Social Science Quarterly*, 84(3), 561–573.

Duch, R. M. and Stevenson, R. T. (2008). *The economic vote: How political and economic institutions condition election results*. Cambridge University Press.

Elfadaly, F. G. and Garthwaite, P. H. (2013). Eliciting dirichlet and connor–mosimann prior distributions for multinomial models. *Test*, 22(4), 628–646.

Elfadaly, F. G. and Garthwaite, P. H. (2017). Eliciting dirichlet and gaussian copula prior distributions for multinomial models. *Statistics and Computing*, 27(2), 449–467.

Erikson, R. S. and Wlezien, C. (2013). Forecasting with leading economic indicators and the polls in 2012. *PS, Political Science & Politics*, 46(1), 38.

Feld, S. L. and Grofman, B. (2010). Puzzles and paradoxes involving averages: An intuitive approach. In *Collective Decision Making*, pages 137–150. Springer.

Garthwaite, P. H., Kadane, J. B. and O'Hagan, A. (2005). Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 100(470), 680–701.

Gayo-Avello, D. (2013). A meta-analysis of state-of-the-art electoral prediction from twitter data. *Social Science Computer Review*, 31(6), 649–679.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410), 398–409.

Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995). Model checking and sensitivity analysis. *Bayesian Data Analysis. London: Chapman & Hall*, pages 161–189.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). Bayesian data analysis chapman & hall. *CRC Texts in Statistical Science*.

Gelman, A., Goel, S., Rivers, D., Rothschild, D. *et al.* (2016). The mythical swing voter. *Quarterly Journal of Political Science*, 11(1), 103–130.

Gelman, A., Katz, J. N. and Tuerlinckx, F. (2002). The mathematics and statistics of voting power. *Statistical Science*, pages 420–435.

Gelman, A., Rubin, D. B. *et al.* (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4), 457–472.

Geweke, J. *et al.* (1991). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*, volume 196. Federal Reserve Bank of Minneapolis, Research Department Minneapolis, MN.

Grant Jr, P. A. (1992). Presidential politics in south dakota, 1936. *South Dakota History*, 22 (3), 261–275.

Greene, J. P. (1993). Forewarned before forecast: Presidential election forecasting models and the 1992 election. *PS: Political Science and Politics*, 26(1), 17–21.

Grignon, F., Rutten, M. and Mazrui, A. (2001). Observing and analysing the 1997 general election: an introduction. *Out for the count: the 1997 general elections and prospects for democracy in Kenya*, pages 1–28.

Groves, R. M. and Lyberg, L. (2010). Total survey error: Past, present, and future. *Public opinion quarterly*, 74(5), 849–879.

Hendrix, C. S. (2010). Measuring state capacity: Theoretical and empirical implications for the study of civil conflict. *Journal of peace research*, 47(3), 273–285.

Hogan, J. M. (1997). George gallup and the rhetoric of scientific democracy. *Communications Monographs*, 64(2), 161–179.

Hogan, M. J. and Hogan, M. J. (1987). *The Marshall Plan: America, Britain and the Reconstruction of Western Europe, 1947-1952*. Cambridge University Press.

Holbrook, T. M. and DeSart, J. A. (1999). Using state polls to forecast presidential election outcomes in the american states. *International Journal of Forecasting*, 15(2), 137–142.

Holbrook-Provow, T. M. (1987). National factors in gubernatorial elections. *American Politics Quarterly*, 15(4), 471–483.

Jackman, S. (2005). Pooling the polls over an election campaign. *Australian Journal of Political Science*, 40(4), 499–517.

Jaidka, K., Ahmed, S., Skoric, M. and Hilbert, M. (2019). Predicting elections from social media: a three-country, three-method comparative study. *Asian Journal of Communication*, 29(3), 252–273.

Jones, R. J. (2002). *Who will be in the White House?: Predicting presidential elections*. Longman Publishing Group.

Kaplan, E. H. and Barnett, A. (2003). A new approach to estimating the probability of winning the presidency. *Operations Research*, 51(1), 32–40.

Katz, D. (1941). The public opinion polls and the 1940 election. *Public Opinion Quarterly*, 5 (1), 52–78.

Kennedy, R., Wojcik, S. and Lazer, D. (2017). Improving election prediction internationally. *Science*, 355(6324), 515–520.

Kernell, S. (2000). Life before polls: Ohio politicians predict the 1828 presidential vote. *PS: Political Science & Politics*, 33(3), 569–574.

Lakeman, E. (1974). February and october.

Lewis-Beck, M. S. and Dassonneville, R. (2015). Comparative election forecasting: Further insights from synthetic models. *Electoral Studies*, 39, 275–283.

Linzer, D. A. (2013). Dynamic bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108(501), 124–134.

Lock, K. and Gelman, A. (2010). Bayesian combination of state polls and election forecasts. *Political Analysis*, 18(3), 337–348.

Lohr, S. L. and Brick, J. M. (2017). Roosevelt predicted to win: Revisiting the 1936 literary digest poll. *Statistics, Politics and Policy*, 8(1), 65–84.

Mahler, S. J. (2017). Theoretical and empirical contributions toward a research agenda for transnationalism 1. In *Transnationalism from below*, pages 64–100. Routledge.

Mavragani, A. and Tsagarakis, K. P. (2016). Yes or no: Predicting the 2015 greferendum results using google trends. *Technological Forecasting and Social Change*, 109, 1–5.

McDonald, D. G., Glynn, C. J., Kim, S.-H. and Ostman, R. E. (2001). The spiral of silence in the 1948 presidential election. *Communication Research*, 28(2), 139–155.

Miller, W. L. and Mackie, M. (1973). The electoral cycle and the asymmetry of government and opposition popularity: an alternative model of the relationship between economic conditions and political popularity. *Political Studies*, 21(3), 263–279.

Moala, F. A. and O'Hagan, A. (2010). Elicitation of multivariate prior distributions: A non-parametric bayesian approach. *Journal of Statistical Planning and Inference*, 140(7), 1635–1655.

Moyo, J. N. (1992). State politics and social domination in zimbabwe. *The Journal of Modern African Studies*, 30(2), 305–330.

Muhati, P. (2014). Is the kenyan public given enough information about presidential opinion polls? *Kenya Studies*, page 30.

Newman, N. (2017). Journalism, media, and technology trends and predictions 2017.

Pearson, K. (1920). Notes on the history of correlation. *Biometrika*, 13(1), 25–45.

Powell Jr, G. B. and Whitten, G. D. (1993). A cross-national analysis of economic voting: taking account of the political context. *American Journal of Political Science*, pages 391–414.

Raftery, A., Lewis, S., Gilks, W., Spiegelhalter, D. and Richardson, S. (1996). Practical markov chain monte carlo.

Rattinger, H. (1991). Unemployment and elections in west germany. *Economics and politics: the calculus of support*, pages 49–62.

Richardson, S. and Best, N. (2003). Bayesian hierarchical models in ecological studies of health–environment effects. *Environmetrics: The official journal of the International Environmetrics Society*, 14(2), 129–147.

Rothschild, D. (2009). Forecasting elections: Comparing prediction markets, polls, and their biases. *Public Opinion Quarterly*, 73(5), 895–916.

Sargent, T. J. (2010). Rational expectations. In *Macroeconometrics and Time Series Analysis*, pages 193–201. Springer.

Scott, A., Smith, T. and Jones, R. G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical Review/Revue Internationale de Statistique*, pages 13–28.

Sheffrin, S. M. *et al.* (1996). *Rational expectations*. Cambridge University Press.

Shirani-Mehr, H., Rothschild, D., Goel, S. and Gelman, A. (2018). Disentangling bias and variance in election polls. *Journal of the American Statistical Association*, 113(522), 607–614.

Sitkoff, H. (1971). Harry truman and the election of 1948: The coming of age of civil rights in american politics. *The Journal of Southern History*, 37(4), 597–616.

Smith, T. W. (1990). The first straw? a study of the origins of election polls. *Public Opinion Quarterly*, 54(1), 21–36.

Spiegelhalter, D. J., Thomas, A., Best, N. and Lunn, D. (2003). Winbugs version 1.4 user manual. *MRC Biostatistics Unit, Cambridge. URL http://www. mrc-bsu. cam. ac. uk/bugs*.

Stroud, T. (1994). Bayesian analysis of binary survey data. *Canadian Journal of Statistics*, 22 (1), 33–45.

Tankard Jr, J. W. (1972). Public opinion polling by newspapers in the presidential election campaign of 1824. *Journalism Quarterly*, 49(2), 361–365.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.

Topping, S. (2004). "never argue with the gallup poll": Thomas dewey, civil rights and the election of 1948. *Journal of American Studies*, 38(2), 179–198.

Venkataramani, M. (1960). Norman thomas, arkansas sharecroppers, and the roosevelt agricultural policies, 1933-1937. *The Mississippi Valley Historical Review*, 47(2), 225–246.

Walther, D. (2015). Picking the winner (s): Forecasting elections in multiparty systems. *Electoral Studies*, 40, 1–13.

Walton, H., Gray, C. V. and McLemore, L. (2017). African american public opinion and thepre-scientific polls: The literary digest magazine's straw-vote presidential polls, 1916-1936. In *The Politics of the Black "Nation" a Twenty-Five-Year Retrospective*, pages 222–243. Routledge.

Wang, W., Rothschild, D., Goel, S. and Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, 31(3), 980–991.

Wanyande, P. (2006). Electoral politics and election outcomes in kenya. *Africa Development*, 31(3), 62–80.

Wattenberg, M. P. (2003). Elections: Was the 2000 presidential election fair? an analysis of comparative and retrospective survey data. *Presidential Studies Quarterly*, 33(4), 889–897.

Webb, W. B. (2010). Opinion polls and science. *Sleep*, 33(7), 865–866.

Wolfers, J. and Leigh, A. (2002). Three tools for forecasting federal elections: Lessons from 2001. *Australian Journal of Political Science*, 37(2), 223–240.

APPENDICES

Appendix I: Beta distribution and Binomial profile plots

```
p = seq(0,1, length=100)
plot(p, dbeta(p, 100, 100), xlab="",ylab="Frequency", type ="l",
col=4,lwd=4)
lines(p, dbeta(p, 10, 10), type ="l", col=2,lwd=4)
#lines(p, dbeta(p, 2, 2), col=1,lwd=4)
lines(p, dbeta(p, 1, 1), col=1,lwd=4)
#legend(0.7,8, c("Be(100,100)","Be(10,10)","Be(2,2)", "Be(1,1)"),
lty=c(1,1,1,1),col=c(4,3,2,1))

#library(arrow)
arrows(0.05, 3, x1 = 0.1, y1 = 1, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
#arrows(0.0, 3, x1 = 0.1, #y1 = 1, length = 0.25, angle = 60,
code = 2, col = par("fg"), lty = par("lty"), lwd = par("lwd"))
text(0.05,3.5,"Beta(2,2)",family = "serif",cex=2)
arrows(0.2, 6, x1 = 0.35, y1 = 1.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.2,6.5,"Beta(10,10)",family = "serif",cex=2)
arrows(0.3, 9, x1 = 0.45, y1 = 4.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.3,9.5,"Beta(100,100)",family = "serif",cex=2)


p = seq(0,1, length=100)
plot(p, dbeta(p, 10, 10),xlim=c(0,1),ylim=c(0,60), xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
lines(p, dbeta(p, 120, 100), type ="l", col=2,lwd=3)
#lines(p, dbeta(p, 2, 2), col=1,lwd=4)
lines(p, dbeta(p, 1, 1), col=1,lwd=4)
#legend(0.7,8, c("Be(100,100)","Be(10,10)","Be(2,2)", "Be(1,1)"),
lty=c(1,1,1,1),col=c(4,3,2,1))

#library(arrow)
arrows(0.05, 3, x1 = 0.1, y1 = 1, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
#arrows(0.0, 3, x1 = 0.1, #y1 = 1, length = 0.25, angle = 60,
code = 2, col = par("fg"), lty = par("lty"), lwd = par("lwd"))
text(0.05,3.5,"Beta(2,2)",cex=2)
arrows(0.2, 6, x1 = 0.35, y1 = 1.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.2,6.5,"Beta(10,10)",cex=2)
arrows(0.3, 9, x1 = 0.45, y1 = 4.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
```

```
text(0.3,9.5,"Beta(100,100)",cex=2)


#install.packages("LearnBayes")
library(LearnBayes)
prior = c( a= 3, b = 27 ) # beta prior
data = c( s = 3, f = 9 ) # s events out of f trials
triplot(prior,data)


p = seq(0,1, length=100)
plot(p, dbeta(p, 100, 100),xlab="",ylab="Frequency", type ="l",
col=4,lwd=4)

#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
lines(p, dbeta(p, 160, 140), type ="l", col=2,lwd=4)
#lines(p, dbeta(p, 2, 2), col=1,lwd=4)
lines(p, dbeta(p, 60,40), col=1,lwd=4)
#legend(0.7,8, c("Be(100,100)","Be(10,10)","Be(2,2)", "Be(1,1)"),
lty=c(1,1,1,1),col=c(4,3,2,1))

#library(arrow)
arrows(0.05, 3, x1 = 0.1, y1 = 1, length = 0.35, angle = 20,
code = 2,col = par("fg"), lty = par("lty"), lwd = 2.5)
#arrows(0.0, 3, x1 = 0.1, #y1 = 1, length = 0.25, angle = 60,
code = 2,col = par("fg"), lty = par("lty"), lwd = par("lwd"))
text(0.05,3.5,"Beta(2,2)",cex=2)
arrows(0.2, 6, x1 = 0.35, y1 = 1.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.2,6.5,"Beta(10,10)",cex=2)
arrows(0.3, 9, x1 = 0.45, y1 = 4.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.3,9.5,"Beta(100,100)",cex=2)


#plot.new()
#plot.window(xlim=c(0,0.6), ylim=c(0,60))
#axis(1)
#axis(2)
#box()
p = seq(0,1, length=100)
plot(p, dbeta(p, 100, 100),ylim=c(0,14),xlab="",ylab="Frequency",
type ="l", col=4,lwd=4)
#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
```

```
#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
lines(p, dbeta(p, 160, 140), lty="dashed", col=2,lwd=4)
#lines(p, dbeta(p, 2, 2), col=1,lwd=4)
lines(p, dbeta(p, 60,40),lty="dotted", col=1,lwd=4)
#legend("topright", c("Be(100,100)","Be(10,10)","Be(2,2)",
"Be(1,1)"),lty=c(1,1,1,1),col=c(4,3,2,1))
arrows(0.3, 8, x1 = 0.43, y1 = 1.5, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.3,9.5,"Beta(100,100)",family = "serif",cex=2)

p = seq(0,1, length=100)
plot(p, dbeta(p, 100, 100),xlim=c(0.35,0.8),ylim=c(0,14),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
#plot(p, dbeta(p, 100, 100),xlim=c(0,0.6),ylim=c(0,100),xlab="",
ylab="Frequency", type ="l", col=4,lwd=4)
lines(p, dbeta(p, 160, 140), type="l", col=2,lwd=4)
#lines(p, dbeta(p, 2, 2), col=1,lwd=4)
lines(p, dbeta(p, 60,40),type="l", col=1,lwd=4)
#legend("topright", c("Prior","Posterior","Likelihood"),lty=c(1,1,1),
col=c("Blue","Red","black"))

arrows(0.4, 5, x1 = 0.43, y1 = 1.5, length = 0.35, angle = 20,
 code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.4,5.5,"Prior",family = "serif",cex=2)
arrows(0.425, 12.5, x1 = 0.52, y1 = 12, length = 0.35, angle = 20,
code = 2, col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.4,12.5,"Posterior",family = "serif",cex=2)
arrows(0.65, 11, x1 = 0.61, y1 = 8, length = 0.35, angle = 20,
code = 2,col = par("fg"), lty = par("lty"), lwd = 2.5)
text(0.65,11.5,"Normalized(likelihood)",family = "serif",cex=2)

#text(0.2, 0.2, "Prior", family = "Times New Roman")

abline(a=6, b=3)
title(main="The Overall Title")
title(xlab="An x-axis label")
title(ylab="A y-axis label")
box()

prior = c( a= 100, b = 100 ) # beta prior
data = c( s = 40, f = 60) # s events out of f trials
triplot(prior,data)
title("BNN")
```

Appendix II: Simulations using R2WinBUGS

```
#load library
library(R2WinBUGS)

#change working directory
setwd("C:\\Work\\JKU\\phd\\jk")
# look at the data
source("mult.dat")
# Define or generate the parameter in the models of WinBUGS
I <- mult$I
J <- mult$J
X <- mult$X
data <- list ("I", "J", "X")

# Define the initial values by using a function

inits <- function(){ # it's a really weird way to write function
        list(alpha = c(NA,rnorm(2, 1,2)), beta = c(NA,rnorm(27,2,1)))
}

inits()

# call WinBUGS
# We first stop on the WinBUGS to see what happened
mult.sim <- bugs(data, inits,
working.directory="C:\\Work\\JKU\\phd\\jk",
model.file = "mult.bugs",
# try inits=NULL
parameters.to.save = c("alpha", "beta", "p","phat"), n.thin=1,
# define the thin rate
# n.thin=1, which means that we keep tracking the samples
# without dropping any of them
# if not setting n.thin, it will drop some terms
n.chains = 3, n.iter = 10000, n.burnin=5000, debug=TRUE,
# burning = 100, n.iter = MCMCsamples(each chain)+n.burin
bugs.directory = "c:/WinBUGS14/")
# debug=TRUE will stop in WinBUGS rather than only
#shows the results

## How to fix multiple initial values
# See what is in the mult.sim
ls(mult.sim)

# results
print(mult.sim, dig=3)
plot(mult.sim)

# MCMC samples
```

```
ls(mult.sim$sims.list)

# mu.theta as example
hist(mult.sim$sims.list$alpha[,1],prob=TRUE)
hist(mult.sim$sims.list$alpha[,2],prob=TRUE)

op <- par()
par(mfrow=c(1,3))
hist(mult.sim$sims.list$phat[,1],prob=TRUE, xlab=expression(p[1]),
main="")
hist(mult.sim$sims.list$phat[,2],prob=TRUE, xlab=expression(p[2]),
main="")
hist(mult.sim$sims.list$phat[,3],prob=TRUE, xlab=expression(p[3]),
main="")


# plot with the prior distribution as line
prior_sample <- rnorm(1000,0,10)
lines(density(prior_sample))

dim(mult.sim$sims.list$alpha)
dim(mult.sim$sims.list$beta)
dim(mult.sim$sims.list$phat)

dim(mult.sim$sims.list$p)

sapply(mult.sim$sims.list$phat, mean)

par(mfrow=c(1,1))
boxplot(mult.sim$sims.list$p[,,1])
boxplot(mult.sim$sims.list$p[,,2])
boxplot(mult.sim$sims.list$p[,,3])

par(mfrow=c(1,3))
ts.plot(apply(mult.sim$sims.list$p[,,1],1,mean),col="red",
lwd=2,ylim=c(0.3,.6),ylab=expression(p[1]))
ts.plot(apply(mult.sim$sims.list$p[,,2],1,mean),col="blue",
lwd=2,ylim=c(.3,.6),ylab=expression(p[2]))
ts.plot(apply(mult.sim$sims.list$p[,,3],1,mean),col="green",
lwd=2,ylim=c(0,.2),ylab=expression(p[3]))


pdf("theta1-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,1],
type="l",ylab=expression(theta[1]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),1],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),1],type="l",
```

```
col="green")
dev.off()

pdf("theta2-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,2],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta3-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,3],
type="l",ylab=expression(theta[3]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),3],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),3],type="l",
col="green")
dev.off()

mean(mult.sim$sims.list$p[,,1])
mean(mult.sim$sims.list$p[,,2])
mean(mult.sim$sims.list$p[,,3])


#density
d <- density(mult.sim$sims.list$phat[1:5000,1])
pdf("theta1-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,1],prob=T,
main="",xlab=expression(theta[1]),col=gray(.5))
lines(d,lwd=2,col="blue")
dev.off()


d <- density(mult.sim$sims.list$phat[1:5000,2])
pdf("theta2-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,2],prob=T,
main="",xlab=expression(theta[2]),col=gray(.5))
lines(d,lwd=2,col="blue")
dev.off()

d <- density(mult.sim$sims.list$phat[1:5000,3])
pdf("theta3-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,3],ylim=c(0,450),prob=T,
main="",xlab=expression(theta[3]),col=gray(.5))
lines(d,lwd=2,col="blue")
```

```
dev.off()


pdf("theta1-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,1],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta2-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,2],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta3-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,3],
type="l",ylab=expression(theta[3]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),3],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),3],type="l",
col="green")
dev.off()
```

Appendix III: Data fit using WinBUGS

```
# Model 1
model
{
        for( i in 1 : 4 ) {
                theta[i] ~ dbeta(alpha[i], sum.alpha[i])
                sum.alpha[i] <-  sum(alpha[1:4])-alpha[i]
                #prior for alpha
                alpha[i] ~  dgamma(0.001,0.001)
                theta.hat[i] <- alpha[i]/sum(alpha[1:4])
        }
}

# Model 2
model
{
        for( j in 1 : nT ) {
                for( i in 1 : nC ) {
        theta[i,j] ~ dbeta(alpha[i,j], sum.alpha[i,j])
        sum.alpha[i,j] <-  sum(alpha[1:4,j])-alpha[i,j]
        #prior for alpha
        alpha[i,j] ~  dgamma(1,1)
        theta.hat[i,j] <- alpha[i,j]/sum(alpha[1:4,j])
                }
        }
}
```

Appendix IV: R2WinBUGS Fitting and Graphics

```
#load library
library(R2WinBUGS)

#change working directory
setwd("D:\\Work\\JKU\\Hold\\phd\\jk")
# look at the data
source("mult.dat")
# Define or generate the parameter in the models of WinBUGS
I <- mult$I
J <- mult$J
X <- mult$X
data <- list ("I", "J", "X")


# Define the initial values by using a function

inits <- function(){
# it's a really weird way to write function
list(alpha = c(NA,rnorm(2, 1,2)), beta = c(NA,rnorm(27,2,1)))
}

inits()

# call WinBUGS
# We first stop on the WinBUGS to see what happened
mult.sim <- bugs(data, inits,
working.directory="D:\\Work\\JKU\\Hold\\phd\\jk",
model.file = "mult.bugs",
# try inits=NULL
parameters.to.save = c("alpha", "beta", "p","phat"), n.thin=1,
# define the thin rate
# n.thin=1, which means that we keep tracking the samples without dropping
# if not setting n.thin, it will drop some terms
n.chains = 3, n.iter = 10000, n.burnin=5000, debug=TRUE,
# burning = 100, n.iter = MCMCsamples(each chain)+n.burin
bugs.directory = "c:/WinBUGS14/")    # debug=TRUE will stop in WinBUGS
#rather than only shows the results


## How to fix multiple initial values
# See what is in the mult.sim
ls(mult.sim)

# results
print(mult.sim, dig=3)
plot(mult.sim)
```

```
# MCMC samples
ls(mult.sim$sims.list)

# mu.theta as example
hist(mult.sim$sims.list$alpha[,1],prob=TRUE)
hist(mult.sim$sims.list$alpha[,2],prob=TRUE)

op <- par()
par(mfrow=c(1,1))

setwd("D:\\Work\\JKU\\Hold\\phd\\jk\\2020-September")

pdf("hist-phat1.pdf")
par(mfrow=c(1,1))
hist(mult.sim$sims.list$phat[,1],prob=TRUE,xlab=expression(p[1])
,main="")
dev.off()

pdf("hist-phat2.pdf")
par(mfrow=c(1,1))
hist(mult.sim$sims.list$phat[,2],prob=TRUE,xlab=expression(p[2])
,main="")
dev.off()

pdf("hist-phat3.pdf")
par(mfrow=c(1,1))
hist(mult.sim$sims.list$phat[,3],prob=TRUE,xlab=expression(p[3])
,main="")
dev.off()

# plot with the prior distribution as line
#prior_sample <- rnorm(1000,0,10)
#lines(density(prior_sample))

dim(mult.sim$sims.list$alpha)
dim(mult.sim$sims.list$beta)
dim(mult.sim$sims.list$phat)

dim(mult.sim$sims.list$p)

sapply(mult.sim$sims.list$phat, mean)

par(mfrow=c(1,3))
boxplot(mult.sim$sims.list$p[,1,1])
boxplot(mult.sim$sims.list$p[,1,2])
boxplot(mult.sim$sims.list$p[,1,3])

boxplot(cbind(mult.sim$sims.list$p[,2,1],
```

```
mult.sim$sims.list$p[,2,2]))

boxplot(mult.sim$sims.list$p[,,1])
boxplot(mult.sim$sims.list$p[,,2])
boxplot(mult.sim$sims.list$p[,,3])

#par(mfrow=c(1,3))
pdf("trace-phat1.pdf")
ts.plot(apply(mult.sim$sims.list$p[,,1],1,mean),col="red",
lwd=2,ylim=c(0.505,.525),ylab=expression(p[1]))
dev.off()

pdf("trace-phat2.pdf")
ts.plot(apply(mult.sim$sims.list$p[,,2],1,mean),col="blue",
lwd=2,ylim=c(.4125,.4325),ylab=expression(p[2]))
dev.off()

pdf("trace-phat3.pdf")
ts.plot(apply(mult.sim$sims.list$p[,,3],1,mean),col="green",
lwd=2,ylim=c(0.0575,.065),ylab=expression(p[3]))
dev.off()

pdf("theta1-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,1],
type="l",ylab=expression(theta[1]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),1],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),1],type="l",
col="green")
dev.off()

pdf("theta2-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,2],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta3-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,3],
type="l",ylab=expression(theta[3]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),3],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),3],type="l",
```

```
col="green")
dev.off()

mean(mult.sim$sims.list$p[,,1])
mean(mult.sim$sims.list$p[,,2])
mean(mult.sim$sims.list$p[,,3])

#summary statistcics

#density

d <- density(mult.sim$sims.list$phat[1:5000,1])
pdf("theta1-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,1],prob=T,
main="",xlab=expression(theta[1]),col=gray(.5))
lines(d,lwd=2,col="blue")
dev.off()


d <- density(mult.sim$sims.list$phat[1:5000,2])
pdf("theta2-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,2],prob=T,
main="",xlab=expression(theta[2]),col=gray(.5))
lines(d,lwd=2,col="blue")
dev.off()

d <- density(mult.sim$sims.list$phat[1:5000,3])
pdf("theta3-density.pdf")
hist(mult.sim$sims.list$phat[1:5000,3],ylim=c(0,450),prob=T,
main="",xlab=expression(theta[3]),col=gray(.5))
lines(d,lwd=2,col="blue")
dev.off()

pdf("theta1-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,1],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta2-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,2],
type="l",ylab=expression(theta[2]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),2],type="l",
col="red")
```

```
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),2],type="l",
col="green")
dev.off()

pdf("theta3-hist.pdf")
par(mfrow=c(1,1))
plot(1:5000,mult.sim$sims.list$phat[1:5000,3],
type="l",ylab=expression(theta[3]),xlab="Iteration")
lines(1:5000,mult.sim$sims.list$phat[5000+(1:5000),3],type="l",
col="red")
lines(1:5000,mult.sim$sims.list$phat[2*5000+(1:5000),3],type="l",
col="green")
dev.off()
```

Appendix V: Diagnostic using CODA

```
library(coda)
setwd(D:/Work/JKU/Hold/phd/jk/docs")
source("coda-add.r") # load the additional functions
chain1=read.coda("coda1.out", "coda1.ind",start=3001,end=12000,
thin=1)
chain2=read.coda("coda2.out", "coda1.ind",start=3001,end=12000,
 thin=1)


# Geweke Diag
geweke.diag(chain1)
geweke.diag(chain2)
geweke.plot(chain1)
geweke.plot(chain2)

# Gelman Diag
gelman.diag(list(chain1,chain2))
gelman.plot(list(chain1,chain2))

# density plot

pdensplot(list(chain1,chain1),"mu",main=expression(mu))
pdensplot(list(chain1,chain1),"theta[1]",
main=expression(theta[1]))


# histogram of the half of the chain
par(mfrow=c(2,2))
hist.chain(chain1,"mu","first",main=expression(mu*" (Chain 1)"))
hist.chain(chain1,"mu","second",main=expression(mu*" (Chain 1)"))

hist.chain(chain2,"mu","first",main=expression(mu*" (Chain 2)"))
hist.chain(chain2,"mu","second",main=expression(mu*" (Chain 2)"))


# qqplot of the first and the second half of the chain
par(mfrow=c(1,1))
qq.chain(chain1,"mu",main=expression(mu))
```