A Bayesian multilevel model for women unemployment in South Africa

Ramarumo VP

(Student No: 14001393)

submitted in fulfilment of the requirements for Masters of Science degree in Statistics

in the

Department of Statistics

School of Mathematical and Natural Sciences

University of Venda, Thohoyandou, Limpopo

South Africa

Supervisor: **Dr. A. Bere**

Co-Supervisor: **Dr. Caston Sigauke**

Date Submitted: August 2, 2021

Abstract

The study is aimed at investigating and explaining the demographic and socio-economic determinants components affecting women unemployment in South Africa. The classical and the Bayesian estimation approach were applied to a multilevel logistic regression (MLR) model. Secondary data acquired from the Demographic and Health survey (DHS) held in South Africa in 2016 was used in the study.

Information criteria revealed that the random intercept model outperformed the MLR model of the null and random coefficient multilevel models. The Intraclass Correlation Coefficient (ICC) proposes that there is an understandable difference in women unemployment level over various provinces of South Africa. The results of the classical MLR and the Bayesian MLR indicate inflated commonness for women unemployment and the chance of being without employment for women was established to decrease with an increase of age, wealth index, and educational attainment.

Keywords: *A Bayesian inference, A multilevel logistic regression, Provincial Variations, unemployment.*

Declaration

I, Ramarumo V.P, [student Number: 14001393], hereby declare that the dissertation titled: "A Bayesian multilevel model for women unemployment in South Africa" for the Master of Science degree in Statistics at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signed (Student): .... Date: 2/8/2021 ....

Acknowledgment

I would like to express my special thanks of gratitude to my supervisor (Dr A. Bere) as well as my co-supervisor (Dr C. Sigauke) for giving me an opportunity to explore the drivers to high women unemployment in South Africa, this has helped me to gain more knowledge in research, and for their constant support thoughtout this period, I am really thankful to them. Secondly I would also like to thank my mother (Langanani A.E.), siblings and colleagues (Ratshilengo M and Nematswerani P) who helped me a lot in finalizing this project within the limited time frame. Special thanks to the CSIR DST-Interbursary Support (IBS) for the Financial support.

Dedication

This work is dedicated to my mother Langanani A.E.

Contents

List of Figures

List of Tables

Abbreviations

MLR	Multilevel Logistic Regression
SADHS	South African Demographic and Health Survey
MCMC	Markov Chain Monte Carlo
GLMM	Generalised Linear Mixed Models
HIV	Human Immunodeficiency Virus
AIDS	Acquired Immuno Deficiency Syndrome
BIC	Bayesian Information Criterion
AIC	Akaike Information Criterion
MLE	Maximum Likelihood Estimate
LMM	Linear Mixed Models
ICC	Intraclass Correlation Coefficient
DHS	Demographic and Health Survey

Chapter 1

Introduction

1.1 Background

A multilevel model is used in analysing data that consist of a hierarchical or clustered structure. It permits the modelling of data measured at different levels at the same time. Such data arise in different disciplines, for example, within educational research whereby learners are embedded in schools, in family studies whereby women are embedded in families (Austin, 2010). Individuals within the same families or clusters often display a significant level of correspondence or uniformity of the effects in contrast to arbitrarily chosen individuals from various families or clusters. Due to the absence of independence of individuals in the same family or cluster, a customary statistical approach might not be suitable in analysing such nested data. This necessitates the use of multilevel models.

The multilevel regression model can also be referred to as hierarchical model, random coefficient model, variance component model (random effect model), and mixed (linear) model. In most cases, it presumes hierarchical data, with one response variable estimated at the low level and the explanatory variables estimated at all existing levels. The estimation of parameters for multilevel models can be done using either the frequentist (classical) approach or a Bayesian approach.

The difference linking the Bayesian inference and frequentist inference is established against that parameters are treated as arbitrary variables within the Bayesian inference. In Bayesian inference we use data to modernise prior insight regarding parameters as well as functions of certain parameters. We are also likely to need model predictions and these are provided as part of the updating process. Prior knowledge about parameters and updated (or posterior) knowledge about them, as well as implications for functionals and predictions, are expressed in terms of densities, (Congdon, 2005). Bayesian methods are more pliable and easy to approximate parameters of composite hierarchical models (Kruschke and Vanpaemel, 2015). When analysing multivariate categorical data, the Bayesian approach has major benefits above a Quasi-likelihood and a likelihood-based classical approach. The use of MCMC (Markov Chain Monte Carlo) methods to acquire the approximation of the exacted updating (posterior) distributions, no requirement to depend on huge sample foundations, excluding about the number of the MCMC iterations, which are likely to increase easily.

In this dissertation, the multilevel logistic regression model will be constructed for unemployment among South African women. A Bayesian approach will be mainly used for parameter estimation.

1.2 Definition of unemployment

Unemployed people are explained via the bureau of labor statistics **?** as individuals who are currently not working, busy looking for employment in the past month and are currently available for employment. Also, individuals who are dismissed for the time being and are on stand-by to be called back to work are included within the unemployment statistics. Unemployed people are also explained by the International Labour Organisation **?** as individuals who are not in labor for pay or profit, who are looking and are ready to begin to work for pay or profit in specified reference periods. Unemployment consists of four groups, namely, structural, demand deficiency, voluntary, and frictional unemployment.

The demand deficiency is the largest source of unemployment that occurs during a recession. This occurs when demand for products or services is low, whereby the company production is reduced and also the workers are being reduced. The frictional unemployment is referred to as workers intermediary jobs. For example, an employee who has currently left or got dismissed and is searching for another job in a financial system in which recession is not taking place. Structural unemployment occurs as soon as there's a mismatch of skills. Such a scenario arises when the skills that the available job demands do not match the skills that the worker has or if the geographical location of the job cannot be reached by the worker. Lastly, voluntary unemployment occurs when the employee decides to leave the employment because it's not financially fulfilling. This study uses the definition given by the SADHS which defined an unemployed person as a person who does not have work, regardless of whether that person was searching for employment or not.

1.3 Problem statement

South African employment market is beneficial to the males than it is to the females and the males are in a higher degree of being in paid jobs than the females, no matter the ethnicity as stated by the Quarterly Labour Force Survey published through the Stats SA 2018. The Quarterly Labour Force Survey published through the Stats SA 2018 further stated that, the formal unemployment rate has been remarkably inflated after 2008. Within the previous 10 years, the rate has increased from 23,2% in the first quarter of 2008 to 27.2% within the 2nd-trimester of 2018. The enlarged unemployment rate, which does not consider job hunt was even bigger and increased from 30.9% within 2008 to 37.2% within 2018. Throughout this period, both of these rates were higher amongst women than men.

The unemployment rate amongst the females was 29.5% within the second trimester of 2018 as compared to 25,3% amid the males, consistent with the formal definition of unemployment. According to the expanded definition, the rate of unemployment amongst women was 7.5 percentage points higher than that of males. There is a need to investigate factors leading to unemployment amongst women in South Africa. A logistic regression model will be used for this purpose.

The Demographic and Health Survey (DHS) data will be employed in this study. A familiar attribute of the DHS data is whereby individuals are clustered in the same families, communities, and provinces. As a result, the multilevel models which are models that grasp into consideration the existing data hierarchies can be used because these models allow for a residual unit at each measure of the hierarchy. Residual components account for the variation due to an unobserved source of variation applicable to that level of the hierarchy. The ignorance of the unobserved heterogeneity can lead to inconsistency of parameter estimates, wrong standard errors, (Nicoletti and Rondinelli, 2010).

Many studies were conducted on unemployment using a logistic regression model, Theodossiou (1998); Viljoen and Dunga (2013); Baah-Boateng (2008); Kingdon and Knight (2004); Fergusson et al. (2001); Lundin et al. (2012); Baah-Boateng et al. (2013); Ford et al. (2010); Kessler et al. (1987); Latif (2010); Drydakis (2015). There are very few studies in which the multilevel structure of the data was considered. Examples are Bertoni and Ricchiuti (2017) and Chikako (2018). The current study will apply a multilevel logistic regression model.

The estimation of parameters may be done by either the classical (frequentist) or Bayesian estimation approaches. Hammer (1997), Page et al. (2013), De Moortel et al. (2018), Wege et al. (2017) used the classical (frequentist) approach in their studies. Few studies were conducted which used a Bayesian estimation approach, Chikako (2018);

Phillips (2004); Kiros and Abebe (2019). The current study will focus on the Bayesian estimation approach and it differs from the above studies since the South African data set will be used. The frequentist approach will also be considered for comparison purposes.

There are many advantages to be derived from the use of Bayesian approach: For big sample proportions, the Bayesian method with non-informative priors give effects of parametric models that are the same as the results from the classical methods. The Bayesian method is more interesting as compared to the classical method as it updates prior information with the data, allows compact samples to be analysed, and strong approximations to be made (Ardia and Hoogerheide, 2009). The Bayesian method relies not on the asymptotic estimation Coles et al. (2001); Bayarri and Berger (2004); Beirlant et al. (2006). Also, the Bayesian modelling structure complies with the likelihood concept which says that for a certain specified piece of information, every pair of feasibility models which possess similar likelihood function gives similar inference (Gelman and Meng, 2004). A Bayesian analysis yields a reasonable modelling structure to find a solution for problems under a state of unpredictability.

1.4 Aim and objectives

1.4.1 Aim

The main aim of this research is to investigate the demographic and the socio-economic causation components of women unemployment in South Africa (through a Bayesian and classical estimation approach applied to a multilevel logistic regression model.

1.4.2 Objectives

The objectives of this study are to

1. Identify demographic and social economic factors leading to women unemployment in South Africa.
2. Investigate the significance of 'province' as a differential factor of unemployment of women in South Africa.
3. Compare the results from a classical and a Bayesian approach.

1.5 Data source

This study will use secondary data acquired from the SADHS 2016. The data give thorough details on marriage, fertility, birth control, infant, child, adult and maternal mortality, maternal and child health, gender, nutrition, malaria, knowledge of HIV/AIDS, employment status, and other sexually transmitted diseases.

1.6 Study variables

The women's unemployment level is the response variable. It specifies if one is employed or unemployed. The independent parameters to be considered as factors of the women unemployment in this study are as follows: province, marital status, age group, literacy status, women educational level, type of place of residence, wealth index, and race.

1.7 Proposed methodology

A Bayesian MLR model will be utilised to evaluate the relationship linking women's unemployment status and selected socioeconomic components. All analysis will be done using R-statistical software (Team et al., 2013).

1.8 Significance of the study

The demographic and socio-economic factors provide knowledge regarding the risk components for women's unemployment. The current research could be utilised as a reference for future research associated with unemployment and others and it may give some guidance to the government and other responsible parties in the formation of laws and strategies for reducing unemployment.

1.9 Dissertation layout

This dissertation is organised into chapters, each chapter is a stand-alone chapter. Chapter 2 gives a literature review on the potential determinants of unemployment and the statistical models used in studying the unemployment. It also covers the definition of unemployment. Chapter 3 covers the detailed methodology used. This covers the multilevel logistic regression models for both the classical and the Bayesian Inferences. Chapter 4 covers the results for the methodology built-in chapter 3 and the discussion of the results. Chapter 5 presents the conclusion, recommendations, and limitations of the research.

Chapter 2

Literature review

2.1 Introduction

The current chapter gives details of what other researchers have done concerning the determinants of unemployment, statistical methods applied to model unemployment, and prior selection in logistic regression.

2.2 Definition of unemployment

The international labour organisation defines an unemployed individual as an individual aged 15 and above, who does not possess any job position and is searching for one ?. This individual should match the following states: (a) currently not working, this means that the person has not been working for a minimal of 60 minutes in the weekly reference, (b) the person should be ready to get hold of employment position in 14 days, (c) diligently looked for a job in the preceding four weeks or have got one commencing in the upcoming trimester. This study uses the definition given by the SADHS which defined an unemployed person as a person who does not have work, regardless of whether that person was searching for employment or not.

2.3 Potential determinants of unemployment

It was reported in Pakistan by Qayyum and Siddiqui (2007) that the unemployment rate was found to be higher among a females than a males. Similar results were reported in South Africa Gyekye and Kyei (2011); Mathebula (2017) and also in Ghana by Abraham et al. (2017). In a similar study conducted by Mathebula (2017) in South Africa, it was established that a Black African woman and Indian woman had greater chances of being unemployed as compared to their male counterparts. The author also confirmed that age was significantly associated with unemployment since the youth unemployment rate was twice that of the adults. In Pakistan, Qayyum and Siddiqui (2007) also concluded that employability advances with work experience, so recent qualified graduates were not taken into the labour markets caused by the high demand for work experience. Dagume and Gyekye (2016) made recommendations such that the South African authority should build up the labour laws and the strategies to permit the youth to gain more working experience and training since youth with some training was absurd to be without the employment than youth without the training.

In South Africa, marital status was established as one of the significant factors associated with unemployment Yakubu (2010); Mathebula (2017); Viljoen and Dunga (2013). The divorced, never married and widowed persons were found less likely to be without employment as compared to the married persons, the reason being that they had little or no prospects of economic dependence Yakubu (2010); Mathebula (2017); Viljoen and Dunga (2013). Similar observations were also reported in Ghana by Abraham et al. (2017). It was reported in South Africa by Yakubu (2010); Mathebula (2017) that the odds of unemployment for individuals with primary and lower education were found to be greater than those for individuals with tertiary education, meaning that those with tertiary education had a greater chance of employment than those with primary and lower educational attainment. Similar observations were also reported in North Cyprus, Turkey by Usman and Sanusi (2016). However, Abraham et al. (2017) from Ghana argue that there was no evidence that suggested that obtaining higher education increased the chance of being employed as the above-mentioned authors suggested. This argument was further supported by Sackey (2005) also from Ghana, the author concluded that there was no notable contrast between a primary and a post-primary educational obtainment as far as labour force engagement is concerned.

Conclusions have been made in Nigeria by Iweagu and Chukwudi (2015) regarding religion as it was found to be one of the significant determinants of female labour participation. The women who reported that they were the Muslims or the Traditionalists had a lower probability of participating in labour force. It was concluded that the regions with more Muslims and Traditionalist would have a lower probability of women participating in labour force. This was expected since those religious groups give priority to the men than to the women in most of the aspects of life. It was also observed in their study that the female's literacy rate for those residing in the countryside areas decreased the chance that a woman would participate in the workforce. The explanation for this was that due to lack of the available jobs in the countryside areas, people are no longer working according to their qualifications and the available jobs don't even require any qualifications.

In Pakistan, Qayyum and Siddiqui (2007) reported a high unemployment rate in the city areas as compared to the countryside areas. The reason being that there was a mismatch between the skills provided and the skills demanded. However, in Mauritius, it was reported by Tandrayen-Ragoobur and Kasseeah (2015) that unemployment was not associated with the place of residence. Whether a person was located in the rural areas or the urban areas never influenced the chance of engaging in the workforce. In the study conducted in Nigeria by Iweagu and Chukwudi (2015) the household size was found to be insignificant in both the urban and the rural areas. The authors found it strange and different from other works conducted, (Viljoen and Dunga, 2013) which showed that the household size was a determinant of female labour participation. It was also discovered in South Africa (Limpopo, Vhembe district) by (Dagume and Gyekye, 2016) that age, gender, marital status, educational attainment were unimportant to determine youth unemployment. The contrasting results in terms of the significance of some factors give justification for research into the determinants of unemployment using data from a variety of settings.

2.4 Statistical models used in studying unemployment

The most widely utilized model to investigate the factors leading to unemployment is the logistic regression model. In South Africa, Dagume and Gyekye (2016), Viljoen and Dunga (2013), Mathebula (2017) used a logistic regression model to establish the socio-economic and demographic factors which influenced unemployment. The same model was used by Iweagu and Chukwudi (2015), Usman and Sanusi (2016), Taamouti and Ziroili (2011), Tandrayen-Ragoobur and Kasseeah (2015), Che and Sundjo (2018), Kessler et al. (1987), Baah-Boateng (2015), Lundin et al. (2012), Fergusson et al. (2001) respectively.

The analyses were made in Pakistan by Qayyum and Siddiqui (2007) using a probit regression model to understand the determinants of female unemployment. This model was also used in Ghana and in South Africa by Sackey (2005), Kingdon and Knight (2004), Lam et al. (2007). The authors like Banerjee et al. (2008) employed a multinomial logit model to conclude that educational level, race, gender, and age were significantly associated with unemployment.

With the use of the multilevel models being rare in South Africa and the other countries as well, regarding unemployment, only one study was found conducted by Bertoni and Ricchiuti (2017) in Egypt. The authors applied a multilevel generalised linear mixed model, logit link function was selected. There is also a rare use of a Bayesian estimation approach. Few studies were conducted in Ethiopia by Kiros and Abebe (2019) and Chikako (2018). However, in a study conducted by Kiros and Abebe (2019) the multilevel structure of the data was not considered. The main contrast linking frequentist and Bayesian inference is the introduction of prior knowledge in probability distributions structure. To attain the posterior distribution of the parameters in a model, and to draw a conclusion regarding the posterior parameters, the prior distribution is included in the model. Kiros and Abebe (2019) used a commonly dispersal non-informative prior, which is a familiar prior for logistic regression coefficients beta. Chikako (2018) used inverted gamma prior to the random effect and normal distribution prior to the fixed effect. Both studies used non-informative priors for the perception that no value is recommended over any other value. The current study will follow the work done by Chikako (2018), whereby the author used a multilevel model for which the parameters have been approximated utilising a Bayesian method.

2.5 Conclusion

Concerning the literature cited above, there was no study that had been done on women unemployment using a Bayesian estimation approach in South Africa. Only a few studies were conducted using the multilevel logistic regression model. The current study will employ the multilevel logistic regression model to model women unemployment in South Africa. Model parameters will be approximated through the Bayesian method.

Chapter 3

Methodology

3.1 Introduction

This chapter covers a more detailed theoretical background of the proposed methodology. The logistic regression model, multilevel model, and parameter estimation approach (both the Classical and the Bayesian approach)

3.2 The logistic regression model

In this section, the theory behind logistic regression modelling is being discussed, following the work by Chatfield et al. (2010). Introduction to the exponential family of distributions and the definitions of the generalized linear models are presented.

Considering a single random variable Y whose probability distribution depends on a single parameter θ, the distribution is part of the exponential family on conditional that it may be denoted as

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)}, \tag{3.1}$$

where a, b, s and t are familiar functions. The symmetric relationship linking y and θ is asserted on conditional that the equation 3.1 is denoted as follows;

$$f(y;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)], \tag{3.2}$$

where $s(y) = \exp[d(y)]$ and $t(\theta) = \exp[c(\theta)]$.

If $a(y) = y$, the distribution is within the canonical (standard) structure and $b(\theta)$ is known as the natural parameter of the distribution. In the presence of other parameters than parameters of concern θ, those parameters are known to be the nuisance parameters that form a portion of the functions a, b, c, and d, are regarded as known.

Now let's assume $E(Y_i) = \mu_i$. Here μ_i exist as a function of θ_i. In the generalised linear model, μ_i is transformed to an extent that $g(\mu_i) = \mathbf{X_i^T}\boldsymbol{\beta}$, here g is known as a monotone, that is a contrasting function named the link function; $\mathbf{X_i}$ is a $p \times 1$ vector of independent variables,

$$\mathbf{X_i} = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix} \tag{3.3}$$

so,

$$\mathbf{X_i^T} = \begin{bmatrix} x_{i1} & \dots & x_{ip} \end{bmatrix} \tag{3.4}$$

and $\boldsymbol{\beta}$ is known as the $p \times 1$ vector of parameters $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$. The vector $\mathbf{X_i}$ is known as the i^{th} column of the design matrix $\mathbf{X}$.

A generalised linear model consists of three components;

1. Dependent parameters $Y_1, \dots, Y_N$ are assumed to share the identical distribution from the exponential family usually referred to as the random component;
2. A set of parameters $\boldsymbol{\beta}$ and explanatory variables $\mathbf{X} = \begin{bmatrix} \mathbf{X_1^T} \\ \vdots \\ \mathbf{x_N^T} \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1p} \\ \vdots & \vdots & \\ x_{N1} & & x_{Np} \end{bmatrix}$ known as the systematic component;

3. A monotone link function g such that $g(\mu_i) = X_i^T\boldsymbol{\beta}$. Here $\mu_i = E(Y_i)$ referred to as a link function.

A logistic regression model is one kind of the generalised linear model utilised when the response variable follows a binomial distribution. The response variable for the current study, employment status was categorised into two categories: employed and unemployed. Let Y_i

$$\mathbf{Y} = \begin{cases} 1, & \text{if a woman is unemployed} \\ 0, & \text{if a woman is employed ,} \end{cases}$$

where $i = 1, \ldots, n$. Let $P(Y_i) = \pi_i$. The general logistic regression model expresses π_i as a function of the parameters through an equation of the form

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \boldsymbol{x_i^T\beta}. \tag{3.5}$$

Equation 3.5 implies that $\pi_i = g(\boldsymbol{x_i^T\beta})$. The maximum likelihood approximates of the parameters β are acquired by maximising

$$\log(\pi_\mathbf{1}, \ldots, \pi_\mathbf{N}; \mathbf{y_1}, \ldots, \mathbf{y_N}) = \left[\sum_{\mathbf{i=1}}^{\mathbf{N}} \mathbf{y_i} \log\left(\frac{\pi_\mathbf{i}}{\mathbf{1}-\pi_\mathbf{i}}\right) + \mathbf{n_i}\log(\mathbf{1}-\pi_\mathbf{i})\log\binom{\mathbf{n_i}}{\mathbf{y_i}}\right] \tag{3.6}$$

3.3 The multilevel model

3.3.1 The linear multilevel model

A multilevel regression model is known as the hierarchical linear model, the random coefficient model, the variance component, and the mixed model. This model presumes the hierarchical data, with the response variable measured at the lowest level and the explanatory variables at all the existing levels. Two levels are considered for the current study, that is, the respondent and the provincial levels. The measurements on the response variable are expressed as $\boldsymbol{Y_{ij}}$, here $\boldsymbol{i}$ is known as the respondent, and $\boldsymbol{j}$ is known as the provincial level. Following the fundamental concepts of the multilevel modelling given by Goldstein (2011), a 2-level multilevel model can be expressed generally as

$$\boldsymbol{y} = \alpha + \beta x + e, \tag{3.7}$$

where $\boldsymbol{j}$ makes reference to the level 2 component and $\boldsymbol{i}$ to the level 1 component. However, equation 3.7 is a one-level model up to this point. To convert equation 3.7 into a real 2-level model, $\boldsymbol{\alpha_j}$ and $\boldsymbol{\beta_j}$ are allowed to be random variables. To make the notations to be consistent, $\boldsymbol{\alpha_j}$ and $\boldsymbol{\beta_j}$ are replaced by $\boldsymbol{\beta_{0j}}$ and $\boldsymbol{\beta_{1j}}$ respectively and we assume that

$$\boldsymbol{\beta_{0j} = \beta_0 + u_{0j}},$$

$$\boldsymbol{\beta_{1j} = \beta_1 + u_{1j}},$$

where $\boldsymbol{u_{0j}}$ and $\boldsymbol{u_{1j}}$ are arbitrary variables with parameters $\boldsymbol{E(u_{0j}) = E(u_{1j}) = 0}$ and $\boldsymbol{var(u_{0j}) = \sigma^2_{u0}}$, $\boldsymbol{var(u_{1j}) = \sigma^2_{u1}}$, $\boldsymbol{cov(u_{0j}, u_{1j}) = \sigma_{u01}}$. Now, equation 3.7 may be given by;

$$\boldsymbol{y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + u_{1j} x_{ij} + e_{0ij})}, \tag{3.8}$$

where $\boldsymbol{var(e_{0ij}) = \sigma^2_{e1}}$. The response $\boldsymbol{y_{ij}}$ is demonstrated as the summation of a fixed proportion and a random portion in the brackets. The fixed portion of equation 3.8 may be expressed generally in matrix form

$$\boldsymbol{E(Y) = X\beta,}$$

with $Y = \{y_{ij}\}$

$$\boldsymbol{E(y_{ij}) = X_{ij}\beta = (X\beta)_{ij}, X = \{X_{ij}\}},$$

where $\{\}$ denotes a matrix, $\boldsymbol{X}$ is the design matrix for the independent variables and $\boldsymbol{X_{ij}}$ is the ij^{th} row of $\boldsymbol{X}$. For this equation 3.8 it is given as

$$\mathbf{X} = \begin{Bmatrix} \mathbf{1} & \mathbf{x_{ij}} \end{Bmatrix} \tag{3.9}$$

In the multilevel regression analysis, two various likelihood functions are commonly utilised: the full maximum likelihood and the restricted maximum likelihood. The restricted maximum likelihood magnifies a likelihood function that is constant for the fixed effects whereas the full maximum likelihood magnifies both the fixed and the random part. The part of variance within the inhabitants defined by the grouping structure is expressed by the ICC $\boldsymbol{\rho}$. $\boldsymbol{\rho}$ is estimated by the model which comprises no independent variables referred to as the intercept-only model

$$\boldsymbol{Y_{ij} = \gamma_{00} + u_{0j} + e_{ij}} \tag{3.10}$$

by the use of the model above, the ICC (Intraclass Correlation Coefficient) $\boldsymbol{\rho}$ is approximated through equation

$$\boldsymbol{\rho = \frac{\sigma^2_{uo}}{\sigma^2_{uo} + \sigma^2_e}}, \tag{3.11}$$

where $\boldsymbol{\sigma^2_{uo}}$ is known as the variance of the 2-unit residuals $\boldsymbol{u_{0j}}$ and $\boldsymbol{\sigma^2_e}$ is known as the variance of level 1 residual $\boldsymbol{e_{ij}}$. The following subsection will be covering the multilevel binary logistic model.

3.3.2 Multilevel binary logistic

Let's denote the response of person $\boldsymbol{i}$ in class j be encoded as $Y_{ij} = 1$ for the responses in the target class (unemployed women) and $Y_{ij} = 0$ for response in other class (employed women). Level one logistic regression model for the likelihood that the person $\boldsymbol{i}$ within-cluster $\boldsymbol{j}$ has a response in the target class (employed women) is

$$P(Y_{ij} = 1) = \frac{\exp(\beta_{0j} + \beta_1 x_{1ij} + \ldots + \beta_p x_{pij})}{1 + \exp(\beta_{0j} + \beta_1 x_{1ij} + \ldots + \beta_p x_{pij})}, \tag{3.12}$$

where x_{pij} is the value of the p^{th} level one predictor variable for person $\boldsymbol{i}$ in group $\boldsymbol{j}$, $\boldsymbol{\beta_{0j}}$ is the intercept for group $\boldsymbol{j}$, and $\boldsymbol{\beta_p}$ is the regression coefficient for $\boldsymbol{x_{pij}}$ in group $\boldsymbol{j}$.

3.3.3 Random intercept model

This model is utilised to evaluate the hidden variety within the general response by presenting the random effects. Within the random intercept model, the only random effect is the intercept which implies these classes vary regarding a common value regarding a response variable, but the link connecting the independent and also the dependent variables can't vary among the classes. A random intercept model indicates the log-odds, for example, the logit of π_{ij} , as a quantity of a linear function of the independent variables. such that,

$$\boldsymbol{\log(\pi_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_{2ij} + \ldots + \beta_k x_{kij}, i = 1, 2, \ldots n, j = 1, 2, \ldots J}, \tag{3.13}$$

where the intercept term $\boldsymbol{\beta_{0j}}$ is presumed to differ randomly and is denoted by the summation of mean intercept $\boldsymbol{\beta_0}$ and group-response random errors $\boldsymbol{U_{0j}}$, such that $\boldsymbol{\beta_{0j} = \beta_0 + U_{0j}}$. This results in

$$\text{logit}(\boldsymbol{P_{ij}}) = \boldsymbol{\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + U_{0j}}, \tag{3.14}$$

where $\boldsymbol{\beta_0 + \sum_{h=1}^{k} \beta_h x_{hij}}$ is referred to as the fixed portion of the model, whereas the last part $\boldsymbol{U_{0j}}$ is known as the random part of the model. The residual $\boldsymbol{U_{0j}}$ assumed to be mutually independent and normally distributed with mean of zero and variance $\boldsymbol{\sigma_0^2}$.

3.3.4 Random coefficient model

This model expands from the random intercept model by permitting the results of individual forecasts to differ randomly over level two, such that, level one slope coefficients are permitted to employ varying values within varying cluster collections, (Chikako, 2018). Within the random coefficient model the intercepts, as well as the slopes, are permitted to vary over the province; Which is denoted as,

$$\log(\pi_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) + \beta_0 + \sum_{h=1}^{k} \beta_h x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{1j} X_{1ij} \tag{3.15}$$

In the current study, the parameter approximation can be done by the maximum likelihood approximation and the Bayesian approximation approach.

3.4 Frequentist parameter estimation

3.4.1 Maximum marginal likelihood approximation

In the current section, a two-level model is considered;

$$g(\pi_{ij}) = x_{ij}^T \beta + Z_{ij}^T u_j, \tag{3.16}$$

where $x_{ij} = p[y_{ij} = 1|u_j], (j = 1, \ldots, J, i = 1, \ldots, n)$ and $u_j \propto N(0, \Omega)$. With independence assumption, the conditional probability of component j (second level) extracts the binomial structure. The log minor probability can be acquired by merging across the random effects and may be given as follows;

$$\ell_j(\beta, \Omega) = \log \int \prod_{i=1}^{n_j} \pi_{ij}^{y_{ij}} (1 - \pi_{ij})^{1-y_{ij}} \phi(u_j, \Omega) du_j, \ldots, \tag{3.17}$$

where $\phi(u_j, \Omega)$ is known as the normal density function $N(0, \Omega)$. The log marginal probability

$$\ell(\beta, \Omega) = \sum_{j=1}^{n} \ell_j(\beta, \Omega), \ldots$$

can be maximized to get β and Ω estimates utilising any quality optimisation procedures. There is a need to do the numerical integration since equation 3.17 is intractable. The integral can be evaluated using the Gauss-Hermite quadrature. The technique is known to work well if the dimension of the integration is small.

To reduce the computational load which is brought by numerical integration requirements, several approaches had been proposed. The approaches involve the marginal quasi-likelihood (MQL) and penalised quasi-likelihood (PQL). The PQL makes use of Laplace's integral estimation. The PQL and MQL are considered as repetition approaches. These procedures require fitting the linear multilevel models conditional on the 1^{st} order Taylor extension for mean function regarding the contemporary fixed portion predictor (MQL) or the contemporary forecast value (PQL). Even though the above-estimated techniques are computational coherent than the maximum likelihood procedures they have got some shortcomings. These procedures may produce results that are biased when the number of simulations is quite large.

3.4.2 The Laplace Approximation

Considering the model $\boldsymbol{g(\pi) = X\beta + Zu}$ with $\boldsymbol{u \propto N(0, \Omega_\theta)}$, here $\boldsymbol{\Omega}$ is known as the $\boldsymbol{q} \times \boldsymbol{q}$ dimension variance-covariance matrix. Even though $\boldsymbol{\Omega}$ remains a big matrix, it's set on through a variable vector $\boldsymbol{\theta}$, of which the proportion is generally compact.

In this section, the interest is in obtaining the estimate of $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ which maximises the probability of variables $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, provided data $\boldsymbol{y}$. The likelihood correspond numerically to the marginal density of $\boldsymbol{y}$, provided $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$;

$$\boldsymbol{f(y|\beta, \theta) = \int_u p(y|\beta, u) f(u|\Omega(\theta)) d\theta}, \tag{3.18}$$

where $\boldsymbol{f(y|\beta, \theta)}$ is known to be the probability mass function of $\boldsymbol{y}$, $\boldsymbol{\beta}$, and $\boldsymbol{u}$ specified, where $\boldsymbol{f(u|\Omega)}$ is the Gaussian probability density at $\boldsymbol{u}$ and $\boldsymbol{p(y|\beta, u)}$ is given as the conditional quantity of $\boldsymbol{y}$. In a case where $\boldsymbol{p(y|\beta, u)}$ is binomial, the above integral do not have a closed form solution. The Laplace approximation is one way to approximate it and find its solution. To get Laplace's estimation to the likelihood $\boldsymbol{L(\beta, \theta|y)}$, the numerical of an integral part in equation 3.18 is replaced with intrinsic 2nd order Taylor sequence at conditional maximum, $\boldsymbol{u(\beta, \theta)}$.

3.5 Model diagnostics and selection

3.5.1 The Deviance

Deviance is one of the widely utilised statistics for the evaluation of the goodness of fit. It is explained as;

$$\boldsymbol{D^2 = 2\{\ln[L_s(\hat{\beta})] - \ln[L_m(\hat{\beta})]\}}, \tag{3.19}$$

with $\ln[L_s(\hat{\beta})]$ defined as the maximised log-likelihood for adapted model and $\ln[L_m(\hat{\beta})]$ is known as a maximised log-likelihood of a full model. Deviance contrast the forecasted values of the fitted model with the ones forecasted by the most absolute model that one could possibly fit. The large value of D^2 shows evidence that there is a lack of fit. D^2 converges asymptotically to a χ^2 distribution under a specific regularity condition, with p degrees of freedom; $D^2 \propto \chi^2(p)$, p is given as the contrast linking the number of variables within the full model and the sum of variables within the considered model. A full model is the biggest feasible model a researcher may fit and it guides to an ideal forecast of the result of interest.

3.5.2 Akaike's Information Criterion (AIC)

The AIC is the extensively utilised procedure for the selection of the model, (Acquah, 2010). The AIC aims to choose a model which reduces the negative likelihood penalised by the aggregate of variables. The AIC is given mathematically by;

$$\boldsymbol{AIC = -2\log(\hat{L}) + 2p}, \tag{3.20}$$

where $\hat{L}$ is known as the probability of the equipped model, and p is the sum of variables within a model. Low AIC values indicate a better fit.

3.5.3 Bayesian Information Criterion (BIC)

The BIC is another method for the selection of a model that measures the establishment linking model fit and complication of the model, (Acquah, 2010). The BIC can be represented mathematically as;

$$\boldsymbol{BIC} = -\mathbf{2}\ln(\hat{\boldsymbol{L}}) + \boldsymbol{p}\ln(\boldsymbol{n}), \tag{3.21}$$

where $\hat{\boldsymbol{L}}$ is the maximized value of the likelihood function, $\boldsymbol{n}$ is the number of recorded observations or sample size and $\boldsymbol{p}$ is the number of approximated parameters by the model.

3.5.4 Diagnostics for multilevel models

(DHARMs) Diagnostics for Hierarchical Regression models, was created by Hartig (2020), it creates readily illustratable residuals of generalised linear (mixed) models which are consistent with values between 0 and 1, and can be illustrated as naturally as the residuals for the linear model. This is attained by a simulation-based method, same as the Bayesian p-value or the parametric bootstrap, which changes the residuals to a consistent scale. Below is the residual plot for the multilevel null model, the random intercept model, and the random coefficient model.

3.6 The Bayesian estimation approach

The difference between Bayesian inference classical inference is that the Bayesian inference treats variables as random parameters and it uses data to incorporate prior knowledge regarding parameters and functions of these parameters. Model predictions are required and are being provided as a portion of the updating strategy. Understanding the prior and the posterior regarding the parameters, as well as the inference for functionals and the predictions, are indicated in terms of densities. To understand the Bayesian inference, one must first understand its basis obtained from basic probability theory, Congdon (2005). Consequently, the conditional probability of $\boldsymbol{A}$ and $\boldsymbol{B}$ is given as

$$\boldsymbol{Pr(A|B) = \frac{Pr(B|A)Pr(A)}{Pr(B)}}.$$

$\boldsymbol{B}$ is replaced with the estimation of $\boldsymbol{y}$, $\boldsymbol{A}$ is replaced with a set of parameter $\boldsymbol{\theta}$ and probabilities are replaced by densities result in the relation

$$\boldsymbol{p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)}}, \tag{3.22}$$

where $\boldsymbol{p(y|\theta)}$ is the likelihood of $\boldsymbol{y}$ under a model and $\boldsymbol{p(\theta)}$ is the prior density or the density of $\boldsymbol{\theta}$ before $\boldsymbol{y}$ is observed. This density expresses the accumulated knowledge about $\boldsymbol{\theta}$. The classical analysis through the maximum likelihood focuses on the likelihood $\boldsymbol{p(y|\theta)}$ without introducing a prior, whereas a fully Bayesian analysis updates the prior information about $\boldsymbol{\theta}$ with the information contained in the data. The denominator $\boldsymbol{p(y) = \int p(y|\theta)p(\theta)d\theta}$ in equation 3.22 defines the marginal likelihood or the prior predictive density of the data and may be set to be an unknown constant $\boldsymbol{c}$. Consequently the posterior inferences about $\boldsymbol{\theta}$ under equation 3.22 can be equivalently stated in the relation

$$\boldsymbol{p(\theta|y) = \frac{p(y|\theta)p(\theta)}{c}}$$

or

$$\boldsymbol{p(\theta|y) \propto p(y|\theta)p(\theta)}$$

which is expressed as posterior is correspondent to the likelihood multiplied by the prior.

3.6.1 The likelihood function

Following a brief description of the likelihood inference for GLMM(Generalised linear mixed model) which is found in Fahrmeir and Tutz (2013), in GLMM it is assumed that the effect y_i is conditionally independent, conditional density for an entity i is written as;

$$f(y_i|\beta, b_i) = \prod_{j=1}^{n_i} f(y_{ij}|\beta, b_i), \tag{3.23}$$

where $f(.)$ is known as a density or probability mass function from the exponential family. The marginal density $f(y_i)$ is determined by integrating the random effects in the conditional density

$$f(y_i) = \int f(y_i|\beta, b)f(b_i|D(\delta))db_i, \tag{3.24}$$

the marginal likelihood for all m entities can be defined as

$$L(\beta, b, \delta) = \prod_{i=1}^{m} \int \prod_{j=1}^{n_j} f(y_{ij}|\beta, b_i)f(b_i|D(\delta))db_i, \tag{3.25}$$

where δ are unknown hyper-parameters which determine the distribution of the random effects covariance matrix $D(\delta)$.

3.7 A Bayesian generalised linear mixed model

The model

The Bayesian GLMMs are an extension of the Bayesian LMM. Bayesian generalized linear mixed models vary from Bayesian linear mixed models in the following ways;

1. The relation linking the response, the parameters, and the random effects are non-linear.

2. The dependent follows a dispersal within the exponential family, not a normal distribution.

Now, assume that the dependents $\{y_{i1}, y_{i2}, \ldots, y_i n_i\}$ in i^{th} group are conditionally independent given the average of the parameters β and the random effects u_i.

Suppose that $y_i = (y_{i1}, y_{i2}, \ldots, y_i n_i)^T$, then a saturated Bayesian GLMM may be given as; $E(y_i|\beta, u_i) = g(X_i\beta + Z_i u_i)$, with $i = 1, \ldots, n_i$, $u_i \propto N(0, \Omega)$, $\beta \propto N(\beta_0, \sigma_0)$, $\Omega \propto W_q^{-1}(\eta, \Psi)$.

Here $g(.)$ is called the inverted link function (in this case, the inverted logit link) and $X_i(n_i \times p)$ and $Z_i(n_i \times q)$ are called the design matrices, $W_q^{-1}(\eta, \Psi)$ is the inverted Wishart dispersal with η and Ψ parameters and the random effects $u_i(1 \times q)$.

Assume that the β_0 and σ_0 hyper-parameters are known. Let $u_i = (u_1, u_2, \ldots, u_q)$, then further presume that the prior distributions are independent, i.e, $f(\beta, \Omega) = f(\beta)f(\Omega)$. The posterior distribution of all parameters can be given as;

$$f(\beta, \Omega, u|y) \propto \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\beta, u_i) f(\beta)\right]\left[\prod_{i=1}^{n} f(u_i|\Omega) f(\Omega)\right]$$

The conditionals for the Bayesian inference are written as;

$$f(\beta|\omega, u, y) \propto \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\beta_j, u_i) f(\beta)\right]$$

$$f(u|\beta, \omega, y) \propto \left[\prod_{i=1}^{n}\prod_{j=1}^{n_i} f(y_{ij}|\beta_j, u_i) f(u_i|\Omega)\right]$$

$$f(\Omega|\beta, u, y) \propto \left[\prod_{i=1}^{n} f(u_i|\Omega) f(\Omega)\right],$$

where,

$$[\Omega|\beta, u, y] \propto W_q^{-1}\left(\eta + E\frac{n}{2}, \Psi + \sum_{i=1}^{n} u_i u_i^T/2\right)$$

The prior distribution

The choice of prior distributions in Bayesian inference is of the greatest significance. Most of the arguments adjoining the utilisation of the Bayesian procedures focus on the prior distribution. The Bayesian estimation can be affected by the choice of a prior $\boldsymbol{f(\theta)}$. A powerful prior possesses more effect on the Bayesian inference. Without any of the prior details, the non-informative prior can be used; $\boldsymbol{f(\theta)} \propto 1$. In the Bayesian inference, the likelihood is given by; $\boldsymbol{L(\theta|y) = f(y|\theta)}$, such that $\boldsymbol{f(\theta|y) \propto L(\theta|y)f(\theta)}$. This implies that for the relationship linking the Bayesian and the likelihood established methods, particularly while a non-informative prior is utilised, the Bayesian inference is reduced to the likelihood inference. Within the process of selecting a prior distribution, given no valid motive to choose a single prior above the other, a conjugate prior distribution can be selected for simplicity.

While using the conjugate prior distribution, the generated posterior distribution is a member of the same family of the distribution as well. The prior distribution $\boldsymbol{f(\theta)}$ is assumed that is conjugate to $\boldsymbol{f(y|\theta)}$ on conditional that the posterior distribution $\boldsymbol{f(\theta|y)}$ belongs to a similar group as the prior distribution $\boldsymbol{f(\theta)}$. Take, for example, the normal (Gaussian family) is an independent conjugate (conjugate of itself), on condition that a normal distribution is utilised, then the derived posterior distribution also belongs to the normal distribution.

Every member within the exponential family possesses a conjugate prior. In most regression models, the multivariate ordinary distribution is chosen for the mean parameters $\boldsymbol{\beta}$ like a prior distribution, presume that $\boldsymbol{\beta \propto N(\beta_0, \Sigma_0)}$ where $\boldsymbol{\beta_0}$ and $\boldsymbol{\sigma_0}$ are regarded as hyper-parameters. The hyper-parameters are also assumed unknown and can lead to additional hierarchy in the model specification, though this is regarded as the last stage in practice. The priors $\boldsymbol{\Sigma_0^{-1} \approx 0}$ or $\boldsymbol{\beta_0 \propto U(-\infty, \infty)}$ are chosen if a non-informative prior is desired for the variance-covariance matrices. The Wishart distribution is normally utilised as a prior distribution.

The Wishart distribution is generalised into 2 distributions, the gamma distribution of multiple dimensions and the χ^2 distribution. The Wishart distribution works well in the approximation of the covariance matrices. For illustration, allow Z to be $n \times p$ matrix, with i^{th} row $Z_i \propto N_p(0, v)$ independently, V is known as $p \times p$ covariance matrix, that is, a positive definite. A probability distribution of $W = Z^T Z$ has a Wishart distribution with n degrees of freedom and is given as $W_p(v, n)$, of which the density function may be given by;

$$f(W) = \frac{\mid W \mid^{(n-p-1)/2}}{2^{np/2} \mid V \mid} \Gamma_p\left(\frac{n}{2}\right) \exp\left(\frac{-1}{2} tr(V^{-1}W)\right), \tag{3.26}$$

where $W > 0$ (positive definite), tr is the trace function, $\Gamma_p(.)$ is the multivariate gamma function denoted as;

$$\Gamma_p\left(\frac{n}{2}\right) = \pi^{\frac{p(p-1)}{4}} \prod_{p}^{j=1} \Gamma\left(\frac{(n+1-j)}{2}\right). \tag{3.27}$$

The wishart distribution $W_p(v, n)$ has a mean nV and a mode $(n-p-1)v$ for $n \geq p+1$. When $p = 1$ and $V = 1$, the Wishart distribution $W_p(V, n)$ correspond with the univariate χ^2 distributions. In a multivariate normal distribution, the Wishart distribution is known as the distribution of the maximum likelihood estimate for the covariance matrix.

The covariance matrix of a multivariate normal distribution has a conjugate prior which is the inverted Wishart distribution explained below. Let's assume that $p \times p$ arbitrary matrix $A \propto W_P(V, n)$, therefore, $B = A^{-1}$ possesses an inverted Wishart distribution indicated by $W_p^{-1}(V^{-1}, n)$ or $IW_p(V^{-1}, n)$ to whom probability density function is written as;

$$f(B) = \frac{\mid V \mid^{\frac{-n}{2}} \mid B \mid^{\frac{-(n+p+1)}{2}} \exp(-tr(V^{-1}B^{-1}))/2}{2^{\frac{np}{2}} \Gamma_p^{(n/2)}} \tag{3.28}$$

then the average of $W_p^{-1}(V^{-1}, n)$ is denoted by $E(B) = \dfrac{V^{-1}}{n-p-1}$. Let $X = (x_1 \ldots x_n)$, with $x_i \propto N_p(0, \sigma)$ assume a prior distribution, $\Sigma \propto W_p^{-1}(\Phi, m)$, posterior distribution is written as;

$$\Sigma|X \propto W_P^{-1}(XX^T + \Phi, m + n).$$

When $p = 1$, the inverted wishart distribution decreases to an inverted gamma distribution.

3.8 Estimation of parameters using MCMC methods

The two most commonly used Markov Chain Monte Carlo methods are the Metropolis-Hastings and the Gibbs sampler. The two algorithms are based on simulating dependent samples of a Markov chain type, which result in the MCMC sampling. The MCMCs are method-based and do not use integration.

The Metropolis-Hastings algorithm

The Metropolis-Hastings method was first introduced by Metropolis et al. (1953) and was then extended further by Hastings (1970). It is known as an MCMC procedure that does not require full conditionals, unlike the Gibbs sampler. In the Bayesian framework, the Metropolis-Hastings method can be implemented in the successive iterative steps.

Let θ be a vector of a parameter of the interest and $f(\theta|y)$ be the posterior distribution, T is the sample size. Concerning θ and $f(\theta|y)$, the method may be stated as;

- Set initial values of $\theta^{(0)}$ for $t = 1, 2, \ldots, T,$ the next steps are repeated.

  Set $\theta = \theta^{(t-1)}$

- New candidate values of θ' are generated from the proposal density distribution $q(\theta'|\theta)$.

Then calculate $\boldsymbol{\alpha = min\left(1, \frac{f(\theta'|y)q(\theta|\theta')}{f(\theta|y)q(\theta'|\theta}\right)}$

- update $\boldsymbol{\theta^{(t)} = \theta'}$ with probability $\boldsymbol{\alpha}$ and $\boldsymbol{\theta^{(t)} = \theta^{(t-1)}}$ with probability $\boldsymbol{1-\alpha}$

There is no need to evaluate the normalising constant $\boldsymbol{f(y)}$ connected in $\boldsymbol{f(\theta|y)}$ since it cancels out in the $\boldsymbol{\alpha}$.

The Gibbs sampler

The Gibbs sampling was first established by Geman and Geman (1984). It is regarded as a randomized algorithm for acquiring a sequence of observations approximated from a specified probability distribution. It produces a Makov chain of samples, with the neighbouring samples correlated.

For illustration purposes, suppose that the interest is in sampling from the posterior $\boldsymbol{p(\theta|y)}$, here $\boldsymbol{\theta}$ is a vector of two parameters, that is, $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$. The sampling procedure is commenced by a starting value for the parameters, i.e $\boldsymbol{\theta^0}$, for $\boldsymbol{\theta_1^0}$ and $\boldsymbol{\theta_2^0}$. The posterior distribution is explored by generating $\boldsymbol{\theta_1^k}$ and $\boldsymbol{\theta_2^k}$, where $\boldsymbol{k = 1, 2, \ldots}$ in successive order. Essentially, given $\boldsymbol{\theta_1^k}$ and $\boldsymbol{\theta_2^k}$ at iteration $\boldsymbol{k}$, the $\boldsymbol{k+1}$ value of each of the parameters is produced under the following iterative strategy;

- Extract $\boldsymbol{\theta_1^{k+1}}$ from $\boldsymbol{p(\theta_1|\theta_2^k, y)}$
- Extract $\boldsymbol{\theta_2^{k+1}}$ from $\boldsymbol{p(\theta_2|\theta_1^{k+1}, y)}$.

Then the Gibbs sampler generates a series of values $\boldsymbol{\theta^k = (\theta_1^k, \theta_2^k)^T, k = 1, 2, \ldots}$ which are dependent and produce a chain. The summary measures, that is, the mean or the mode from the chain estimate the accurate posterior measures.

3.9 Convergence diagnostics for a Makov chain

The use of a Markov chain techniques is rooted in the property that the produced chain eventually generates a sample from the posterior distribution and that the summary measure determined consistently estimates the corresponding true posterior measures. The idea is for the chain to converge to a stationary distribution, i.e, the target posterior distribution. The main aim of the convergence techniques in an MCMC algorithm is to check how near the process is to the real posterior distribution.

In an MCMC algorithm, the convergence is an asymptotic characteristic which suggests that the distribution of $\boldsymbol{\theta^k}$, i.e $\boldsymbol{p_k(\theta)}$ converges to the target distribution $\boldsymbol{p(\theta|y)}$ as $\boldsymbol{k} \longrightarrow \infty$, $\boldsymbol{k}$ is known as the number of iterations. The evaluation of the convergence of a chain involves the assessment of the convergence of the marginal posterior distribution by examining how well the chain is mixing or moving around the parameter space. Statistical and graphic tests can be used to check if convergence is convenient through convergence diagnostics. Primarily, the diagnostics are used to examine for the static of the chain and confirm the validity of the posterior summary measures.

Different convergence diagnostics are available and for the current study, trace plots, Geweke plots, and the Heidelberger-Welch (HW) are being utilised. The trace plots are plots of iteration number against the value of the draw of the parameter at each iteration. The trace plots are generated for each parameter independently and the assessments are done univariately. A stationary chain forms the informal "thick pen", Gelfand et al. (1990). A trace plot that portrays the dependence of the chain on its opening state by disclosing an upwards or downwards trend suggests gross deviations from stationarity. The Geweke diagnostics for a Markov chain was first implemented by Geweke (1992). It is based on a comparison of the averages of an initial and last part of the chain using a significant test. Assume that there are n values of θ^k assumed to be $i.i.d$ and suppose that they are broken into two parts: the initial part (A) with n_A elements and the last part (B) with n_B elements with posterior means $\bar{\theta}_A$ and $\bar{\theta}_B$ respectively. A Z-test can be used to compare the means based on

$$Z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}, \tag{3.29}$$

where s_A^2 and s_B^2 are known as the classical estimates of the respective variances. Still, the elements of a Markov chain are dependent, consequently, another estimator of the variance is needed. A spectral density concept is utilised based on a time series approach. To make sure that $\bar{\theta}_A$ and $\bar{\theta}_B$ are independent. Geweke (1992) recommended taking for A the initial 10% of the iterations, $n_A = \frac{n}{10}$, and for B the last 50%, $n_B = \frac{n}{2}$ to generate a distance between the two parts. Therefore, if the ratios $\frac{n_A}{n}$ and $\frac{n_B}{n}$ are fixed with $(n_A + n_B)|n < 1$, thus, it is known that Lesaffre and Lawson (2012) and Brooks and Roberts (1998),

$$Z = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} \rightarrow^d N(0,1) \tag{3.30}$$

as $n \rightarrow \infty$. This result is used to test the null hypothesis of equal location. If $|Z|$ is large, the null hypothesis is rejected, this indicates that the chain has not converged by iteration k. The Heidelberger-Welch (HW) diagnostic was first proposed by Heidelberger and Welch (1983). It is an automated test for checking the stationarity of the chain and to further assess whether the length of the chain is adequate to ensure the desired accuracy for the posterior means of the parameters.

To decide either to accept or reject the null hypothesis that the chain is from a stationary distribution, the Cramer-von mises test statistic is used. Two steps are involved in the test: checking stationarity and determining accuracy. The first step is based on say, N iterations. The test statistic is calculated, after that the null hypothesis of stationarity is either accepted or rejected. If rejected, the first 10% of the chain is thrown away and another test statistic is calculated on the 90% remaining. The process goes on till the null hypothesis is accepted or if half of the chain is thrown away at the stage in which the chain fails the test and it needs to be run longer. In the second step, the part of the chain not thrown away is considered and the half-width of the $(1-\alpha)\%$ credible interval is calculated around the mean. A threshold value, say, ϵ is determined and if the half-width and the mean is lower than ϵ, then the chain passes the test, otherwise it must be run longer.

3.10 Integrated nested laplace approximation

The integrated nested Laplace approximation (INLA) is an alternative to MCMC sampling which was introduced by Rue et al. (2009). The INLA is an approximate Bayesian approach. This method carries out Bayesian inference over a wide grade of latent Gaussian models. The Gaussian models are models of an effect variable y_i which presume independence depending on certain fundamental latent field ξ and a vector of parameters θ. This method directly estimates the posterior of interest with an accurate expression contrasted to the MCMC methods that sample out of a posterior distribution.

The INLA method aims at approximating the minor posterior for the latent variables and also hyperparameters of this Gaussian latent model is denoted as

$$f(\xi_i|\boldsymbol{y}) = \int f(\xi_i|\boldsymbol{\theta}, \boldsymbol{y}) f(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{3.31}$$

$$f(\boldsymbol{\theta_j}|\boldsymbol{y}) = \int f(\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{\theta} \tag{3.32}$$

The fundamental of the estimation belongs to the union of Laplace approximation to the thorough condition $f(\boldsymbol{\theta}|\boldsymbol{y})$ and $f(\xi_i|\boldsymbol{\theta}, \boldsymbol{y})$, for $i = 1, 2, \ldots, n$ and computational integration procedures to merge away hyperparameters θ. To obtain the minor posteriors within equation 3.31 and equation 3.32, the estimation requires the following 3 procedures. The 1^{st} procedure is to estimate the saturated posterior $f(\boldsymbol{\theta}|\boldsymbol{y})$. To perform this, $f(\boldsymbol{\xi}|\boldsymbol{\theta}, \boldsymbol{y})$ is approximated with a multivariate Gaussian density $\tilde{f}_G(\boldsymbol{\xi_i}|\boldsymbol{\theta}, \boldsymbol{y})$ estimated on its mode. Therefore, the posterior density of θ is estimated by utilising the Laplace estimation

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{f(\boldsymbol{\xi_i}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{f}_G(\boldsymbol{\xi_i}|\boldsymbol{\theta}, \boldsymbol{y})} |\xi = \xi * (\theta), \tag{3.33}$$

where $\xi * (\theta)$ is known as the mode of the saturated conditional of ξ for a specified θ. No accurate clarification is applicable for $\xi * (\theta)$, an analytical method such as the Newton-Raphson algorithm may be utilised. In the 2nd procedure, the Laplace estimation of the full conditionals is computed $f(\xi_i|\boldsymbol{y}, \boldsymbol{\theta})$ for chosen values of $\boldsymbol{\theta}$. The chosen values of $\boldsymbol{\theta}$

should be selected precisely as will be utilised as assessment marks in the computational integration enforced to acquire the posterior minors of ξ_i in equations equation 3.31 and equation 3.32. The density $f(\xi_i|\boldsymbol{y}, \boldsymbol{\theta})$ is estimated using Laplace estimation explained by;

$$f_{LA}(\xi_i|\theta, y) \propto \frac{f(\boldsymbol{\xi_i}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{f}_G(\boldsymbol{\xi_{-1}}|\boldsymbol{\xi_i}\boldsymbol{\theta}, \boldsymbol{y})}|\boldsymbol{\xi_{-1}} = \boldsymbol{\xi_{-1}} * (\boldsymbol{\xi_i}, \boldsymbol{\theta}), \tag{3.34}$$

where ξ_{-1} indicates the vector $\boldsymbol{\xi}$ with i^{th} element excluded, $\tilde{f}_G(\boldsymbol{\xi_{-1}}|\boldsymbol{\xi_i}\boldsymbol{\theta}, \boldsymbol{y})$ is known as the Gaussian approximation of $f(\boldsymbol{\xi_{-1}}|\boldsymbol{\xi_i}\boldsymbol{\theta}, \boldsymbol{y})$, treating ξ_i as fixed and $\boldsymbol{\xi_{-1}} = \boldsymbol{\xi_{-1}} * (\boldsymbol{\xi_i}, \boldsymbol{\theta})$ is the mode of $f(\boldsymbol{\xi_{-1}}|\boldsymbol{\xi_i}\boldsymbol{\theta}, \boldsymbol{y})$. The last step includes joining the 2 saturated posteriors acquired in the preceding procedures and the minor densities of ξ_i and θ_j acquired by merging away from the applicable expression. The estimation of the minor latent variables may be acquired by the terms

$$f(\xi_i|\boldsymbol{y}) = \int f(\xi_i|\boldsymbol{y}, \boldsymbol{\theta}) f(\boldsymbol{\theta}, \boldsymbol{y}) d\theta \approx \Sigma_k \tilde{f}(\xi_i|\boldsymbol{\theta_k}, \boldsymbol{y}) \tilde{f}(\boldsymbol{\theta_k}|\boldsymbol{y}) \Delta_k \tag{3.35}$$

which is estimated utilising computational integration on coordination of grid marks for $\boldsymbol{\theta}$, with area masses Δ_k for $k = 1, 2, \ldots, K$.

Chapter 4

Results

4.1 Univariate analysis

This section presents information on the demographic and the socioeconomic characteristics of the survey respondents such as age, education, place of residence, marital status, employment, and wealth status. This information is useful for understanding the factors that affect women unemployment in South Africa. A total number of 8,514 women aged 15-49 were interviewed in the SADHS 2016 with the standard individual questionnaires. The black African is the largest self-reported population group, making up 86.4% of the women, and the other categories, White, Coloured, Indian/Asian, and Other constituted small percentages. The Coloured, the Indian/Asian, and the Other categories were combined to form the 'Other' category as shown in Table 4.1 below.

Table 4.1: Distribution of the respondents by race.

Race	frequencies	percentage
Black/African	7359	86.4
Other	941	11.1
White	214	2.5
Total	8514	100

Table 4.2 indicates that about 77.3% of the South African women have at least some secondary education and about 10.4% of the women have gone beyond secondary school. Only 2.2% of the women have not attended school at all. Almost all women aged 15-49 (97.8%) in South Africa have attained some education.

Table 4.2: Summary of the respondents educational attainment.

Educational level	frequencies	percentage
No education	190	2.2
Primary	862	10.1
Secondary	6581	77.3
Higher	881	10.4
Total	8514	100

The household scores were specified based on the quantity and types of consumer goods they possess, varying from a television to a bicycle or a car, and the housing characteristics like the water supply, the toilet provision, and the floor materials. Table 4.3 represents the distribution of the household inhabitants by the wealth index. The richest class constitutes about 10% while the poorest, poorer, middle, and richer constitute about 20% each.

Table 4.3: Summary of respondants wealth index.

Wealth index	frequencies	percentage
Poorest	1763	20.71
Poorer	1865	21.91
Middle	1956	22.97
Richer	1733	20.35
Richest	1197	14.06
Total	8514	100

South Africa has divided into nine regions or provinces and the largest percentage of the respondents lived in the Kwazulu-Natal (16%), followed by Limpopo (13%). Only 7.7% of the respondents lived in the Western Cape province. The complete distribution is given in Table 4.4 below.

Table 4.4: Distribution of respondents per region.

Region	frequencies	percentage
Western cape	656	7.7
Eastern cape	1041	12.2
Northern cape	718	8.4
Free state	854	10.0
Kwazulu-Natal	1360	16.0
North west	863	10.1
Gauteng	863	10.1
Mpumalanga	1054	12.4
Limpopo	1105	13.0
Total	8514	100

Table 4.5 below shows that two-thirds of the women aged 15-49 (67.8%) have reported that they were not working and about 32.2% of the women responded that they were employed.

Table 4.5: Summary of respondants employment status.

Employed	frequencies	percentage
Yes	2740	32.2
No	5774	67.8
Total	8514	100

The majority of South African women stay in urban areas. The data indicates that over half (56.4%) of the respondents were residing in the urban areas and about (43.6%) of the respondents were in the rural areas.

Table 4.6: Distribution of the respondents by the type of residence.

Residence type	frequencies	percentage
Urban	4805	56.4
Rural	3709	43.6
Total	8514	100

The respondents who went through to attain higher education were assumed to be literate. The remaining respondents, shown a typed sentence to read aloud, were regarded literate if they were able to read all or portion of the sentence. The illiterate category is a combination of the categories, 'Cannot read at all', 'No card with required language' and the literate category is a combination of categories, 'Able to read', 'Can only read part of the sentence' and 'Can read all of the sentence'. For the category 'blind or visually impaired', there was no indication of whether the blind persons had a reading support device to enable them to read, hence, we cannot state whether they were able to read or not. Table 4.7 below indicates that almost all of the women aged 15-49 (96.1%) in South Africa were literate.

Table 4.7: Respondents literacy staus.

literacy status	frequencies	percentage
Blind	8	0.1
Illiterate	324	3.8
Literate	8182	96.1
Total	8514	100

Table 4.8 shows that six in ten of the women aged 15-49 (60.3%) reported that they were single or have never been in a union (never married or lived together with a partner as though married) and one-thirds of the women (33.4%) reported that they were married or stay with a partner as though they were married at the time of the survey. Very few women (6%) were either divorced or widow and the two categories were combined to form one category.

Table 4.8: Respondents marital status.

Marital status	frequencies	percentage
Single	5134	60.3
Married	2841	33.4
Divorced/widow	539	6.3
Total	8514	100

Table 4.9 indicates the frequency of the respondents in each of the five-year age groups. These results show that the proportion of the respondents in each age group decrease with an increase in age.

Table 4.9: Distribution of respondents by age group.

Age group	frequencies	percentage
15-19	1505	17.7
20-24	1408	16.5
25-29	1397	16.4
30-34	1295	15.2
35-39	1032	12.1
40-44	964	11.3
45-49	913	10.7
Total	8514	100

Table 4.10 below represents the women's exposure to the use of the internet. The women who were exposed to the internet 'less than once a week' and those exposed 'at least once a week' were grouped to form the category 'Weekly' and those women who responded that they use it 'almost every day' were grouped as 'daily' users. The internet is one of the critical tools through which most information is shared. Since on the questionnaire of the data set used there was no question asked the respondents on the steps taken to search for employment, the study assumes that the women might have used this tool to search for jobs. The results on the table below indicate that over half of the women responded that they were never exposed to the use of the internet. Only 42.5% of the women responded that they were exposed to the internet on 'weekly' and 'daily' basis.

Table 4.10: Respondents' use of internet.

Frequency of internet use	Count	percentage
Never	4892	57.5
Weekly	1153	13.5
Daily	5469	29.0
Total	8514	100

4.2 Multilevel logistic regression analysis

In this section, the multilevel logistic regression model was constructed using the lme4 package Bates et al. (2015) and the function glmer was also used to compute multilevel logistic regression model from a GLMM perspective for women unemployment on the demographic and socioeconomic variables. The fixed effects considered are the women's demographic and socioeconomic variables. Two levels were considered, the first level being the individual level and the second level being the provincial level.

4.2.1 Multilevel Logistic Regression Model (NULL)

The Multilevel Logistic Regression null equation is given as

$$\text{logit}(\boldsymbol{P_{ij}}) = \boldsymbol{\beta_0} + \boldsymbol{U_{0j}} \tag{4.1}$$

The empty model constructed below consists of no independent variables and may be regarded as a parametric variation of estimating diversity among the provinces concerning women's unemployment status. To acquire the approximation of how much of the difference in unemployment among women aged between 15-49 years is due to the provincial level components, it is helpful to see the intra-region correlation coefficient (ICC). The ICC was found to be 0.04547 that estimates the proportion of variance of women unemployment, which is between provinces, not in provinces. This means that around 4.55% of the variance in the women unemployment is due to the variation across (between) the provinces. Whereas the remaining 95.45% is due to the individual level, i.e., within the province differences. The ICC is non-zero which justifies the use of the multilevel approach to the analysis.

Table 4.11: Results Multilevel Logistic Regression Model (NULL).

Fixed part	Estimate	Std.Error	P-value
β_0(intercept)	-0.72075	0.07651	$< 0.00001 ***$
ICC	0.04547		

4.2.2 Random Intercept Multilevel Logistic Regression Model

The random intecept multilevel logistic regression equation is given as;

$$\log(\pi_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) = \beta_{0j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \ldots + \beta_k x_{kij}, i = 1, 2, \ldots n, j = 1, 2, \ldots J \quad (4.2)$$

Table 4.12: Results of a Random Intercept Multilevel Logistic Regression Model.(MLR estimation approach)

Covariates	Estimate	Std.error	P-value	95% CI
Intercept	-4.763912	0.795919	$< 0.00001***$	(-6.324 -3.204)
Education(Primary)	-0.005362	0.200765	0.978693	(-0.399 0.388)
Education(Secondary)	0.146400	0.200367	0.464985	(-0.246 0.539)
Education(Higher)	0.794384	0.217608	$0.000262***$	(0.368 1.221)
Residence type(Rural)	-0.130704	0.068790	0.057428 .	(-0.266 0.004)
Age group(20-24)	2.148082	0.192019	$< 0.00001***$	(1.772 2.524)
Age group(25-29)	3.210070	0.188198	$< 0.00001***$	(2.841 3.579)
Age group(30-34)	3.665441	0.189963	$< 0.00001***$	(3.293 4.038)
Age group(35-39)	3.894134	0.193079	$< 0.00001***$	(3.516 4.273)
Age group(40-44)	4.032322	0.195216	$< 0.00001***$	(3.650 4.415)
Age group(45-49)	4.107958	0.197904	$<< 0.00001***$	(3.720 4.496)
Race(Other)	0.221931	0.096820	$0.021893*$	(0.032 0.412)
Race(White)	0.271852	0.172270	0.114554	(-0.066 0.609)
Wealth index(Poorer)	0.183331	0.085577	$0.032169*$	(0.016 0.351)
Wealth index(Middle)	0.192080	0.087513	$0.028172*$	(0.021 0.364)
Wealth index(Richer)	0.263954	0.097716	$0.006908**$	(0.072 0.455)
Wealth index(Richest)	0.515605	0.116074	$< 0.00001***$	(0.288 0.743)
Internet use(weekly)	0.552972	0.087281	$< 0.00001***$	(0.382 0.724)
Internet-use(Daily)	0.674433	0.072859	$< 0.00001***$	(0.532 0.817)
Marital-status(Married)	-0.197552	0.059709	$0.00938**$	(-0.315 -0.018)
Marital(Divorced/widow)	0.397486	0.102843	$0.00011***$	(0.196 0.599)
Literacy-status(illiterate)	-0.027767	0.766045	0.971085	(-1.529 1.474)
Literacy status (literate)	0.216488	0.752491	0.773581	(-1.258 1.691)
Random effects				
$var(u_{0j})$	0.01811			

4.2.3 Random Coefficient Logistic Regression Model

The random coefficient logistic regression model equation is given by

$$\log(\pi_{ij}) = \log\left(\frac{P_{ij}}{1-P_{ij}}\right) + \beta_0 + \sum_{h=1}^{k}\beta_h x_{hij} + U_{0j} + \sum_{h=1}^{k} U_{1j}X_{1ij} \tag{4.3}$$

while constructing the random intercept model above, only the intercept was allowed to vary across the regions by fixing the independent variables. The relationship between the independent and the dependent covariates differs between groups (in this case is the regions or the provinces). In this model, the random slope for women's educational level was contained, meaning that the effect of the coefficient educational level was allowed to vary from province to province. The estimate of the random effects for the intercepts and the slopes differ significantly, meaning that there is a significant difference in the results of the educational level of a woman and this variable vary significantly over the provinces.

Table 4.13: Results of a Random Coefficient Multilevel Logistic Regression Model.

Covariates	Estimate	Std.error	P-value	95% CI
Intercept	-4.70635	0.80542	5.12< 0.00001 * **	(-6.285 -3.128)
Education(Primary)	-0.02675	0.20863	0.897965	(-0.436 0.382)
Education(Secondary)	0.14052	0.23320	0.546779	(-0.317 0.598)
Education(High)	0.79110	0.25332	0.001791 * *	(0.295 1.288)
Residence type(rural)	-0.14781	0.06964	0.033796 *	(-0.284 -0.011)
Age group(20-24)	2.14427	0.19199	< 0.00001***	(1.768 2.521)
Age group(25-29)	3.20860	0.18817	< 0.00001***	(2.840 3.577)
Age group(30-34)	3.66111	0.18994	< 0.00001***	(3.289 4.033)
Age group(35-39)	3.89286	0.19306	< 0.00001***	(3.514 4.271)
Age group(40-44)	4.02900	0.19523	< 0.00001***	(3.646 4.412)
Age group(45-49)	4.10925	0.19796	< 0.00001***	(3.721 4.497)
Race(Other)	0.25173	0.10234	0.013904 * *	(0.051 0.452)
Race(White)	0.28186	0.17262	0.102507	(-0.056 0.620)
Wealth index(Poorer)	0.18445	0.08606	0.032089*	(0.016 0.353)
Wealth index(Middle)	0.20005	0.08806	0.023092*	(0.027 0.373)
Wealth index(Richer)	0.26112	0.09808	0.007759 * *	(0.069 0.453)
Wealth index(Richest)	0.50752	0.11627	< 0.00001***	(0.280 0.735)
Internet use(weekly)	0.55199	0.08731	< 0.00001***	(0.381 0.723)
Internet-use(Daily)	0.67568	0.07286	< 0.00001***	(0.533 0.818)
Marital-status(Married)	-0.19599	0.05979	0.001046 * *	(-0.313 -0.079)
Marital(Divorced/widow)	0.39659	0.10317	0.000121 * **	(0.194 0.599)
Literacy-status(illiterate)	-0.10870	0.76446	0.886931	(-1.607 1.390)
Literacy status (literate)	0.1731	0.75010	0.823494	(-1.303 -1.637)
Random effects				
$var(u_{0j})$	0.167			
$var(u_{1j})$	0.246			
$corr(u_{1j}, u_{0j})$	-0.02			

4.2.4 Model comparison

The comparison was made based on the AIC and the BIC values, the results show that the random intercept model has the least AIC and BIC compared to the comparing models. Consequently, based on these results, conclusions may be made that the random intercept model fits best to the data.

Table 4.14: Results indicating the best fit model.

Fitted model	Multilevel null model	Random intercept model	Random coefficient model
-log likelihood	-5318.8	-4334.8	-4331.6
Deviance	10637.6	8669.6	8663.3
AIC	10641.6	8711.6	8723.3
BIC	10655.7	8859.6	8934.8

4.3 Diagnostics for multilevel models

Below is the residual plot for the multilevel null model, the random intercept model, and the random coefficient model.

4.3.1 The multilevel null model

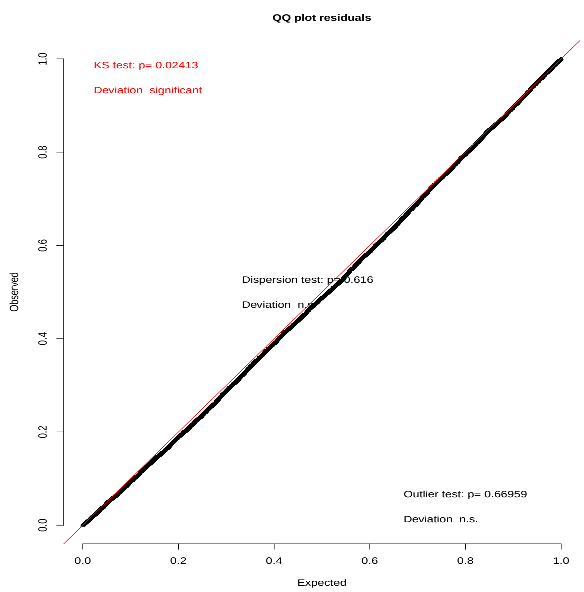


Figure 4.1: The residual plot for the multilevel null model.

The residuals are useful for determining if individual points are not well fit by the model. There is no deviation from the expected normal line, the line looks straight and therefore pretty normal and suggests that the assumption of normality is not violated.

4.3.2 The multilevel random intercept model

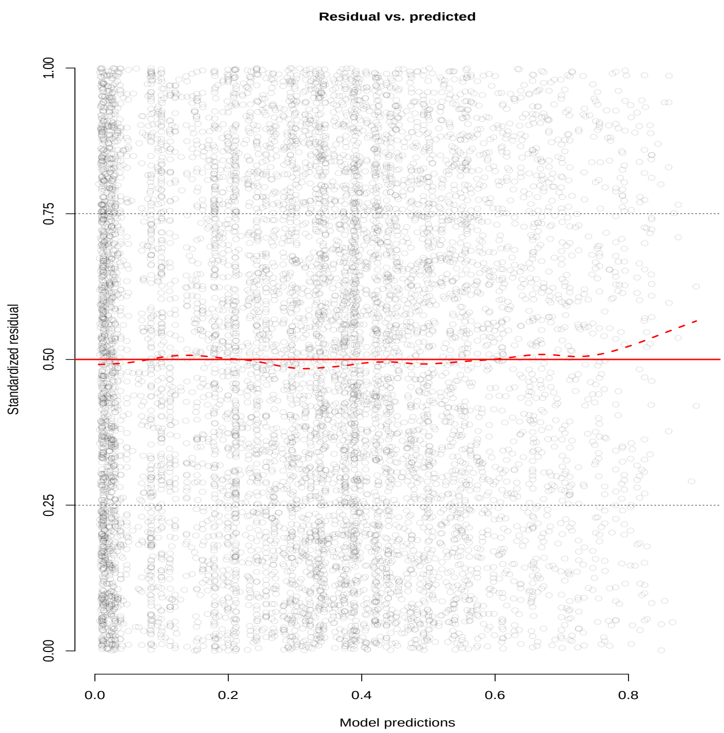


Figure 4.2: The residual plot for the multilevel random intercept model.

There is a minority of deviation from the anticipated normal line to the tail, but generally, the line looks aligned and consequently pretty normal and indicates that the assumption of normality is not breached.

4.3.3 The multilevel random coefficient model

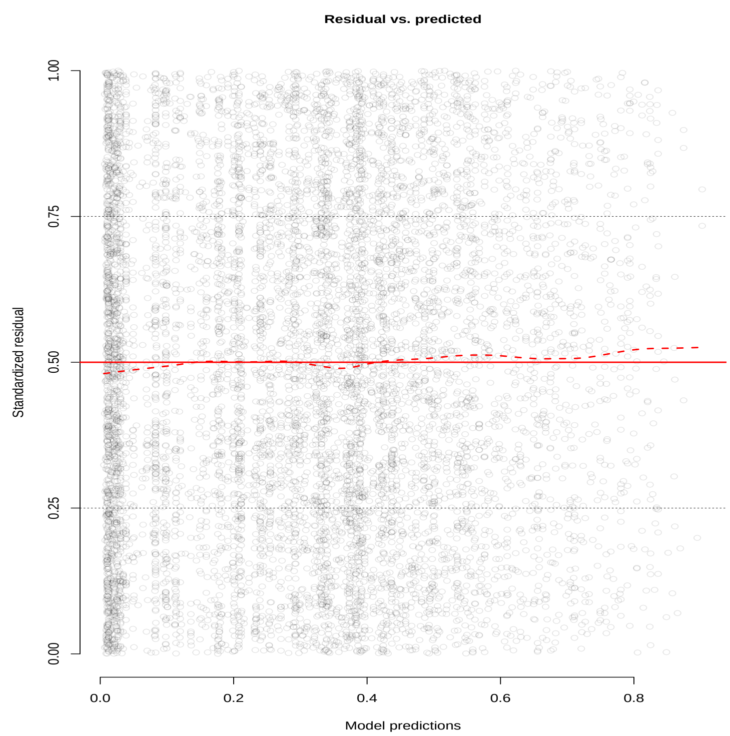


Figure 4.3: The residual plot for the multilevel random coefficient model.

There exist a few slight divergence from the anticipated normal line to the tail also and the overall line appears straight and normal and it indicates that indeed the assumption of normality was not breached.

4.3.4 Interpretation of the best fit mode1

The results in Table 4.12 display the estimated variance of the random effect at the provincial level, $\boldsymbol{\sigma^2 = 0.01811}$. Since the variance is greater than zero, it is an indication that there exists a provincial level effect. The multilevel analysis results displayed in the Table also showed that education, type of residence, age group, race, wealth index, internet use, and marital status were significant determinants and were also the factors contributing to the variation of the women unemployment among South African regions.

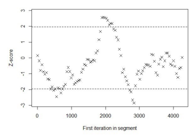
The results indicate that the odds of unemployment for a woman with primary and secondary education do not differ significantly from those of a woman with no education. For a woman with higher education, the odds ratio of unemployment is $\boldsymbol{e^{0.794384} = 2.21}$ times that of a woman with no education. Generally the odds of unemployment decrease with increasing education. These results are similar to those found by (Yakubu, 2010; Mathebula, 2017).

The results also revealed that the odds ratio of unemployment for a woman residing in the rural areas is $\boldsymbol{e^{-0.130704} = 0.88}$ times that of a woman residing in the urban areas. These results are in accord with those reported in Pakistan by (Qayyum and Siddiqui, 2007). The reason for this might be that there is a mismatch between the skills provided and the skills demanded. Furthermore, the women in the rural areas might be taking any available jobs regardless of their qualifications due to the scarcity of available jobs in rural areas. However, these results differ from those found in Mauritius by Tandrayen-Ragoobur and Kasseeah (2015) which revealed that unemployment was not associated with the place of residence.

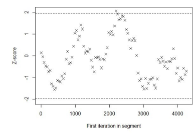
The odds of unemployment of a woman aged 20-24 up to 45-49 were significantly different from that of a woman aged 15-19. The reason for this might be that those women aged 15-19 are still at school and are not seeking employment. Another finding of this study indicates that the race of a woman has a significant contribution to unemployment. The odds ratio of unemployment for a White woman does not differ significantly from that of a black woman. The odds ratio of unemployment of a woman in the race category 'Other' is $\boldsymbol{e^{0.069883} = 1.3}$ times that of a black woman. The odds ratio of unemployment of a woman in the wealth index category 'Poorer', 'Middle', 'Richer' and 'Richest' differs significantly from that of a woman in the wealth index category ́Poorest©.

The odds of unemployment of a woman who uses the internet 'weekly' and 'daily' are $\boldsymbol{e^{0.552972} = 1.74}$ and $\boldsymbol{e^{0.674433} = 1.96}$ times that of a woman who never uses the internet respectively. The reason for these results might be that more job opportunities are being advertised on internet. Since the women were not asked which steps they took to search for jobs, the study assumes that internet use might be one of the steps used to search for employment. This study also revealed that the odds of unemployment of the married and the divorced/widow are $\boldsymbol{e^{-0.197552} = 0.82}$ and $\boldsymbol{e^{0.397486} = 1.49}$ times that of a woman who was never married. The literacy status of a woman was not significant and this result differ from those found in Ethiopia by Chikako (2018).

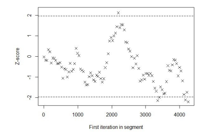
Figure 4.4 gives the Geweke plots for some selected parameters. As a rule of thumb, a significant proportion of Z-scores outside the two-standard deviation bands suggest a chain that has not converged by iteration k. The results in Figure 4.4 indicate that all the Z-scores fall within the two-standard deviation bands for the parameters educational level 'primary', marital status 'widow/divorced' and race group 'other' whereas there is a negligible proportion for the Z-scores under the intercept, educational levels 'secondary and higher', internet-use, marital status 'married' and race group 'White'. The chain has converged by iteration 4000.



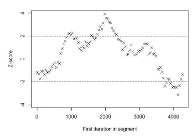
(a) Intercept



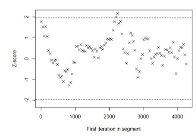
(b) Education-primary



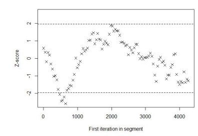
(c) Education-secondary



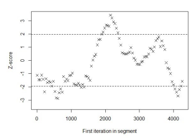
(d) Education-higher



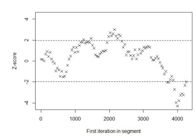
(e) Race(Other)



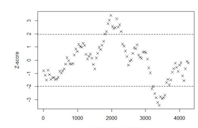
(f) Race(White)



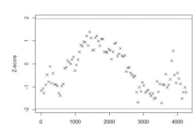
(g) Internet-use(weekly)



(h) Internet-use(daily)



(i) Marital status(married)



(j) Marital-status(w/d)

Figure 4.4: Geweke plots for some choosen coefficients from the posterior distribution.

The Heidelberger-Welch diagnostic test was also performed. The results of a test with $\epsilon = 0.1$ indicated that all of the parameters have passed both the stationary test and the half-width test, showing that the chain was run sufficiently long.

4.4 Results of a Bayesian Multilevel Logistic Regression Analysis

In this section, SADHS data were analysed using Bayesian GLMMS, in R statistical windows software. In R statistical software, the MCMCglmm Hadfield (2010) package was utilized. The MCMCglmm package utilizes the Gibbs sampler. The total amount of iterations were 250 000. Within previous sections, it was discussed that in Bayesian inference, prior distributions for parameters are used. Within the analysis, the multivariate prior distribution was utilised for the mean parameters (fixed effects) and an inverted Wishart prior with a Cauchy parameter expansion was utilised for the variance of random effects.

4.4.1 Results of Bayesian Null Logistic Regression Model.

Table 4.15 gives the estimate of the fixed and random effects, together with the residuals for the marginal posterior distribution obtained by the random draws of samples from the joint distribution together with the 95% credibility intervals(CI) and the P-values. The model was fitted in MCMCglmm using 10000 iterations, with burning of 100 and a thining of 10.

For the random effects, the results give the provincial level variability. The province to province variability accounts for a lower variation (0.111) in the women's unemployment status. The corresponding estimated intra-correlation coefficient is 0.074. This implies that about 7.4% of the total variability in women's unemployment is due to differences across provinces, and the remaining unexplained 92.6% is due to individual differences.

Table 4.15: Random and fixed effects for the Bayesian null multilevel model.

Coefficient	Posterior mean	95% cred.int
Fixed effects		
Intercept	-0.909	(-1.202 -0.653)
Random effects		
Province	0.111	(0.015 0.265)
Residual	1.388	(0.481 2.845)

4.4.2 Results of a Bayesian Random Intercept Logistic Regression Model.

The package MCMCglmm by Hadfield (2010) in R was used in this study to compute the posterior distribution of the parameters using the MCMC algorithm. This algorithm is iterative and it is based on the proposal, at each step, of a new value for a given parameter as a function of the other parameters in the model. In the MCMCglmm package, the prior argument took a list of the three elements specifying the priors for fixed effects, the random effects, and the residuals. For the fixed effects, a multivariate normal prior distribution was specified with mean vector $\boldsymbol{\mu}$ and a covariance matrix $\boldsymbol{V}$. For the random effects, a non-informative inverse-gamma prior distribution was used (is a special case of the inverse Wishart distribution) parameterized by $\boldsymbol{v}$ and $\boldsymbol{V}$ as proposed by Hadfield (2014). Specifically, the function MCMCglmm in the MCMCglmm package was used to establish the marginal posterior distribution by extracting random samples from the joint distribution of the prior and the likelihood of the data. To intensify the movement of the chain through the parameter space the package uses a combination of the Gibbs sampling and the Metropolis-Hastings updates, see Hadfield (2010).

Table 4.16 gives the estimates of the parameters for the marginal posterior distribution obtained by the random draws of samples from the joint distribution together with the 95% credibility intervals (CI) and the p-values for each parameter estimate. The model was fitted in MCMCglmm using 250000 iterations, with burning of 100 and a thinning of 10. Table 4.17 presents the variance components for the random effects for the province and the residuals. Both the classical and the Bayesian inference yield similar results. However, from the classical and the Bayesian multilevel models, the Bayesian multilevel analysis is better since it provides the posterior distributions of the parameters.

The results in Table 4.16 indicate that the women's unemployment status depends on the respondent's age, level of education, place of residence, race, economic status (measured by wealth index), marital status, and frequency of using the internet. These results indicate that age is a powerful predictor of women's unemployment status. The women in age group 30-34 (OR:74.52, 95% CI:36.89-156.80), 35-39 (OR:100.48, 95% CI:48.62-205.61), 40-44 (OR:119.22, 95% CI:53.95-253.66) and 45-49(OR:131.24, 95% CI:58.85-286.29) have higher odds ratio(OR) and are more likely to be employed than those aged 15-19. The results also show that the odds of the women unemployment status are slightly lower for the women with primary educational level (OR:1.04, 95% CI:0.62-1.77) and more than twice higher (OR:2.85, 95% CI:1.64-5.32) for the women with higher educational level as compared to those with no education. The chance of being employed does differ from province to province, (OR:0.85, 95% CI:0.72-0.99) women in the countryside are implausible to be unemployed as compared to those in the towns or cities. The odds of unemployment for a white woman and a woman in the 'Other' race category are (OR:1.33, 95% CI:1.05-1.71) and (OR:1.42, 95% CI:0.92-2.21) respectively as compared to those of a black woman. The results further indicate that the odds of unemployment are lower (OR:0.78, 95% CI:0.68-0.91), for married and (OR:1.67, 95% CI:1.30-2.25), for the divorced/widow as compared to the single/never married. A woman who uses the internet weekly or daily is more than two times higher to be employed (OR:2.03, 95% CI:1.50-2.29) and (OR:2.32, 95% CI:1.85-2.59) as compared to a woman who never used the internet. A woman in the 'richest' category is more likely to be employed (higher odds) (OR:1.93, 95% CI:1.45-2.65) as compared to the poorest woman. The results displayed in Table 4.17 show that there is a low province to province variability, $\sigma^2_{prov} = 0.047, 95\%CI(0.004 - 0.116)$ and individual to individual variability, residual, $\sigma^2_{residual} = 1.542, 95\%CI(0.540 - 2.861)$. This gives a low ICC of 0.030.

Table 4.16: The marginal posterior distribution for the fixed effects.

Coefficients	Posterior mean	95% cred.int	P-value
Fixed effects			
Intercept	-5.773	(-8.095 -3.667)	$< 0.001**$
Education(Primary)	0.038	(-0.480 0.572)	0.885
Education(Secondary)	0.230	(-0.286 0.743)	0.394
Education(Higher)	1.049	(0.496 1.671)	$< 0.001**$
Residence type(Rural)	-0.166	(-0.317 -0.005)	0.026*
Age group(20-24)	2.431	(1.920 3.002)	$< 0.001**$
Age group(25-29)	3.747	(3.105 4.399)	$< 0.001**$
Age group(30-34)	4.311	(3.608 5.055)	$< 0.001**$
Age group(35-39)	4.610	(3.884 5.326)	$< 0.001**$
Age group(40-44)	4.781	(3.988 5.536)	$< 0.001**$
Age group(45-49)	4.877	(4.075 5.657)	$< 0.001**$
Race(Other)	0.286	(0.053 0.536)	0.026*
Race(White)	0.338	(-0.084 0.792)	0.125
Wealth index(Poorer)	0.233	(0.031 0.448)	0.020*
Wealth index(Middle)	0.258	(0.031 0.486)	0.026*
Wealth index(Richer)	0.349	(0.121 0.609)	0.008**
Wealth index(Richest)	0.658	(0.374 0.974)	$< 0.001**$
Internet use(weekly)	0.707	(0.403 0.828)	$< 0.001**$
Internet-use(Daily)	0.840	(0.614 0.953)	$< 0.001**$
Marital-status(Married)	-0.241	(-0.374-0.096)	$< 0.001**$
Marital(Divorced/widow)	0.513	(0.262 0.813)	$< 0.001**$
Literacy-status(illiterate)	-0.021	(-2.037 1.883)	0.968
Literacy status(literate)	0.279	(-1.648 2.167)	0.798

Table 4.17: Variance components for the random effects

Coefficient	Posterior mean	95% cred.int
province	0.047	(0.004 0.116)
residual	1.542	(0.540 2.861)

4.5 Convergence diagnostics for a markov chain

Figures 4.5 to 4.11 below give the trace plots for all of the parameters of the posterior distribution obtained by the MCMC algorithm as explained in chapter 3, section 3.6 above. All the trace plots do not display any significant upward or downward trend along with the iterations and the density plots also show almost symmetrical distributions. In particular, the trace plots exhibit the so-called "thick pen" as reported by Gelfand et al. (1990). This is indicative of insignificant deviations from stationarity and the MCMC algorithm may be regarded to have converged.

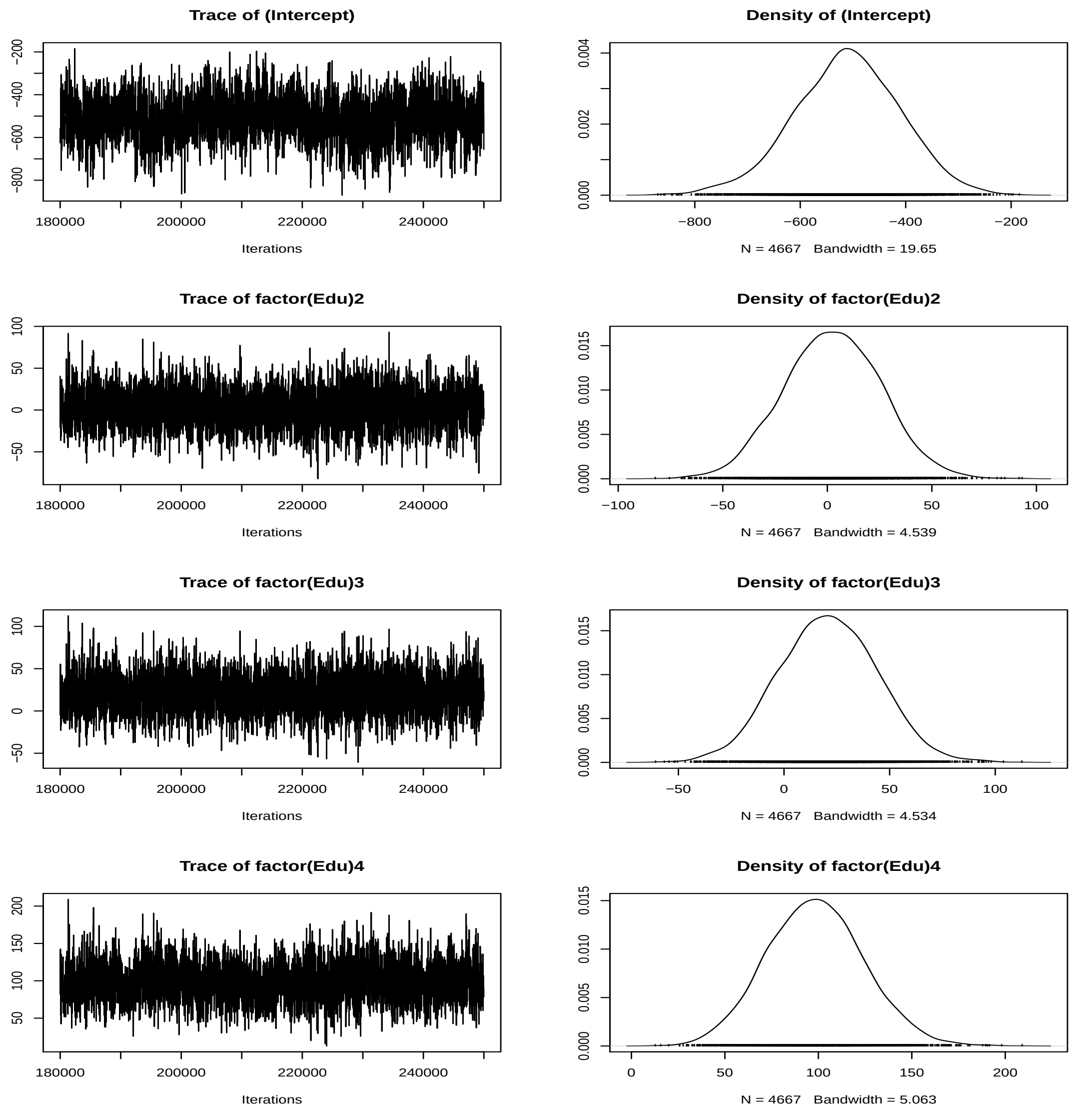


Figure 4.5: Trace plot for fixed effects: intercept and educational level.

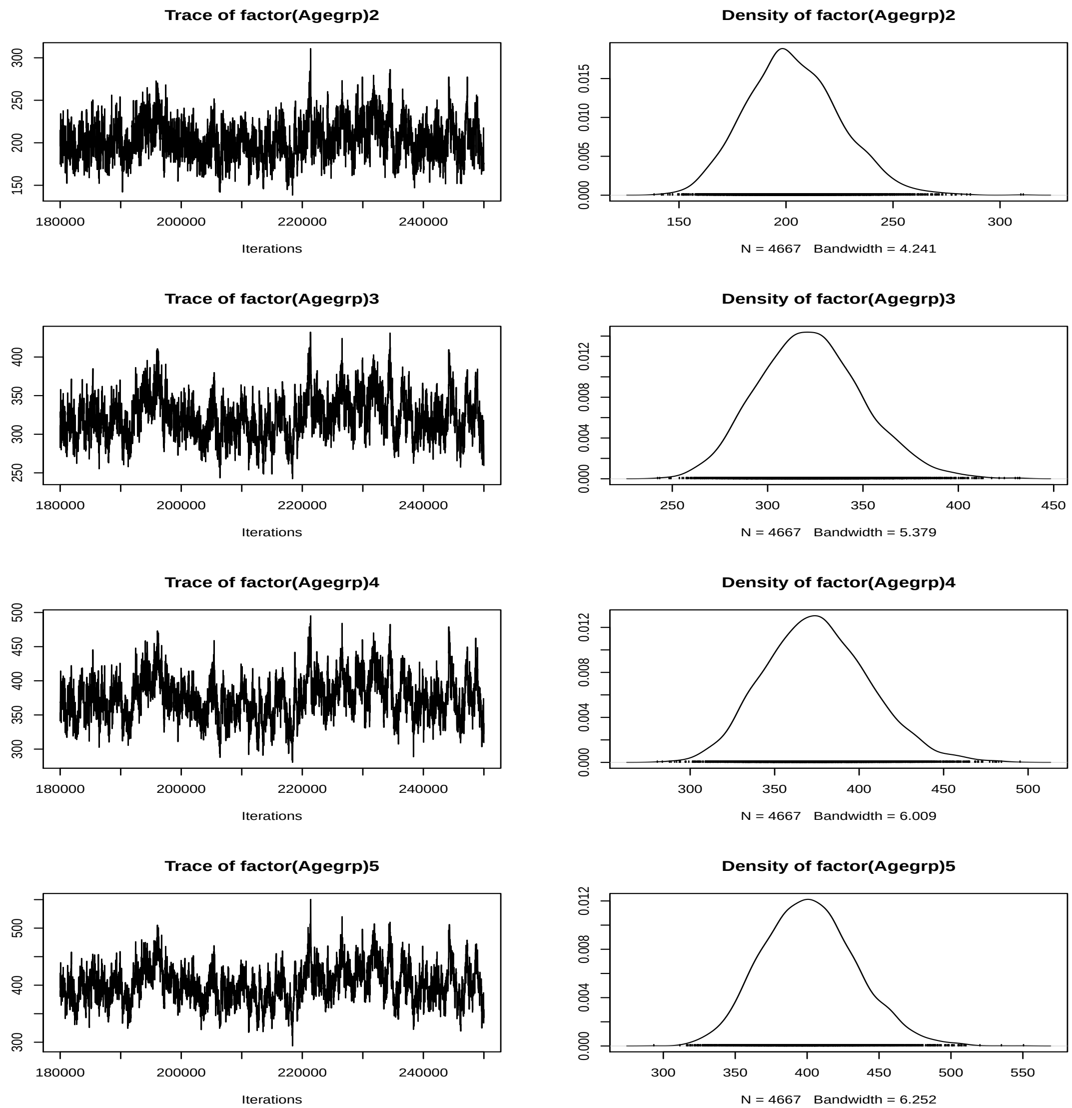


Figure 4.6: Trace plot for fixed effects: Age group.

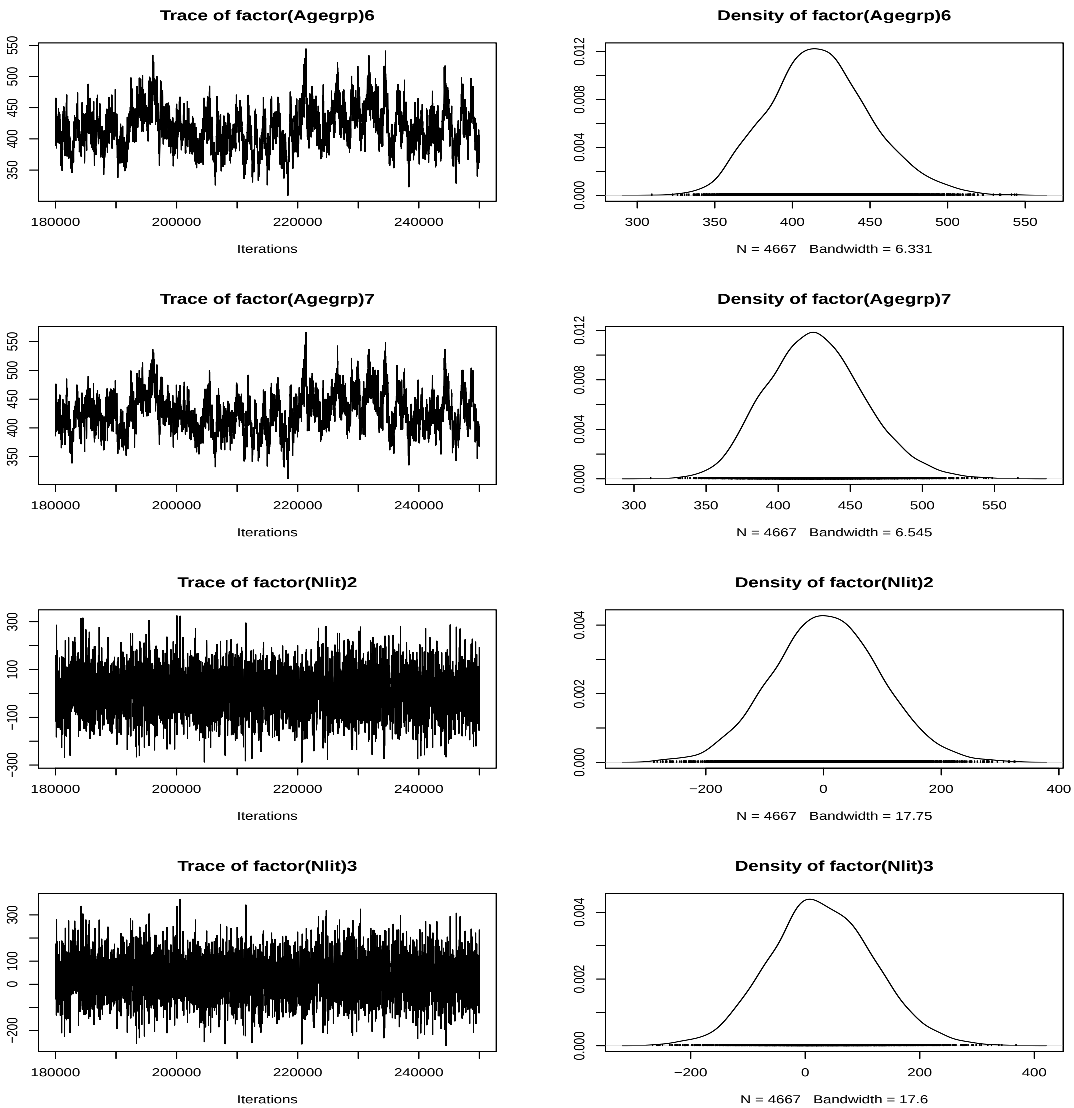


Figure 4.7: Trace plot for fixed effects: Age group and literacy status.

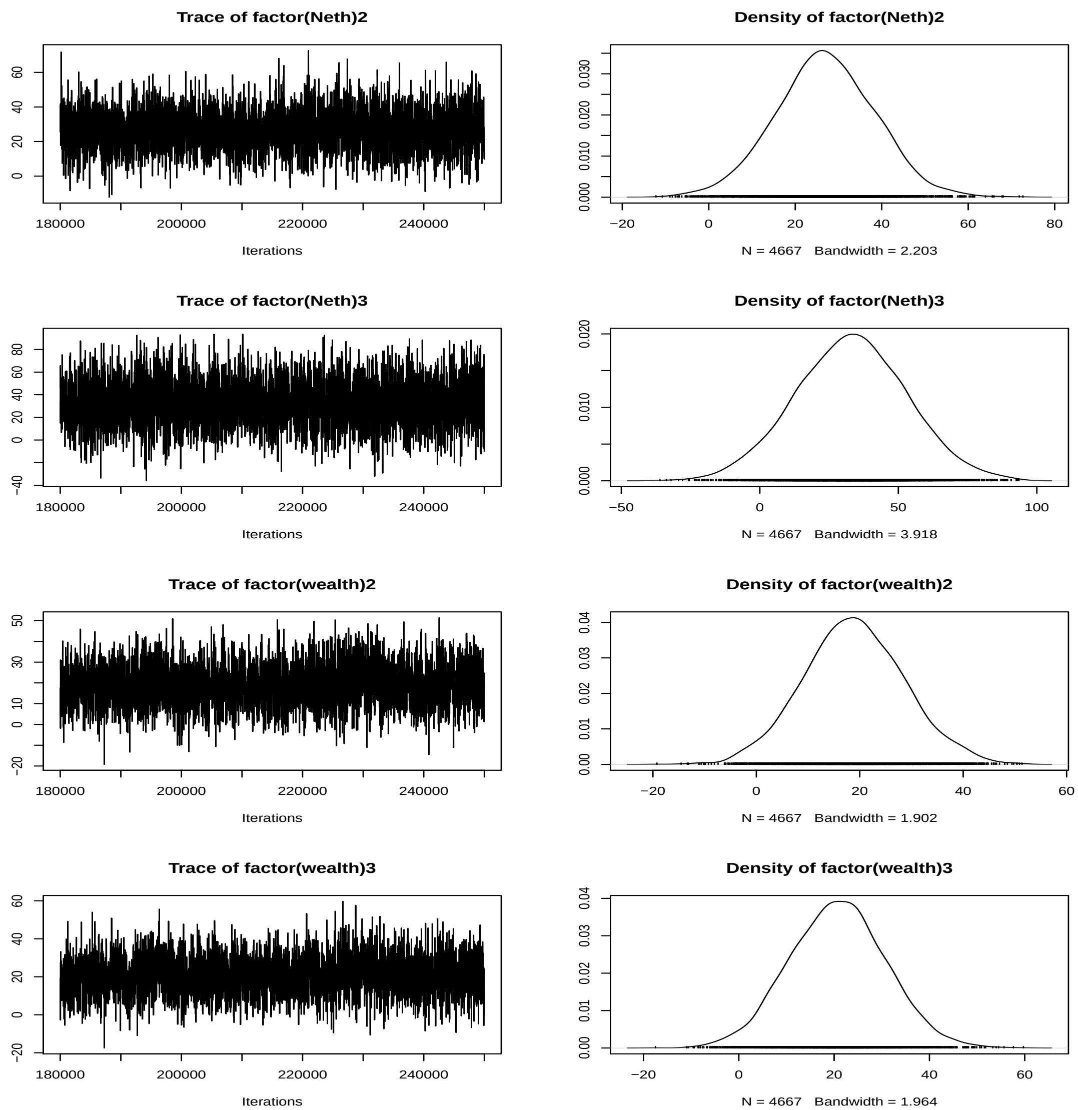


Figure 4.8: Trace plot for fixed effects: Race and wealth index.

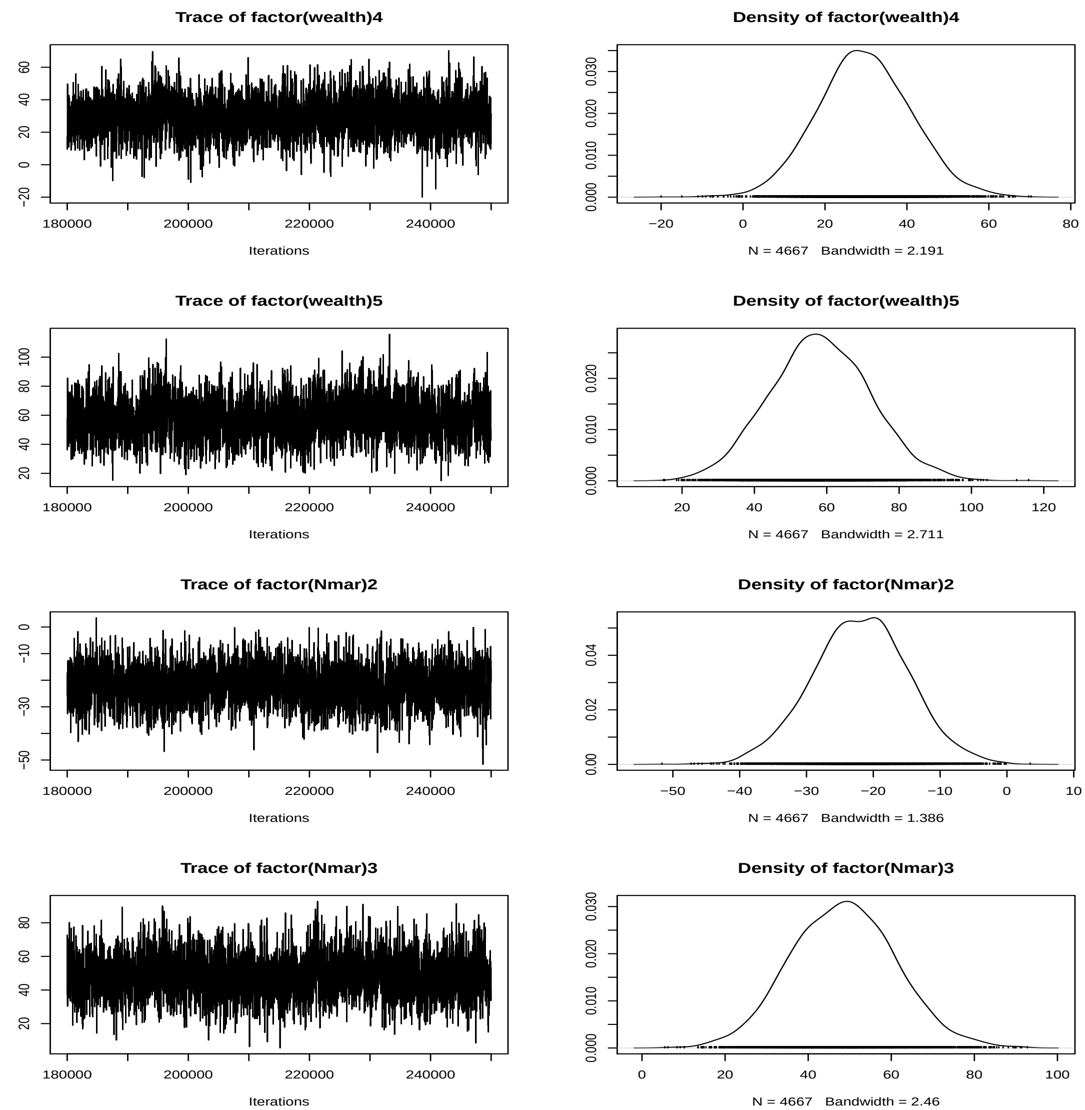


Figure 4.9: Trace plot for fixed effects: Marital status and wealth index.

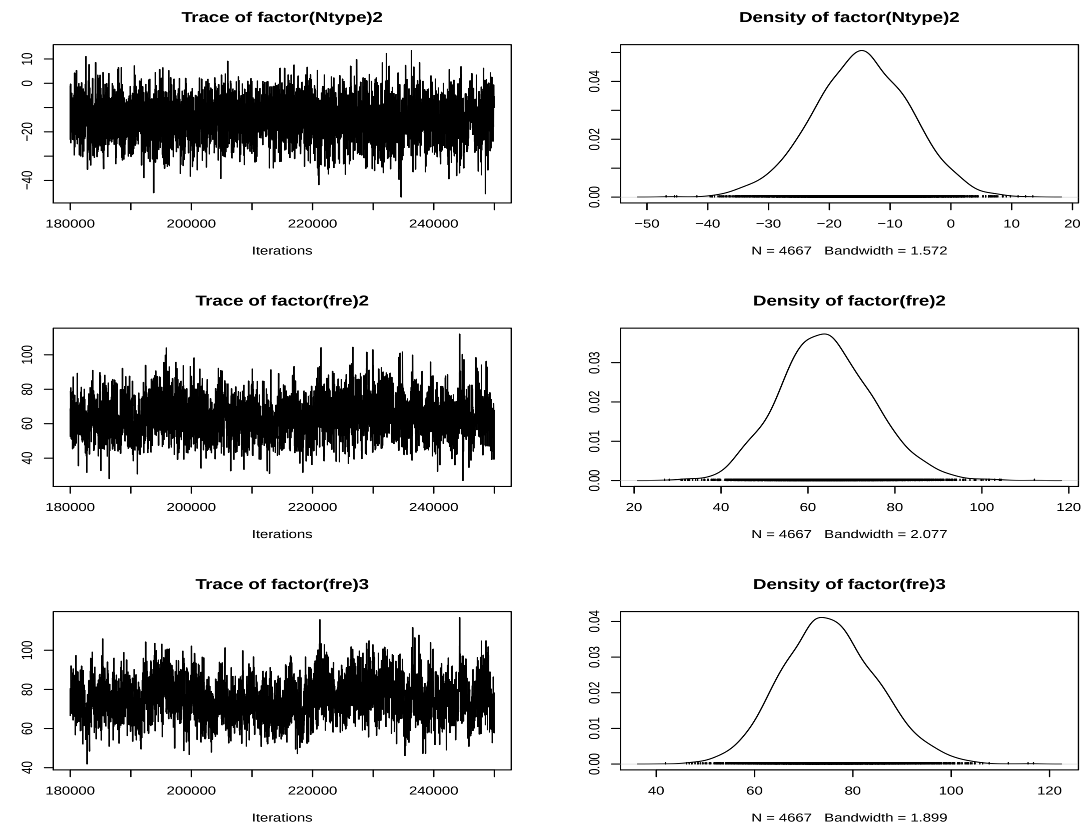


Figure 4.10: Trace plot for fixed effects: Type of residence and internet usage.

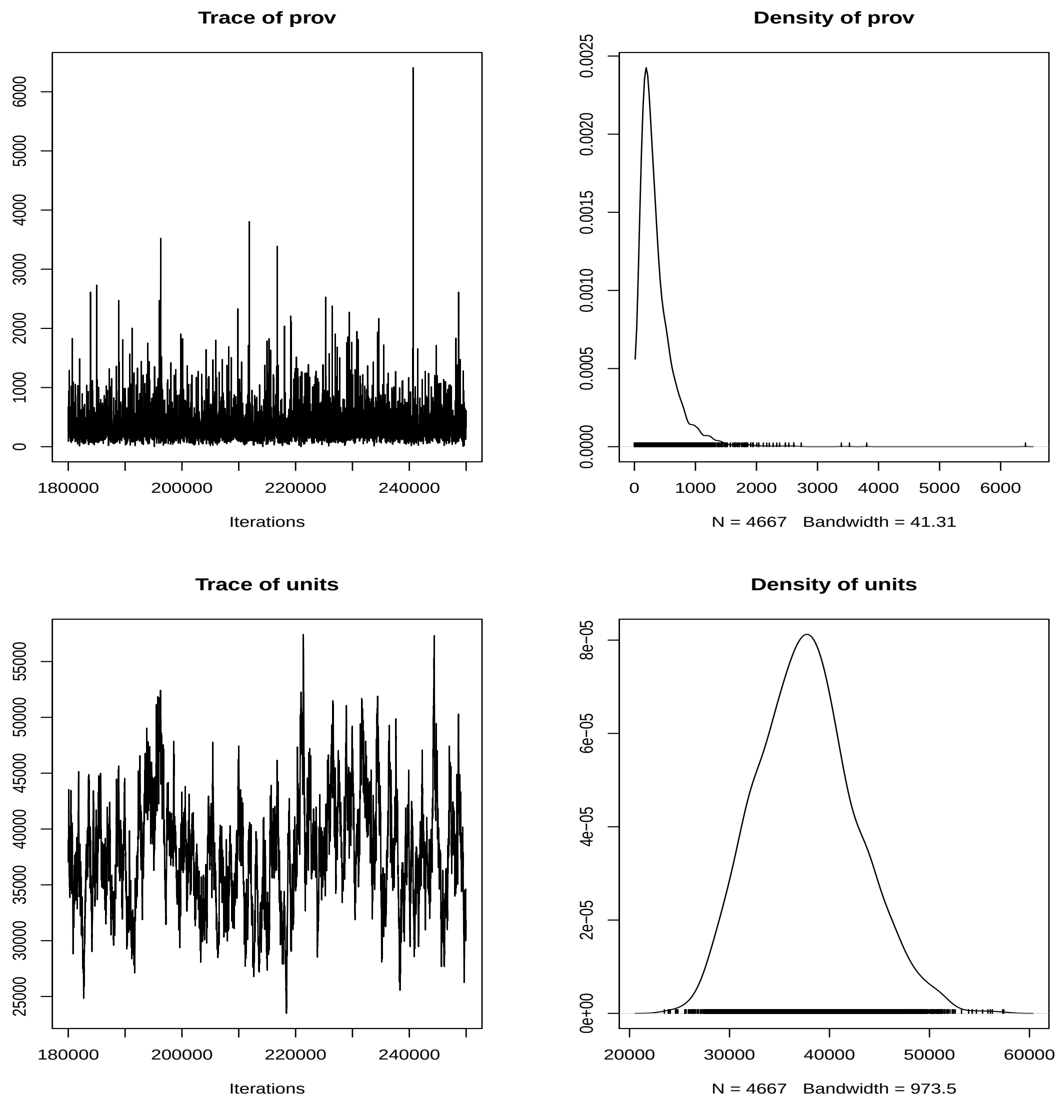


Figure 4.11: Trace plot for random effects: province and residuals.

4.6 Summary and discussion of the results

This study aimed at investing the relationship between a woman's employment status and different explanatory variables using SADHS data. The SADHS data exhibit a multi-layered structure which requires methods that are proficient in considering the underlying processes that are responsible for the observed data structure. The multi-level models are appropriate for describing sources of variability in a response variable across various levels in multilevel and clustered data. A Bayesian approach was adopted to compute a multilevel logistic regression model from a GLMM perspective to explain the variation in women unemployment. In the Bayesian inference, the maximum likelihood estimation methods of the parameter estimation are not tractable, hence the iterative methods based on the MCMC simulations were utilized. The posterior summary measures in the form of the means and the credibility intervals were determined from depicting random samples from the posterior distribution. The prior distributions for the fixed effects were assumed to be multivariate normal parameterized with mean $\boldsymbol{\mu}$ and the variance-covariance matrix $\boldsymbol{V}$, whereas the prior distributions for the random effects were presumed to be the non-informative inverse-gamma. It was found that a woman's employment status is dependent on one's age group, marital status, educational attainment, wealth index, place of residence type, race, and internet-use. The combination of the multilevel modelling and the Bayesian inference intensified the identification of risk factors for women's unemployment status.

Chapter 5

Conclusion and recommendation

5.1 Conclusion

In this project, the women's unemployment status was being examined using both the frequentist and the Bayesian multilevel models to discover the extent to which unemployment status is associated with different explanatory variables. The SADHS data was collected via multistage sampling, that is, the data has some levels or groupings. The nesting in the data is in a way that the individuals are nested within the families or the households and the households are nested in the clusters or provinces. This study was restricted to two levels, the level 1 being individuals and the level two being the provinces, that is, the households was ignored as a level for simplicity. The resultant variable of this study is binary, and due to the existence of the random effects or the levels, the GLMMs and the multilevel models were considered to be suitable in the analysis of the data. The multilevel models vary from the ordinary GLMMs in the sense that they include predictors at individual and group levels to reduce the unexplained variation in each level, (Gelman and Hill, 2007). In seeking to identify key drivers of women's unemployment, it was revealed that women's unemployment is quite prevalent in South Africa with a rate of 65.8%. The majority of the unemployed women were found in the age group 15-19 years. This may be due to that a woman in the age group 15-19 is not most affected by unemployment since some can be still at school and not seeking employment. This study also found that the unemployment status of a woman is less likely for a woman who has secondary and above educational attainment as compared to a woman who is not educated. Unemployment was found to decrease with an increased educational attainment. This study found that age group, marital status, educational attainment, internet-use, residence type, race, and wealth index were associated with the women's unemployment status.

The methodology revealed that the multilevel random intercept model outperformed the null model and the random coefficient model in fitting the data and in explaining the variations of woman unemployment status across the provincial levels of South Africa. Furthermore, from the random intercept model, the general variance of the constant term was found to be statistically significant, indicating the existence of the difference in unemployment status of the women among provinces of South Africa. This implies that the woman with the same qualities in two different provinces has various unemployment status, that is, there is a clear provincial effect.

5.2 Recommendations

The results acquired in this study are of greater importance to policymakers because of the negative effects of unemployment on the decline of the yield, on society, and the psychological wellbeing of the unemployed and close family members. To construct policies to standardise the rising problem of unemployment in South Africa, it is important not only to understand the effect of the improvement on the incidence of women unemployment but also on the span of the unemployment and on the probability of exiting unemployment and how it differs with demographic and economic characteristics. This study suggests that policies that encourage educational attainment should be formulated and the creation of more jobs opportunity should be implemented. Also, for the government to take measures of action in supporting the poor and bringing rapid economic growth at the national level. Last but not least, community-based development interventions giving priority to the poor households to participate in the labor market, health facility, education, and access to job areas. The policies focusing on the reduction of poverty should also be implemented to reduce unemployment amongst South African women.

5.3 Limitation

This study used secondary data obtained from the Demographic and Health Survey data conducted in South Africa in 2016. The potential limitation of the current study comes from the use of secondary data which often leaves the researcher with limited control over the data collection process. For the current project, a major limitation of using secondary is that some of the variables such as the type of training and steps taken to search work were not contained in the data set used and these variables would have helped to understand the prevalence of women unemployment status in South Africa. This study only analysed data on unemployment only for women and was not able to explore the link of unemployment between men and women.

5.4 Future work

Further work can be done on the sensitivity analysis for the bayesian multilevel model estimates to the prior distribution and also on missing data as the dataset of this study contained no missing data. Lastly, further work can also be done on the contribution of the culture of unemployment in South Africa.

References

A. Abraham, F. Ohemeng, and W. Ohemeng. Female labour force participation: evidence from Ghana. *International Journal of Social Economics*, 44(11):1489–1505, 2017.

H. Acquah. Comparison of akaike information criterion (aic) and bayesian information criterion (bic) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2(1):001–006, 2010.

J. Aldrich et al. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical science*, 12(3):162–176, 1997.

D. Ardia and L. Hoogerheide. Bayesian estimation of the garch (1, 1) model with student-t innovations. 2009.

P. Austin. Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures. *The international journal of biostatistics*, 6(1), 2010.

W. Baah-Boateng. Employment generation for poverty alleviation. *Poverty Reduction Strategies in Action: Perspectives and Lessons from Ghana*, page 223, 2008.

W. Baah-Boateng. Unemployment in Ghana: a cross sectional analysis from demand and supply perspectives. *African Journal of Economic and Management Studies*, 6 (4):402–415, 2015.

W. Baah-Boateng, Y. Ansu, and J. Amoako-Tuffour. Mapping of country information on employment, unemployment and policy initiatives. *Report submitted to African Centre for Economic Transformation (ACET), for Knowledge Platform Development Policies*, 2013.

A. Banerjee, S. Galiani, S. Levinsohn, Z. McLaren, and I. Woolard. Why has unemployment risen in the new South Africa? 1. *Economics of Transition*, 16(4):715–740, 2008.

D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015. doi: 10.18637/jss.v067.i01.

M. Bayarri and J. Berger. The interplay of bayesian and frequentist analysis. *Statistical Science*, pages 58–80, 2004.

J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of extremes: theory and applications*. John Wiley & Sons, 2006.

E. Bertoni and G. Ricchiuti. A multilevel analysis of unemployment in Egypt. *LABOUR*, 31(4):494–514, 2017.

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

C. Chatfield, J. Zidek, and J. Lindsey. An introduction to generalized linear models. 2010.

G. Che and F. Sundjo. Determinants of female labour force participation in Cameroon. *International Journal of Applied Economics, Finance and Accounting*, 3(2):88–103, 2018.

T. Chikako. Multilevel modelling of determinants of youth unemployment in urban

Ethiopia: Bayesian estimation approach. *International Journal on Data Science and Technology*, 4(2):67, 2018.

S. Coles, J. Bawa, L. Trenner, and P. Dorazio. *An introduction to statistical modeling of extreme values*, volume 208. Springer, 2001.

P. Congdon. *Bayesian models for categorical data.* John Wiley & Sons, 2005.

M. Dagume and A. Gyekye. Determinants of youth unemployment in South Africa: evidence from the Vhembe district of Limpopo province. 2016.

D. De Moortel, P. Hagedoorn, C. Vanroelen, and S. Gadeyne. Employment status and mortality in the context of high and low regional unemployment levels in Belgium (2001–2011): A test of the social norm hypothesis across educational levels. *PloS one*, 13(2):e0192526, 2018.

N. Drydakis. The effect of unemployment on self-reported health and mental health in Greece from 2008 to 2013: a longitudinal study before and during the financial crisis. *Social Science & Medicine*, 128:43–51, 2015.

L. Fahrmeir and G. Tutz. *Multivariate statistical modelling based on generalized linear models.* Springer Science & Business Media, 2013.

D. Fergusson, L. Horwood, and L. Woodward. Unemployment and psychosocial adjustment in young adults: causation or selection? *Social science & medicine*, 53(3): 305–320, 2001.

E. Ford, C. Clark, S. McManus, J. Harris, R. Jenkins, P. Bebbington, T. Brugha, H. Meltzer, and S. Stansfeld. Common mental disorders, unemployment and welfare benefits in England. *Public health*, 124(12):675–681, 2010.

A. Gelfand, S. Hills, A. Racine-Poon, and A. Smith. Illustration of bayesian inference in normal data models using gibbs sampling. *Journal of the American Statistical Association*, 85(412):972–985, 1990.

A. Gelman and J. Hill. *Data analysis using regression and multilevelhierarchical models*, volume 1. Cambridge University Press New York, NY, USA, 2007.

A. Gelman and X.-L. Meng. *Applied Bayesian modeling and causal inference from incomplete-data perspectives*. John Wiley & Sons, 2004.

A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian Data Analysis*. 2004.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6):721–741, 1984.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments. *Bayesian statistics*, 4:641–649, 1992.

H. Goldstein. *Multilevel statistical models*, volume 922. John Wiley & Sons, 2011.

K. Gyekye and K. Kyei. Determinants of unemployment in Limpopo province in South Africa: exploratory studies. *Journal of Emerging Trends in Economics and Management Sciences*, 2(1):54–61, 2011.

J. Hadfield. Mcmcglmm: Markov chain monte carlo methods for generalised linear mixed models. *Tutorial for MCMCglmm package in R*, 125, 2010.

J. Hadfield. Mcmcglmm course notes (accessed august 7, 2014). https://cran. r-project. org/web/packages/mcmcglmm/vignettes. *CourseNotes. pdf*, 2014.

T. Hammer. History dependence in youth unemployment. *European Sociological Review*, 13(1):17–33, 1997.

F. Hartig. *DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models*, 2020. URL http://florianhartig.github.io/DHARMa/. R package version 0.3.3.0.

W. Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

P. Heidelberger and P. D. Welch. Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144, 1983.

H. Iweagu and N. Chukwudi. Determinants of female labour force participation in Nigeria: The rural/urban dichotomy. *Journal of Economics and Sustainable Development*, 6(10):212–219, 2015.

R. Kessler, J. House, and J. Turner. Unemployment and health in a community sample. *Journal of health and social behavior*, pages 51–59, 1987.

G. Kingdon and J. Knight. Unemployment in South Africa: The nature of the beast. *World development*, 32(3):391–408, 2004.

K. Kinney, S and D. Dunson. Fixed and random effects selection in linear and logistic models. *Biometrics*, 63(3):690–698, 2007.

H. Kiros and A. Abebe. Statistical modeling of women employment status at Harari region urban districts: Bayesian approach. *Annals of Data Science*, pages 1–14, 2019.

J. Kruschke and W. Vanpaemel. Bayesian estimation in hierarchical models. *The Oxford handbook of computational and mathematical psychology*, pages 279–299, 2015.

D. Lam, M. Leibbrandt, and C. Mlatsheni. Education and youth unemployment in south africa. 2007.

E. Latif. Crisis, unemployment and psychological wellbeing in Canada. *Journal of Policy Modeling*, 32(4):520–530, 2010.

E. Lesaffre and A. B. Lawson. *Bayesian biostatistics*. John Wiley & Sons, 2012.

A. Lundin, I. Lundberg, P. Allebeck, and T. Hemmingsson. Unemployment and suicide in the stockholm population: a register-based study on 771,068 men and women. *Public health*, 126(5):371–377, 2012.

J. Marin, R. Diez, and j. v. n. p. y. p. Insua, DR. Bayesian methods in plant conservation biology.

I. Mathebula. *Determinants of unemployment and earnings in South Africa*. PhD thesis, University of Limpopo, 2017.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21 (6):1087–1092, 1953.

C. Nicoletti and C. Rondinelli. The (mis) specification of discrete duration models with unobserved heterogeneity: a monte carlo study. *Journal of Econometrics*, 159(1): 1–13, 2010.

S. O'brien and D. Dunson. Bayesian multivariate logistic regression. *Biometrics*, 60(3): 739–746, 2004.

A. Page, A. Milner, S. Morrell, and R. Taylor. The role of under-employment and unemployment in recent birth cohort effects in australian suicide. *Social science & medicine*, 93:155–162, 2013.

S. Phillips. The expanded public works programme (epwp). In *Presentation to the Conference on Overcoming Under-Development in South Africa's Second Economy, Pretoria*, 2004.

W. Qayyum and R. Siddiqui. Causes of youth unemployment in Pakistan [with comments]. *The Pakistan Development Review*, pages 611–621, 2007.

H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

H. Sackey. Female labour force participation in Ghana: The effects of education. 2005.

M. Taamouti and M. Ziroili. Individual determinants of female labor participation in Morocco. Technical report, mimeo, 2011.

V. Tandrayen-Ragoobur and H. Kasseeah. Youth unemployment in Mauritius: The ticking bomb. *Urban Youth Unemployment in Eastern and Southern Africa*, page 177, 2015.

R. C. Team et al. R: A language and environment for statistical computing, 2013.

I. Theodossiou. The effects of low-pay and unemployment on psychological well-being: a logistic regression approach. *Journal of health economics*, 17(1):85–104, 1998.

O. Usman and A. Sanusi. Education and labor force participation of women in North Cyprus: Evidence from binomial logit regression model. 2016.

D. Viljoen and S. Dunga. Determining the factors that influence female unemployment in a South African township. *International Journal of Social Sciences and Humanity Studies*, 5(1):63–72, 2013.

N. Wege, P. Angerer, and J. Li. Effects of lifetime unemployment experience and job insecurity on two-year risk of physician-diagnosed incident depression in the German working population. *International journal of environmental research and public health*, 14(8):904, 2017.

Y. Yakubu. Factors influencing female labor force participation in South Africa in 2008. *The African Statistical Journal*, 11:85–104, 2010.

Appendices

Before fitting any model with the data, combination of variables was done

LITERACY STATUS

```
Nlit<-NULL
for(i in 1:length(dataset$literacy)){
if((dataset$literacy[i]=="Blind/visually impaired")){
}else if((dataset$literacy[i]=="Cannot read at all")
|(dataset$literacy[i]=="No card with required language")){
Nlit[i]<-2
}else{Nlit[i]<-3}
}
Nlit <-factor(Nlit,levels=c(1, 2,3),
labels=c("Blind","Illiterate","Literate"))
```

MARITAL STATUS

```
Nmar<-NULL
for(i in 1:length(dataset$maritals)){
if((dataset$$maritals[i]=="Never in union")){
Nmar[i]<-1
}else if((dataset$$maritals[i]==
"Currently in union/living with a man")){
```

```
Nmar[i]<-2
}else if((dataset$$wealth[i]=="Richer")
|(dataset$$wealth[i] =="Richest")){
}else{Nmar[i]<-3}
}
Nmar <-factor(Nmar,levels=c(1, 2,3),
labels=c("Single","Married","Divorced/widow"))

ETHNICITY
Neth<-NULL
for(i in 1:length(dataset$Race)){
if((dataset$$Race[i]=="Black/African")){
Neth[i]<-1
} else if((dataset$Race[i]=="Coloured")|(dataset$$Race[i]==
"Indian/Asian")|(dataset$$Race[i]=="Other")){
Neth[i]<-2
}else{Neth[i]<-3}
}
Neth <-factor(Neth,levels=c(1:3),labels=c("Black/African",  "Other","White"))

INTERNET USE
fre<-NULL
for(i in 1:length(dataset$frequency)){
if((dataset$frequency[i]=="Not at all")){
fre[i]<-1
}else if((dataset$frequency[i]=="Less than once a week")
|(dataset$frequency[i]=="At least once a week")){
fre[i]<-2
}else{fre[i]<-3}
```

```
}
fre <-factor(fre,levels=c(1, 2,3),
labels=c("Never","weekly","Daily"))

FITTING THE NULL MODEL
model1<-glmer(currentwork~1+(1|prov),family=
binomial,data=dataset)
summary(model1)
confint(model1)
ICC(outcome = "currentwork",group = "prov" ,data = dataset)

RANDOM INTERCEPT LOGISTIC REGRESSION
model3<-glmer((currentwork)~(Edu)+(type)+(Agegrp)+
(Neth)+(NWEL)+(fre)+(Nlit)+(Nmar)+(1|prov),
data=dataset,family=binomial(link="logit"),
control = glmerControl(optimizer = "bobyqa"))
summary(model3)

RANDOM COEFFICIENT LOGISTIC REGRESSION
model2<-glmer(currentwork~Edu+type+Agegrp+Neth+NWEL+
fre+Nlit+Nmar+(Edu|prov),data=dataset,
family = binomial(link="logit"),control =
glmerControl(optimizer = "bobyqa"))
summary(model2)

BEST FIT MODEL
anova(model1, model2, model3)

BAYESIAN NULL MODEL
```

```
prior = list(R = list(V = 2, nu = 0.004), G = list(G1 =
 list(V = 2, nu = 0.004),
G2 = list(V = 2, nu = 0.004)))
model4 <- MCMCglmm(currentwork ~ 1, random = ~prov,
                    rcov = ~idh(prov):units, data = newdataset,
                     prior = prior, verbose = FALSE)
summary(model4)

BAYESIAN RANDOM INTERCEPT MODEL
prior = list(R = list(V = 1, nu = 0.002), G = list(G1 = list(V = 1, nu = 0.002),
                                                   G2 = list(V = 1, nu = 0.002)))
model5<-(MCMCglmm(currentwork~(Edu)+(Agegrp)+
                        (Nlit)+(Neth)+(NWEL)+(Nmar)+(type)+(fre)
                        , random = ~prov, data=newdataset,
                      verbose=FALSE,prior= prior,nitt=10000,
                      burnin=100,thin=10,family= "categorical"))
prior = list(R = list(V = 2, nu = 0.004), G = list(G1 = list(V = 2, nu = 0.004),
            G2 = list(V = 2, nu = 0.004)))
             plot(model5)
              summary(model5)
```