Hierarchical Multiple Linear Regression: A Comparative Analysis of Classical and Bayesian Approach

By

Mzwandile Lindokuhle Gabela

212524545

A dissertation submitted in fulfilment of the academic requirement of

Master of Science in Statistics

School of Mathematics, Computer Science and Statistics

College of Agriculture, Engineering and Science

Supervisor: Dr Siaka Lougue

November 2017

Disclaimer

This document describes work undertaken as part of a Masters Research Programme of Study at the University of KwaZulu-Natal (UKZN). All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the institution.

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Statistics, School of Mathematics, Statistics and Computer Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville campus, South Africa. The research was financially supported by DELTA Scholarship.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.

29 NOV 2017

Mzwandile Lindokuhle Gabela (Student) — Date

29 NOV 2017

Dr Siaka Lougue (Supervisor) — Date

Abstract

Low birth weight is a problem in Africa due to its contribution to high infant mortality. Most studies on low birth weight have neglected the use of Bayesian methods in analysis of medical data. This study aims to investigate the risk factors of low birth weight in Malawi. Malawi is a country in the sub-Saharan region which is characterized by infant and child mortality of 12%. Inferences made in this study are only based on Malawi demographic and health survey data for the three years 2000, 2004 and 2010. The year 2010 Malawi Demographic and Health Survey data is used to make classical and Bayesian analysis for the study. The years 2000 and 2004 data are used to set up the prior information for Bayesian approach to the analysis. Data will be analysed using descriptive and inferential statistics. Hierarchy is taken into consideration since some of the risk factors are known to be hierarchical. The hierarchical multivariate linear regression analysis is done in a comparative procedure of Classical and Bayesian approach. The study shows that the age of the mother, birth order number of a child, region, work during pregnancy and HIV status of the mother are significant determinants of birth weight. In the comparison of classical and Bayesian approach it was found that all the variables that were significant in the Bayesian approach were also significant in the classical approach and the opposite was not true. The use of the Bayesian prior in the analysis gave more realistic results on factors of weight at birth.

Key Words:

Underweight at birth, MCMC algorithm, Bayesian linear regression, Bayesian multilevel linear regression, Bayesian and Classical techniques, Bayesian hierarchical linear regression

Contents

List of Figures

List of Tables

Acknowledgements

This work was supported through the DELTAS Africa Initiative. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107754/Z/15/Z- DELTAS Africa Sub-Saharan Africa Consortium for Advanced Biostatistics (SSACAB) programme] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

I would like to express immeasurable appreciation and deepest gratitude to the following persons; Dr. Siaka Lougue, supervisor and statistician, for his support, advice, guidance, valuable comments and suggestions. Sharing his knowledge has helped in the analysis of data and its statistical computations. The author is most grateful for having such an advisor who has given endless help to finish this project.

The financial assistance from DELTA scholarship which Professor Henry Godwell Mwambi (in the division of School of Mathematics, Statistics and Computer Science) assisted with is hereby acknowledged. Finally yet importantly, my family who has been the motive behind my hard work. I will make you proud.

Chapter 1: Introduction

Low Birth Weight (LBW) has been defined by the World Health Organization (WHO) as weight less than 2.5kg, although between 2.5kg and 4.5kg are considered normal. Across the world, 20 million low birth weights are recorded annually, and this equates to 15.5% of all births (WHO, 2004). Approximately every ten seconds, an infant from a developing country dies from a disease or infection that can be attributed to low birth weight (Judith & Laura, 2000). This study aimed to investigate risk factors of low birth weight in Malawi using the Malawi Demographic and Health Survey data for the years 2010, 2004 and 2000. The data for the year 2004 and year 2000 is used to develop the prior information for the Bayesian approach.

Birth weight is a dominant predictor of infant's growth and survival. A child born with a low birth weight begins life immediately at a disadvantage and faces extremely poor survival rates. Death of a child is commonly a result of several risk factors (Black, Morris and Bryce, 2003). The data on child's birth weight is often under-reported because more than 40% of infants born in developing countries are not weighed at birth (WHO, 2004). According to UNICEF and WHO, half of low birth weight children are in the South Central Asia where more than a quarter of all children born weigh less than 2.5kg at birth. This represents 27% of all new births with LBW. Sub-Saharan Africa has the second highest incidence of low birth weight, which is 15%. Malawi is part of sub-Saharan Africa with the latest low birth weight incidence at 12%. Although birth weight is an important predictor of subsequent health outcomes in rural or developing countries, the role of the factors in this study is not understood completely as birth weight risk influence. Determining whether these factors are the risk influences for birth weight or not can help reduce the incidence of low birth weight and maximize infant health in Malawi. Despite the importance of weight at birth, there are insufficient advanced statistical studies on the subject. Most existing studies of weight at birth use classical statistical methods and are usually limited to descriptive or bivariate analysis level (Ngwira, A., & Stanley, C. C., 2015).

Bayesian methods are gaining momentum in statistical data analysis. Based on the importance of the subject in public health and the need for deep statistical analyses, this study was initiated to contribute to better understanding of the problems of low weight at birth by improving statistical analysis technique. More specifically, the comparison of Bayesian and Classical statistics is made on factors of weight at birth. The independent variables for this study include: which technique best analyse weight at birth, age of a mother, child’s birth order number, gender of a child, mother’s level of education, if the mother was working during pregnancy, antenatal visits for pregnancy and the HIV status of the mother. Since the advent of Markov Chain Monte Carlo (MCMC) and the improvement in speed memory of computers methods in the early 1990s, Bayesian methods have been extended to a large and growing number of applications. Aside from underlying philosophical differences of these two approaches, many readers will be comforted in finding that Bayesian and non-Bayesian analysis often agree to some extent. There are two instances where this is always true. Firstly, when the included prior information is uninformative (there are several ways of providing this), summary statements from Bayesian inferences will match most frequent point estimates (Bill, 1999). Secondly, when the data is very large, the form of the prior information used does not matter and there is an agreement again (Denis, 2005). Other circumstances exist in which Bayesian and non-Bayesian statistical inferences may lead to similar results. It is therefore important to investigate the factors associated with infant and child mortality in Malawi using a different approach of comparing two data analysis results. Furthermore, the factors affecting low weight at birth of a child are known to be hierarchical, occurring at both individual-level and the regional-level with complex interactions. Most previous studies have focused only on individual-level factors of low birth weight and few studies have considered the hierarchical nature of the problem. This research will add to the existing literature on infant and child mortality in Malawi.

This thesis is structured into four chapters:

1. An Introduction to the Topic
2. Methods
3. Results
4. Discussion and Conclusion.

Chapter 2: Literature Review

2.1 Context of the Study

The study mainly focuses on factors of weight at birth of a child in Malawi. This chapter discusses the determinants of child birth weight to be studied and also the population growth in Malawi. These discussed factors are general findings from previous authors to the topic. To incorporate the results to this study, a comparison of Bayesian and classical statistical models has been developed to find which factors best affects weight at birth of a child in Malawi.

Table 1: 2012 Population in Malawi (Malawi Population Data Sheet, 2012)

	Population 2012 (Millions)		
	Total	Women	Men
TOTAL	14.8	7.7	7.1
Rural	12.5	6.5	6.0
Urban	2.3	1.2	1.1
North	1.9	1.0	0.9
Central	6.3	3.2	3.1
South	6.7	3.5	3.2

Table 1 shows that there was a large proportion of people that are based in rural areas than in urban arears. Women had the largest population than men in all regions.

Figure 1 below shows that the population of Malawi is growing faster in the Southern districts than in the Northern regions.

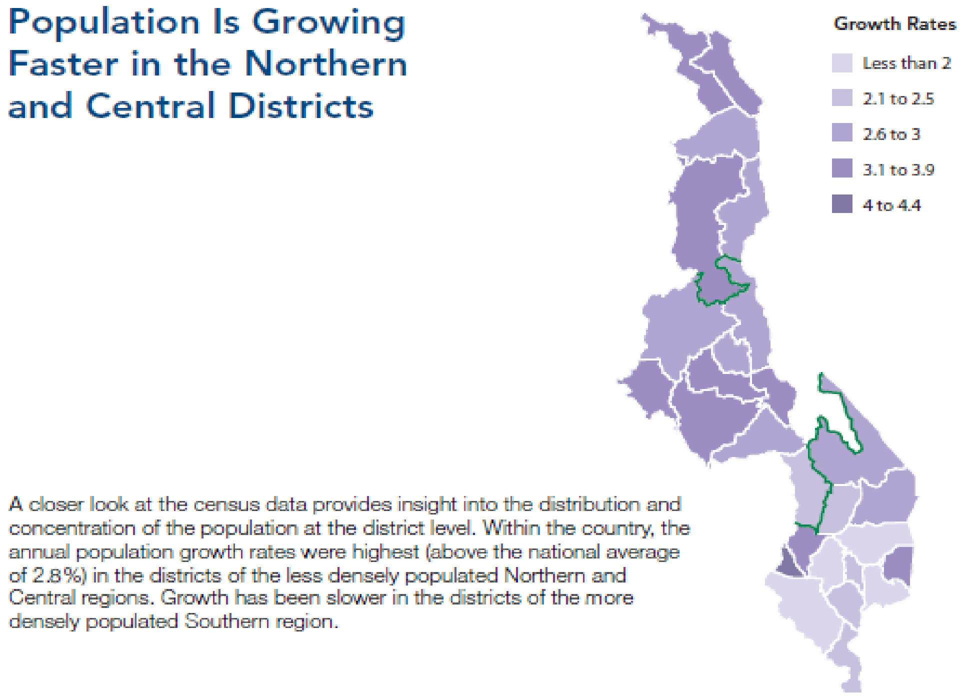


Figure 1: Population growth in Regions of Malawi (Malawi Population Data Sheet, 2012)

2.2 Factors of Low Birth Weight

Low birth weight can bring about significant general problems and impacts on a child's growth. Previous studies on the topic show that there are several factors that are the cause of an infant being underweight at birth (Black, R. E., Morris, S. S., & Bryce, J., 2003). These risk factors include age of a mother, child's birth order number, region of birth, gender of a child, mother's educational level, if the mother of a child was working or not during pregnancy, antenatal visits during pregnancy and HIV status of a mother. Weight at birth of a child is an important indicator of the child's health (Institute of Medicine, 1985). To understand these factors further, a review of previous studies is summarized for each the above-mentioned factors of underweight children.

The age of a mother at the time of delivery is an important aspect in this study and one of the factors that might have a negative effect on a child's weight at birth. In making inferences about the effects of these factors on weight at birth of a child, the study focuses mainly on mothers with ages of between 15 and 49 years. It is common in Malawi to have children born out of wedlock. Overall, the median age for first birth for women in Malawi is 18.9 years. Seven percent of women who are between ages 25 and 49 years gave birth by the age of 15 across age groups. The percentage of women who gave birth by exactly the age of 15 is 7% or higher among women aged 35-49, around 5% among women aged 20-34, and less than 2% among women aged 15-19. Previous studies have shown that mothers who are younger than 15 years of age have the lowest weight at birth results (Reichman and Padilla, 1997). Approximately 11% of all births worldwide are to mothers who are aged 15-19 years old (Gibbs et al., 2012). These young mothers are associated with higher risk of having adverse birth outcomes than older mothers (Okosun et al., 2000). Gibbs et al. (2012) conducted a study which showed that there is a dose-response relationship between maternal age and low birth weight that decreased in magnitude as maternal age increased. On the other hand, mother's level of education is one of the important significant effects that has both positive and negative results on the birth weight of a child. In Malawi, the proportion of people who have never attended school is higher for females than for males across all age groups, and the proportion with some primary education is about the same among men (65%) and women (64%). However, more males have attended or completed secondary education than women (17% compared to 11%), (Malawi DHS, 2010).

According to Silvestrin et al. (2013), education is the strongest socio-economic predictor of health status when considered alone and the most important determinant of birth weight in a population. Previous studies have shown that mothers who are better educated are at an advantage of not being the victims of risks associated with low birth weight of their children (Ngwira, A., & Stanley, C. C., 2015). The lower the education level, the greater the vulnerability of delivering a baby with a low birth weight or having an underweight child. There is a 33% protection effect against low birth weight for women that have higher education and a 9% higher probability of having a low birth weight child if the mother has not finished high school Silvestrin et al. (2013).

The variables "age" and "education level" of a mother are independent variables that have a relationship for this study. Young mothers generally attain a lower education level than mothers in normal and older age groups. This is because the average age of high school women in Malawi is 18 years. Education plays an important role in pregnancy and ties directly to many other risk factors. The Malawi Demographic and Health Survey (2004) estimated 16.3% of adult women and 10.1% of adult men had no formal education, while 12.2% of women and 18.8% of men had secondary education or higher.

Malawi is divided into three administrative regions: The Northern, Central and Southern regions. Birth weight has been studied across the three regions. These regions differ in many regards on factors such as majority ethnic group, wealth and education status. Based on education level in Malawi, the median number of years of school completed is highest for women from the Northern Region at 6.8 years, followed by 4.7 years in the Southern Region, and 4.3 years in the Central Region (Malawi DHS, 2010). Based on mother's age in Malawi, the proportion of teenagers who have started childbearing is highest in the Southern Region (29%) and the Northern Region (28%) when compared with the Central Region (22%) (Malawi DHS, 2010). Due to regional differences, the birth weight of a child is affected differently by each factor.

Human Immunodeficiency Virus (HIV) status of a mother is one of the factors that might have a significant effect on the birth weight of a child. This is because HIV can infect infants during pregnancy, during labour and delivery, and after delivery when the infant may become infected with HIV present in breast milk (Bettercare, 2016). An HIV/AIDS (UNAIDS) 2010 global report stated that there were 920,000 adults and children living with HIV in Malawi in 2009 (UNAIDS, 2010). In 2015 Malawi's HIV prevalence shows that 10.3% of the population was living with HIV. Malawi accounted for 4% of the total number of people living with HIV in sub-Saharan Africa which then makes this variable an important one for the model. Major factors in the transmission of HIV in Malawi are poverty, low literacy levels, high rates of casual and transactional unprotected sex in the general population, (particularly among youth between the ages of 15 and 24 years), low levels of male and female condom use, cultural and religious factors, and stigma and discrimination (UNAIDS, 2010).

Research has shown that mothers who stopped smoking during pregnancy had an increase in the weight of their babies at delivery, in comparison to mothers who smoked for the entirety of their pregnancy (Horta et al., 1997). The prevalence of smoking tobacco, particularly cigarettes, increases for pregnant Black mothers with age (Rich-Edwards et al., 2003). Income and socio-economic status is another factor. Poverty is usually the culprit and the physical demands on these mothers to earn wages can contribute to poor foetal growth (WHO, 2004). Mothers in deprived socio-economic conditions frequently have low birth weight babies (WHO, 2004).

Chapter 3: Methods

3.1 Data

In this study, the data utilised for the analysis is the year 2010 Malawi Demographic and Health Survey data (2010 MDHS). This survey is a large, nationally-representative sample survey conducted by the National Statistical Office (NSO) with Ministry of Health Community Sciences Unit (CHSU). The primary objectives of the 2010 Malawi Demographic and Health survey are to provide reliable and updated information. The methodology of the MDHS survey is based on a stratified sampling frame that was used to collect information on demographic issues relevant to Malawi. All the data for the years 2010, 2004 and 2000 was collected using the same methodology. Analysis for both classical and Bayesian approach focussed on the 2010 data. The year 2004 and 2000 data is only utilized to build an informative prior for the Bayesian approach to the analysis. The sample for the 2010 MDHS was designed to provide population and health indicators at the national, regional, and district levels. The overall data collected using the sampling frame was stratified into 27 districts of Malawi. For each district, the sampling frame was further stratified under the categories “urban” and “rural” areas. The overall data collected from the MDHS focused on almost all factors affecting child’s weight at birth (less than average, average and higher than average) in Malawi. This study only focuses on children with birth weight of 2.5kg or less, provided that there were no missing observations in the observed explanatory variables. Predictors included in this study were identified from the literature review and also by their availability.

The total number of households that were successfully interviewed were 24 825, yielding a response rate of 98% in the overall data. The number of eligible women was between 23 748 and 23 020 and these were successfully interviewed, yielding a response rate of 97%. Eligible men numbered 7783 and those that were successfully interviewed numbered 7175, yielding a response rate of 92%. The response variable is the child's birth weight in grams. The final sample size consisted of 13 087 mothers who had answered all the questions for the variables used in the analysis. The age of the mother was restricted to between 15 and 49 years of age. Subjects with missing observations for any of the studied variables in Malawi data were excluded from the study. This procedure is applied to the year 2000, 2004 and 2010 data. MDHS data for the years 2000 and 2004 is used to set up the prior information needed for Bayesian technique implementation. The following programs were used to do the data analysis:

- R-Studio used to fit: Classical Multiple Linear Regression and Classical Multilevel Linear Regression
- WinBUGS used to fit: Bayesian Multiple Linear Regression and Bayesian Multilevel Linear Regression for both informative and non-informative priors.

For the overall weight at birth data, percentages of missing observations in the data show that variables such as mother's weight, mother's height, and place of delivery should be removed from the model. This is due to high percentages of missing observations. The remaining independent variables are then used for this study, in both Classical and Bayesian linear regression.

3.2 Classical Modelling Approach

Classical Linear Regression

Classical linear regression is a statistical model used in predicting future values of the target (dependent) variable based on the behaviour of a set of explanatory factors (independent variables).

3.2.1 Multiple Linear Regression

Regression Analysis

Linear regression model assumes that the relationship between the dependent variable Z and the r-vector of explanatory variables $\boldsymbol{X}$ is linear. The model can be organized into vectors and matrices:

$$\boldsymbol{Z} = \boldsymbol{X\beta} + \boldsymbol{\epsilon} \tag{3.21}$$

Where $\mathbf{Z} = (Z_1, Z_2, \dots, Z_n)'$ is the vector of the dependent variable and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_r)'$ is a vector of regression coefficients associated with the vector $\boldsymbol{X}$ of covariates:

$$\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_r)' = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1r} \\ 1 & X_{21} & & X_{2r} \\ \vdots & & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nr} \end{pmatrix},$$

For $i \in \{1, \dots, n\},\ k \in \{1, \dots, r\}$

$\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)'$ is the random error.

The model assumptions are

$$E(\boldsymbol{\epsilon}) = \mathbf{0}, \qquad Var(\epsilon_i) = \sigma^2, \qquad Cov(\epsilon_i, \epsilon_{i'}) = 0 \quad \forall i \neq i'$$

A $n \times n$ variance-covariance matrix for the random errors and for $\boldsymbol{Z}$ is

$$Cov(\boldsymbol{\epsilon}) = E(\boldsymbol{\epsilon\epsilon}') = \sigma^2 \boldsymbol{I}.$$

$$E(\boldsymbol{Z}) = \boldsymbol{X\beta}, \quad Cov\,(\boldsymbol{Z}) = \boldsymbol{\sigma}^2 \boldsymbol{I}$$

Since we have not made any assumptions about the distribution of $\boldsymbol{Z}$ or $\boldsymbol{\epsilon}$, **Least Squares Estimation is** then introduced, this is an approach used to estimate the vector $\boldsymbol{\beta}$ which minimizes the sum of squares residuals:

$$(\boldsymbol{Z}-\boldsymbol{X\beta})'(\boldsymbol{Z}-\boldsymbol{X\beta}) \quad (3.22)$$

That is, we find $\frac{\partial}{\partial \boldsymbol{\beta}} = 0$

$$\frac{\partial}{\partial \boldsymbol{\beta}}(\boldsymbol{Z}-\boldsymbol{X\beta})'(\boldsymbol{Z}-\boldsymbol{X\beta})$$

$$\boldsymbol{X}'\boldsymbol{X}\hat{\beta}-\boldsymbol{X}'\boldsymbol{Z}=0$$

$$\hat{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} \quad (3.23)$$

Thus the predicted values are $\hat{Z}=X\hat{\beta}$ on which we can find the residuals that are computed as:

$$\hat{\epsilon}=\boldsymbol{Z}-\widehat{\boldsymbol{Z}}=\boldsymbol{Z}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}=\boldsymbol{Z}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}=(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Z}$$

$\boldsymbol{H}=\boldsymbol{Z}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ Is called the ‘hat’ matrix.

The residual sums of squares (or error sums of squares) is

$$\hat{\epsilon}'\hat{\epsilon}=\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{H})'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Z}=\boldsymbol{Z}'(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{Z}$$

We can partition variability in z into variability due to changes in predictors and variability due to random noise (effects other than the predictors). The sum of squares decomposition is:

$$\sum_{j=1}^{n}(z_j-\bar{z})^2=\sum_{j}(\hat{z}-\bar{z})^2+\sum_{j}\hat{\epsilon}^2$$

The coefficient of multiple determination is

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST}$$

R^2 Indicates the proportion of the variability in the observed responses that can be attributed to changes in the predictor variables.

Properties of Estimators and Residuals

Under the general regression model described earlier in equation (3.21) we use equation (3.22) results for $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}$ to estimate residuals as:

$$E(\hat{\beta}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{Z}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\beta}$$

$$Cov(\hat{\beta}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}X'Cov(\boldsymbol{Z})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

For residuals; $E(\hat{\epsilon}) = 0, \quad Cov(\hat{\epsilon}) = (\boldsymbol{I} - \boldsymbol{H})\sigma^2, \quad E(\hat{\epsilon}'\hat{\epsilon}) = (n - r - 1)\sigma^2.$

An unbiased estimator of σ^2 is

$$s^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{n - (r + 1)} = \frac{\boldsymbol{Z}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Z}}{n - r - 1} = \frac{SSE}{n - r - 1} \tag{3.24}$$

Now if we assume that the $n \times 1$ vector $\epsilon \sim N_n(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, then it follows that

$Z \sim N_n(\boldsymbol{X\beta}, \sigma^2\boldsymbol{I})$

$\hat{\beta} \sim N_n(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$

$\hat{\beta}$ is distributed independent of $\hat{\epsilon}$ and furthermore

$$\hat{\epsilon} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Z} \sim N(\boldsymbol{0}, \sigma^2(\boldsymbol{I} - \boldsymbol{H}))$$

$$(n - r - 1)s^2 = \hat{\epsilon}'\hat{\epsilon} \sim \sigma^2\chi^2_{n-r-1}$$

3.2.2 Hierarchical Multiple Linear Regression

Hierarchical Linear Model

This section considers multilevel linear models, that is, a class of models that combine the advantages of the mixed-model ANOVA with its flexible modelling of fixed and random effects, and regression. However, the term hierarchical linear model captures two defining features of the models.

- Firstly, the data appropriate for such models are hierarchically structured, with first-level units nested within second-level units, second-level units nested within third-level units, and so on.
- Secondly, the parameters of such models may be viewed as multilevel structures. The investigator may specify a level-one model, the parameters of which characterize linear relationship occurring between level-one units. These parameters are then viewed as varying across level-two units as a function of level-two characteristics.

What distinguishes the multilevel model is that the *random factors are nested and never crossed.* However *fixed factors* can be crossed with random factors (or with each other) and random factors may be nested within fixed factors. The data can be unbalanced at any level. Both discrete and continuous predictors can be specified as having random effects, and these random effects are allowed to co-vary. A good example of a hierarchical structure is an educational system where students are “clustered” or grouped within classes.

When an effect in the statistical model is considered as being random, we mean that we wish to draw conclusions about the population from which the observed units were drawn, rather than about these particular units themselves. The first decision concerning random effects in specifying a multilevel model is the choice of the levels of analysis. These levels can be, for example, individuals, classrooms, schools, organizations, and neighbourhoods. Formulated generally, a level is a set of units, or equivalently a system of categories, or a classification factor in a statistical design. In statistical terminology, a level in a multilevel analysis is a design factor with random effects. In addition, the assumption is made that the random effects are uncorrelated with the explanatory variables.

The Fixed Effects Model is a statistical model that represents the observed quantities in terms of explanatory variables that are treated as if the quantities were non-random. Fixed effect regression is important because data often falls into categories and when that happens, you will want to control for characteristics of those categories that might affect the dependent variable. To reflect this in the model, assume they are in groups indexed by j, and in there are n_j individuals in group j.

Table 2: Factors at each hierarchical level that affect the students' grade, for the classroom example.

Hierarchical Level	**Example of Hierarchical Level**	**Example Variables**
Level-1	Student Level	Gender Intelligent Quotient (IQ) Socio-economic status Self-esteem rating Behavioural conduct rating Breakfast consumption
Level-2	Classroom Level	Class size Homework assignment load Teaching experience Teaching style
Level-3	School Level	Schools' geographic location Annual budget

Hierarchical models

Random variability only occurs as a "within subjects" effect, for level-2 independent variable the model is

$$Z_{ij} = \beta_{0j} + \beta_j^* X_{ij} + \varepsilon_{ij} \tag{3.25}$$

For any individual, the model becomes:

$$Z_j = \beta_{0j} + \beta_j^* X_j + \varepsilon_j \tag{3.26}$$

- Z_{ij} is the value of the dependent variable for an individual i (level 1) and j refers to level 2 component or region in this case.
- $\beta_{0j} = (\beta_0 + \mu_{0j})$ is the intercept on the dependent variable in region j.
- $\beta_j^* = (\beta_{1j}, \dots, \beta_{rj})$ is the slope for the relationship in region j (Level 2) between individual level predictors and the dependent variable.
- β_0 is the intercept of the regression line with the Z-axis when X=0
- ε_{ij} is the random errors of prediction.

$$Z_j = \left(Z_{1j}, Z_{2j}, \dots, Z_{nj}\right)' \qquad X_j = \begin{pmatrix} X_{1j1} & \cdots & X_{1j,r} \\ X_{2j1} & & X_{2j,r} \\ \vdots & \ddots & \vdots \\ X_{nj1} & \cdots & X_{nj,r} \end{pmatrix}$$

$$X_j = \begin{pmatrix} X_{1j1} & \cdots & X_{1j,r} \\ X_{2j1} & & X_{2j,r} \\ \vdots & \ddots & \vdots \\ X_{nj1} & \cdots & X_{nj,r} \end{pmatrix}$$

For $i \,\&\, j \in \{1, \dots, n\},\ k \in \{1, \dots, r\}$

Level One and Level Two Models:

For the Level One model, the intercept and the slopes becomes

$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{\mathrm{j}} + u_{0j}$

$\beta_{1j} = \gamma_{10} + u_{1j}$

Thus final Level Two model is

$$y_{ij} = \gamma_{00} + \gamma_{01} W_j + u_{0j} + X_j(\gamma_{10} + u_{1j}) + \varepsilon_{ij} \tag{3.27}$$

- γ_{00} Refers to the overall intercept which is the grand mean of the scores on the dependent variables across all the groups when all the predictors are equal to zero.
- W_j Refers to the predictor for the level 2 model.
- γ_{01} Refers to the overall regression coefficient, or the slope between the level 2 predictor and dependent variable.
- u_{0j} Refers to random error component for the deviation of the intercept of a group from overall intercept.
- γ_{10} Refers to the overall regression coefficient or the slope between level 1 predictor and dependent variable.
- u_{1j} Refers to the error component for the slope, that is, the deviation of the group slopes from the overall slope.

Types of Models:

- **Random intercepts model**: this model has two parts, a fixed part (which is the intercept and the product of the coefficient of explanatory variable with explanatory variable) and it has the random part, intercepts are allowed to vary

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + \varepsilon_{ij} \tag{3.28}$$

This is obtained from combining the variance components model and the single level model, that is:

Variance components model: $Y_{ij} = \beta_0 + u_j + \varepsilon_{ij}$

$$u_j \sim N(0, {\sigma_u}^2), \varepsilon_{ij} \sim N(0, {\sigma_\varepsilon}^2)$$

Single level model- $Z_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + \varepsilon_{ij}$

Table 3 Parts of Models

Fixed part $(\boldsymbol{\beta_0} + \boldsymbol{\beta_1 X_{ij}})$	**Random part** $(\boldsymbol{u_j} + \boldsymbol{\varepsilon_{ij}})$
Parameters that we estimate are the coefficients β_0, β_1,…	Parameters that we estimate are the variances ${\sigma_u}^2$and ${\sigma_\varepsilon}^2$

For this model the intercepts are allowed to vary, and therefore the scores on the dependent variable for each individual observation are predicted by the intercept varies across the group. It also provides the information about the intraclass correlation.

- **Random Intercepts Model**: In this model the varying of intercepts is allowed, and therefore, the scores on the dependent variable for each individual observation are predicted by the intercept that varies across the groups. The assumption for such a model is that the slopes are fixed (the same across different contexts). This model also provides information about intraclass correlations, which are helpful in determining that the multilevel models are required in the first place.

- **Random Slopes Model**: A model in which the varying of slopes is allowed, and therefore, slopes are not the same across the groups. The assumption is based on the intercept that they are fixed (the same across different contexts)
- **Random Intercepts and Slopes Model**: A model that includes both random intercept and random slopes is likely the most realistic type of model, although it is the most complex. In this model, both intercepts and slopes are allowed to vary across groups, meaning that they are different in different contexts.

A good hierarchical model should meet the following assumptions:

- Linearity: The relationship between variables is linear.
- Normality: Assumes that the error terms at every level of the model are normally distributed.
- Homoscedasticity: Assumes that the variance around the regression line is similar for all the determinants of low birth weight.
- Independence of Observations: Assumes that the observations have no relationship between groups.

Multilevel models assume that the Level 1 and Level 2 residuals are uncorrelated and that the errors (as measured by the residuals) at the highest level are uncorrelated.

3.3 Bayesian Modelling Approach

3.3.1 Concept of Bayesian

The Bayesian technique aims to determine the posterior distribution using the likelihood function and the prior probability derived from a statistical model for the observed data. Bayesian inference computes the posterior probability according to Bayes' theorem:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} \qquad (3.31)$$

Where $\boldsymbol{A}$ is any hypothesis that has a probability that may be affected by the data and $\boldsymbol{B}$ corresponds to the new data that was not used in developing the prior probability.

Both Bayesian methods and Classical methods have advantages and disadvantages, and there are some similarities. At the point when the sample size is large, Bayesian inferences often gives the results for parametric models that are fundamentally the same as the outcomes created by the classical methods. Below are advantages and disadvantages of using Bayesian methods according to Berger (1985, Sections 4.1 and 4.12).

Advantages of Bayesian Methods

- Bayesian methods uses both sources of information, (the prior information about the process and the information about the process contained in the data). They provide a real and principled way of combing the data and the prior information within a solid decision theoretical framework. You can incorporate previous information about a parameter and form a prior distribution about the future analysis. The previous posterior distribution can be used as a prior when new observations become available and follow from Bayes' theorem.

- It provides inferences that are conditional on the data and are exact, without reliance on asymptotic approximation. Both small and large sample inferences proceed in the same manner. In addition, parameters are estimated directly.

- It obeys the Likelihood Principle. If two distinct sampling designs yield proportional likelihood functions for θ, then inferences of about θ should be the same for both designs. Classical Inferences do not obey the Likelihood Principle.

- It provides interpretable answers for direct probability statements about the parameters, such as "the true parameter θ has a probability of 0.95 of falling in a 95% credible interval."

- Bayesian statistics has a single tool, Bayes' theorem, which is used in all situations. This contrasts to frequentist procedures, which require many different tools.

- The Bayesian Method provides a setting for a wide range of models, such as hierarchical models and missing data problems, MCMC, along with other numerical methods.

- Elimination of nuisance parameters is conceptually straightforward, and is also easy due to advances in Bayesian computing. This convenience is a result of Bayesian analysis being a logically, simple and easy approach. Below are two examples where frequentist answers are not unique;

  Example 1: Confidence interval for the difference between two normal means ($\beta = \mu_1 - \mu_2$), when variance are unknown.

  Example 2: Confidence interval for $\beta = \mu/\sigma$ using a sample from a normal distribution.

Disadvantages of Bayesian Methods

There are also disadvantages to using Bayesian methods:

- The use of Bayesian methods does not tell you how to select the prior. There is no correct approach to pick a prior. Bayesian inferences require skills to translate subjective prior beliefs into a mathematically-formulated prior. If you do not proceed with caution, you can generate misleading results.
- The priors can heavily influence the posterior distribution, which can practically make it difficult to convince subject matter experts who do not agree with the validity of the chosen prior.
- Bayesian Methods often come with a high computational cost, especially in models with a large number of parameters. The posterior distribution of a parameter is exact, given the likelihood function and the priors, while simulation-based estimates of posterior quantities can vary due to the random number generator used in the procedures.

The essential difference between Bayesian and Frequentist statistics is in how probability is used. Frequentist statistics use probability only to model certain processes broadly described as “sampling”. Bayesian statistics use probability more widely to model both sampling and other kinds of uncertainty. Bayesian approaches formulate the problem differently. Instead of saying that the parameter simply has one (known) true value, a Bayesian method says that the parameter’s value is fixed but has been chosen from some probability distribution known as the prior distribution.

3.3.2 Prior Distribution

The prior distribution is the most important aspect of Bayesian inference and represents the information about an uncertain parameter θ. The prior distribution is combined with the probability distribution of new data (*likelihood function*) to yield the posterior distribution.

The key issues in setting up a prior distribution are:

- What information is going into the prior distribution;
- The properties of the resulting posterior distribution.

With well-identified parameters and large sample sizes, reasonable choices of prior distributions will have a small effect on posterior inferences. The prior distribution becomes more important if the sample size is small or available data provides indirect information about the parameters of interest. Usually, the models can be hierarchical, so that clusters of parameters have shared prior distributions which themselves can be estimated from the data. The data to be used in WinBUGS software for both multiple linear and multilevel models are modeled using the same predictors as classical statistics to allow comparison. WinBUGS is a useful computational tool that fits complicated Bayesian models using MCMC methods. Two situations of Bayesian models were considered:

i. Bayesian Non-Informative Prior: Normal distribution with large variance is taken as the distribution of interest to ensure that the prior does not affect the posterior that much.

ii. Bayesian Informative Prior: In this study, a normal distribution is considered as prior and the parameters are determined by computations made on surveys of years 2000 and 2004.

3.3.2.1 Developing a Prior Distribution for the Low Birth Weight Model

The parameters of the informative prior for the low birth weight model are obtained as follows;

- The parameter (beta coefficient) of the prior is obtained by taking the average of the variables mentioned in section 2.2 for the years 2000 and 2004 and the variance component of the prior is obtained using the pooled variance of the years 2000 and 2004.

$$s^2_{pooled} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad , \quad \beta_{average} = \frac{\beta_{2000} + \beta_{2004}}{2} \tag{3.32}$$

Where

$$s_1^2 = \frac{SE(2000)}{\sqrt{n_1}} \quad , \quad s_2^2 = \frac{SE(2004)}{\sqrt{n_2}} \tag{3.33}$$

- When the coefficient is significant for only one of the years 2000 or 2004: we considered the coefficient and variance for the year where it was significant.

- Finally, it is important to mention that the rules put in place to obtain the coefficient and variance of the informative prior were not blindly applied. We used our own thinking (judgment) to assess the priors. In the situation where the prior was following a Gamma distribution, the following formula was used to transform the mean and variance to the parameters of the Gamma distribution:

$$\alpha = \frac{\bar{x}}{s^2} \quad \text{and} \quad \beta = \frac{s^2}{\bar{x}} \tag{3.34}$$

Where $\bar{x} = \frac{\bar{x}_{2000} + \bar{x}_{2004}}{2}$ and $s^2 = \frac{S^2{}_{2000} + S^2{}_{2004}}{2}$

Rejection Region

WinBUGS provide the credible interval and not a p value. The credible interval is the interval containing 95% of the posterior samples generated by the MCMC algorithm.

WinBUGS software uses Gibbs sampler algorithm as MCMC technique. To be more precise, the numbers (posterior sample simulated) generated by MCMC are ordered in an ascending way. The lower bound of CI is the value above the first 2.5% simulated numbers while the upper bound is the value below the last 2.5% simulated numbers.

3.3.3 The Likelihood Function

A Likelihood Function is a function of parameters of a classical statistic model given data. In statistical inferences, the likelihood is an important factor in methods of estimating a parameter from a set of statistics. The Likelihood Function is used once data is available to describe a function of a parameter for a given outcome. The likelihood of a parameter value θ, given outcomes x, is equal to the probability assumed for those observed outcomes given those parameter values, i.e.

$$\mathcal{L}(\theta|\, x) = P(\, x|\theta). \tag{3.36}$$

The Likelihood Function for discrete and continuous distribution is defined differently as:

$$\mathcal{L}(\theta|\, x) = \begin{cases} P_\theta(X = x) & \text{discrete} \\ f_\theta(x) & \text{continuous} \end{cases} \tag{3.37}$$

Where x is the outcome of the random variable X and f is a density function of a random variable X following a continuous probability distribution that depends on a parameter θ.

For example, if the data set for n observations, that is, $X_1, X_2, \dots, X_n$ is independent and identically distributed Poisson (λ) then a gamma(α, β) prior on λ is a conjugate prior.

Then the likelihood is given by:

$$\mathcal{L}(\lambda|\, \mathbf{x}) = P_\lambda(X = x) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum x_i}}{\prod_{i=1}^{n}(x_i!)}$$

(3.38)

3.3.4 The Posterior Distribution

The posterior probability is the probability of the parameters $\boldsymbol{\theta}$ given the evidence $\boldsymbol{X}$**:** $p(\boldsymbol{\theta}|\boldsymbol{X})$. It contrasts with the likelihood, which is the probability of the evidence given the parameters: $p(\boldsymbol{X}|\boldsymbol{\theta})$**.** For example, if we have a prior belief that the probability function is $p(\theta)$ and observations x with the likelihood $p(x|\theta)$, then the posterior probability is defined as:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

(3.39)

Thus Posterior probability ∝ Likelihood x Prior probability

Suppose that we have an unknown parameter μ for which the prior belief can be express in terms of the normal distribution, so that $\mu \sim N(\mu_0, \sigma_0^2)$

Where μ_0 and σ_0^2 are known. The prior distribution is given by;

$$f(\mu) = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}}$$

(3.41)

The likelihood function of observation x given μ is given by;

$$f(x|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

(3.42)

And hence using equation (3.41) and equation (3.42) the posterior distribution of μ given that we have an observation x is;

$$f(\mu|x) = \frac{f(\mu)f(x|\mu)}{\int_{-\infty}^{\infty} f(\mu)f(x|\mu)d\mu} = \frac{f(\mu)f(x|\mu)}{f(x)} \propto f(\mu)f(x|\mu)$$

$$\propto \frac{1}{\sqrt{2\pi\sigma_0^2}} exp\left\{\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\right\} * \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}$$

$$\propto \frac{1}{2\pi\sqrt{\sigma^2\sigma_0^2}} exp\left\{\frac{-\mu^2+2\mu\mu_0-{\mu_0}^2}{2\sigma_0^2} - \frac{x^2-2\mu x+\mu^2}{2\sigma^2}\right\}$$

$$\propto \text{const} * \exp\left\{\frac{-\mu^2\sigma^2+2\mu\mu_0\sigma^2-{\mu_0}^2\sigma^2-\sigma_0^2x^2+2\mu\sigma_0^2x-\mu^2\sigma_0^2}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2(\sigma^2+\sigma_0^2)+2\mu(\mu_0\sigma^2+\sigma_0^2x)-(\sigma_0^2x^2+{\mu_0}^2\sigma^2)}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-\mu^2+2\mu\frac{\mu_0\sigma^2+\sigma_0^2x}{\sigma^2+\sigma_0^2}-\left(\frac{\sigma_0^2x\ +\mu_0\ \sigma^2}{\sigma^2+\sigma_0^2}\right)^2}{\frac{2\sigma_0^2\sigma^2}{\sigma^2+\sigma_0^2}}\right\} * exp\left\{-\frac{\sigma_0^2x^2+{\mu_0}^2\sigma^2}{2\sigma_0^2\sigma^2}\right\}$$

$$\propto \exp\left\{\frac{-(\mu-\frac{\mu_0\sigma^2+\sigma_0^2x}{\sigma^2+\sigma_0^2})^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+\sigma_0^2}}\right\}$$

Letting

$$\sigma_1^2 = \frac{\sigma_0^2\sigma^2}{\sigma^2+\sigma_0^2} = \frac{1}{\sigma^{-2}+{\sigma_0}^{-2}}$$

$$\mu_1 = \frac{\mu_0\sigma^2+\sigma_0^2x}{\sigma^2+\sigma_0^2} = \frac{\mu_0\sigma^{-2}+\sigma^{-2}x}{\sigma^{-2}+{\sigma_0}^{-2}} = \sigma_1^2(\mu_0{\sigma_0}^{-2}+x\sigma^{-2})$$

So that

$$\sigma_1^{-2} = \sigma^{-2} + \sigma_0^{-2}$$

$$\mu_1\sigma_1^{-2} = \mu_0\sigma_0^{-2} + x\sigma^{-2}$$

And hence

$$f(\mu|x) \propto \exp\left\{\frac{-(\mu - \mu_1)^2}{2\sigma_1^2}\right\}$$

From which it follows that

$$f(\mu|x) = \frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left\{\frac{-(\mu - \mu_1)^2}{2\sigma_1^2}\right\} \tag{3.43}$$

Then the posterior density is

$$\Rightarrow \ \mu|x \sim N(\mu_1, \sigma_1^2)$$

This is the posterior distribution of a Normal distribution with a sample size of $n = 1$.

Markov Chain Monte Carlo

Markov Chain Monte Carlo (MCMC) is a technique for simulating draws that are slightly dependent and are approximately from a posterior distribution. The draws simulated are used to calculate the quantities of interest for the posterior distribution. The successive random selections form a Markov Chain, the stationary distribution of which is the target distribution.

In Bayesian Statistics there are generally two MCMC algorithms that are used, namely, the Gibbs Sampler and the Metropolis-Hasting Algorithm.

1. **Gibbs Sampler**

Suppose we have a joint distribution $p(\theta_1, \dots, \theta_k)$ that we want to sample from (for example, a posterior distribution). Gibbs Sampler can then be used to sample from the joint distribution if we knew the **full conditional** distributions for each parameter, which is the distribution of parameter conditional on the known information and all other parameters: $p(\theta_j|\theta_{-j}, y)$.Let's suppose that we are interested in sampling from the posterior $p(\boldsymbol{\theta}|\boldsymbol{y})$, where $\boldsymbol{\theta}$ is a vector of three parameters, $\theta_1, \theta_2, \theta_3$.

The steps to a Gibbs Sampler are:

1. Pick a vector of starting values $\theta^{(0)}$.

2. Start with any θ (order does not matter). Draw a value $\theta_1^{(1)}$ from full conditional $p(\theta_1|\theta_2^{(0)}, \theta_3^{(0)}, \boldsymbol{y})$.

3. Draw a value $\theta_2^{(1)}$ from the full conditional $p(\theta_2|\theta_1^{(1)}, \theta_3^{(0)}, \boldsymbol{y})$. Note that the updated value $\theta_1^{(1)}$is used.

4. Draw the value $\theta_3^{(1)}$ from the conditional $p(\theta_3|\theta_1^{(1)}, \theta_2^{(1)}, \boldsymbol{y})$. Using both updated values.

5. Draw $\theta^{(2)}$ using $\theta^{(1)}$ and continually using the most updated values.

6. Repeat until we get *M* draws, with each draw being a vector $\boldsymbol{\theta}^{(t)}$.

7. Optional burn-in and/or thinning.

Gibbs Sampler has the following strengths:

- It is an easy algorithm to think about.
- It exploits the factorization properties of the joint probability distribution.
- No difficult choices need to be made to tune the algorithm.

However, sampling from full conditional distributions is considered difficult and sometimes impossible, which is the weakness of the Gibbs Sampler.

2. Metropolis-Hastings Algorithm

Metropolis-Hastings Algorithm is a Markov Chain Monte Carlo (MCMC) method for obtaining a sequence of random samples from a probability distribution for which direct sampling is difficult. This algorithm aims to construct Markov Chain $Y^{(t)}$ with stationary distribution $f(y)$. At some time t, it generates next value $Y^{(t+1)}$ in the following two steps:

- *Proposal step:* Sample Y from the proposal distribution,

  $Z \sim q(z|Y^{(t)})$.

- *Acceptance step:* With probability

$$\alpha\left(Y^{(t)}, Z\right) = min\left\{1, \frac{f(Z)}{f(Y^{(t)})}\frac{q(Y^{(t)}|Z)}{q(Z|Y^{(t)})}\right\}$$

  Set $Y^{(t+1)} = Z$ (acceptance) and otherwise set $Y^{(t+1)} = Y^{(t)}$ (rejection).

Suppose we want to sample from the posterior distribution

$\gamma(\theta|Y) = \frac{f(Y\,|\,\theta)\gamma(\theta)}{f(Y)}$ with $f(Y) = \int f(Y|\theta)\,\gamma(\theta)\,d\theta$.

then

$$\frac{\gamma(\theta'|Y)}{\gamma(\theta|Y)} = \frac{f(Y|\theta')\gamma(\theta')}{f(Y|\theta)\gamma(\theta)}$$

i.e. the normalizing constant is not required to run the algorithm. Usually the proposal distribution q is chosen in such a way that it is easy to sample from it. If the proposal distribution is symmetrical then $q(z|y) = q(y|z)$ then the Metropolis Algorithm becomes:

$$\alpha\left(Y^{(t)}, Z\right) = min\left\{1, \frac{f(Z)}{f(Y^{(t)})}\right\}$$

Where the proposal state Z with higher probability are always accepted. In the process, change to state with lower probability possible with α. Sometimes Random-Walk Metropolis has a special case where $q(z|y) = q(|z - y|)$.

3.3.5 Hierarchical Multiple Linear Regression

Hierarchical (accurate term is multilevel) data structures are regularly encountered in the social and behavioural sciences meaning that the data collected often represents different levels of aggregation for the subjects of study.

The starting point is the basic form of multilevel models; these models take the standard restriction that the estimated coefficients are constant across the individual cases by specifying the levels of additional effects. They start with a standard linear model specification indexed by subjects and a first level of grouping. The form for a model with a single explanatory variable is:

$$Y_{ij} = \beta_{j0} + \beta_{j1}X_{ij} + \epsilon_{ij} \tag{3.44}$$

Now add a second level to the model that explicitly nests effects within groups and index these groups $j = 1$ to J:

$$\beta_{j0} = \gamma_{00} + \gamma_{10}Z_{j0} + u_{j0}$$

$$\beta_{j1} = \gamma_{01} + \gamma_{11}Z_{j1} + u_{j1},$$

Where all individual level variation is assigned to groups producing department level residuals: $u_{j1}$and u_{j0}. These Z_{j1} are context level variables in that their effect is assumed to be measured at the group level rather than at the individual level. Finally, after substitution the level-2 model becomes:

$$Y_{ij} = \gamma_{00} + \gamma_{01}X_{ij} + \gamma_{10}Z_{j0} + \gamma_{11}X_{ij}Z_{j1} + u_{j1}X_{ij} + u_{j0} + \epsilon_{ij} \tag{3.45}$$

For the data matrices, X_{ij} for individual i in cluster j, and Z_j for the cluster j, there are four canonical models that are listed in table 4 below.

Table 4 Canonical Models

Pooled	$Y_{ij} = \alpha + \boldsymbol{X'}_{ij}\boldsymbol{\beta} + \boldsymbol{Z'}_j\gamma + e_{ij}$
Fixed Effect	$Y_{ij} = \alpha_j + \boldsymbol{X'}_{ij}\boldsymbol{\beta} + e_{ij}$
Random Effect	$Y_{ij} = \alpha_j + \boldsymbol{Z'}_j\gamma + e_{ij}$
Random Intercepts and Random Slope	$Y_{ij} = \alpha_j + \boldsymbol{X'}_{ij}\boldsymbol{\beta} + \boldsymbol{Z'}_j\gamma + e_{ij}$

3.3.6 Basic Structure of the Bayesian Hierarchical Model

As usual, the central interest is to generate the posterior distribution from the product of the likelihood and the prior:

$$\pi(\theta|\mathbf{X})\ \alpha\ L(\theta|\mathbf{X}) * p(\theta) \tag{3.46}$$

Now suppose that the parameter θ is conditional on another unknown parameter λ, which has its own distribution, the posterior becomes:

$$\pi(\theta|\mathbf{X})\ \alpha\ L(\theta|\mathbf{X}) * p(\theta|\lambda)p(\lambda) \tag{3.47}$$

If the form $p(\theta|\lambda)$ and $p(\lambda)$ are cooperative then this system can be quite straightforward. This extension complicates the derivation of the posterior above to the point where the MCMC tools are required to get marginal distribution of θ. Here $P(\theta)$ is the prior distribution, $p(\theta|\lambda)$ in turn is now conditional on another parameter that has its own prior, $p(\lambda)$ called a hyperprior, which can also have hyper parameters if desired.

Inference for either parameter of interest can be obtained through looking at the respective marginal densities:

$$\pi(\theta|\mathbf{X}) = \int_{\lambda} \pi(\theta, \lambda|\mathbf{X})\, d\lambda$$

$$\pi(\lambda|\mathbf{X}) = \int_{\theta} \pi(\theta, \lambda|\mathbf{X})\, d\theta,$$

Where the other parameter has been integrated out of the joint distribution. If we continue this procedure of adding hyper priors, beginning with making λ conditional on another parameter, $p(\lambda|\psi)$, and adding new hyper prior distribution, $p(\psi)$, to the calculation of posterior distribution then this results to:

$$\pi(\theta, \lambda, \psi|\mathbf{X})\ \alpha\ L(\theta|\mathbf{X}) * p(\theta|\lambda)p(\lambda|\psi)\ p(\psi)$$

Now ψ are the highest-level parameters and therefore the only hyper prior that is unconditional. As a simple example, consider data that are normally distributed with known variance and unknown mean, $X_i \sim N(\mu_i\,,\ \sigma^2)$, $i = 1,2, \dots n,$ for an unknown constant μ we assume random μ_i , $i = 1, \dots, n$ that are drawn independently from a common, also normal, distribution: $\mathcal{N}(m_\mu, s_\mu^2)$. The model can be summarized as:

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix}, \sigma_0^2\mathbf{I}\right) \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} \sim \mathcal{N}(m_\mu, s_\mu^2)$$

The X_i values are assumed to be generated by underlining means in normal specification, these means are of common distribution with fixed hyper parameters $m_\mu \text{and } s_\mu^2$. This is the similar idea as Non-Bayesian Multilevel Models above, except that the second stage is given in the form of a common prior distribution for the unknown μ_i.

Chapter 4: Results

In the following, we illustrate all models described in previous chapters by applying them to the data set from Malawi. Before fitting of the models, this chapter outlines the descriptive summaries of Malawi data. At the end of this chapter, all the fitted models are compared highlighting that the Bayesian approach gives better results than classical approach.

4.1 Multiple Linear Regression

4.1.1. Descriptive Summaries

Distribution of child weights by year

Table 5 below shows the total number of children in each birth weight group for the years 2000, 2004 and 2010. This study focuses only on children with birth weight that is less than 2.5kg.

Table 5: Distribution of child weights by year

Child's weight at birth (kg)	2010	2004	2000
<2.5	2315	909	909
2.5-4.5	18440	10333	10333
>4.5	7906	6135	6654

Means and Standard deviations of child weights by underweight factors and year

The first results of this study presented in the descriptive Table 6 below show the average weight at birth (in grams), standard deviation and number of observations in each category for all underweight factors in this study. This result concerns only underweight (less than 2.5kg) children. There seems to be no big mean differences in child's birth weight for mother's age in all of three years. In year 2010, the youngest mothers (aged 15-19 years) had children with average birth weight of 2.19kg and oldest mothers (aged 40-49 years) had children with an average birth weight of 2.20kg. Results in Table 6 also show that the average weight at birth is affected by the gender of a child, that is, female children have a slightly higher average birth weight than male children in all three years. There is a proportional average relationship between average birth weight and education level of a mother. Results in Table 6 show that mothers who have secondary or tertiary education have children with a higher average birth weight than mothers with no education at all. Few mothers were HIV positive compared to mothers who were HIV negative in all three years. The average weight at birth of a child with a mother who is HIV positive is less than that of a child with a mother who is HIV negative in all three years.

In this study, more mothers were working during pregnancy than mothers who were not working. In the year 2010, the average birth weight of a child with a working mother was smaller than that of a child whose mother was not working. The variable of whether the mother was working or not had little impact on the results.

There seem to be a larger spread or variability in the values of the young mothers (aged 15-19) than older mothers (aged 40-49). For all the factors in the study, there were no big spread differences in total average variability for the year 2010 when compared to the years (2000 and 2004). The overall averages for all the variables are almost the same for all three years - approximately 2.1kg. Since the comparison is between Classical Linear Regression and Bayesian Linear Regression, further and sufficient analysis will be made using the inferential statistics. In other words, by Multiple Linear Regression, Multilevel Linear Regression, Bayesian Multiple Linear Regression and Bayesian Multilevel Linear Regression.

Table 6: Means and Standard deviations of child weights by underweight factors and year

	2010			2004			2000		
Mother's age	**Mean**	**Standard deviation**	**Nobs**	**Mean**	**Standard deviation**	**Nobs**	**Mean**	**Standard deviation**	**Nobs**
15-19	2185.29	340.64	204	2134.84	401.81	64	2243.36	302.62	122
20-24	2170.57	381.30	666	2196.72	361.69	332	2266.37	327.88	409
25-29	2173.74	368.06	613	2170.30	349.78	234	2254.22	363.61	271
30-34	2198.22	336.51	393	2168.49	387.89	138	2271.30	276.81	161
35-39	2167.03	371.12	273	2125.57	365.70	88	2256.80	345.71	103
40-49	2202.19	333.10	166	2235.66	331.74	53	2340.74	255.32	81
Child's gender									
Male	2174.91	357.58	1068	2166.58	389.82	432	2254.85	329.16	530
Female	2175.78	370.14	1247	2185.79	340.08	477	2275.83	320.90	617
Region									
Northern	2197.49	374.94	478	2209.23	325.22	163	2233.54	358.72	195
Central	2160.34	350.26	837	2121.48	354.44	254	2270.59	307.81	407
Southern	2177.40	370.44	1000	2194.35	379.22	492	2274.48	324.23	545
Mother's education level									
no education	2152.31	367.97	325	2167.50	354.59	206	2289.36	323.01	264
Primary	2164.15	373.18	1615	2174.24	357.30	558	2250.43	331.21	756
Secondary+	2243.73	311.81	375	2198.97	405.28	145	2311.36	281.98	127
Mother's working status									
Not working	2190.19	356.55	581	2149.89	397.16	378	2250.13	347.62	415
Previous year	2138.99	427.06	421	2197.37	256.26	38	2305.66	214.31	53
Working	2180.50	344.91	1313	2195.59	344.13	493	2272.84	317.24	679
Mother's HIV status									
Positive	2124.43	384.99	397	2132.67	375.25	138	2113.21	325.59	132
Negative	2185.65	358.81	1903	2111.33	358.24	771	2311.22	314.25	945
Total average	2177.00	362.07		2170.35	359.77		2265.59	314.10	

4.2 Multiple Linear and Hierarchical (Multilevel) Regression

4.2.1 Multiple Linear regression

In this section the classical linear model in equation 3.21 is fitted for the years 2010, 2004 and 2000 to obtain the estimates, standard error and p-values for all the variables in the study. These results are interpreted based on the output and will later be used in comparison of classical and the Bayesian approaches.

Multiple Linear Regression Output for the year 2010

Table 7: Year 2010 linear regression output for the low birth weight model in Malawi

Covariates	Estimate	Std.Error	tvalue	Pr(>|t|)
(Intercept)	2133.61	51.485	41.442	< 2e-16 ***
Mother's age (Ref:15-19)				
20-24	50.25	29.882	1.681	0.0928
25-29	81.57	33.627	2.426	0.0153 *
30-34	120.60	40.218	2.999	0.0028 **
35-39	111.08	49.482	2.245	0.0249 *
40-49	108.59	58.475	1.857	0.0635
Child's birth order number	-16.55	6.705	2.468	0.0137 *
Region (Ref: Northern)				
Central	-47.99	27.32	-1.757	0.0792
Southern	-13.74	27.656	-0.497	0.6193
Child's gender (Ref: Male)				
Female	-2.94	17.045	-0.172	0.8631
Education level (Ref: No education)				
Primary	-2.50	25.314	-0.099	0.9212
Secondary +	46.85	32.346	1.448	0.1477
Mother's working status (Ref: Not working)				
Worked (past year)	-51.48	26.056	-1.976	0.0484 *
Working	-13.98	20.342	-0.687	0.4922
Antenatal visits	5.45	2.978	1.831	0.0672
Mother's HIV Status (Ref: positive)				
Negative	81.89	21.688	3.776	0.00016 ***

Signif. codes : 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '' 1
Residual standard error: 344000 on 1574 degrees of freedom
Multiple R-squared: 0.03164, Adjusted R-squared: 0.02242
F-statistic: 3.429 on 15 and 1574 DF, p-value: 8.833e-06

The age of the mother, the child's birth order number, the mother's working status and the mother's HIV Status are significant factors determining the birth weight of a child in year 2010.

The model explains approximately 2.24% of variation in child's weight at birth if we consider all the data variables in the model. Since the p-value is small, the hypothesis that all the model coefficients are zero is rejected at 5% level of significance (p-value = $8.833e^{-6}$)
The residual standard error for the model is 344 000 on 1 574 degrees of freedom. This gives an idea of how far observed child's weight at birth (*Y* values) are from the predicted or fitted child's weight at birth ($\hat{Y}$ values), that is, residual or error $e = Y - \hat{Y}$. The intercept of 2133.605 is the estimated mean *Y* (dependent variable) value when all the X's (independent variables) are zero. Mothers aged between 25 and 39 years have children who have better weight at birth compared to mothers between the ages 15 and 19 years. This is based on the significant variables in the above output. When adjusting or controlling all other variables in the model, -16.55 is the rate of change in birth order number for a unit change in child's weight at birth. This means that birth weight increases with a low birth order number of a child i.e. mothers with few children are in lesser danger of having underweight babies compared to mothers with more children. For categorical variables, an appropriate reference category is selected to compare all the other categories to that reference category. The average difference in child's weight at birth between HIV positive and HIV negative mothers group is 81.89. This means that the weight at birth of a child from a mother who is not HIV infected is higher than that of an HIV positive mothers. The average difference in child's weight at birth between mothers who were working during their pregnancy and those who were not working is -51.476. Using the Pearson method, the correlation between gender of a child and mother's age is 0.0394 and this is significant.

There is a positive correlation of 0.0394 between the gender of a child and mother's age. The collinearity between these independent variables directly interpret the slope as the effect of each variable on child's weight at birth while controlling the other variable. This low correlation between the gender of a child and the age of the mother suggests that these two effects are not bounded together.

Table 7: Year 2010 Multiple Linear Regression Confidence Intervals

Covariates	**2.50%**	**97.50%**
(Intercept)	2032.62	2234.59
Mother's age (Ref:15-19)		
Age 20-24	-8.37	108.86
Age 25-29	15.61	147.52
Age 30-34	41.71	199.49
Age 35-39	14.02	208.14
Age 40-49	-6.11	223.29
Child's birth order number	-29.70	-3.40
Region (Ref: Northern)		
Central	-101.58	5.60
Southern	-67.99	40.50
Child's gender (Ref: Male)		
Female	-36.37	30.49
Education level (Ref: No education)		
Primary education	-52.16	47.15
Secondary+	-16.60	110.29
Mother's working status (Ref: Not working)		
Worked (past year)	-102.58	-0.37
Working	-53.87	25.92
Antenatal visits	-0.39	11.29
Mother's HIV status (Ref: Positive)		
Negative	39.35	124.43

A confidence interval is a range of values within which it is believed that a certain parameter (often the mean) will fall with a certain degree of confidence. The percentage of the confidence is denoted by (1 – α)100% where α is 0.05. A confidence interval can be used to test hypothesis for the model, if the confidence interval contains the value of the unknown parameter as hypothesized under H_0 , then H_0 would not be rejected in a two-sided hypothesis test with corresponding α. If the confidence interval does not contain the value of the unknown parameter as hypothesized under H_0 , then H_0 would be rejected in a two-sided hypothesis test with the corresponding α. The result of the above determination is:

- If the hypothesized value does not fall in the confidence interval, then there is a very small chance that the value can be a true value for the unknown variable, so H_0 will be rejected.
- If the hypothesized values falls in the confidence interval, then there is a very good chance that the value can be a true value for the unknown variable, so H_0 is definitely possible and will not be rejected.

In a 95% confidence interval, the true slope for mother's age (20-24 years) is between -8.36 and 108.86 i.e. (-8.36, 108.86) and is insignificant at 0.05 level. Based on the hypothesized value, most of the confidence intervals include zero for the model fitted, therefore the null hypothesis is not rejected at 0.05 level of significance. The variables mother's age, birth order number, mother's HIV status and mother's employment status are significant, as the confidence interval does not include zero. This result conforms to the multiple linear regression output above. Therefore, it can be concluded that the use of confidence interval gives similar results to that of linear regression.

Multiple Linear Regression Output for the year 2004

When the same model as the above is fitted for the year 2004 data we get the following parameter estimates.

Table 9: Year 2004 Multiple linear regression output

Covariates	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	2157.9	79.446	27.162	< 2e-16 ***
Mother's age (Ref:15-19)				
Age 20-24	112.185	54.883	2.044	0.04140 *
Age 25-29	122.568	60.28	2.033	0.04248 *
Age 30-34	154.453	69.422	2.225	0.02648 *
Age 35-39	135.04	81.88	1.649	0.09964
Age 40-49	296.983	93.493	3.177	0.00157 **
Child's birth order number	-31.044	10.125	-3.066	0.00227 **
Region (Ref: Northern)				
Central region	17.78	42.698	0.416	0.67727
Southern region	37.093	39.474	0.94	0.34778
Child's gender (Ref: Male)				
Female	18.527	28.506	0.65	0.51601
Education level (Ref: No education)				
Primary education	-31.234	37.864	-0.825	0.40977
Secondary+	25.206	51.568	0.489	0.62518
Mother's working status (Ref: Not working)				
Worked (past year)	-15.61	80.955	-0.193	0.84716
Working	69.209	29.564	2.341	0.01958 *
Antenatal visits	-4.934	3.072	-1.606	0.10879
Mother's HIV Status (Ref: positive)				
Negative	-13.682	28.496	-0.48	0.63133

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 335800 on 573 degrees of freedom
Multiple R-squared: 0.04675, Adjusted R-squared: 0.02179
F-statistic: 1.873 on 15 and 573 DF, p-value: 0.02324

The age of the mother, the child's birth order number and mother's working status are significant independent variables for the model using the year 2004 data. The model explains the approximately 2.18% of variation in child's weight at birth if we consider all the data variables in the model using 2004 data. Since the p-value is small, the hypothesis that all the model coefficients are zero is rejected at 5% level of significance (p-value = 0.02179) . For the final inferences in the multiple linear regression analysis, the results are not approximately the same for the 2010 and 2004 data sets, even though the model is the same. The conclusions made are valid at 5% level of significance for the data used in the study.

Multiple Linear Regression Output for the year 2000

When the same model as the above is fitted for the year 2000 data we get the following parameter estimates.

Table 10: Year 2000 Multiple linear regression

Covariates	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	2150.518	66.535	32.322	<2e-16 ***
Mother's age (Ref:15-19)				
Age 20-24	30.254	39.398	0.768	0.443
Age 25-29	16.75	46.858	0.357	0.721
Age 30-34	63.129	57.213	1.103	0.27
Age 35-39	51.737	70.812	0.731	0.465
Age 40-49	134.347	87.299	1.539	0.124
Child's birth order number	-8.87	9.071	-0.978	0.328
Region (Ref: Northern)				
Central region	62.479	38.911	1.606	0.109
Southern region	59.326	38.534	1.54	0.124
Child's gender (Ref: Male)				
Female	30.274	23.563	1.285	0.199
Education level (Ref: No education)				
Primary education	-15.721	29.169	-0.539	0.59
Secondary +	16.539	45.798	0.361	0.718
Mother's working status (Ref: Not working)				
Worked (past year)	74.714	61.93	1.206	0.228
Working	25.091	25.778	0.973	0.331
Antenatal visits	3.565	6.648	0.536	0.592
HIV Status (Ref: positive)				
Negative	27.138	23.754	1.142	0.254

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 324700 on 740 degrees of freedom
Multiple R-squared: 0.01708, Adjusted R-squared: -0.002849
F-statistic: 0.857 on 15 and 740 DF, p-value: 0.6134

The model output has no significant variables for the year 2000 data, which is not the good regression results compared to other years (2010 and 2004).

4.2.2 Hierarchical (Multilevel) Regression

Below is the multilevel linear regression output executed using the linear and nonlinear mixed effects model (nlme) in R package. Multilevel models are appropriate for a particular kind of data structure where units are nested within groups and where we want to model the group structure of the data. For the weight at birth model, the mixed model has region treated as random effects of the model and all other variables are fixed. The birth weight regression results below are analysed using the MDHS data for the years 2000, 2004, and 2010.

Multilevel Regression Output for the year 2010

Linear mixed-effects model fit using the restricted (or residual) maximum likelihood (REML)

Table 81: Year 2010 Multilevel Regression Output

Covariate	Estimate	Std. Error	DF	t-value	p-value
(Intercept)	2147.51	50.06	1574	42.89	0.0000
Mother's age (Ref=15-19)					
20-24	-1.42	31.71	1574	-0.044	0.9643
25-29	56.06	35.54	1574	1.577	0.1149
30-34	94.97	42.61	1574	2.228	0.0260
35-39	105.89	51.99	1574	2.036	0.0418
40-49	104.74	60.34	1574	1.735	0.0828
Child's birth order number	-20.81	6.972	1574	-2.98	0.0029
Child's gender (Ref: Male)					
Female	6.68	17.99	1574	0.371	0.7105
Education level (Ref: No education)					
Primary education	-3.52	27.61	1574	-0.127	0.8986
Secondary +	45.58	35.31	1574	1.290	0.1970
Mother's working status (Ref: Not working)					
Worked (past year)	-55.74	27.42	1574	-2.032	0.0422
Working	-19.46	21.66	1574	-0.898	0.3692
Antenatal visits	6.19	3.319	1574	1.864	0.0624
Mother's HIV status (Ref: HIV positive)					
Negative	71.58	23.33	1574	3.067	0.0022

The description of the random effects shows the measure of variance at the different levels in the design expressed as a standard deviation of 23.607. This shows that there was a variation between regions and between the residuals in the model. The variables age of a mother, child's birth order number, mother's working status and mother's HIV status are significant variables for the model using 2010 data. Both the standard error and parameter estimates of the multilevel linear regression model seem to differ from the Multiple Linear Regression model output for the 2010 data. The estimates of fixed effects are in the model output; this means that the average differences across observations are controlled. The average weight at birth of a child when all other variables are zero is 2147.51. The weight at birth of a child increases with a decrease in birth order number of a child. The average difference in child's weight at birth between HIV positive and HIV negative mothers is 71.58. This means that the weight at birth of a child with a mother who is not HIV-infected is better than that of a child with an HIV positive mother.

Table 12: Correlations for the Year 2010 Multilevel Linear Regression

	Intr	fctr2	fct3	fct4	fctr5	fctr6	bord	fc(4)2	fctr(d)1	fctr(d)2	fctr(w)1	fctr(w)2	m14
fctr(agegrp)2	-0.415												
fctr(agegrp)3	-0.328	0.689											
fctr(agegrp)4	-0.242	0.611	0.718										
fctr(agegrp)5	-0.187	0.534	0.673	0.725									
fctr(agegrp)6	-0.176	0.479	0.630	0.696	0.721								
bord	-0.228	-0.164	-0.423	-0.580	-0.678	-0.692							
fctr(b4)2	-0.164	0.004	-0.024	-0.043	-0.012	-0.012	0.006						
fctr(educgrp)1	-0.533	0.027	0.027	0.046	0.066	0.101	0.111	0.012					
fctr(educgrp)2	-0.463	-0.050	-0.123	-0.112	-0.072	-0.042	0.277	0.011	0.700				
fctr(work)1	-0.242	0.027	0.049	-0.002	0.016	-0.004	0.012	-0.069	-0.002	0.041			
fctr(work)2	-0.278	-0.005	-0.042	-0.089	-0.062	-0.079	0.039	-0.037	-0.038	0.038	0.539		
m14	-0.218	0.002	-0.037	-0.014	-0.010	-0.019	0.011	-0.011	-0.016	0.003	-0.020	0.014	
fctr(status)2	-0.369	-0.012	0.054	0.076	0.050	0.093	-0.053	-0.018	-0.032	-0.053	0.012	0.033	-0.027

```
Standardized Within-Group Residuals:
 Min         Q1          Med         Q3          Max
-5.6156719   -0.4895505 0.2142568   0.7658392 1.3737537
Number of Observations: 1590
Number of regions: 3
```

Key of the variables listed in table 12 is attached in the Appendix.

The results above show the number of observations to be 1590 in three separate groups. Correlation among the fixed effects variable is used to access multicollinearity. The results show that the predictors are not related, (with the expected exception of the categories of class). Therefore, multicollinearity is not a concern.

Multilevel Regression Output for the year 2004

Linear mixed-effects model fit using the restricted (or residual) maximum likelihood (REML)

Table 13: Year 2004 Multilevel linear regression output

Covariates	Value	Std.Error	DF	t-value	p-value
(Intercept)	2156.60	70.64	573	30.525	0.0000
Mother's age (Ref=15-19)					
20-24	99.09	54.07	573	1.832	0.0674
25-29	107.13	60.24	573	1.778	0.0759
30-34	123.01	70.49	573	1.744	0.0815
35-39	74.70	82.57	573	0.904	0.3660
40-49	224.98	96.61	573	2.328	0.0202
Child's birth order number	-20.23	10.75	573	-1.881	0.0605
Child's gender (Ref: Male)					
Female	25.12	29.30	573	0.857	0.3915
Mother's education level (Ref: No education)					
Primary education	-44.35	38.34	573	-1.156	0.2479
Secondary+	32.79	52.34	573	0.626	0.5312
Mother's working status (Ref: Not working)					
Worked (past year)	13.39	82.46	573	0.162	0.8710
Working	55.41	30.09	573	1.841	0.0661
Antenatal visits	-5.35	2.54	573	-2.102	0.0360
Mothers' HIV status (Ref: Positive)					
Negative	6.61	29.42	573	0.224	0.8223

Mother's age and antenatal visits are the significant variables for this model output. 2 156.60 is the average weight at birth of a child when all other variables are zero. The average weight at birth of a child with mothers who is aged 40-49 years is 224.98 higher than that of a child with a mother aged 15-24 years. The model output for the year 2004 has 589 number of observations and three groups, which is a lower number of observations compared to year 2010. The random effects estimates for the three regions are mentioned in table13 below.

Table 13: Random effects estimate

Region	Intercept
Northern	-0.0750
Central	-7.7185
Southern	7.7936

Multilevel Regression Output for the year 2000

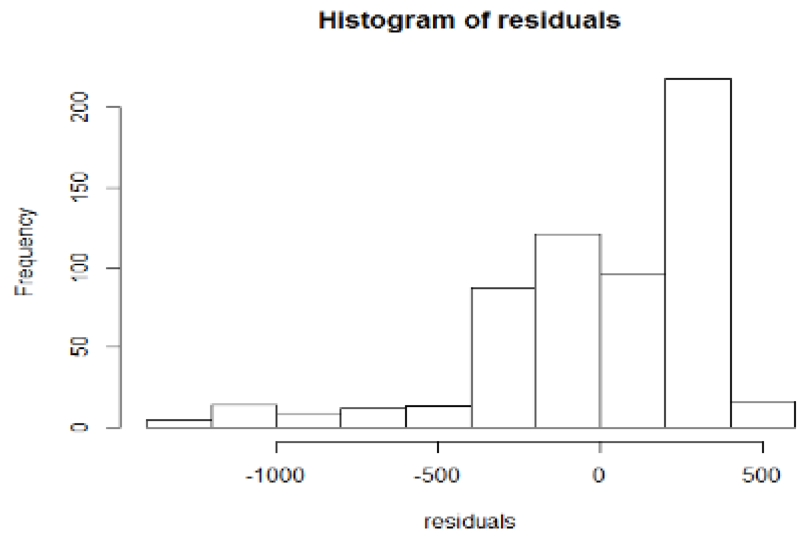
Linear mixed-effects model fit using the restricted (or residual) maximum likelihood (REML)

Table 14: Year 2000 Multilevel regression output

Covariates	Value	Std.Error	DF	t-value	p-value
(Intercept)	2200.184	54.289	740	40.527	0.0000
Mother's age (Ref=15-19)					
20-24	25.355	38.620	740	0.656	0.5117
25-29	-5.065	45.890	740	-0.110	0.9121
30-34	38.405	56.891	740	0.675	0.4999
35-39	32.233	71.033	740	0.454	0.6501
40-49	129.155	87.263	740	1.480	0.1393
Child's birth order number	-6.100	9.2797	740	-0.657	0.5111
Child's gender (Ref: Male)					
Female	42.178	23.659	740	1.783	0.0750
Mother's education level (Ref: No education)					
Primary	-9.840	29.910	740	-0.329	0.7423
Secondary+	41.209	45.027	740	0.915	0.3604
Mother's working status (Ref: Not working)					
Worked (past year)	64.648	61.535	740	1.050	0.2938
Working	29.058	25.499	740	1.139	0.2548
Antenatal visit	2.129	6.7152	740	0.317	0.7513
Mother's HIV status (Ref: Positive)					
Negative	27.698	23.871	740	1.160	0.2463

For a good model, the residuals should be normally distributed around a mean zero. The histogram of residuals plot below shows that residuals are almost negatively skewed for the weight at birth model.

Figure 2: Histogram of Residuals



The histogram of residuals seem to be negatively skewed.

4.3 Bayesian Multiple Linear Regression Output

One of the most important Bayesian aspects for analysis is to check whether the Markov Chains have indeed reached a stable equilibrium distribution, i.e., have *converged*. The time series plots, kernel density plots, and autocorrelation plots and Gelman-Rubin statistic set for each simulated parameter are generated as follows for the convergence diagnostics. For each Markov Chain in these models, we have a sample of more than 200 000 random draws from the joint posterior distribution of all the parameters in the model.

4.3.1 Non-informative Bayesian Linear Regression

This section covers the Bayesian linear regression with non-informative Bayesian prior for the years 2010, 2004 and year 2000. To choose these priors, a Normal distribution with large variance was used. These results will then be compared with classical results.

Table 9: Year 2010 Non-Informative Bayesian Linear Regression Output

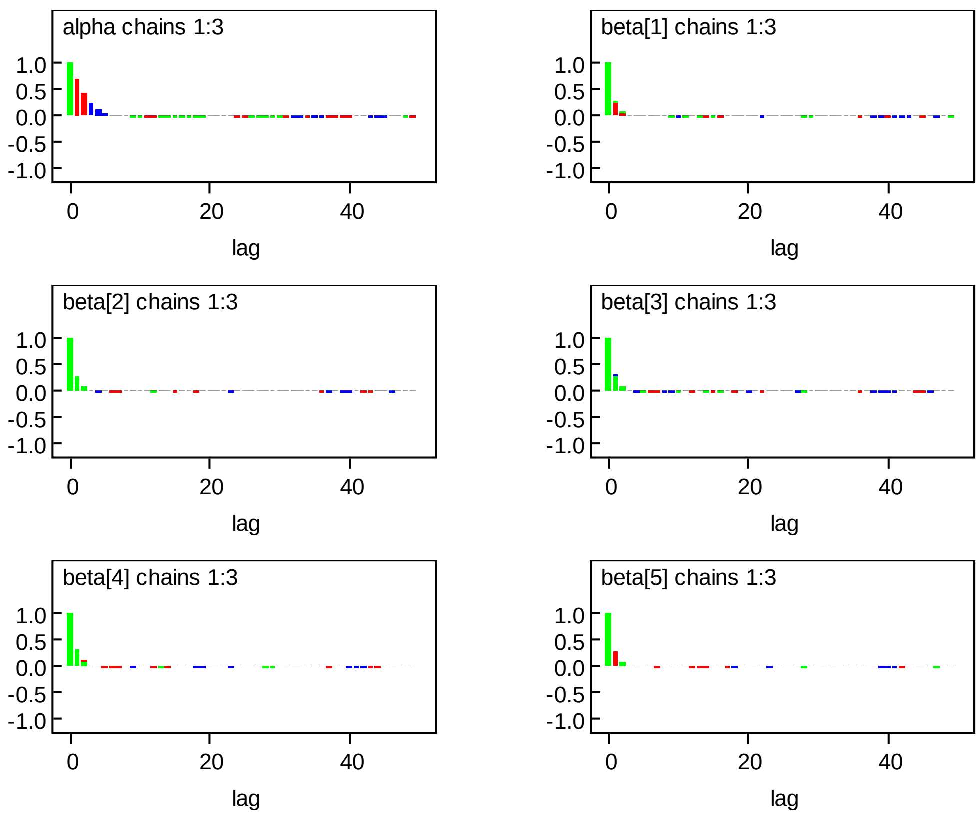
Node	Mean	Sd	MC error	2.5%	Median	97.5%	start	Sample
Alpha	2131.0	52.33	0.4139	2028.0	2131.0	2233.0	1000	59730
Mother's age (Ref=15-19)								
20-24	9.386	31.19	0.1577	-51.61	9.499	70.29	1000	59730
25-29	64.01	34.69	0.1903	-3.827	63.97	131.9	1000	59730
30-34	100.2	41.57	0.2223	18.27	100.2	181.7	1000	59730
35-39	107.1	50.51	0.2786	7.619	107.1	205.6	1000	59730
40-49	107.0	58.45	0.3178	-8.32	107.0	221.8	1000	59730
Child's birth order no.	-18.38	6.84	0.0381	-31.88	-18.39	-4.896	1000	59730
Region (Ref=Southern)								
Central region	-47.55	25.1	0.1154	-96.98	-47.5	1.467	1000	59730
Northern region	-33.04	24.56	0.1166	-81.16	-32.96	14.99	1000	59730
Child's gender (Ref=Male)								
Female	9.538	18.0	0.07674	-25.63	9.539	45.01	1000	59730
Education (Ref=No education)								
Primary education	60.09	35.24	0.188	-8.944	60.02	129.0	1000	59730
Secondary+	8.793	27.62	0.1558	-45.55	8.794	63.03	1000	59730
Employment status (Ref=No work)								
Worked (past year)	-48.78	27.19	0.1205	-102.1	-48.76	4.684	1000	59730
Work (currently)	-12.04	21.6	0.09175	-54.35	-12.06	30.18	1000	59730
Antenatal visits	6.93	3.332	0.01357	0.3974	6.928	13.5	1000	59730
Mother's HIV Status (Ref=Positive)								
Negative	80.53	23.41	0.1112	34.74	80.44	126.4	1000	59730
tau.e	7.89E-6	2.8E-7	1.13E-9	7.35E-6	7.89E-6	8.45E-6	1000	59730

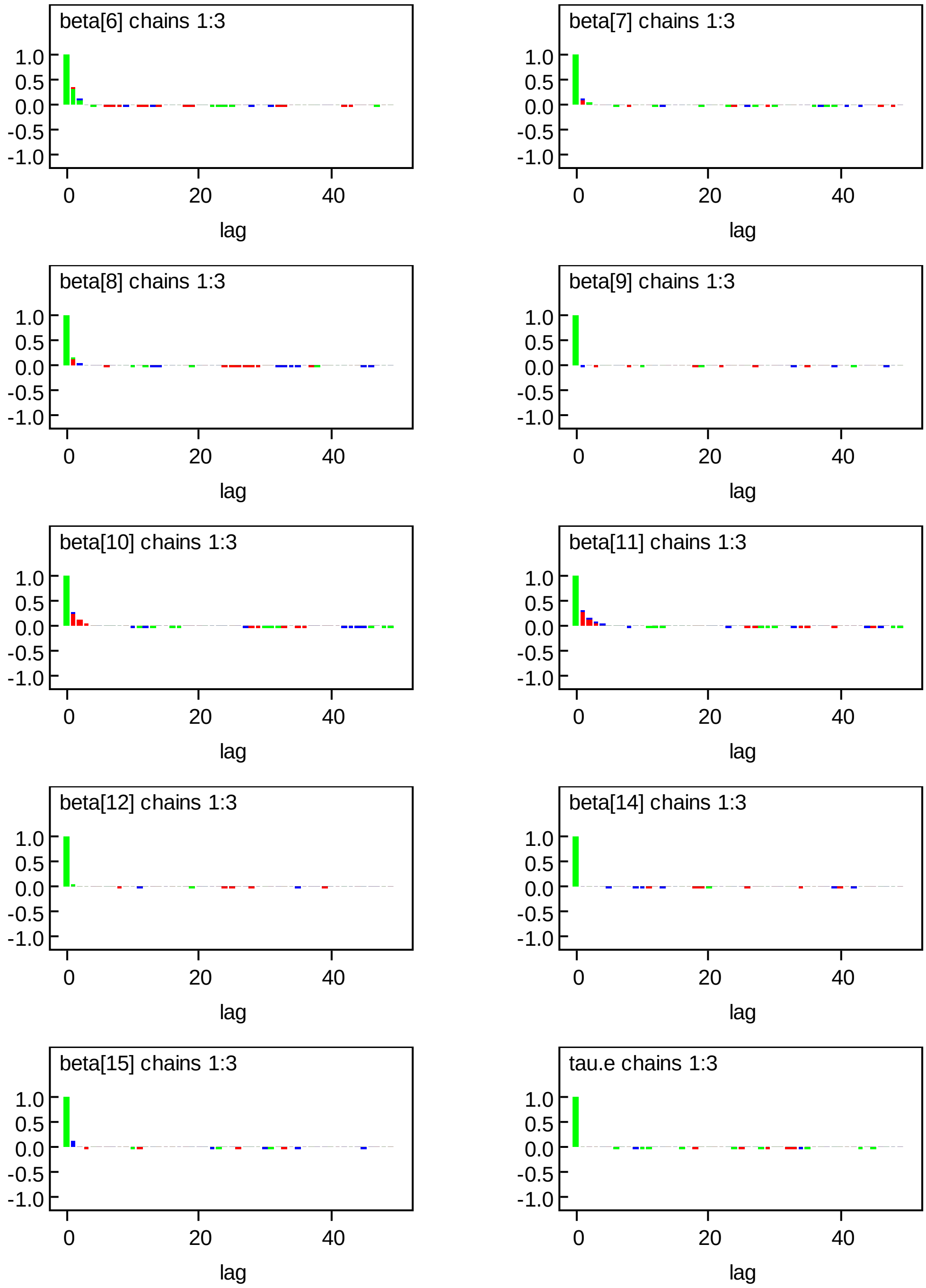
The node statistics give the formal parameter estimates. The non-informative Bayesian linear regression model above shows the mean for estimates, standard error and confidence intervals, which are used to make regression inferences about the estimates since the output has no p-values. A burn of 1000 followed by a further 59 730 updates gave the parameter estimates in Table 14. The variables mother's age, antenatal visits for pregnancy and HIV status of a mother are significant variables for the year 2010 non-informative Bayesian linear regression model.

Autocorrelation Function

The autocorrelation function refers to a pattern of serial correlation in the chain, where sequential draws of a parameter, say beta [5] (the mothers age group between 40 and 49 years old), from the conditional distribution are correlated. If the parameters in a model are highly correlated, this causes autocorrelation.

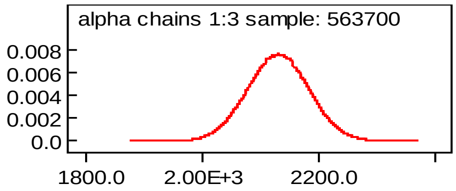
Figure 3: Year 2010 non-informative Bayesian autocorrelation function

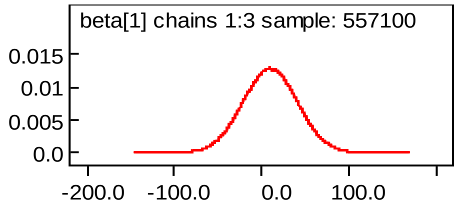


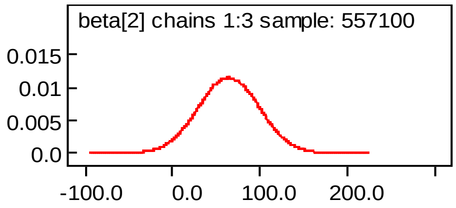


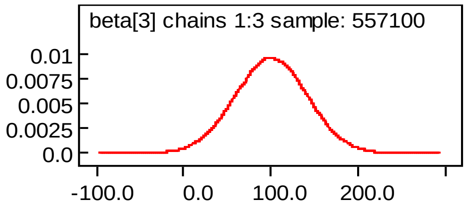
For this model, the chains are hardly autocorrelated at all. This is good, as our sample contains more information about the parameters than when successive draws are correlated.

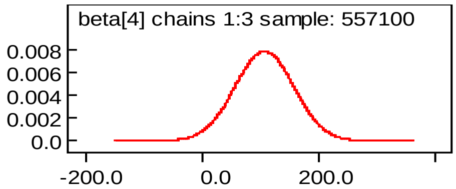
Figure 4: Year 2010 non-informative Bayesian Kernel density

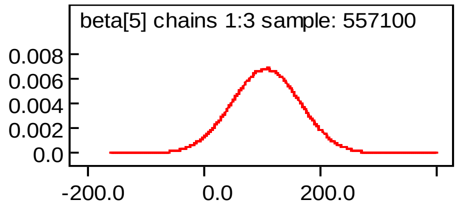

alpha chains 1:3 sample: 563700
0.008
0.006
0.004
0.002
0.0
1800.0
2.00E+3
2200.0

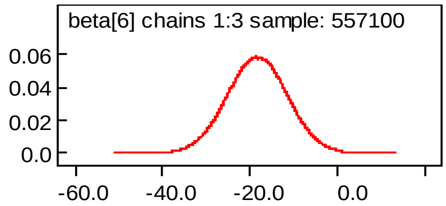

beta[1] chains 1:3 sample: 557100
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0
100.0

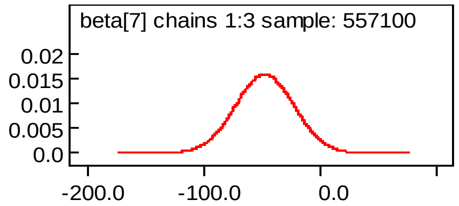

beta[2] chains 1:3 sample: 557100
0.015
0.01
0.005
0.0
-100.0
0.0
100.0
200.0

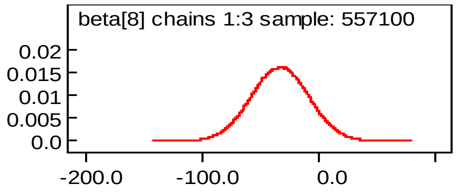

beta[3] chains 1:3 sample: 557100
0.01
0.0075
0.005
0.0025
0.0
-100.0
0.0
100.0
200.0

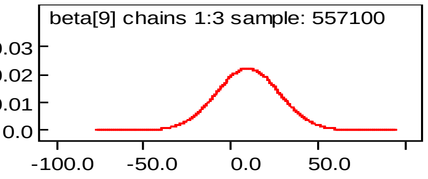

beta[4] chains 1:3 sample: 557100
0.008
0.006
0.004
0.002
0.0
-200.0
0.0
200.0

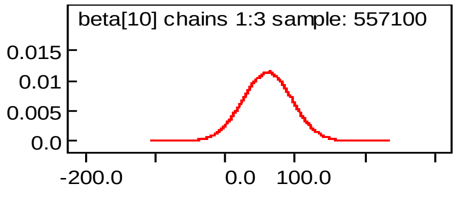

beta[5] chains 1:3 sample: 557100
0.008
0.006
0.004
0.002
0.0
-200.0
0.0
200.0

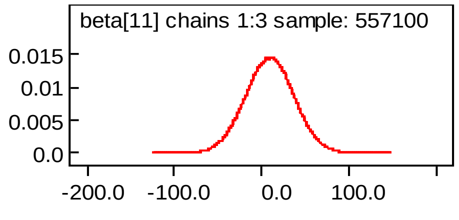

beta[6] chains 1:3 sample: 557100
0.06
0.04
0.02
0.0
-60.0
-40.0
-20.0
0.0

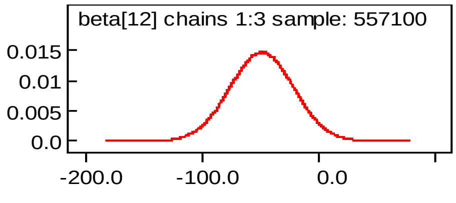

beta[7] chains 1:3 sample: 557100
0.02
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0

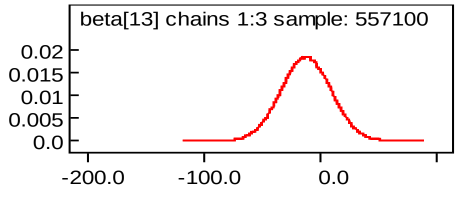

beta[8] chains 1:3 sample: 557100
0.02
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0

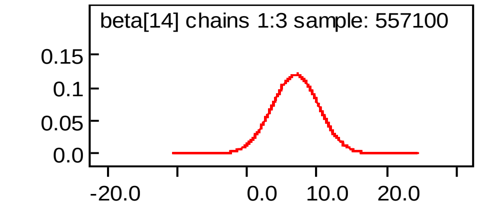

beta[9] chains 1:3 sample: 557100
0.03
0.02
0.01
0.0
-100.0
-50.0
0.0
50.0

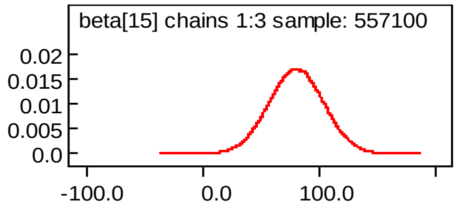

beta[10] chains 1:3 sample: 557100
0.015
0.01
0.005
0.0
-200.0
0.0
100.0

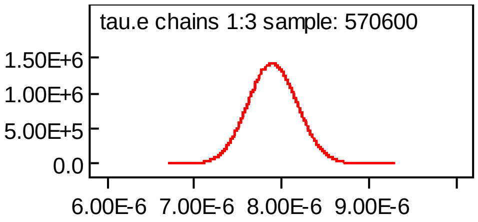

beta[11] chains 1:3 sample: 557100
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0
100.0


beta[12] chains 1:3 sample: 557100
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0


beta[13] chains 1:3 sample: 557100
0.02
0.015
0.01
0.005
0.0
-200.0
-100.0
0.0

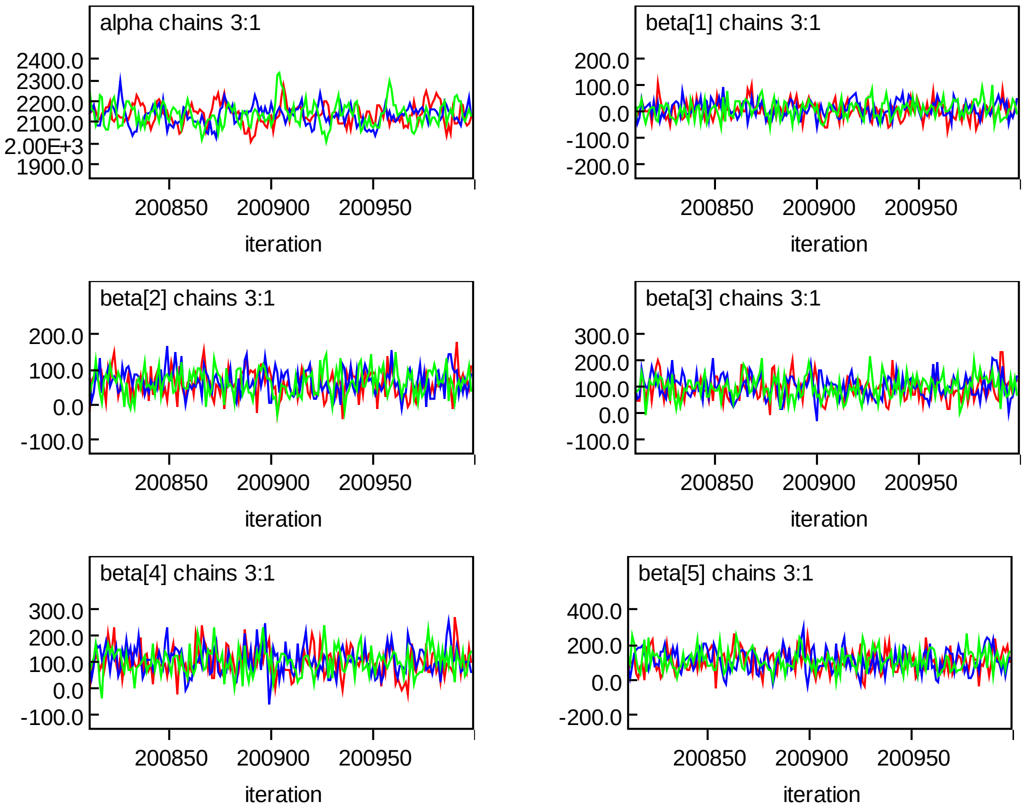


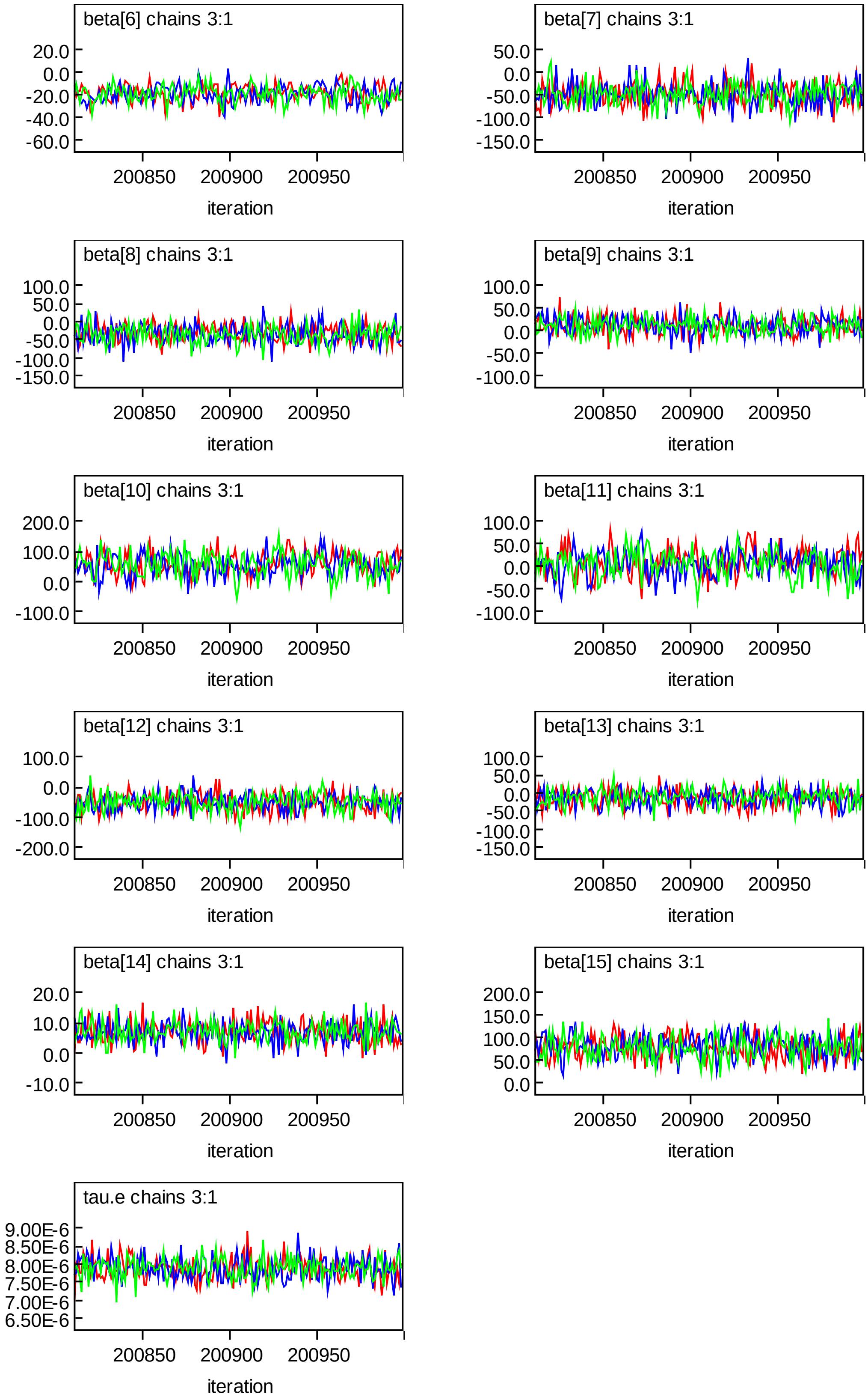




The assumption that the density function of a posterior distribution is normally distributed is met for all the variables of the non-informative Bayesian model.

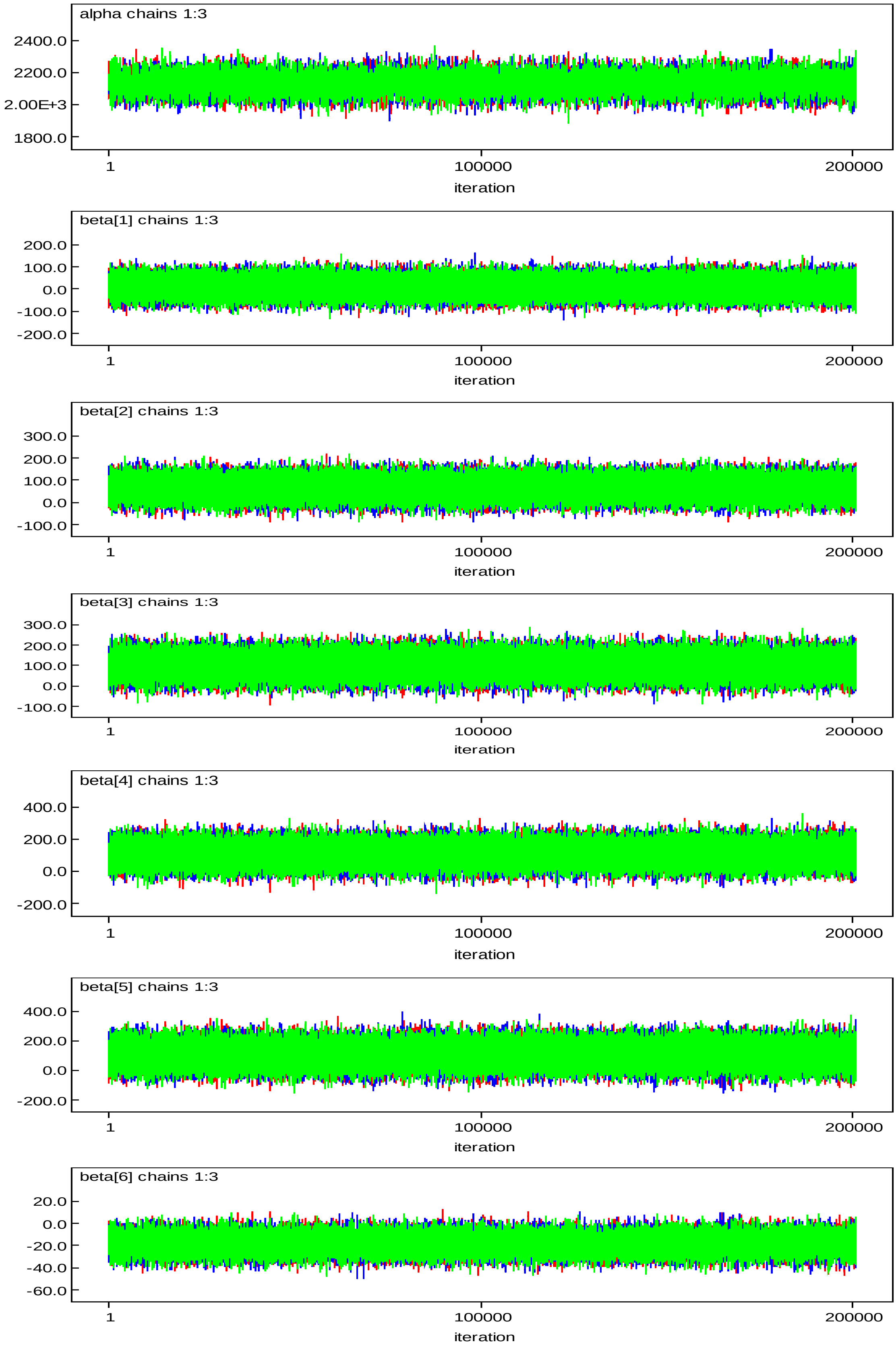
Figure 5: Year 2010 non-informative Bayesian dynamic trace

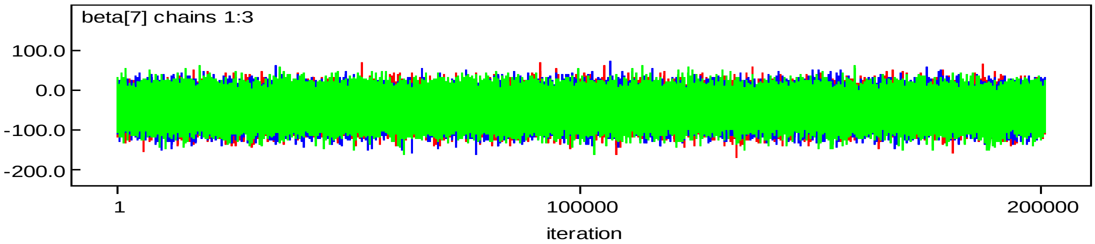


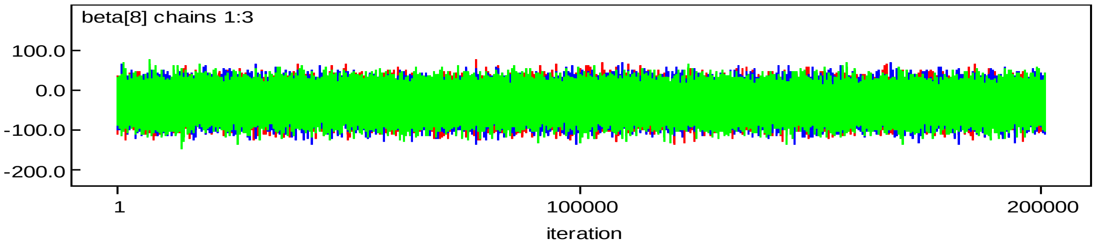


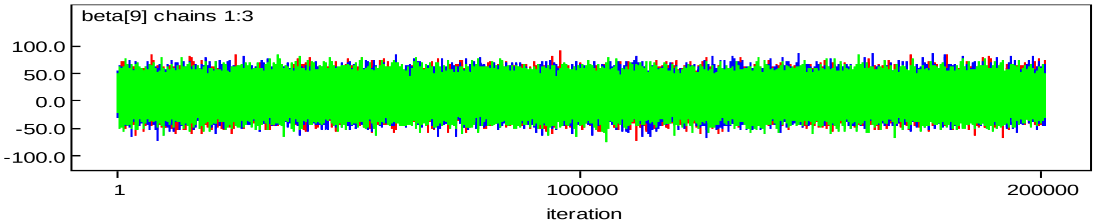
The trace provides a moving time series plot of the posterior draws for each selected parameter in the model. This model gives a smoothly moving dynamic trace of the Markov Chains.

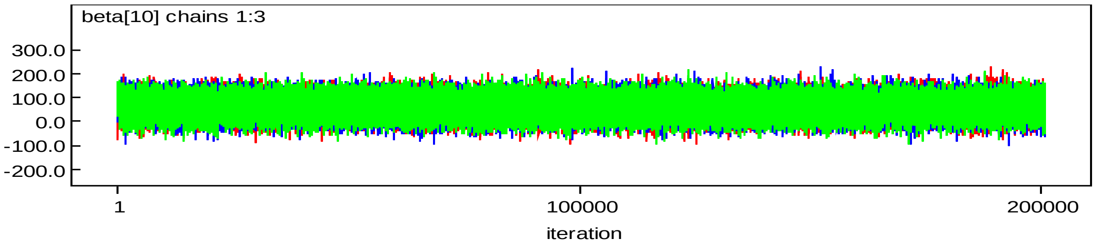
Figure 6: Year 2010 Non-Informative Bayesian Time Series

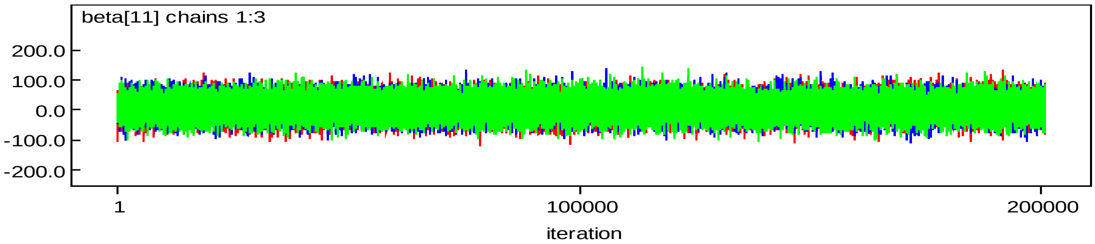


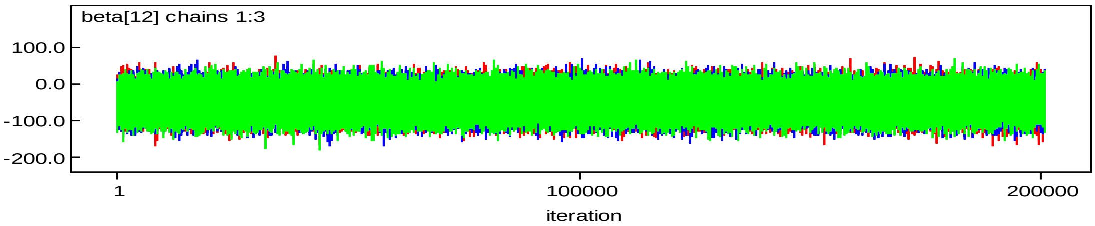

beta[7] chains 1:3
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration

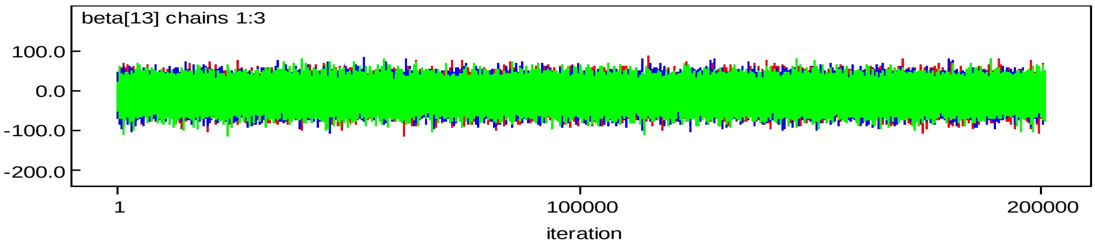

beta[8] chains 1:3
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration

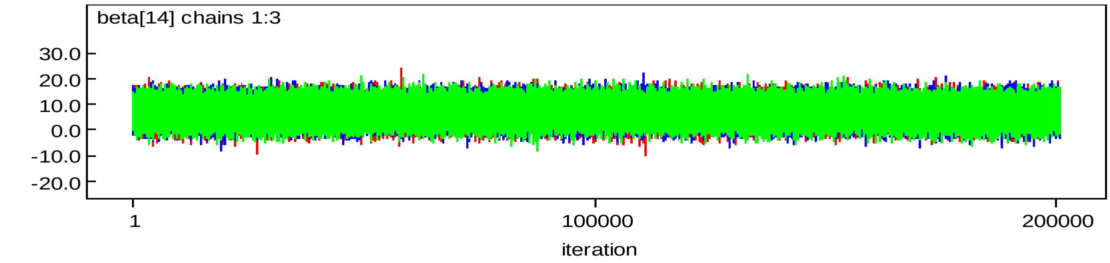

beta[9] chains 1:3
100.0
50.0
0.0
-50.0
-100.0
1
100000
200000
iteration

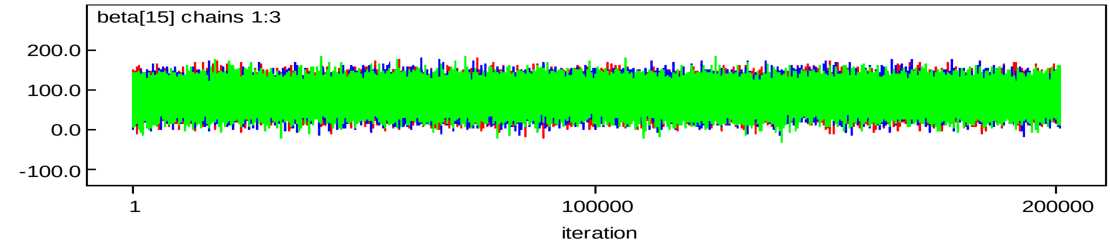

beta[10] chains 1:3
300.0
200.0
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration

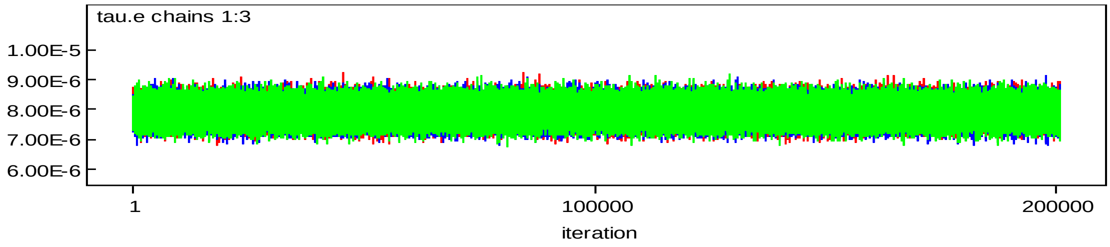

beta[11] chains 1:3
200.0
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration


beta[12] chains 1:3
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration


beta[13] chains 1:3
100.0
0.0
-100.0
-200.0
1
100000
200000
iteration







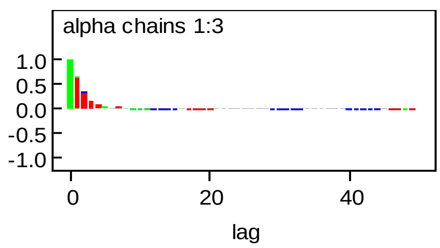
Visual inspection of the time series plot produced suggest that the Markov chains have converged.

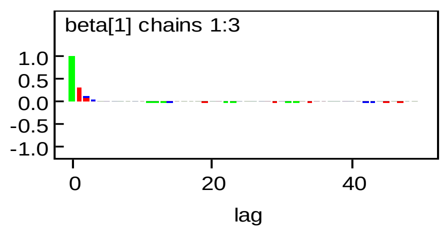
Table 10: Year 2004 Non-Informative Bayesian Linear Regression Output

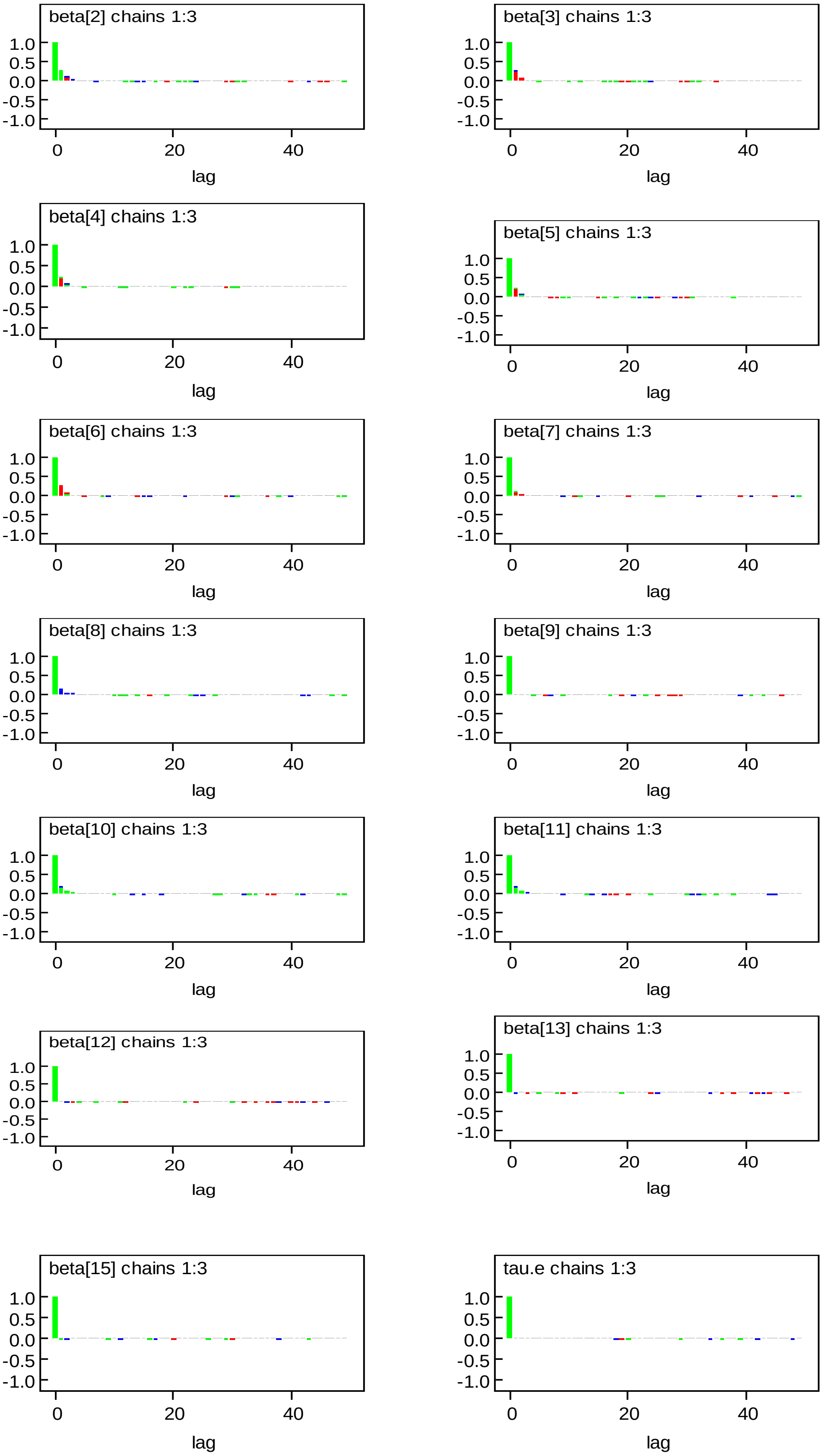
node	mean	sd	MC error	2.5%	median	97.5%	start	Sample
alpha	2034.0	77.29	0.56	1883.0	2034.0	2185.0	1000	65700
Mothers' age (Ref=15-19)								
20-24	128.7	51.38	0.264	28.74	129.0	229.5	1000	65700
25-29	129.1	56.64	0.2879	17.81	128.9	239.5	1000	65700
30-34	138.7	65.99	0.3086	9.444	138.3	269.6	1000	65700
35-39	88.01	77.32	0.3559	-64.26	88.28	239.9	1000	65700
40-49	219.3	90.14	0.4223	43.19	219.5	397.8	1000	65700
Childs' birth order no.	-11.91	10.4	0.0512	-32.28	-11.93	8.604	1000	65700
Region (Ref: Southern)								
Central	-1.267	44.9	0.2038	-89.18	-1.273	87.2	1000	65700
Northern	53.64	40.07	0.1909	-24.97	53.76	133.1	1000	65700
Child's gender (Ref: Male)								
Female	35.43	29.29	0.1112	-22.21	35.46	92.81	1000	65700
Education level (Ref: No education)								
Primary	78.51	52.15	0.2608	-22.84	78.56	181.3	1000	65700
Secondary+	-8.222	38.16	0.1876	-83.1	-8.231	66.55	1000	65700
Woking status (Not working)								
Worked (past year)	20.02	79.84	0.299	-136.2	20.22	177.3	1000	65700
Working	62.7	29.98	0.1141	3.627	62.75	121.3	1000	65700
Antenatal visits	-5.043	2.556	0.0097	-10.07	-5.042	-0.0448	1000	65700
Mother's HIV status (Ref: HIV positive)								
HIV negative	12.32	29.55	0.1163	-45.12	12.22	70.33	1000	65700
tau.e	8.089E-6	4.784E-7	1.95E-9	7.179E-6	8.078E-6	9.053E-6	1000	65700

The variables age of a mother, mother's working status and antenatal visits are significant parameters for this model when using the 95% confidence interval. The autocorrelation function, density function and time series for this parameter are shown below.

Figure 7: Year 2004 Non-Informative Bayesian Dynamic Trace

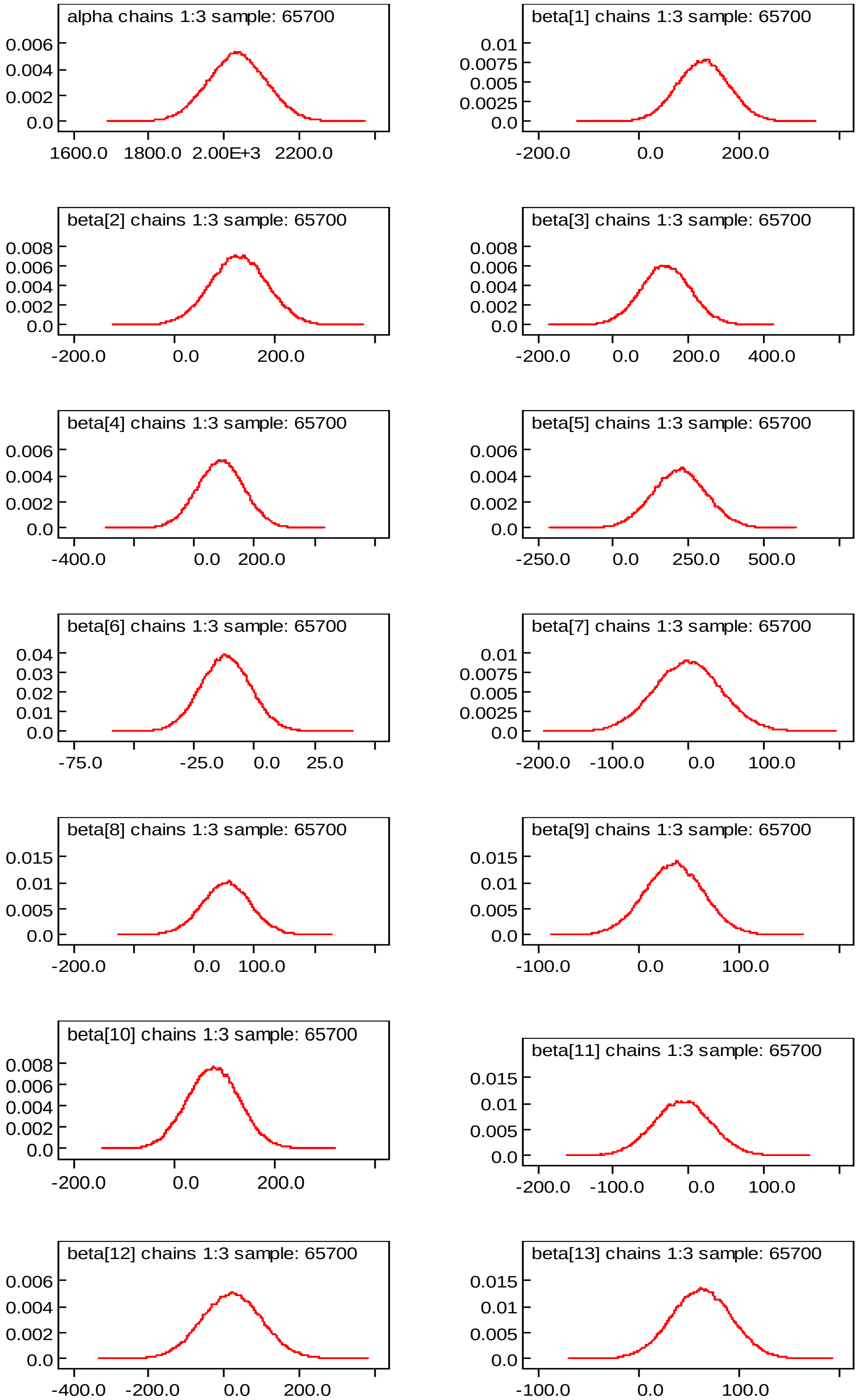


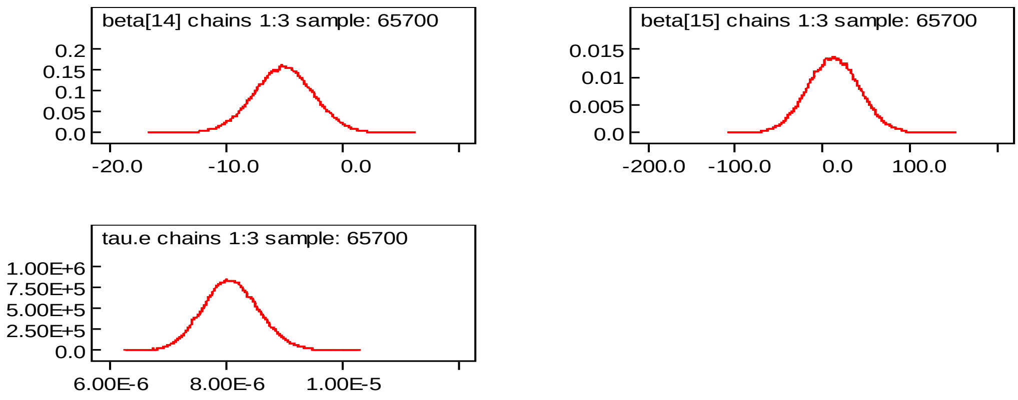




The chains are hardly correlated for the year 2004 data.

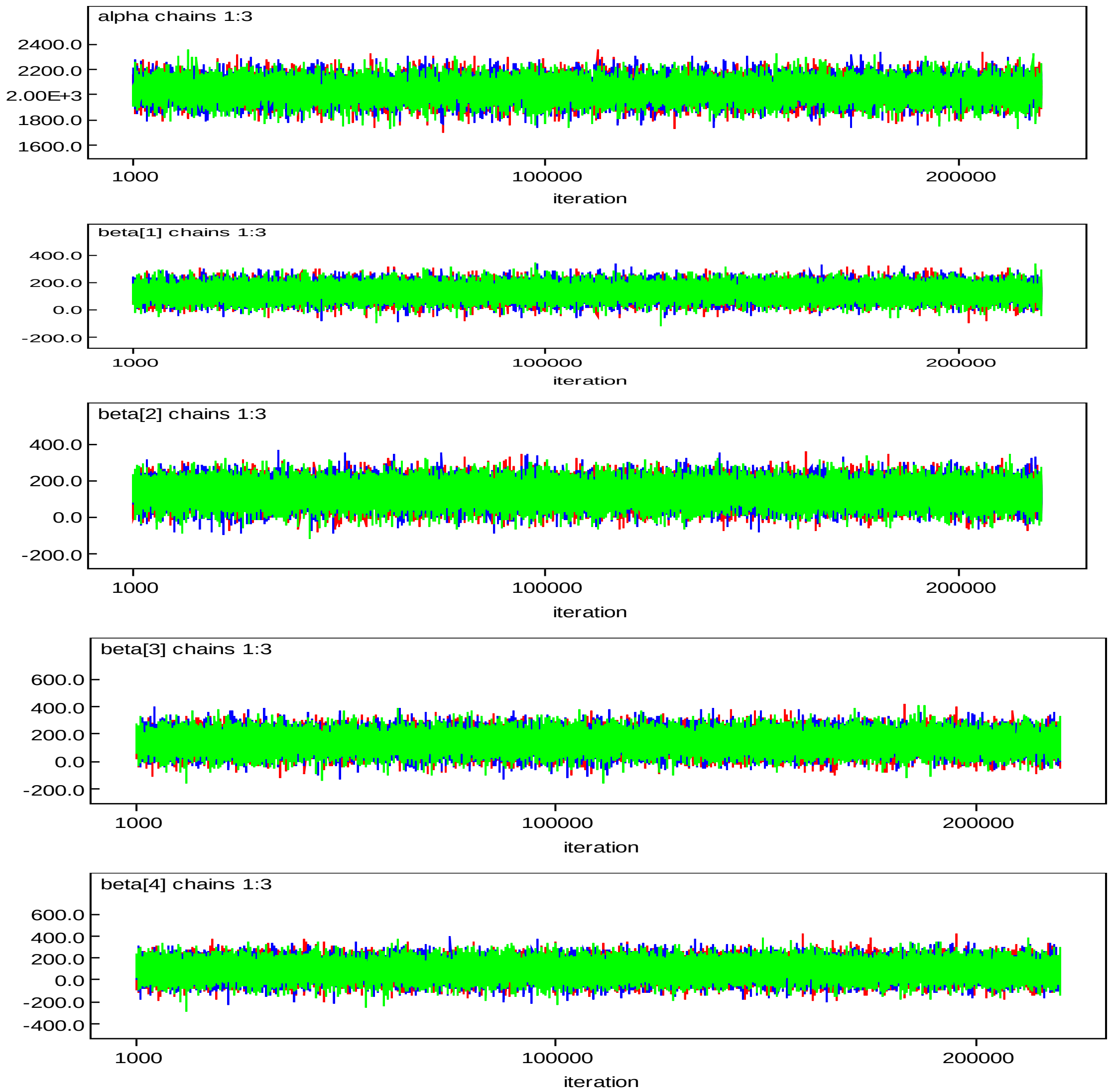
Figure 8: Year 2004 Non-Informative Bayesian Kernel Density

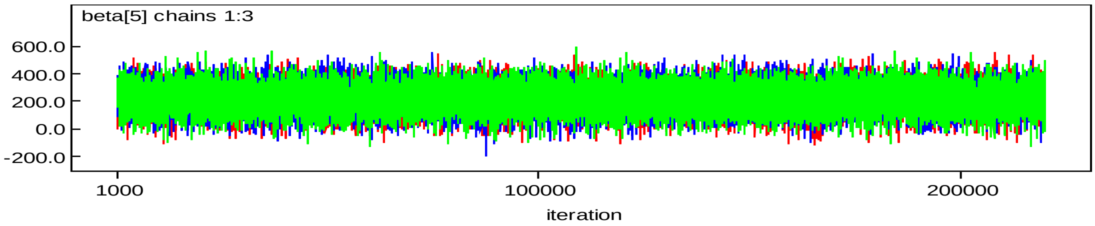


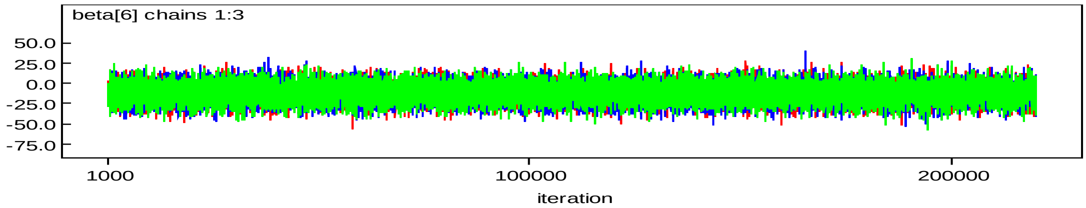


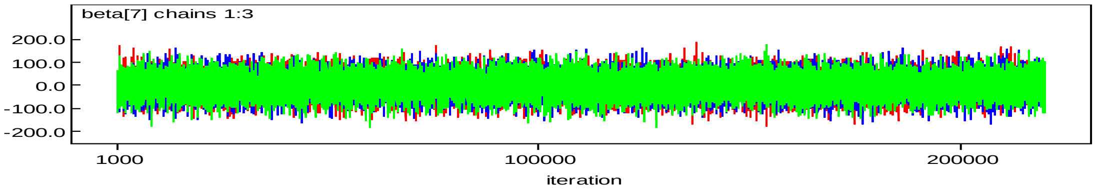
The assumption that the density function of a posterior distribution is normally distributed is met for all the variables of the Non-Informative Bayesian Model.

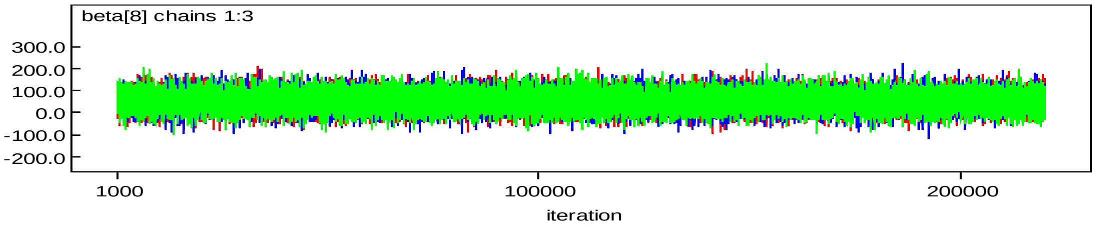
Figure 9: Year 2004 Non-Informative Bayesian Time Series Plot

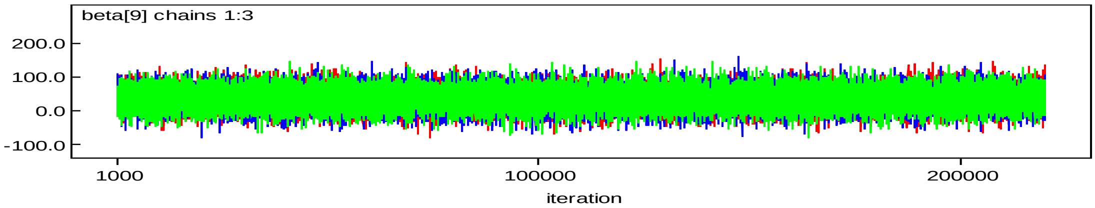


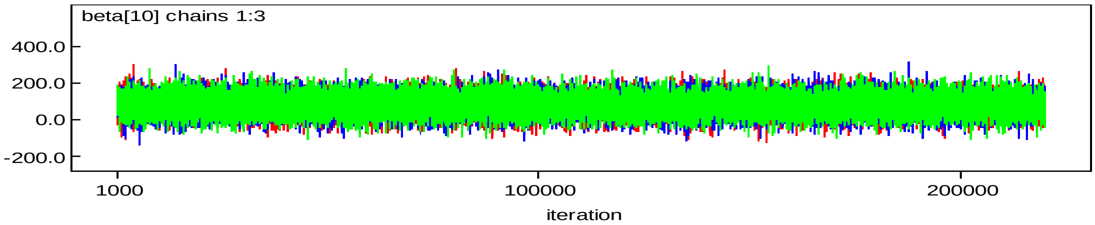

beta[5] chains 1:3
600.0
400.0
200.0
0.0
-200.0
1000
100000
200000
iteration

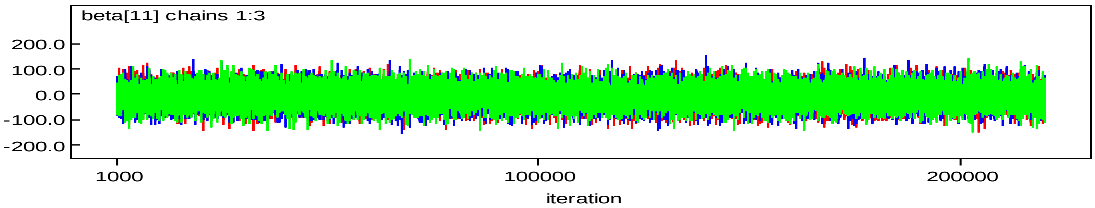

beta[6] chains 1:3
50.0
25.0
0.0
-25.0
-50.0
-75.0
1000
100000
200000
iteration

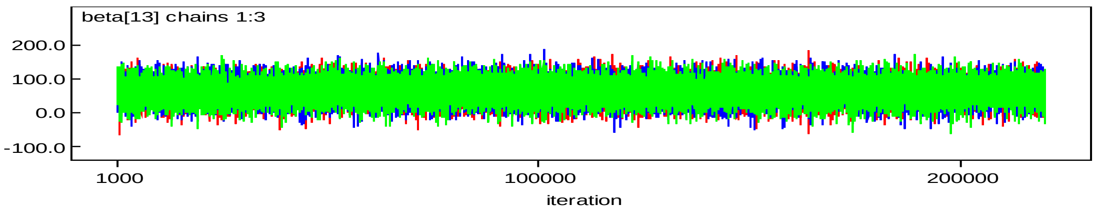

beta[7] chains 1:3
200.0
100.0
0.0
-100.0
-200.0
1000
100000
200000
iteration

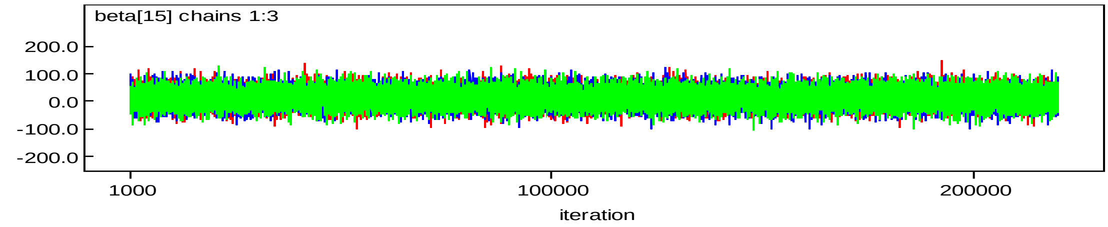

beta[8] chains 1:3
300.0
200.0
100.0
0.0
-100.0
-200.0
1000
100000
200000
iteration


beta[9] chains 1:3
200.0
100.0
0.0
-100.0
1000
100000
200000
iteration


beta[10] chains 1:3
400.0
200.0
0.0
-200.0
1000
100000
200000
iteration


beta[11] chains 1:3
200.0
100.0
0.0
-100.0
-200.0
1000
100000
200000
iteration


beta[13] chains 1:3
200.0
100.0
0.0
-100.0
1000
100000
200000
iteration



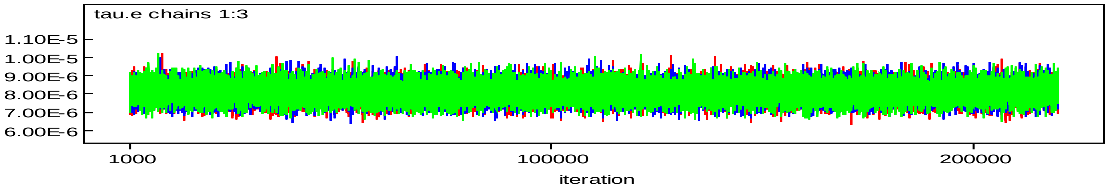
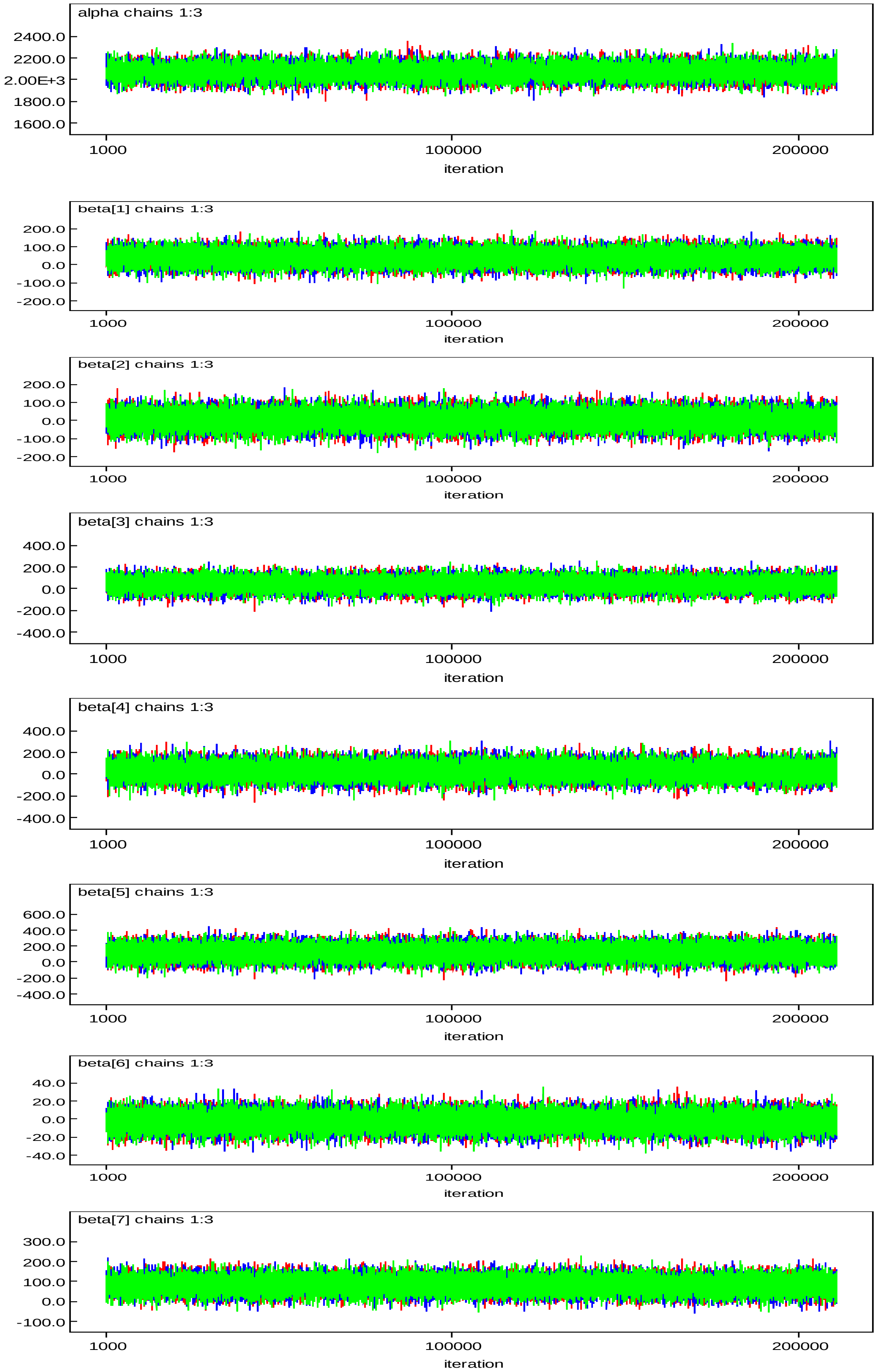


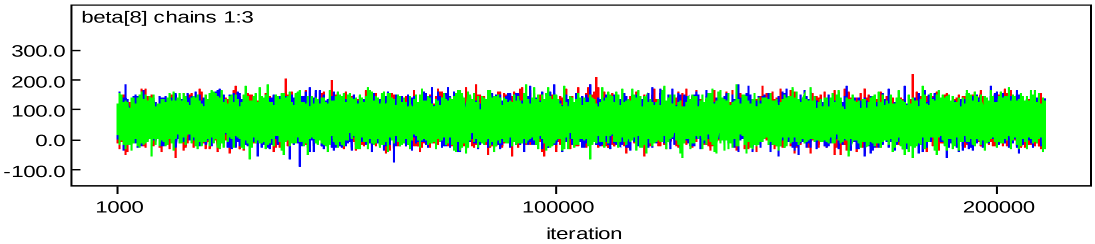
Table 11: Year 2000 Non-Informative Bayesian Linear Regression Output

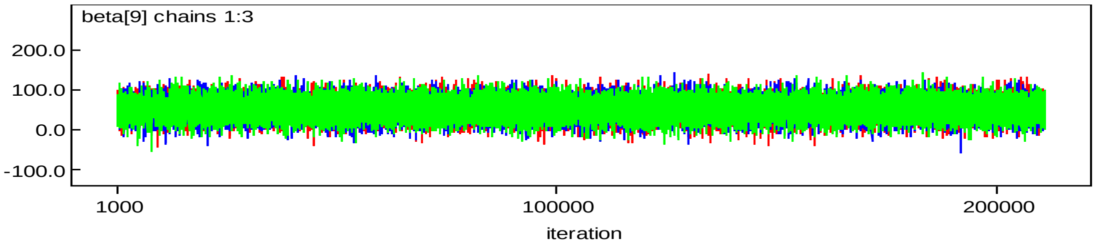
node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	2071.0	63.05	0.4506	1948.0	2072.0	2195.0	1000	63000
Mothers' age (Ref=15-19)								
20-24	39.69	37.68	0.1771	-34.14	39.75	113.5	1000	63000
25-29	4.2	44.11	0.2043	-82.1	4.292	90.23	1000	63000
30-34	40.06	54.24	0.253	-66.03	40.08	145.4	1000	63000
35-39	30.02	67.8	0.3319	-102.9	30.14	162.0	1000	63000
40-49	118.6	82.34	0.4004	-42.97	118.8	279.4	1000	63000
Childs' birth order number	-2.831	8.917	0.04651	-20.23	-2.829	14.64	1000	63000
Region (Ref: Southern)								
Central	81.1	35.88	0.1741	11.1	81.24	151.4	1000	63000
Northern	64.18	34.41	0.1703	-3.049	64.1	131.6	1000	63000
Child's gender (Ref: Male)								
Female	51.65	23.62	0.09579	5.44	51.64	97.86	1000	63000
Education level (Ref: No education)								
Primary	65.82	44.45	0.1872	-21.56	66.13	152.5	1000	63000
Secondary+	13.56	29.97	0.1352	-45.39	13.59	72.3	1000	63000
Woking status (Not working)								
Worked (past year)	80.38	60.85	0.227	-39.6	80.31	199.4	1000	63000
Working	38.38	25.39	0.1017	-11.5	38.41	88.11	1000	63000
Antenatal visits	6.802	6.787	0.03189	-6.514	6.792	20.12	1000	63000
Mother's HIV status (Ref: HIV positive)								
Negative	29.36	23.89	0.08945	-17.36	29.39	76.0	1000	63000
tau.e	9.625E-6	5.023E-7	2.076E-9	8.668E-6	9.616E-6	1.063E-5	1000	63000

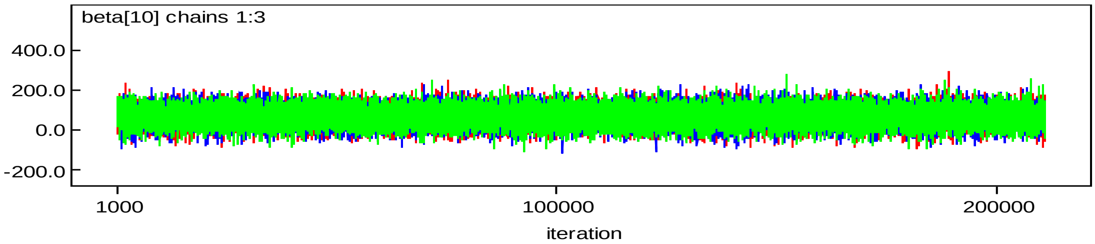
The region and gender of a child are significant at 95% confidence interval - therefore these predictors have an influence in child's weight at birth for the year 2000.The respective dynamic trace, autocorrelation, density function and time series are:

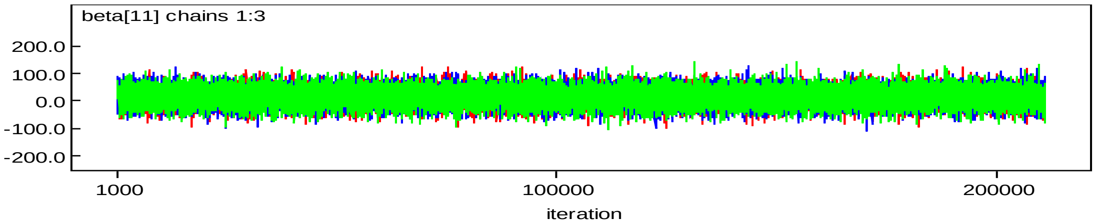
Figure 10: Year 2000 Non-Informative Bayesian Time Series Plot

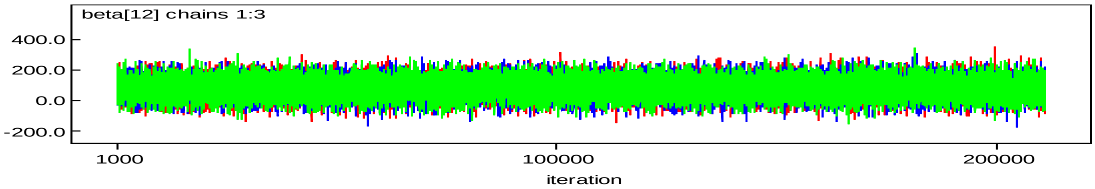


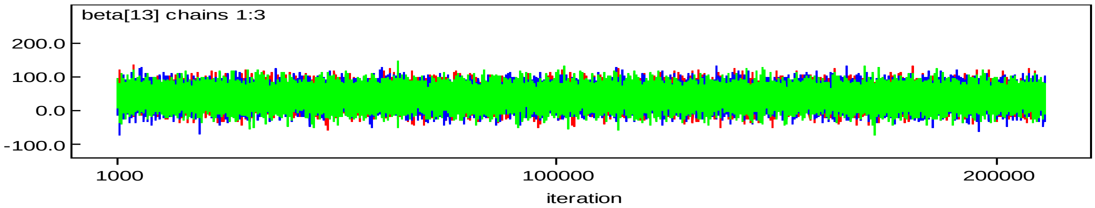

beta[8] chains 1:3
300.0
200.0
100.0
0.0
-100.0
1000
100000
200000
iteration

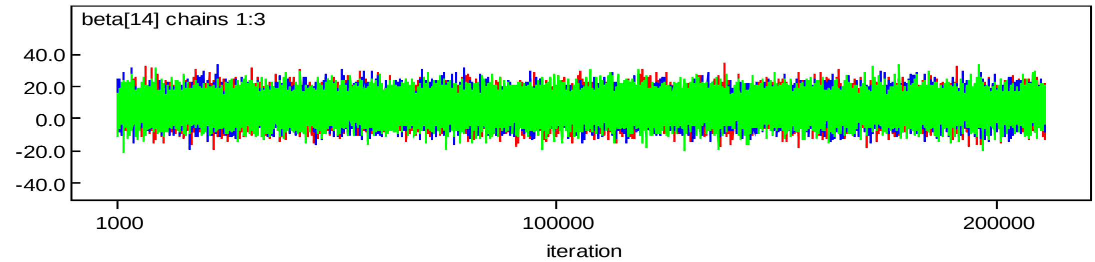

beta[9] chains 1:3
200.0
100.0
0.0
-100.0
1000
100000
200000
iteration


beta[10] chains 1:3
400.0
200.0
0.0
-200.0
1000
100000
200000
iteration


beta[11] chains 1:3
200.0
100.0
0.0
-100.0
-200.0
1000
100000
200000
iteration


beta[12] chains 1:3
400.0
200.0
0.0
-200.0
1000
100000
200000
iteration


beta[13] chains 1:3
200.0
100.0
0.0
-100.0
1000
100000
200000
iteration


beta[14] chains 1:3
40.0
20.0
0.0
-20.0
-40.0
1000
100000
200000
iteration

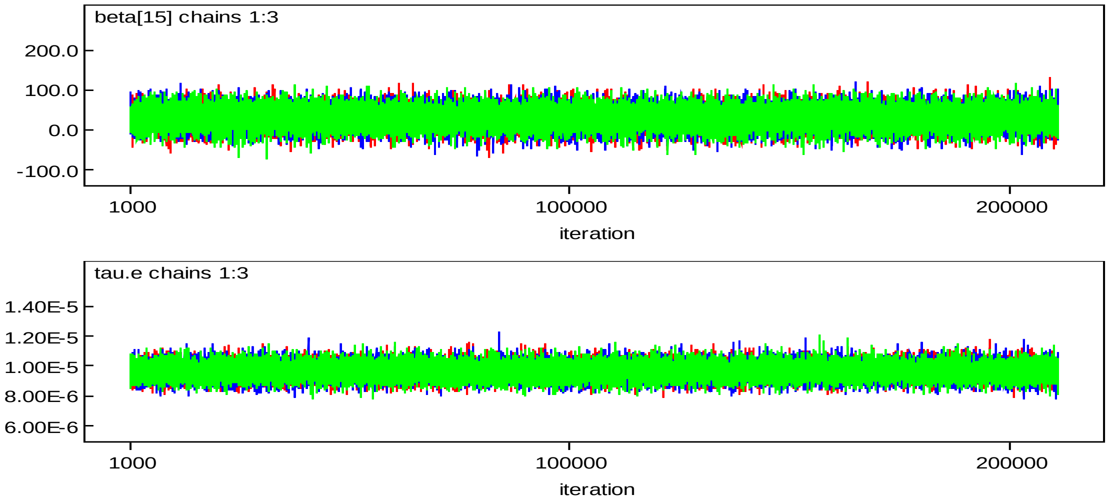
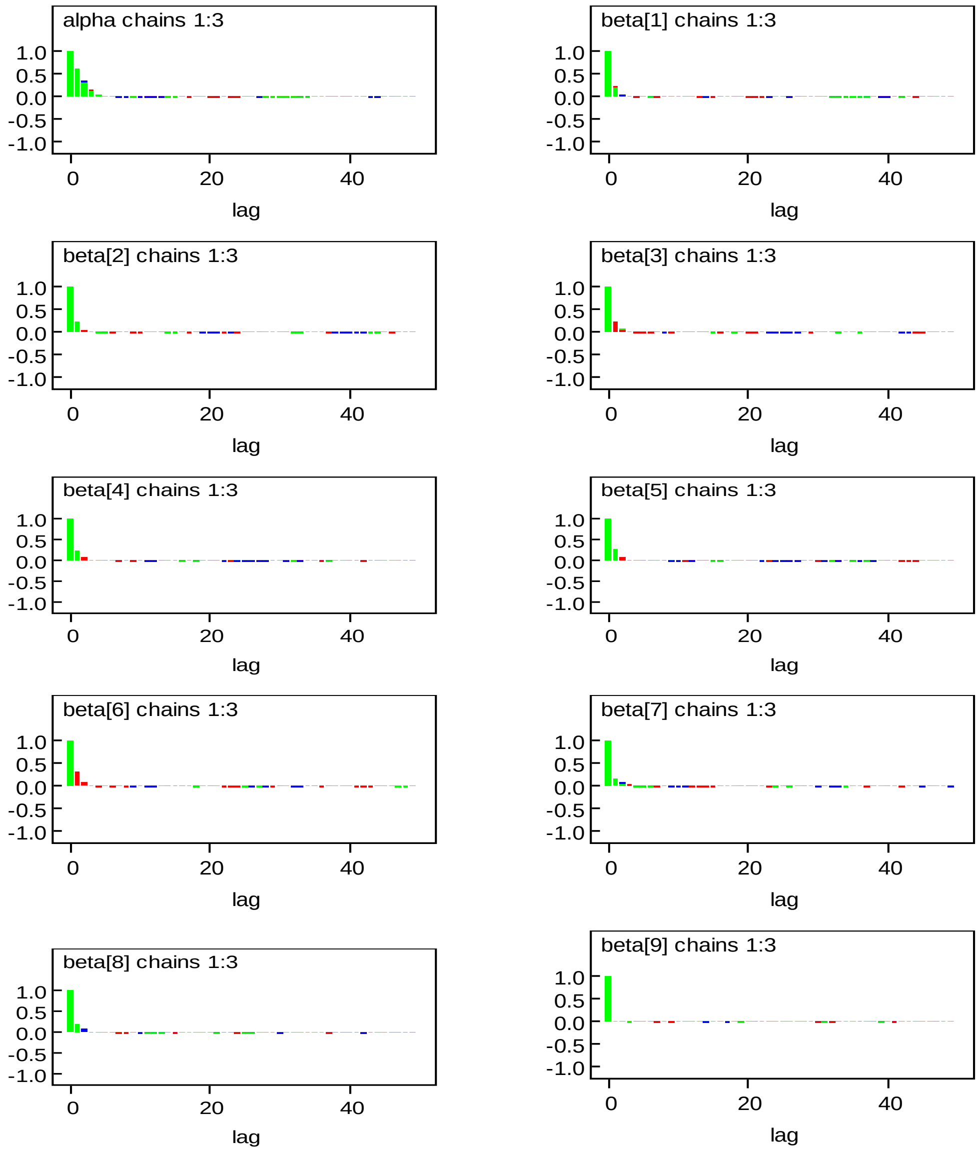
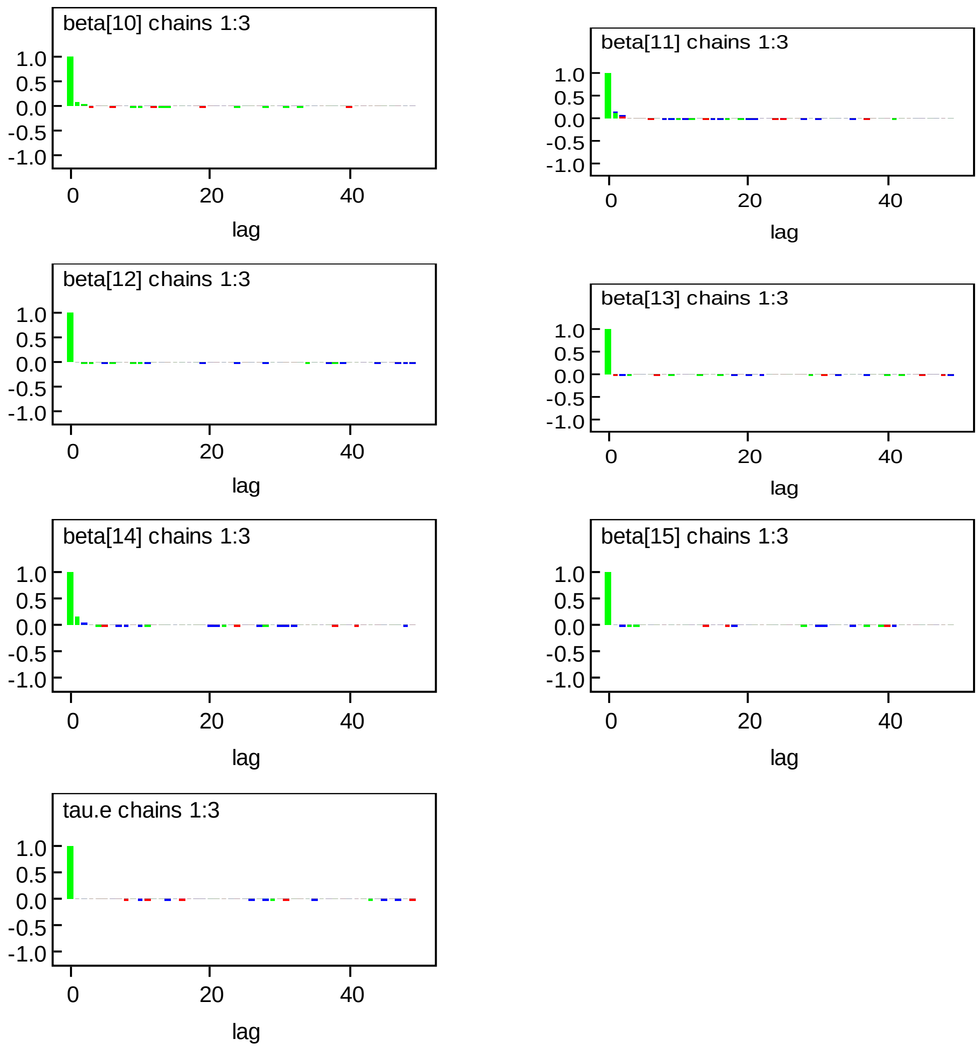


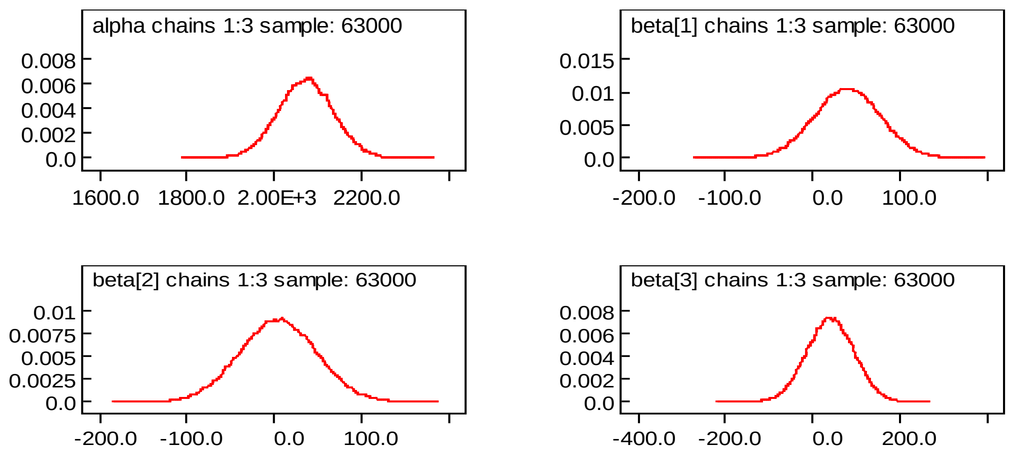
Figure 11: Year 2000 Non-Informative Bayesian Autocorrelation Function

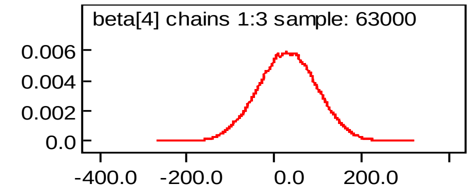


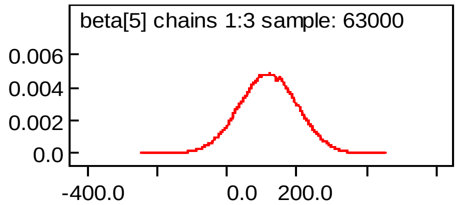


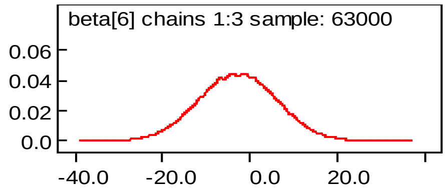
The year 2000 data shows that chains are hardly correlated.

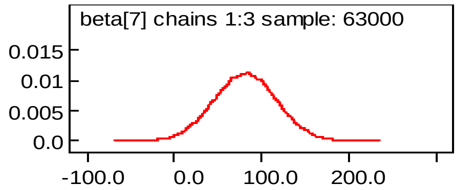
Figure 12: Year 2000 Non-Informative Bayesian Kernel Density

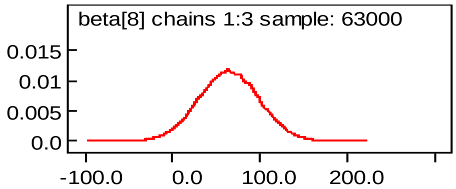


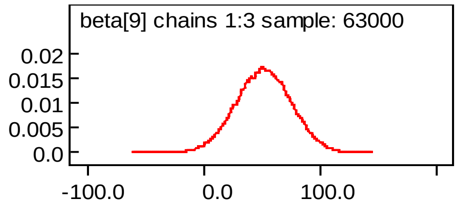

beta[4] chains 1:3 sample: 63000
0.006
0.004
0.002
0.0
-400.0 -200.0 0.0 200.0

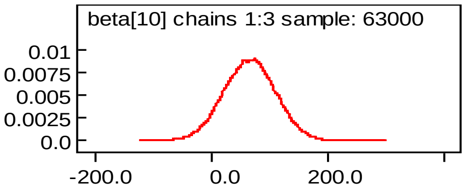

beta[5] chains 1:3 sample: 63000
0.006
0.004
0.002
0.0
-400.0 0.0 200.0

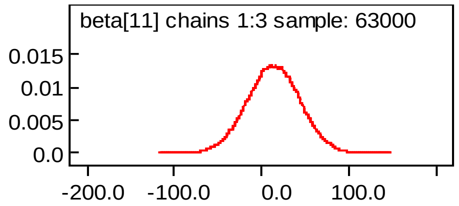

beta[6] chains 1:3 sample: 63000
0.06
0.04
0.02
0.0
-40.0 -20.0 0.0 20.0

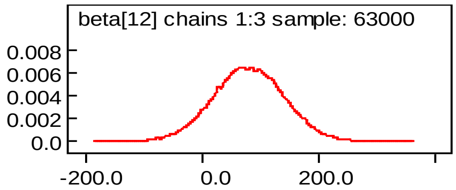

beta[7] chains 1:3 sample: 63000
0.015
0.01
0.005
0.0
-100.0 0.0 100.0 200.0

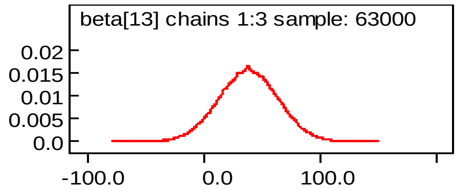

beta[8] chains 1:3 sample: 63000
0.015
0.01
0.005
0.0
-100.0 0.0 100.0 200.0

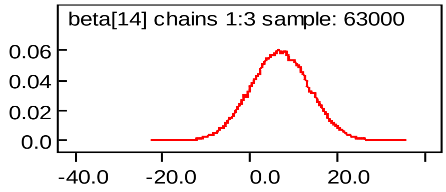

beta[9] chains 1:3 sample: 63000
0.02
0.015
0.01
0.005
0.0
-100.0 0.0 100.0

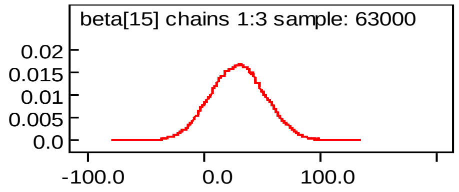

beta[10] chains 1:3 sample: 63000
0.01
0.0075
0.005
0.0025
0.0
-200.0 0.0 200.0


beta[11] chains 1:3 sample: 63000
0.015
0.01
0.005
0.0
-200.0 -100.0 0.0 100.0


beta[12] chains 1:3 sample: 63000
0.008
0.006
0.004
0.002
0.0
-200.0 0.0 200.0


beta[13] chains 1:3 sample: 63000
0.02
0.015
0.01
0.005
0.0
-100.0 0.0 100.0


beta[14] chains 1:3 sample: 63000
0.06
0.04
0.02
0.0
-40.0 -20.0 0.0 20.0


beta[15] chains 1:3 sample: 63000
0.02
0.015
0.01
0.005
0.0
-100.0 0.0 100.0

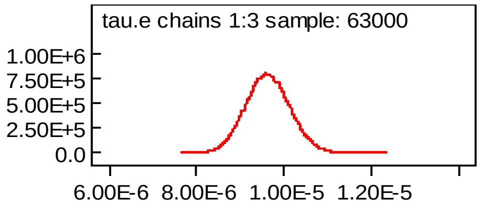
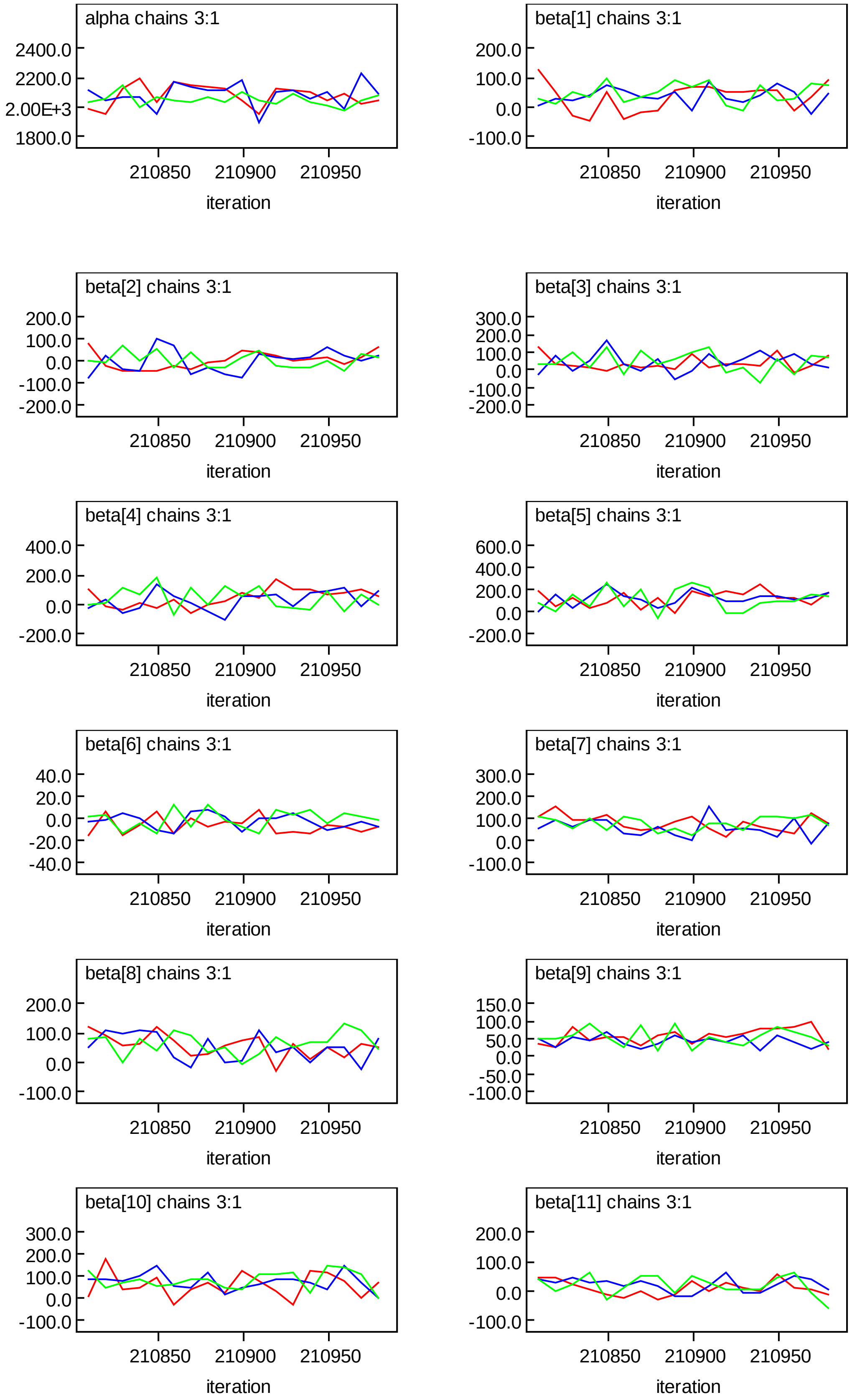
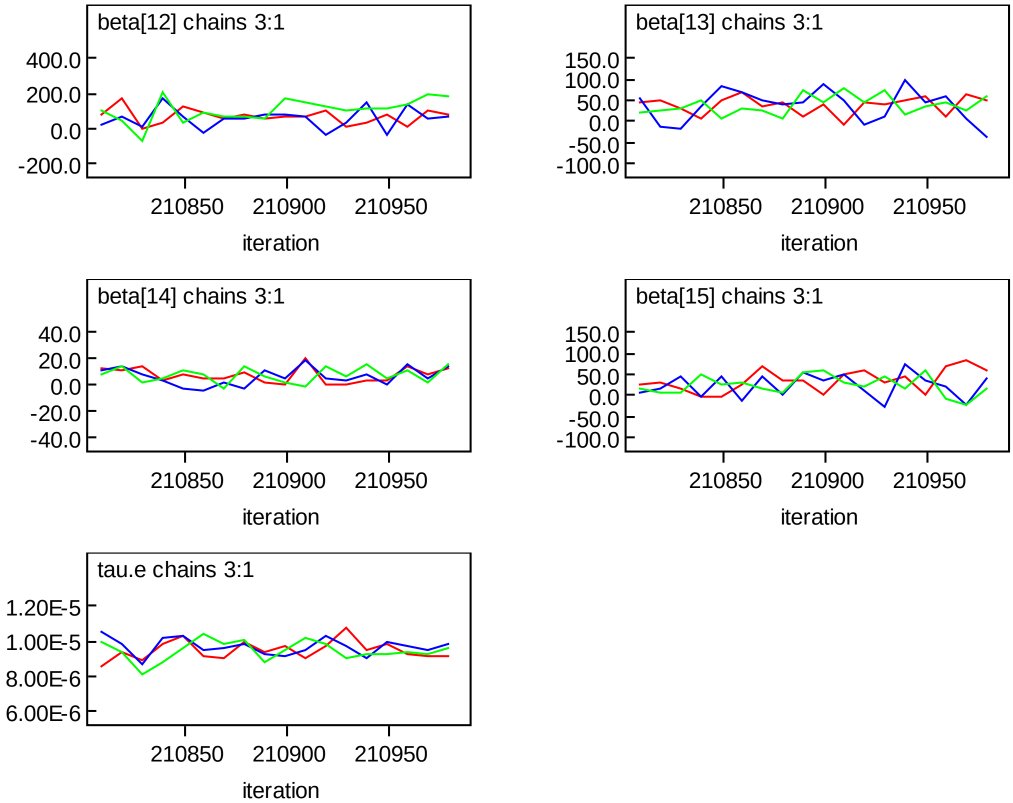

tau.e chains 1:3 sample: 63000
1.00E+6
7.50E+5
5.00E+5
2.50E+5
0.0
6.00E-6 8.00E-6 1.00E-5 1.20E-5

Figure 13: Year 2000 Non-Informative Bayesian Dynamic Trace





4.3.2 Bayesian Linear Regression with Informative Priors

There are three different type of priors considered for Bayesian Linear Regression Models with informative priors, namely mixed Bayesian priors, pure Bayesian priors and pure hierarchical Bayesian priors.

1. **Mixed Bayesian Priors**- These are priors that are calculated using the classical linear regression model output to compute the beta coefficient and variance components to be used as priors of the Gamma distribution function and WinBUGS software.

The classical linear regression output is used to calculate the beta coefficients and variance components for mother's ages between 30 and 34 years old. Using equations 3.32 and 3.33 we get;

- $s_{2000}^2 = \frac{SE(2000)}{\sqrt{n_{2000}}} = \frac{57.21}{\sqrt{161}} = 4.51$

- $s_{2004}^2 = \frac{SE(2004)}{\sqrt{n_{2004}}} = \frac{69.42}{\sqrt{138}} = 5.91$

- $s^2_{pooled} = \frac{(n_{2000}-1)s^2_{2000}+(n_{2004}-1)s^2_{2004}}{n_{2000}+n_{2004}-2}$

  $= \frac{(138-1)(5.91\ )+(161-1)(\ 4.51)}{138+161-2}$

  $=5.16$

- $\beta_{average} = \frac{\beta_{2000}+\beta_{2004}}{2}$

  $= \frac{154.45+63.132}{2}$

  $= 108.79$

Gamma distribution priors are calculated using equation 3.34.

- $\alpha = \frac{\bar{x}}{s^2}$ $\qquad\qquad$ $\beta = \frac{s^2}{\bar{x}}$

  $= \frac{\left(\frac{2157.9+2150.52}{2}\right)}{\left(\frac{2.64+1.96}{2}\right)}$ $\qquad\qquad$ $= \frac{\left(\frac{2.64+1.96}{2}\right)}{\left(\frac{2157.9+2150.52}{2}\right)}$

  $= 936.61$ $\qquad\qquad$ $= 0.0011$

Table 12: Summary for Mixed Bayesian Priors for years 2000 and 2004.

Variable Name	n2004	n2000	Se2004	Se2000	s2004^2	s2000^2	β2010	β2004	β2000	S^2	B
Intercept	909	1147	79.45	66.54	2.64	1.96	2133.61	2157.9	2150.52	2.26	2154.21
Mother's age											
20-24	332	409	54.88	39.4	3.01	1.95	50.245	112.19	30.25	2.42	71.22
25-29	234	271	60.28	46.56	3.94	2.83	81.57	122.57	16.75	3.34	69.66
30-34	138	161	69.42	57.21	5.91	4.51	120.6	154.45	63.13	5.15	108.79
35-39	88	103	81.88	70.81	8.73	6.98	111.08	135.04	51.74	7.78	93.39
40-49	53	81	93.49	87.3	12.84	9.70	108.59	296.98	134.35	10.94	215.665
Child's gender											
Female	477	617	28.51	23.56	1.31	0.95	-2.94	18.53	30.27	1.10	24.4
Region											
Central	254	407	42.7	38.91	2.68	1.93	-47.99	17.78	62.48	2.22	40.13
Southern	492	545	39.47	38.53	1.78	1.65	-13.74	37.09	59.33	1.71	48.21
Educational level											
Primary	558	756	37.86	29.17	1.60	1.06	-2.504	-31.23	-15.72	1.29	-23.475
Secondary/tertiary	145	127	51.57	45.8	4.28	4.06	46.85	25.21	16.54	4.18	20.875
Worked											
Past year	38	53	80.96	61.93	13.13	8.51	-51.476	-15.61	74.71	10.43	29.55
Working	493	679	29.56	25.78	1.33	0.99	-13.98	69.21	25.09	1.13	47.15
HIV status											
Negative	771	945	28.5	23.75	1.03	0.77	81.89	-13.68	27.138	0.89	6.729
Birth order no.	909	1147	10.125	9.07	0.34	0.27	-16.55	-31.04	-8.87	0.30	-19.955
Antenatal visits	909	1147	3.07	6.65	0.10	0.20	5.45	-4.93	3.56	0.15	-0.685

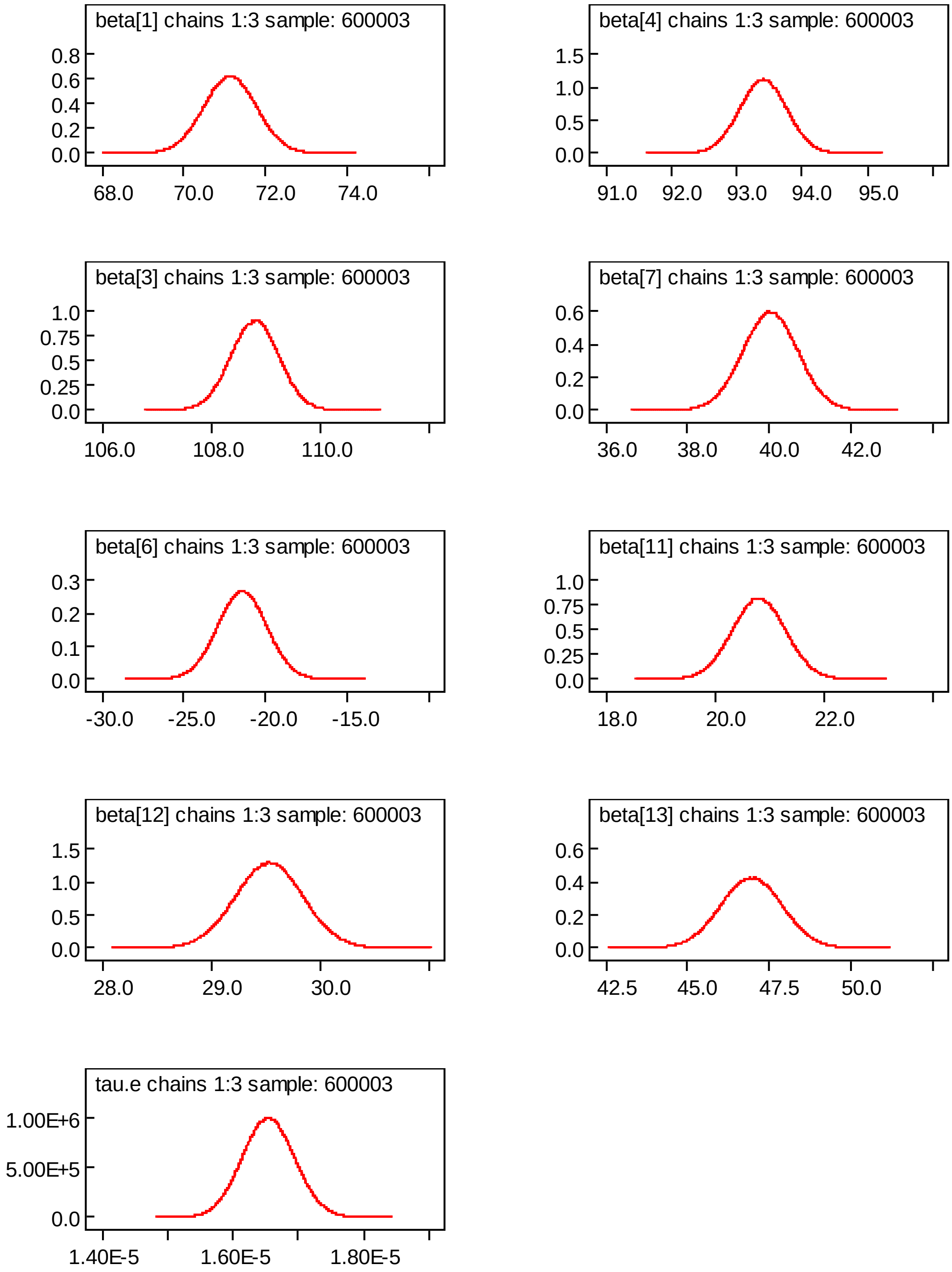
Table 17 shows the mixed Bayesian priors calculated using the classical linear regression output in table 9 and table 10. The data for the years 2000 and 2004 was used to calculate the variance components and β- coefficients displayed in table 17 and will be used as mixed Bayesian priors to construct a posterior distribution for the year 2010.

Table 13: Node statistics for Bayesian linear regression with mixed Bayesian priors.

node	mean	sd	MC error	2.5%	median	97.5%	Start	sample
alpha	2071.0	10.19	0.02569	2051.0	2071.0	2091.0	1000	600003
Mother's age (Ref=15-19)								
20-24	71.13	0.6424	8.468E-4	69.87	71.13	72.39	1000	600003
25-29	69.72	0.5465	7.19E-4	68.65	69.72	70.79	1000	600003
30-34	108.8	0.4407	5.706E-4	107.9	108.8	109.7	1000	600003
35-39	93.4	0.3583	4.606E-4	92.7	93.4	94.1	1000	600003
40-49	215.7	0.3025	3.86E-4	215.1	215.7	216.2	1000	600003
Child's birth order no.	-21.44	1.5	0.003029	-24.38	-21.44	-18.5	1000	600003
Region (Ref=Southern)								
Central region	40.02	0.6706	8.677E-4	38.71	40.02	41.34	1000	600003
Northern region	48.12	0.7638	0.00101	46.62	48.12	49.61	1000	600003
Child's gender (Ref=Male)								
Female	24.3	0.9497	0.001237	22.43	24.3	26.16	1000	600003
Education level (Ref=No education)								
Primary education	-23.17	0.8792	0.001168	-24.9	-23.18	-21.45	1000	600003
Secondary+	20.81	0.4894	6.359E-4	19.84	20.81	21.76	1000	600003
Employment status (Ref: Not working)								
Worked (past year)	29.53	0.3096	3.782E-4	28.93	29.53	30.14	1000	600003
Work (currently)	46.95	0.9367	0.001217	45.11	46.95	48.79	1000	600003
Antenatal visits		1.689	0.003621	0.1305	3.446	6.751	1000	600003
Mother's HIV Status (Ref=Positive)								
Negative	7.084	1.057	0.001424	5.014	7.084	9.16	1000	600003
tau.e	1.654E-5	3.977E-7	4.966E-10	1.577E-5	1.654E-5	1.733E-5	1000	600003

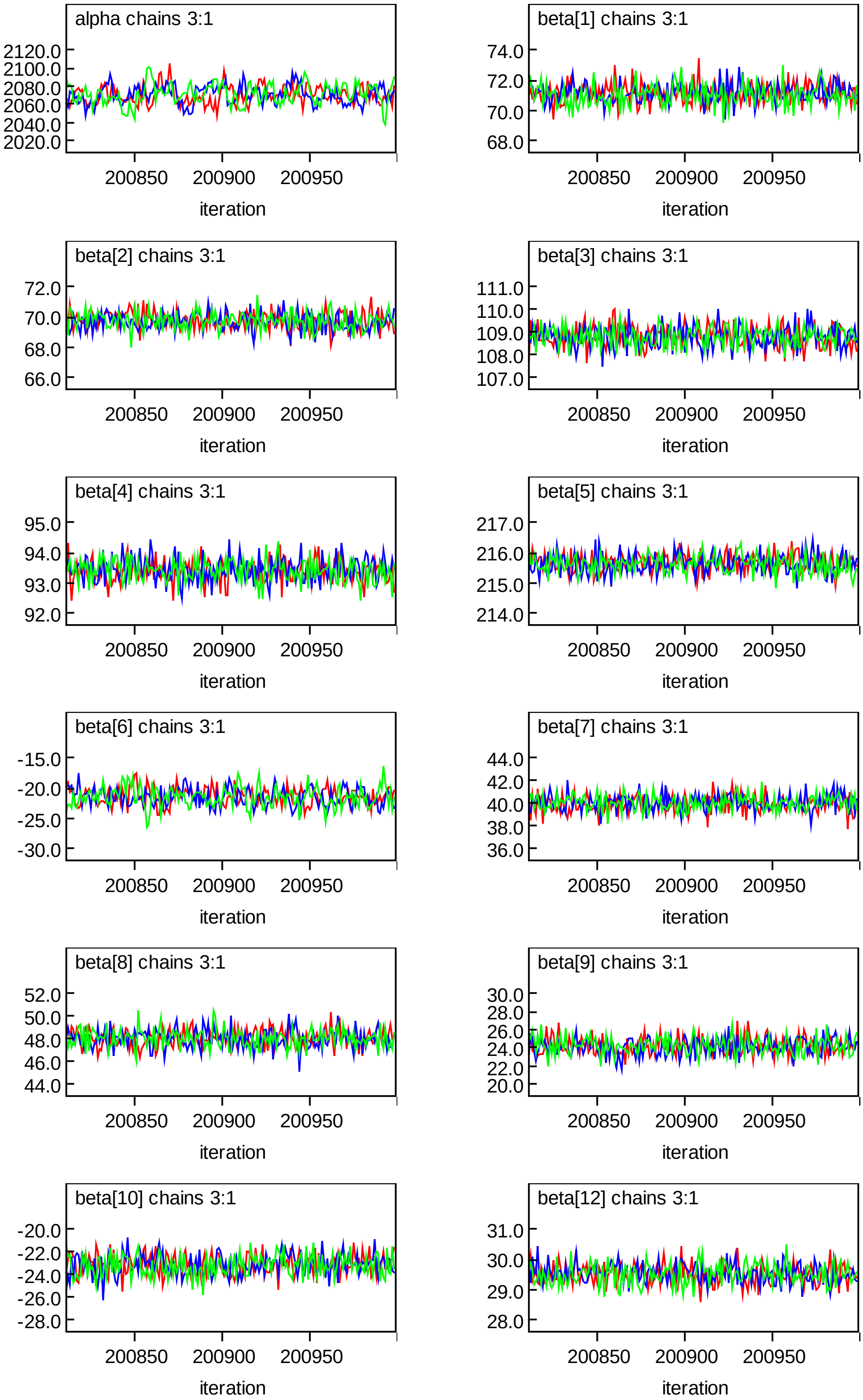
From the node statistics in Table 18, the Bayesian linear regression with informative mixed Bayesian priors shows that the variables, age of mother, child's birth order number, region, gender of a child, educational level, employment status, antenatal visits and mother's HIV status are the significant parameters in the model. The second mother's age group with ages between 25 and 39 years has children who have greater weight at birth compared to mothers between the ages 15 and 19 years based on the significant variables in the above output. The average difference in child's weight at birth between the HIV positive and the HIV negative mothers group is 7.084. This means that the weight at birth of a child with a mother who is not HIV infected is higher when compared to that that of a child with an HIV positive mother.

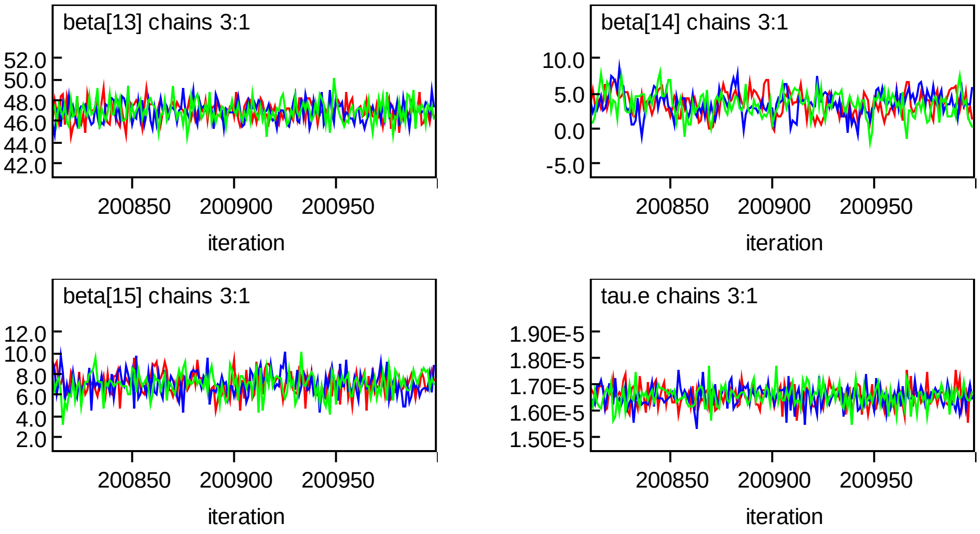
Figure 14 Density Function for Informative Bayesian Linear Regression with Mixed Bayesian Priors.



The assumption that the density function of a posterior distribution is normally distributed is met for all the variables of the informative Bayesian model.

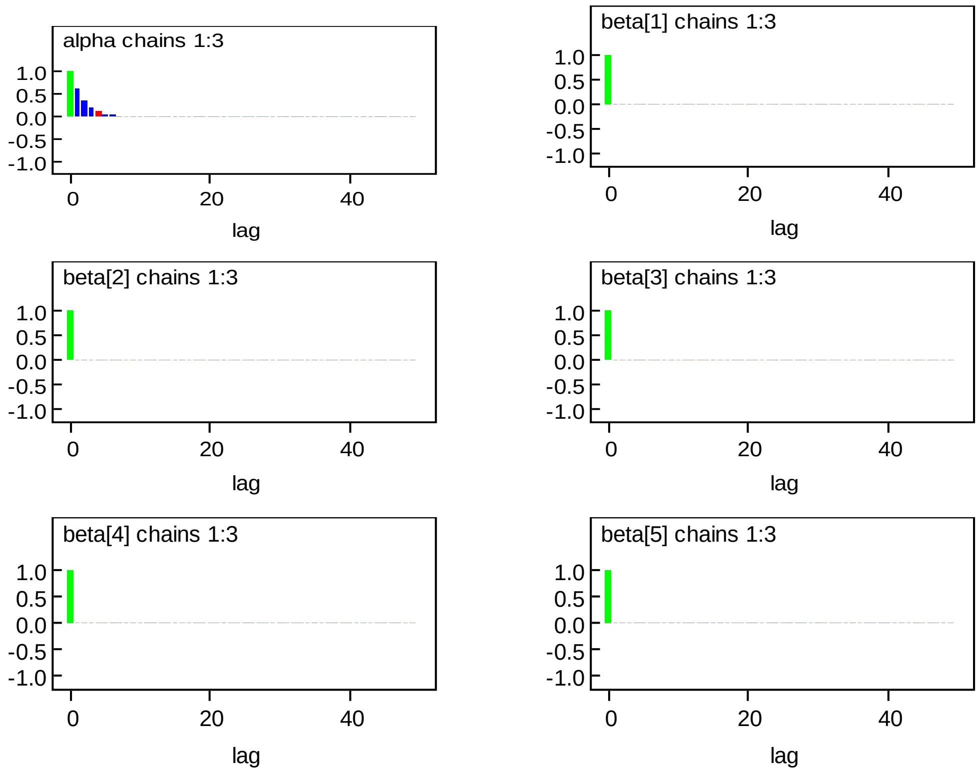
Figure 15: Dynamic Trace for Informative Bayesian Linear Regression with Mixed Bayesian Priors

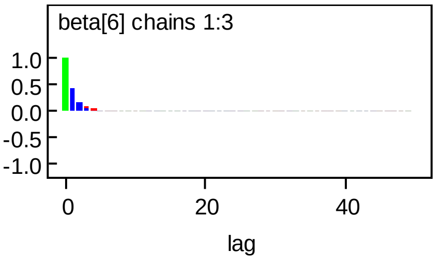


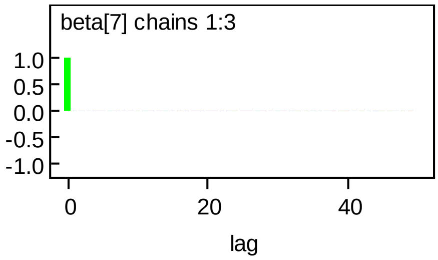


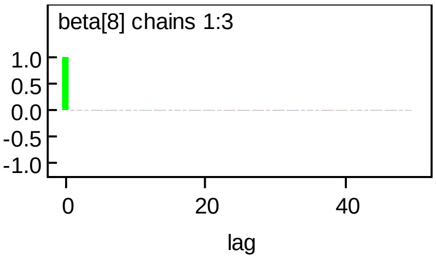
This model gives a smoothly-moving dynamic trace of the Markov Chains.

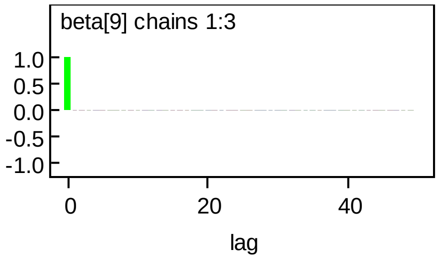
Figure 16: Autocorrelation Function for Informative Bayesian Linear Regression with Mixed Bayesian Priors

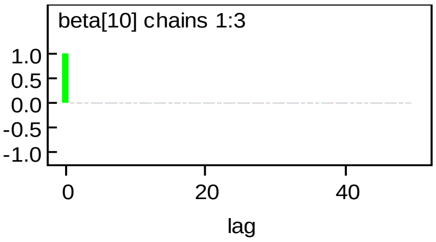


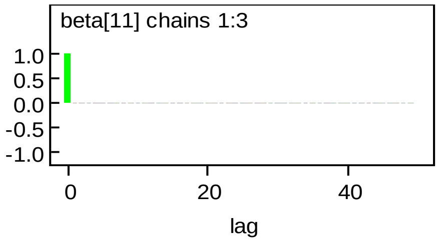

beta[6] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

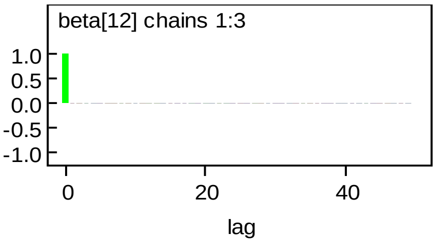

beta[7] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

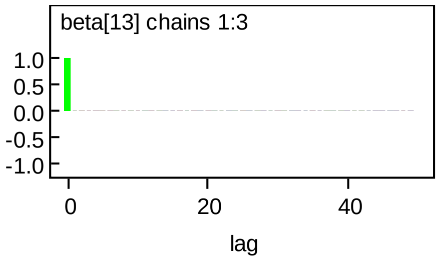

beta[8] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

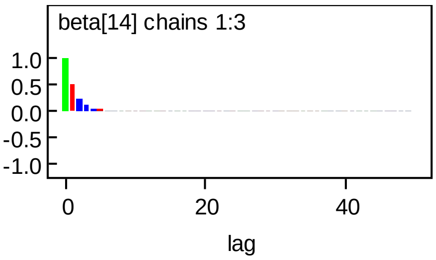

beta[9] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

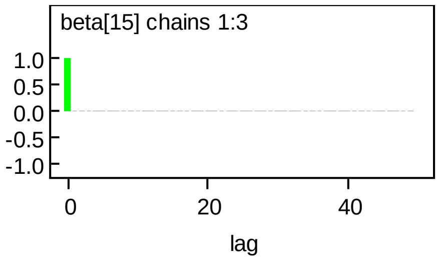

beta[10] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[11] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[12] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[13] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[14] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag


beta[15] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

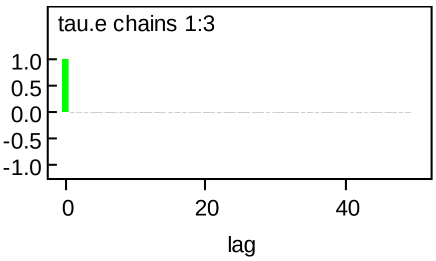
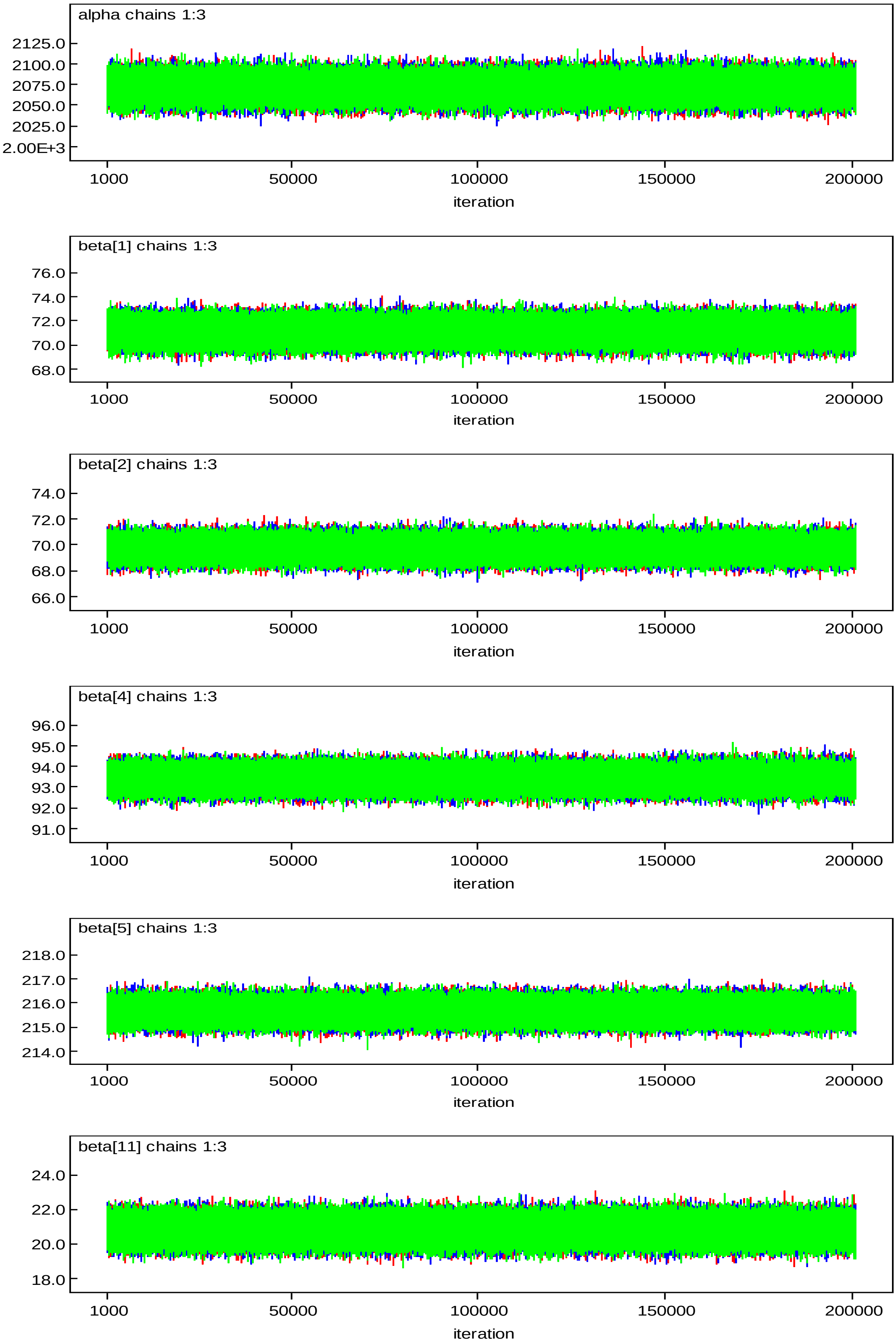
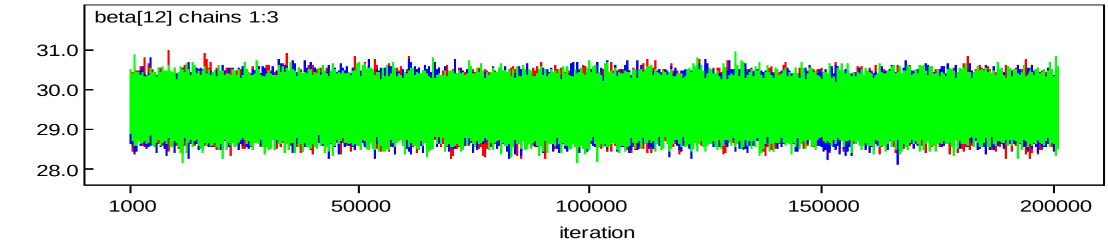
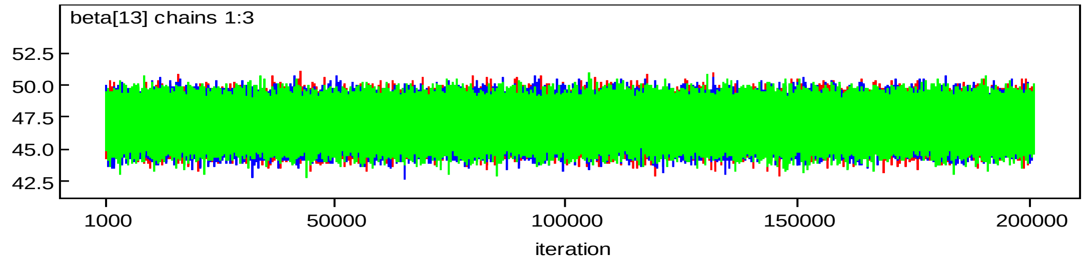
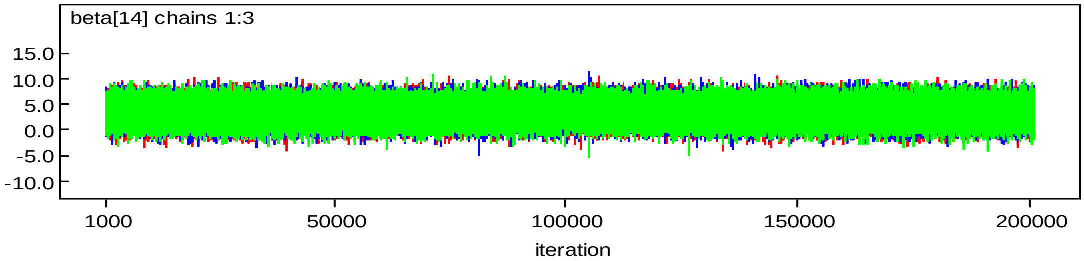

tau.e chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

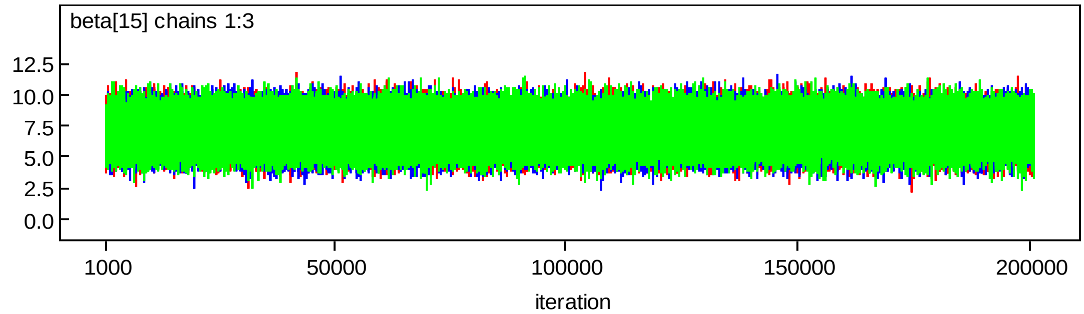
Figure 17: Time Series for Informative Bayesian Linear Regression with Mixed Bayesian Priors

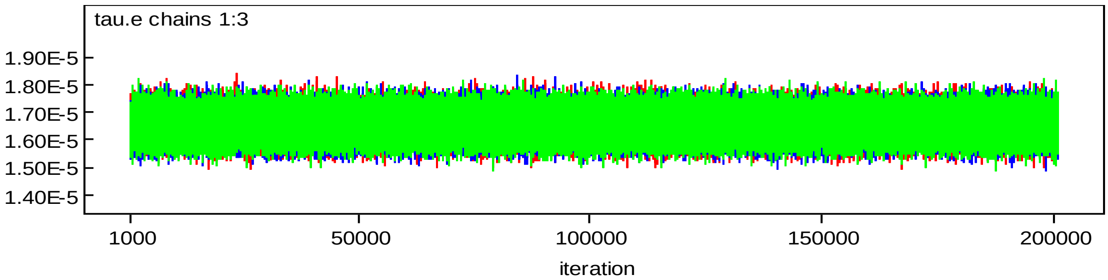



beta[12] chains 1:3
31.0
30.0
29.0
28.0
1000
50000
100000
150000
200000
iteration


beta[13] chains 1:3
52.5
50.0
47.5
45.0
42.5
1000
50000
100000
150000
200000
iteration


beta[14] chains 1:3
15.0
10.0
5.0
0.0
-5.0
-10.0
1000
50000
100000
150000
200000
iteration


beta[15] chains 1:3
12.5
10.0
7.5
5.0
2.5
0.0
1000
50000
100000
150000
200000
iteration


tau.e chains 1:3
1.90E-5
1.80E-5
1.70E-5
1.60E-5
1.50E-5
1.40E-5
1000
50000
100000
150000
200000
iteration

2. **Pure Bayesian Prior**: These are priors calculated using the non-informative Bayesian linear regression output to compute priors of the beta coefficient and variance components. The calculated Bayesian priors are then used as priors of the Gamma distribution function in WinBUGS program.

Table 14: Summary for Pure Bayesian Priors

Variable Name	n2004	n2000	Se2004	Se2000	s2004^2	s2000^2	β2010	β2004	β2000	S^2	B
Intercept	909	1147	77.29	63.05	2.56	1.86	2131	2034	2071	2.17	2052.5
Mother's age											
20-24	332	409	51.38	37.68	2.82	1.86	9.33	128.7	39.69	2.29	84.195
25-29	234	271	56.64	44.11	3.70	2.68	64.01	129.1	4.2	3.15	66.65
30-34	138	161	65.99	54.24	5.62	4.27	100.2	138.7	40.06	4.89	89.38
35-39	88	103	77.32	67.8	8.24	6.68	107.1	88.01	30.02	7.40	59.015
40-49	53	81	90.14	82.34	12.38	9.15	107	219.3	118.6	10.42	168.95
Child's gender											
Female	477	617	29.29	23.62	1.34	0.95	9.54	35.43	51.65	1.12	43.54
Region											
Central	254	407	44.9	35.88	2.82	1.78	-47.55	-1.27	81.1	2.18	39.915
Southern	492	545	40.07	34.41	1.81	1.47	-33.04	53.64	64.18	1.63	58.91
Educational level											
Primary	558	756	52.15	29.97	2.21	1.09	8.79	78.51	13.56	1.56	46.035
Secondary/tertiary	145	127	38.16	44.45	3.17	3.94	60.09	-8.22	65.82	3.53	28.8
Worked											
Past year	38	53	29.98	60.85	4.86	8.36	-48.78	20.02	80.38	6.91	50.2
Working	493	679	62.7	25.39	2.82	0.97	-12.04	62.7	38.38	1.75	50.54
HIV status											
Negative	771	945	29.55	23.89	1.06	0.78	80.53	12.32	29.36	0.91	20.84
Birth order no.	909	1147	2.56	8.92	0.08	0.26	-18.32	-11.91	-2.83	0.18	-7.37
Antenatal visits	909	1147	29.55	6.79	0.98	0.20	6.93	-5.04	6.8	0.55	0.88

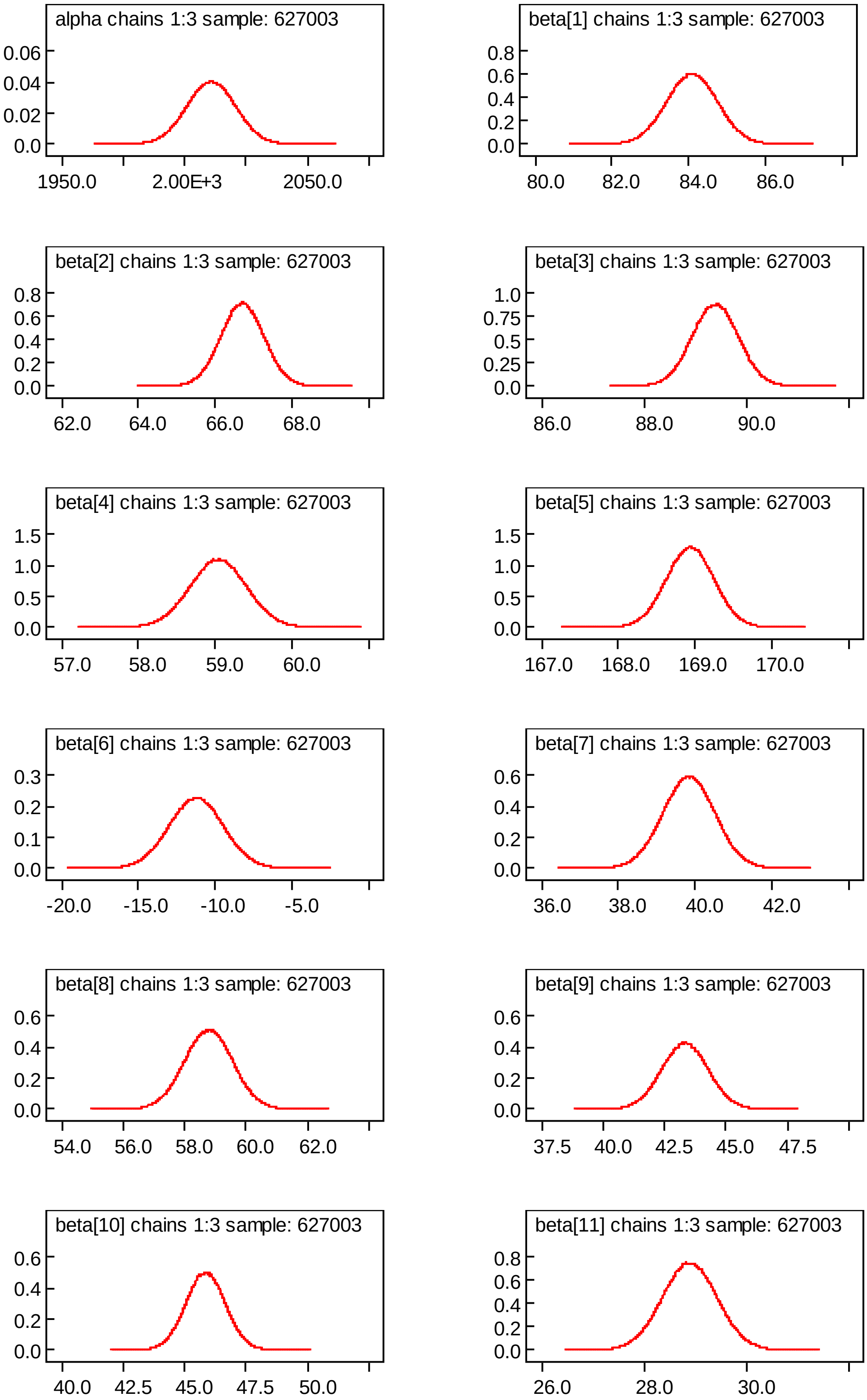
Table 19 shows the Pure Bayesian priors that were calculated using the non-informative Bayesian linear regression output in table 15 and table 16. The data for the years 2000 and 2004 was used to calculate the variance components and β- coefficients displayed in table 19 and will be used as the pure Bayesian priors to construct a posterior distribution for the year 2010.

Table 15: Node Statistics for Informative Bayesian Multiple Linear Regression with Pure Bayesian Priors

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	2011.0	9.911	0.02399	1991.0	2011.0	2030.0	1000	627003
Mother's age (Ref=15-19)								
20-24	84.08	0.6604	8.382E-4	82.79	84.08	85.37	1000	627003
25-29	66.7	0.5627	7.206E-4	65.6	66.7	67.8	1000	627003
30-34	89.39	0.4523	5.568E-4	88.5	89.39	90.27	1000	627003
35-39	59.04	0.3673	4.682E-4	58.32	59.04	59.76	1000	627003
40-49	168.9	0.3099	3.839E-4	168.9	168.9	169.5	1000	627003
Child's birth order no.	-11.25	1.76	0.003843	-14.7	-11.25	-7.79	1000	627003
Region (Ref=Southern)								
Central region	39.83	0.6768	8.578E-4	38.5	39.83	41.16	1000	627003
Northern region	58.78	0.7822	0.001015	57.24	58.78	60.31	1000	627003
Child's gender (Ref=Male)								
Female	43.33	0.9411	0.001212	41.48	43.33	45.17	1000	627003
Education status (Ref=No education)								
Primary education	45.85	0.7992	0.001049	44.28	45.85	47.41	1000	627003
Secondary+	28.89	0.5325	6.731E-4	27.84	28.89	29.93	1000	627003
Employment status (Ref=No work)								
Worked (past year)	50.16	0.3803	4.652E-4	49.42	50.16	50.91	1000	627003
Work (currently)	50.43	0.7535	9.522E-4	48.96	50.43	51.91	1000	627003
Antenatal visits		1.164	0.001992	0.157	2.44	4.722	1000	627003
Mother's HIV Status (Ref=Positive)								
Negative	-5.456	2.331	0.003302	-10.02	-5.458	-0.881	1000	627003
tau.e	1.63E-5	3.946E-7	4.7E-10	1.561E-5	1.63E-5	1.76E-5	1000	627003

The variables age of a mother, child's birth order number, region, gender of a child, education status, mother's employment status, antenatal visits during pregnancy and HIV status of a mother are significant for the Bayesian linear regression with informative pure priors.

Figure 18: Density Function for Informative Bayesian Linear Regression with Pure Bayesian Priors.



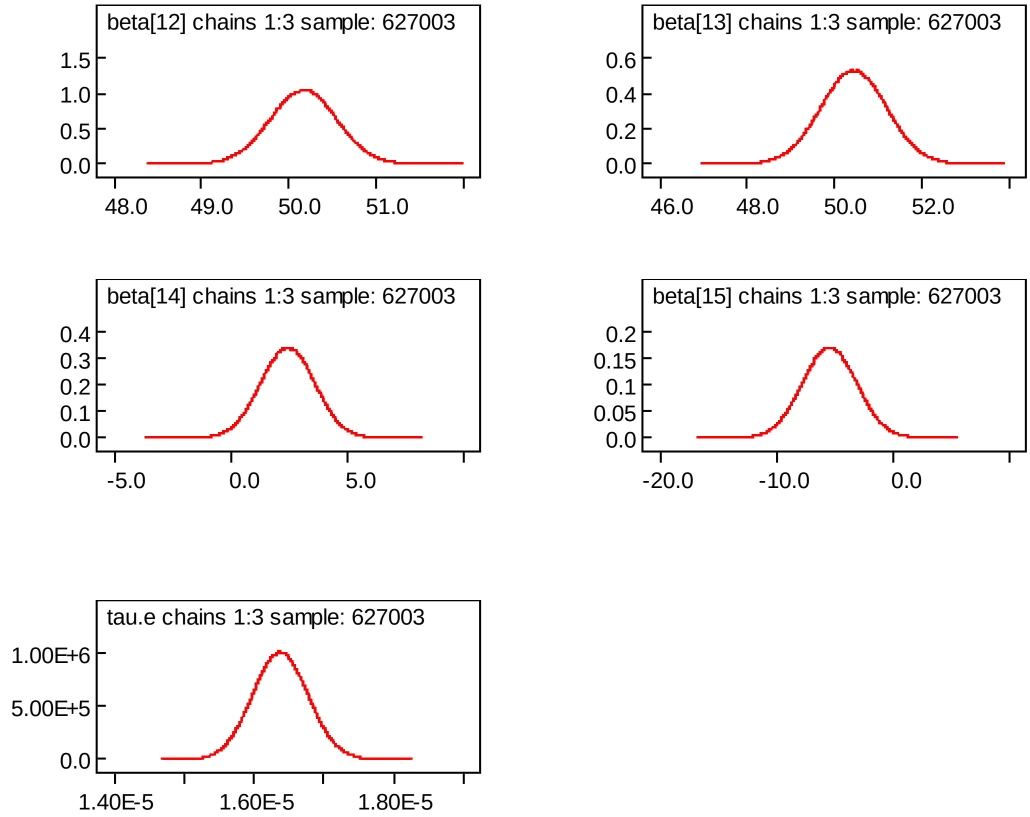
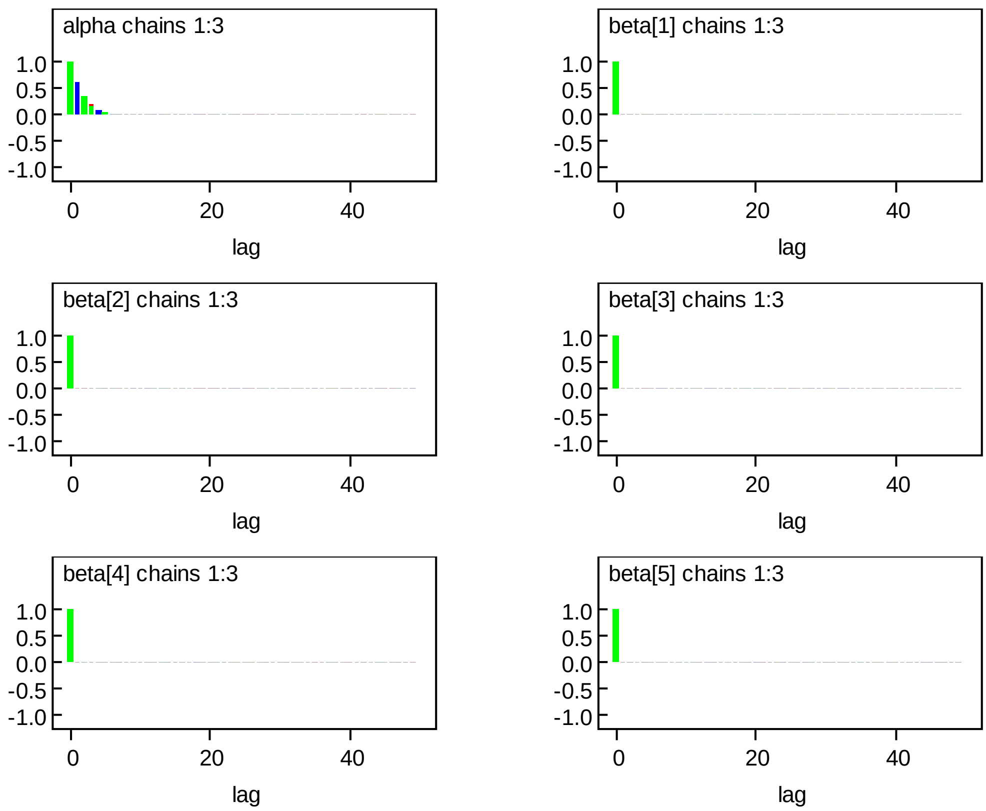


Figure 19: Autocorrelation Function for Informative Bayesian Linear Regression with Pure Bayesian Priors.



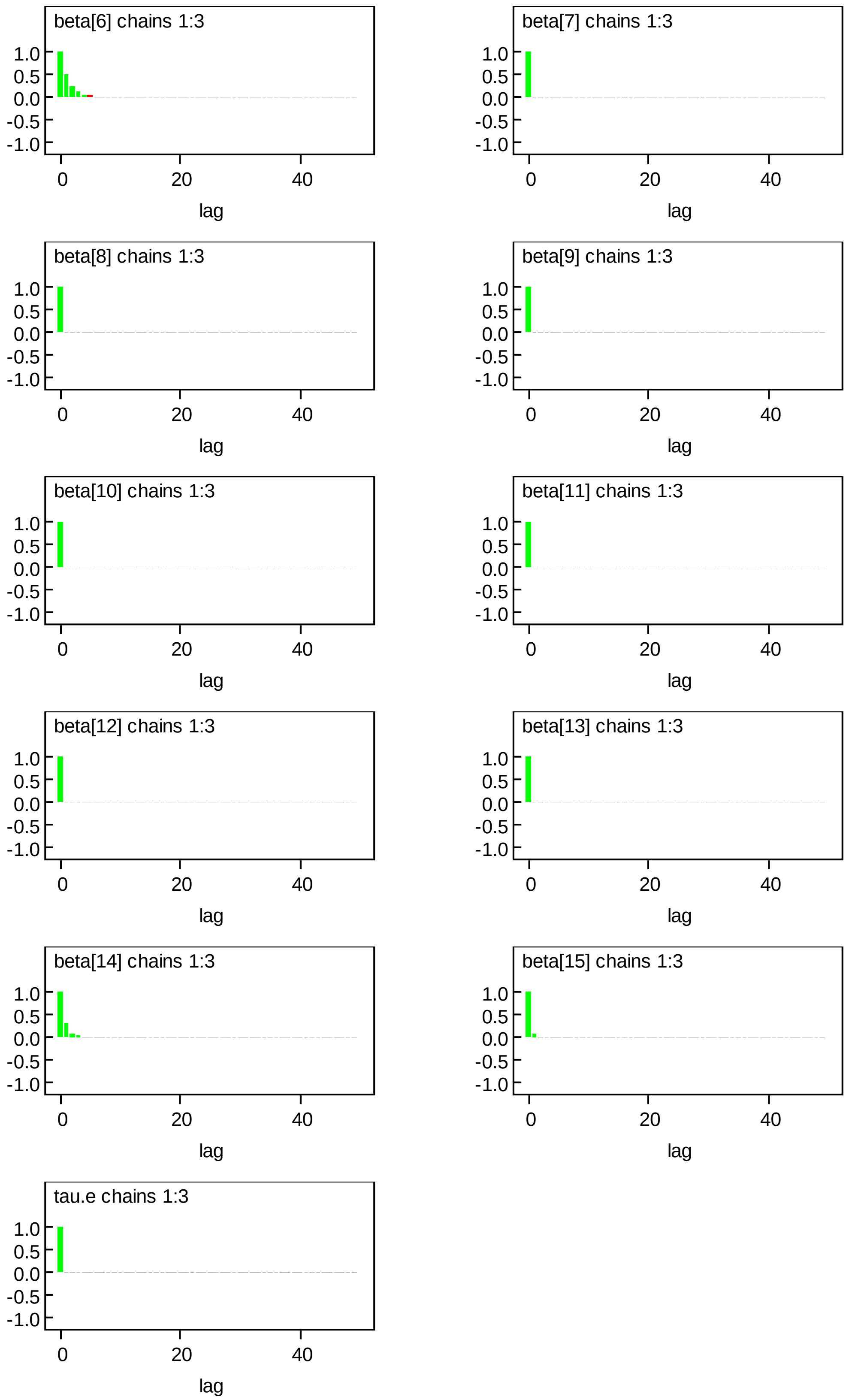
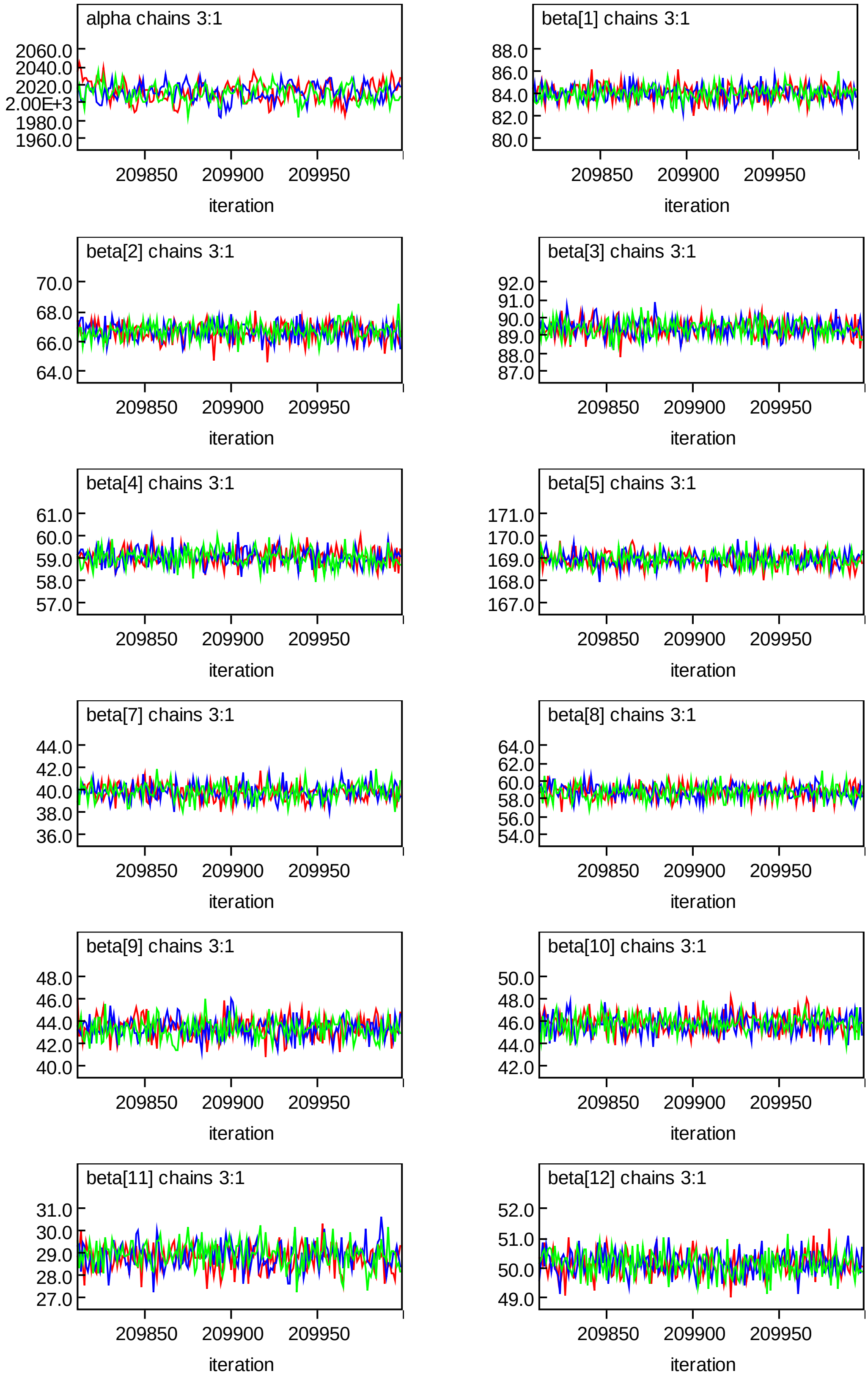

beta[6] chains 1:3
beta[7] chains 1:3
beta[8] chains 1:3
beta[9] chains 1:3
beta[10] chains 1:3
beta[11] chains 1:3
beta[12] chains 1:3
beta[13] chains 1:3
beta[14] chains 1:3
beta[15] chains 1:3
tau.e chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

Figure 20: Dynamic Trace for Informative Bayesian Linear Regression with Pure Bayesian Priors



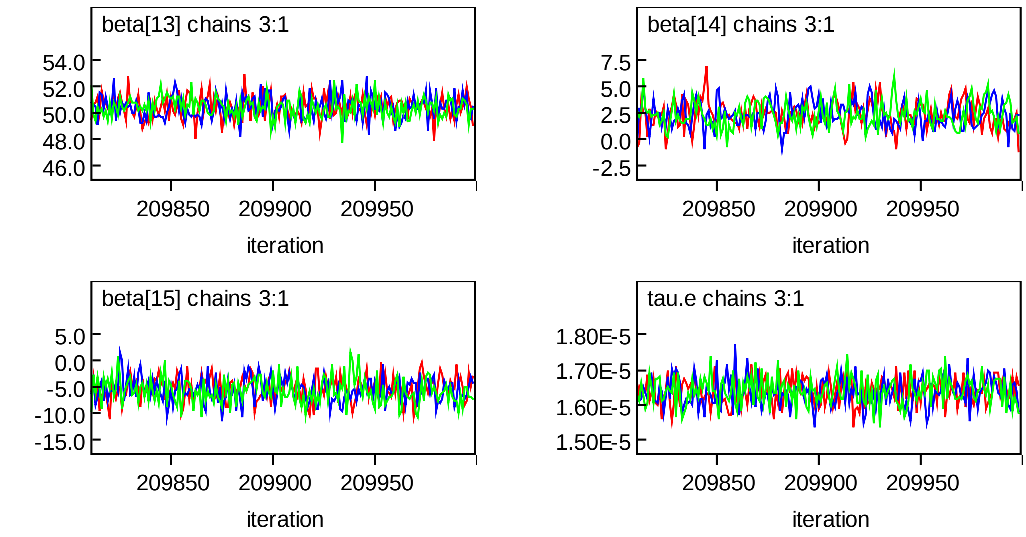
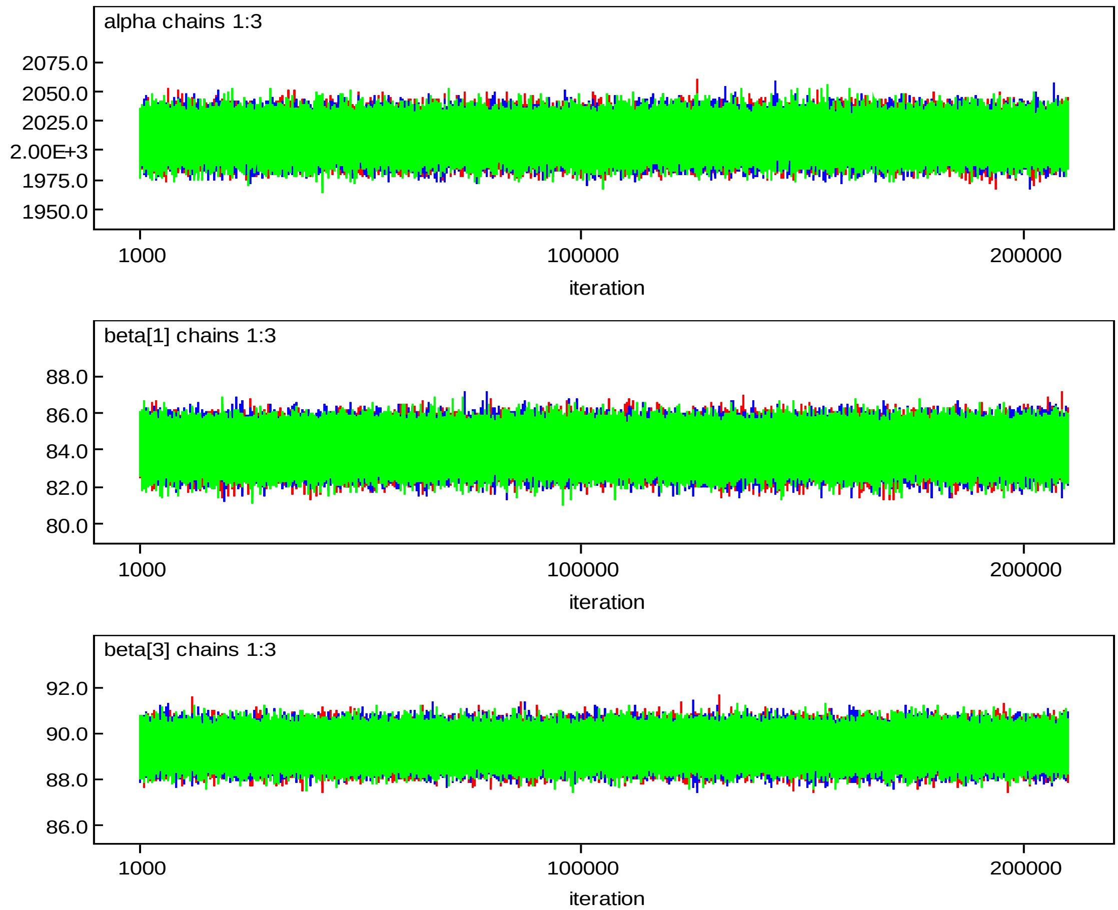
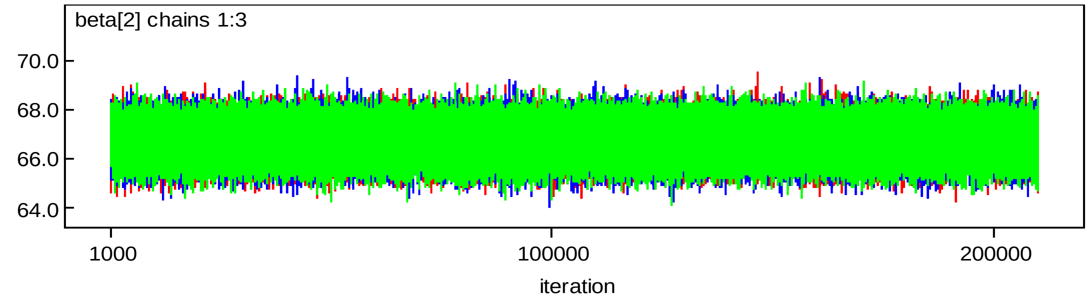
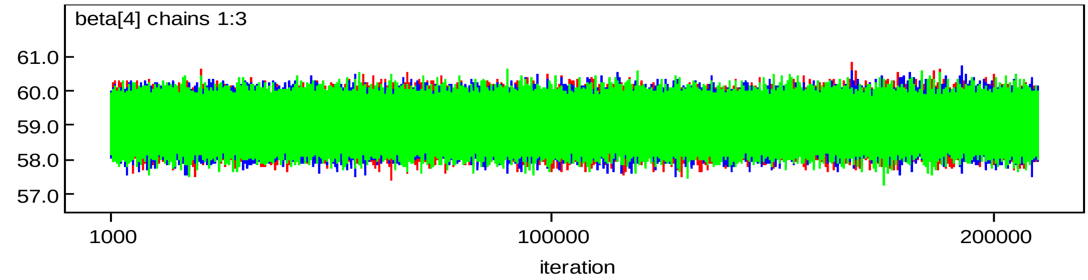
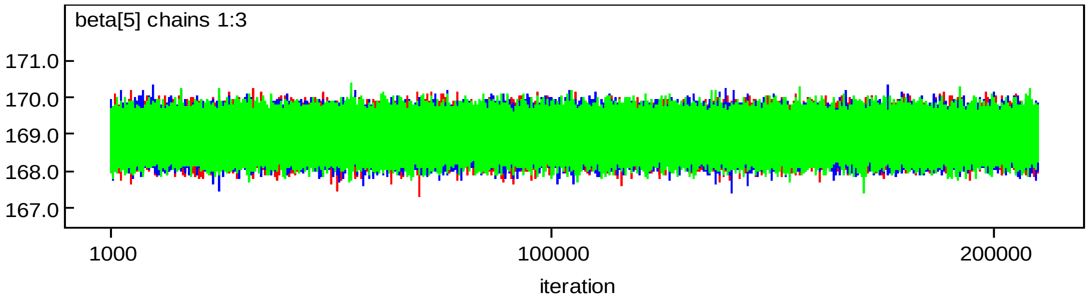


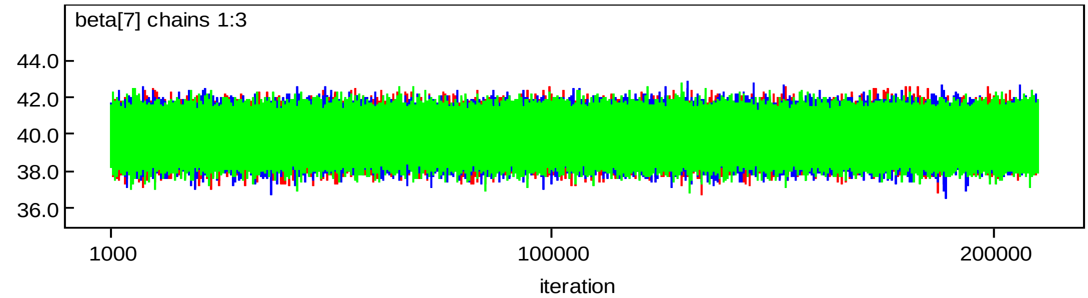
Figure 21: Time Series Plot for Informative Bayesian Linear Regression with Pure Bayesian Priors

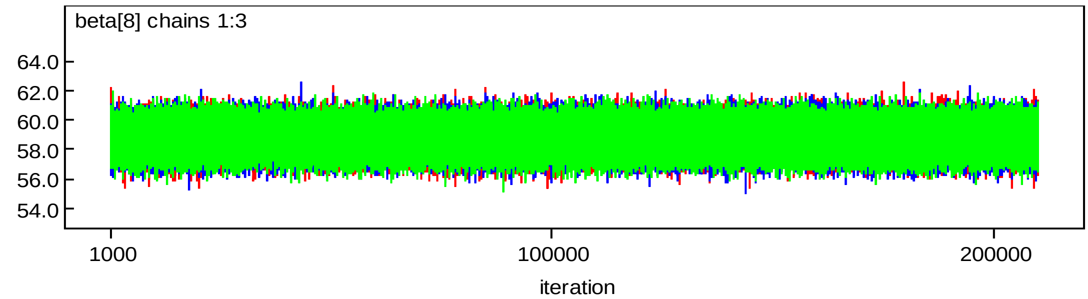


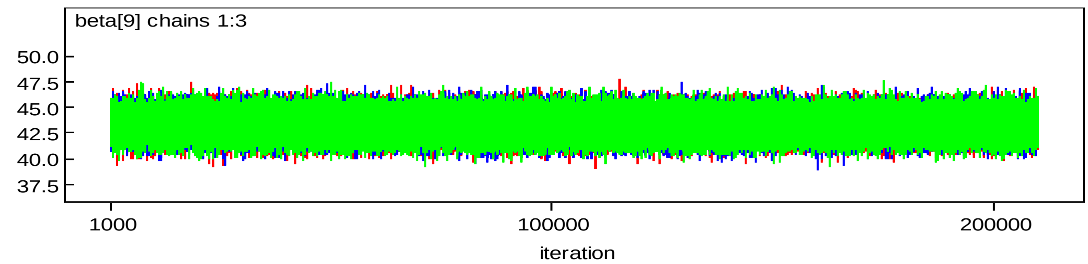

beta[2] chains 1:3
70.0
68.0
66.0
64.0
1000
100000
200000
iteration

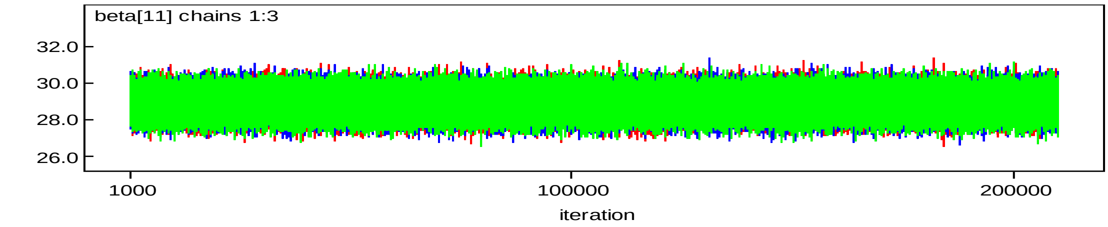

beta[4] chains 1:3
61.0
60.0
59.0
58.0
57.0
1000
100000
200000
iteration

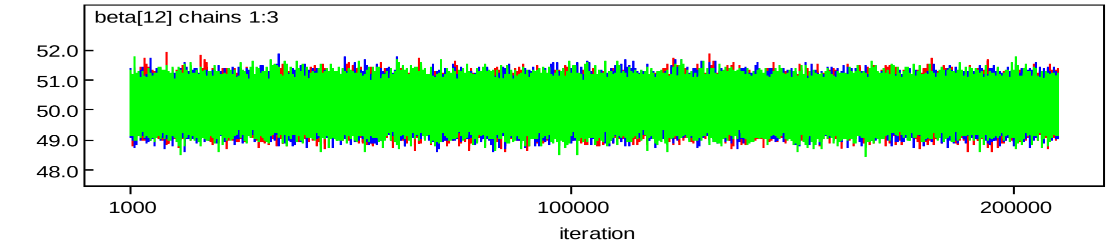

beta[5] chains 1:3
171.0
170.0
169.0
168.0
167.0
1000
100000
200000
iteration

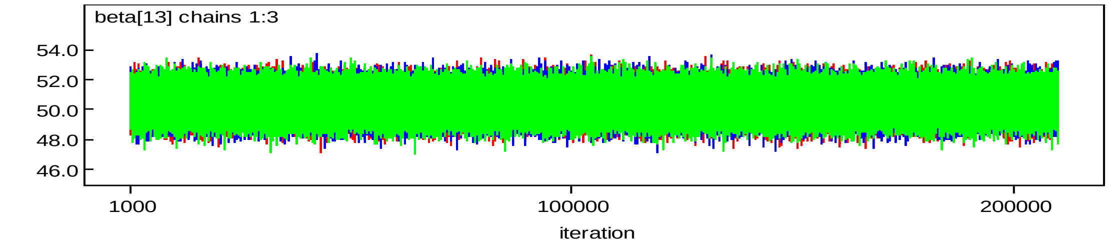

beta[7] chains 1:3
44.0
42.0
40.0
38.0
36.0
1000
100000
200000
iteration

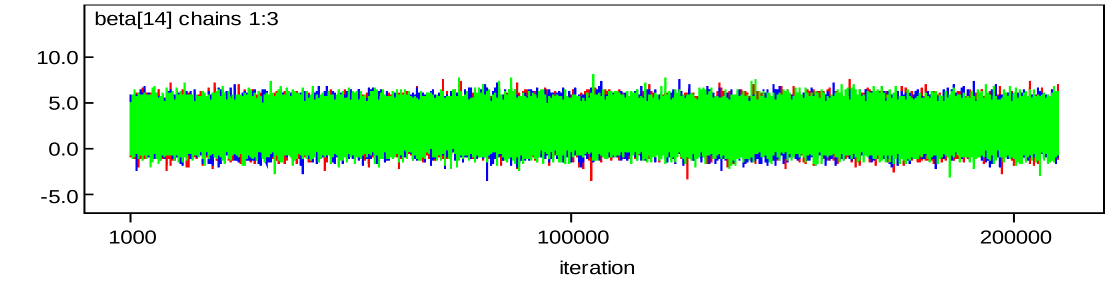

beta[8] chains 1:3
64.0
62.0
60.0
58.0
56.0
54.0
1000
100000
200000
iteration

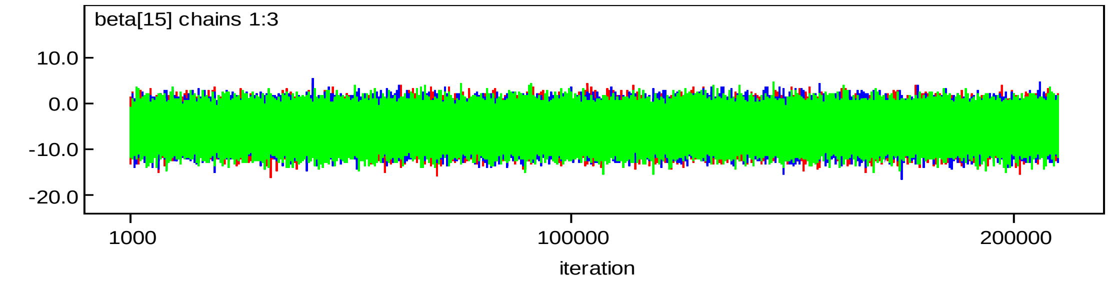

beta[9] chains 1:3
50.0
47.5
45.0
42.5
40.0
37.5
1000
100000
200000
iteration

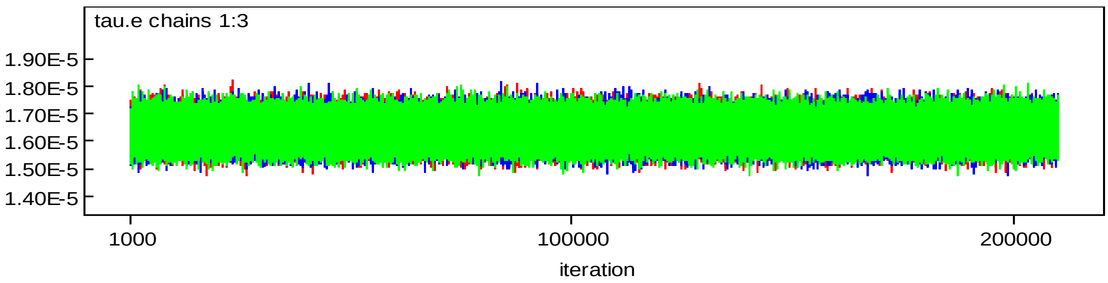

beta[11] chains 1:3
32.0
30.0
28.0
26.0
1000
100000
200000
iteration


beta[12] chains 1:3
52.0
51.0
50.0
49.0
48.0
1000
100000
200000
iteration


beta[13] chains 1:3
54.0
52.0
50.0
48.0
46.0
1000
100000
200000
iteration


beta[14] chains 1:3
10.0
5.0
0.0
-5.0
1000
100000
200000
iteration


beta[15] chains 1:3
10.0
0.0
-10.0
-20.0
1000
100000
200000
iteration


tau.e chains 1:3
1.90E-5
1.80E-5
1.70E-5
1.60E-5
1.50E-5
1.40E-5
1000
100000
200000
iteration

3. **Pure Hierarchical Bayesian Priors**: These are priors that are calculated using the informative Bayesian hierarchical linear regression output to compute the beta coefficient and variance components. The calculated Bayesian priors are then used as priors of the Gamma distribution function in WinBUGS program.

Table 16: Summary for Pure Hierarchical Bayesian Priors

Variable Name	n2004	n2000	Se2004	Se2000	s2004^2	s2000^2	β2010	β2004	β2000	S^2	B
Intercept	590	757	338.9	429.2	13.95	15.60	145.9	160.8	196.3	14.88	178.55
Mother's age											
20-24	207	258	51.6	37.73	3.59	2.35	-1.57	90.91	26.9	2.90	58.905
25-29	132	156	56.78	44.19	4.94	3.54	56.39	94.16	-7.13	4.18	43.515
30-34	94	107	65.9	54.16	6.80	5.24	90.64	108.7	32.74	5.97	70.72
35-39	57	70	77.37	67.34	10.25	8.05	99.88	63.26	24.66	9.03	43.96
40-49	43	62	90.1	82.14	13.74	10.43	96.77	194.2	116.3	11.78	155.25
Child's gender											
Female	320	404	29.28	23.62	1.64	1.18	4.11	24.01	43.54	1.38	33.775
Educational level											
Primary	356	488	38.3	30.02	2.03	1.36	-7.54	-37.91	-3.37	1.64	-20.64
Secondary/tertiary	100	90	52.14	44.85	5.21	4.73	45.08	40.36	48.95	4.98	44.655
Worked											
Past year	21	32	13.51	60.68	2.95	10.73	5.51	13.51	70.65	7.68	42.08
Working	323	462	53.85	25.44	3.00	1.18	1.1	53.85	31.7	1.93	42.775
HIV status											
Negative	331	441	29.36	23.88	1.61	1.14	71.86	4.35	25.88	1.34	15.115
Birth order no.	590	757	10.41	8.95	0.43	0.33	-19.61	-17.2	-4.93	0.37	-11.065
Antenatal visits	590	757	2.55	6.78	0.10	0.25	6.07	-5.43	3.66	0.18	-0.885

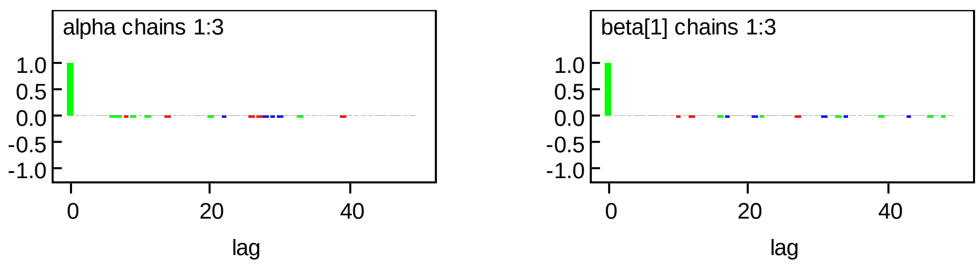
Table 21 shows the summary of the Pure Hierarchical Bayesian priors calculated using the Informative Bayesian linear regression output in table 15. The data for the years 2000 and 2004 was used to calculate the variance components and β- coefficients displayed in table 19 and will be used as pure Bayesian priors to construct a posterior distribution for the year 2010.

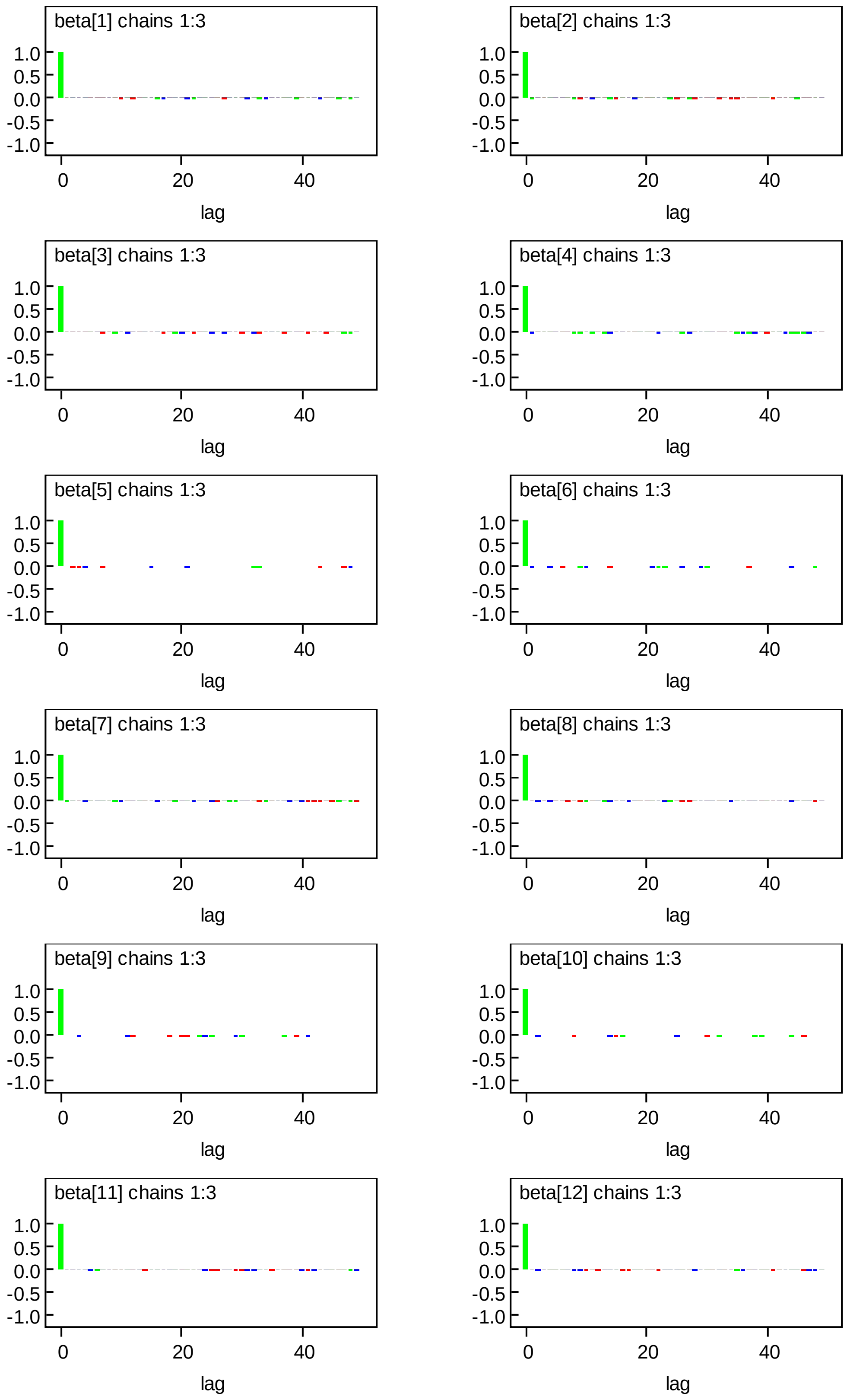
Table 17: Node Statistics for Bayesian Hierarchical Regression with Informative Priors (pure Bayesian priors)

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
alpha	178.6	0.2584	9.539E-4	178.0	178.6	179.1	1000	60000
Mother's age (Ref=15-19)								
20-24	58.86	0.5877	0.002461	57.71	58.86	60.01	1000	60000
25-29	43.54	0.489	0.001861	42.59	43.54	44.5	1000	60000
30-34	70.73	0.4087	0.001605	69.93	70.72	71.52	1000	60000
35-39	43.96	0.3346	0.001374	43.31	43.97	44.62	1000	60000
40-49	155.3	0.2909	0.001219	154.7	155.3	155.8	1000	60000
Child's birth order no.	-12.13	1.516	0.006579	-15.1	-12.13	-9.156	1000	60000
Child's gender (Ref=Male)								
Female	33.72	0.8594	0.003246	32.04	33.73	35.41	1000	60000
Education status (Ref=No education)								
Primary education	-20.51	0.7882	0.003197	-22.04	-20.5	-18.95	1000	60000
Secondary+	44.61	0.4478	0.001818	43.73	44.61	45.9	1000	60000
Employment status (Ref=No work)								
Worked (past year)	42.06	0.3618	0.001469	41.35	42.06	42.77	1000	60000
Work (currently)	42.78	0.7182	0.002987	41.36	42.78	44.19	1000	60000
Antenatal visits		1.933	0.007768	-2.28	1.522	5.313	1000	60000
Mother's HIV Status (Ref=Positive)								
Negative	15.19	0.8654	0.003416	13.49	12.49	16.89	1000	60000
beta.reg[1]	1953.0	22.0	0.08971	1910.0	1953.0	1995.0	1000	60000
beta.reg[2]	1889.0	17.41	0.07306	1855.0	1889.0	1923.0	1000	60000
beta.reg[3]	1908.0	16.08	0.06454	1876.0	1908.0	1939.0	1000	60000
tau.e	7.751E-6	2.752E-7	1.132E-9	7.222E-6	7.748E-6	8.3E-6	1000	60000
tau.reg	2.741E-7	2.234E-7	8.8E-10	1.978E-8	2.168E-7	8.511E-7	1000	60000

All variables are significant under Bayesian hierarchical regression with informative priors which suggest that this is a good model.

Figure 22: Autocorrelation Function for Pure Hierarchical Bayesian Priors




beta[1] chains 1:3
beta[2] chains 1:3
beta[3] chains 1:3
beta[4] chains 1:3
beta[5] chains 1:3
beta[6] chains 1:3
beta[7] chains 1:3
beta[8] chains 1:3
beta[9] chains 1:3
beta[10] chains 1:3
beta[11] chains 1:3
beta[12] chains 1:3
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

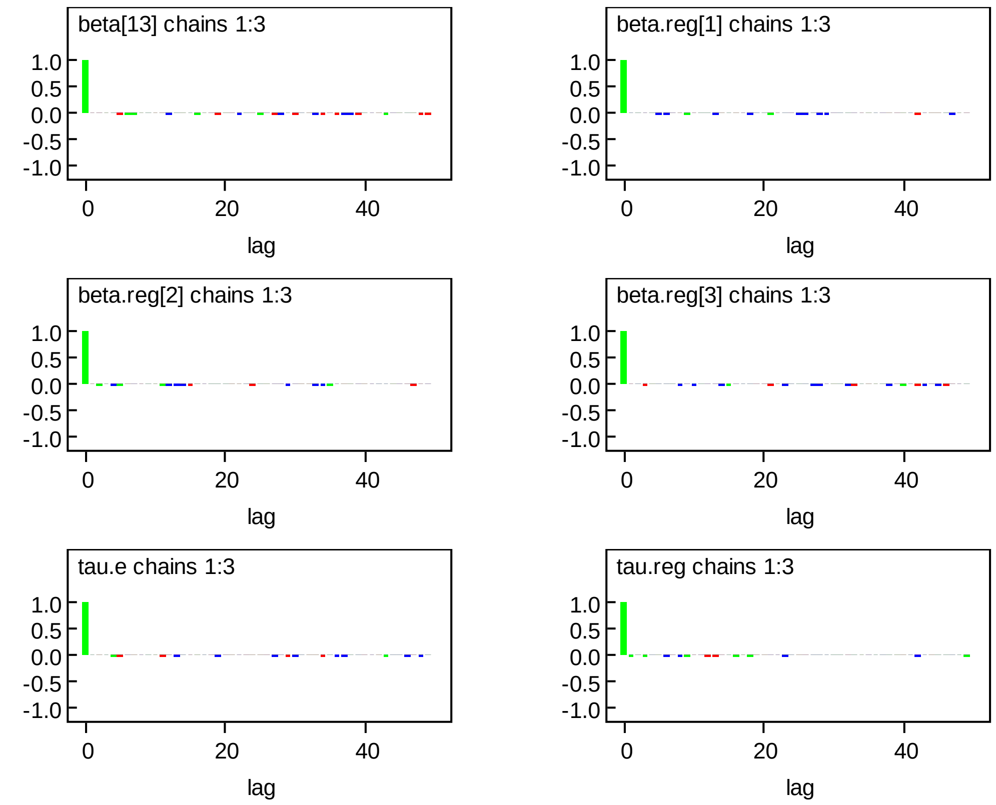
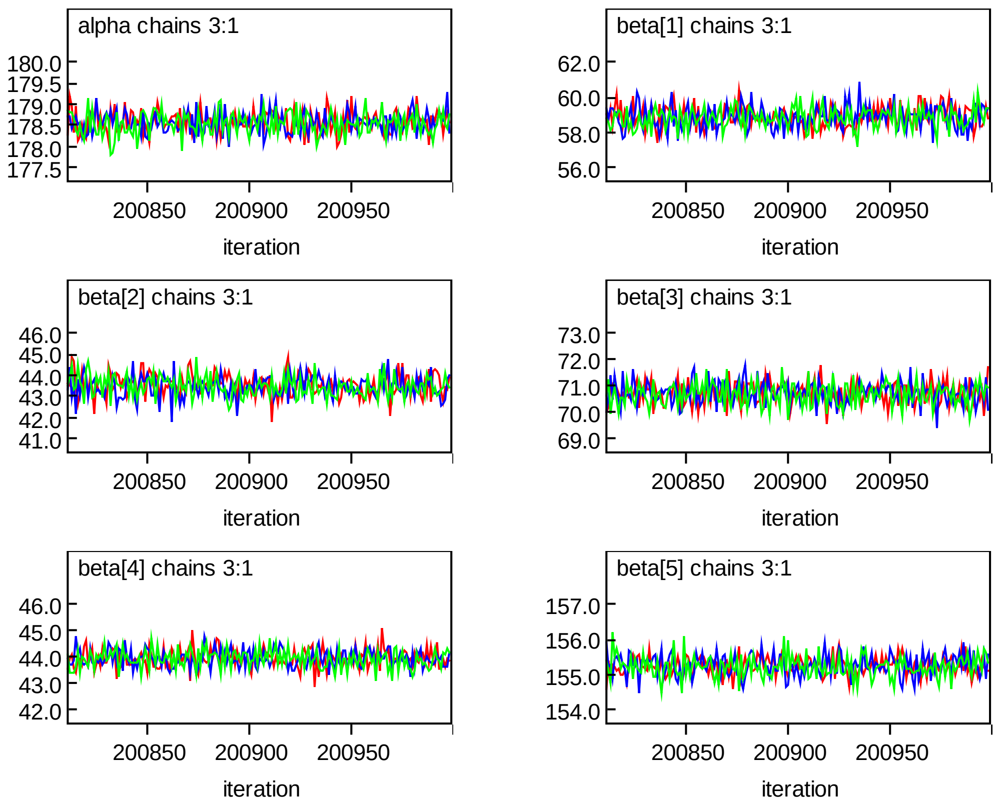
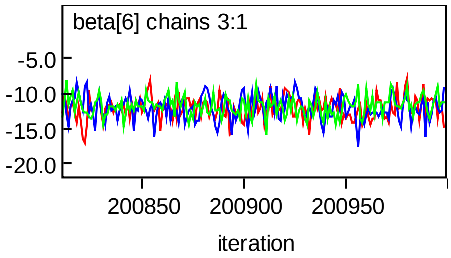
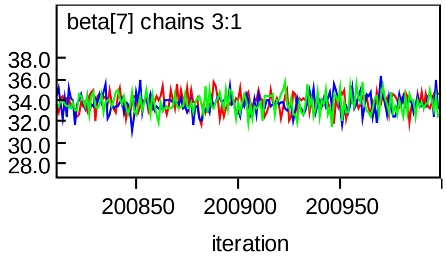
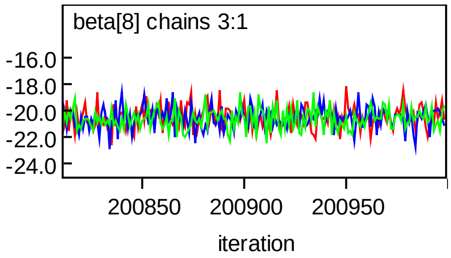


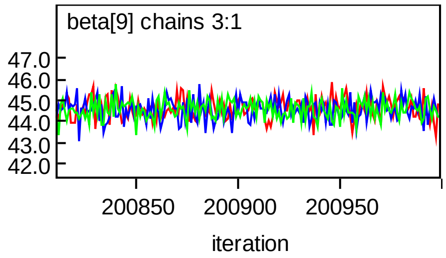
Figure 23: Dynamic Trace for Pure Hierarchical Bayesian Priors

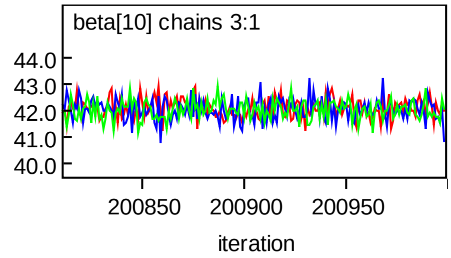


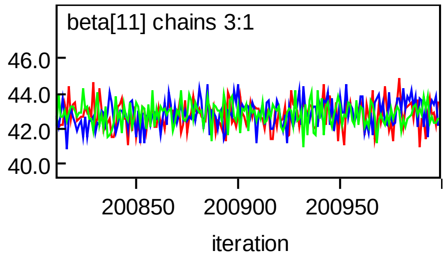

beta[6] chains 3:1
-5.0
-10.0
-15.0
-20.0
200850
200900
200950
iteration

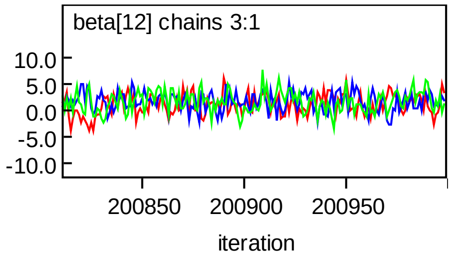

beta[7] chains 3:1
38.0
36.0
34.0
32.0
30.0
28.0
200850
200900
200950
iteration

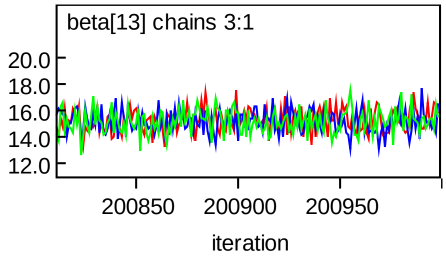

beta[8] chains 3:1
-16.0
-18.0
-20.0
-22.0
-24.0
200850
200900
200950
iteration

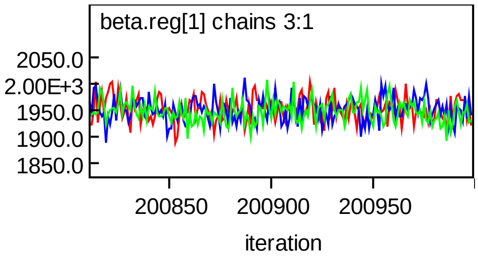

beta[9] chains 3:1
47.0
46.0
45.0
44.0
43.0
42.0
200850
200900
200950
iteration

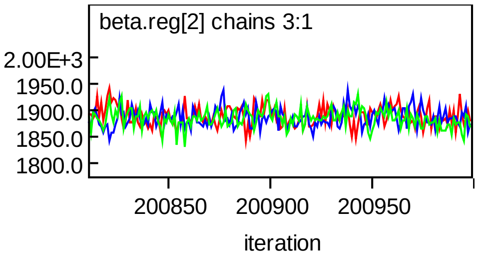

beta[10] chains 3:1
44.0
43.0
42.0
41.0
40.0
200850
200900
200950
iteration


beta[11] chains 3:1
46.0
44.0
42.0
40.0
200850
200900
200950
iteration


beta[12] chains 3:1
10.0
5.0
0.0
-5.0
-10.0
200850
200900
200950
iteration


beta[13] chains 3:1
20.0
18.0
16.0
14.0
12.0
200850
200900
200950
iteration


beta.reg[1] chains 3:1
2050.0
2.00E+3
1950.0
1900.0
1850.0
200850
200900
200950
iteration


beta.reg[2] chains 3:1
2.00E+3
1950.0
1900.0
1850.0
1800.0
200850
200900
200950
iteration

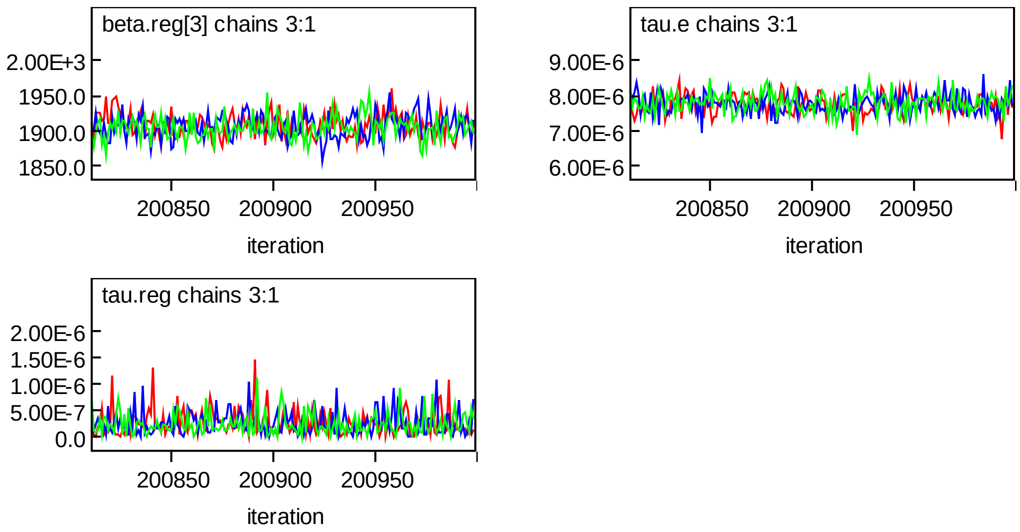
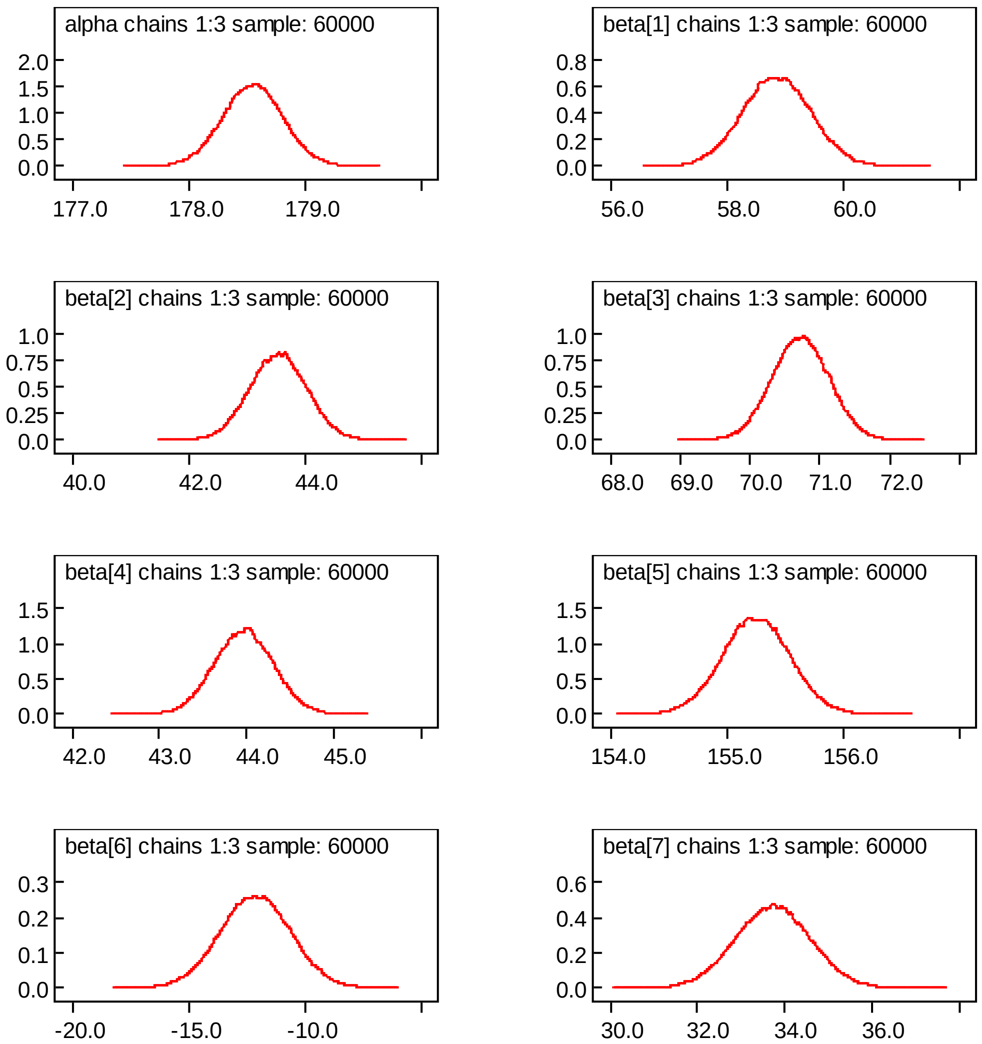
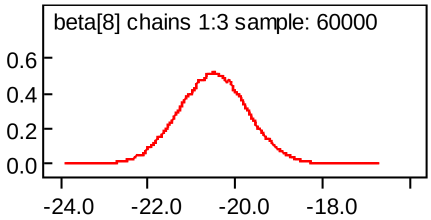
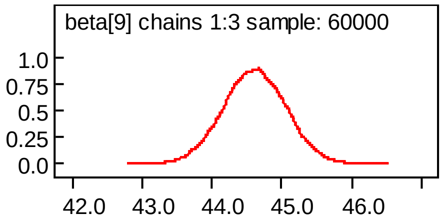
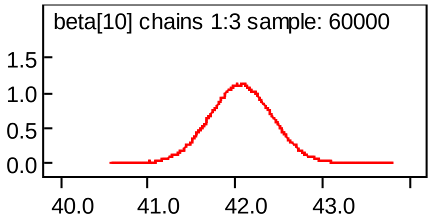


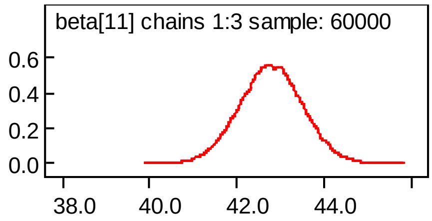
Figure 24: Kernel Density for Pure Hierarchical Bayesian Priors

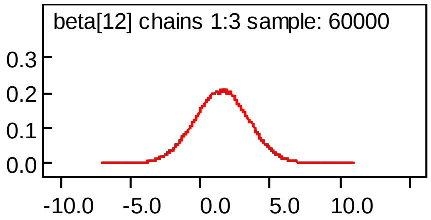


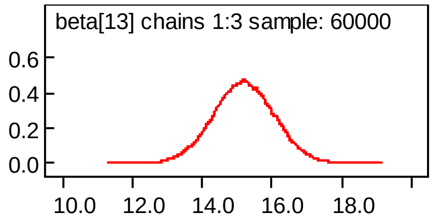

beta[8] chains 1:3 sample: 60000
0.6
0.4
0.2
0.0
-24.0
-22.0
-20.0
-18.0

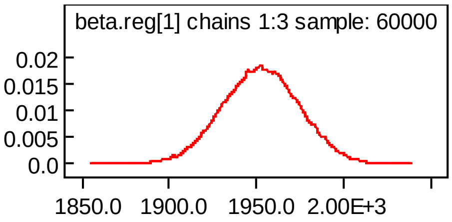

beta[9] chains 1:3 sample: 60000
1.0
0.75
0.5
0.25
0.0
42.0
43.0
44.0
45.0
46.0

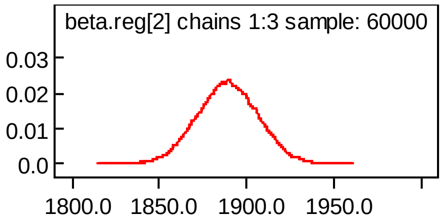

beta[10] chains 1:3 sample: 60000
1.5
1.0
0.5
0.0
40.0
41.0
42.0
43.0

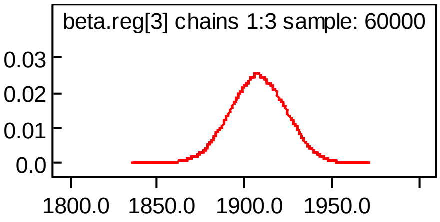

beta[11] chains 1:3 sample: 60000
0.6
0.4
0.2
0.0
38.0
40.0
42.0
44.0

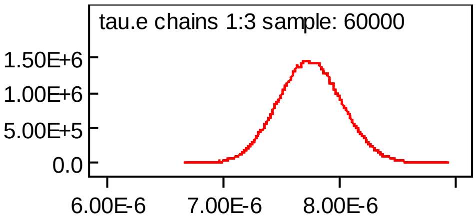

beta[12] chains 1:3 sample: 60000
0.3
0.2
0.1
0.0
-10.0
-5.0
0.0
5.0
10.0


beta[13] chains 1:3 sample: 60000
0.6
0.4
0.2
0.0
10.0
12.0
14.0
16.0
18.0


beta.reg[1] chains 1:3 sample: 60000
0.02
0.015
0.01
0.005
0.0
1850.0
1900.0
1950.0
2.00E+3


beta.reg[2] chains 1:3 sample: 60000
0.03
0.02
0.01
0.0
1800.0
1850.0
1900.0
1950.0


beta.reg[3] chains 1:3 sample: 60000
0.03
0.02
0.01
0.0
1800.0
1850.0
1900.0
1950.0


tau.e chains 1:3 sample: 60000
1.50E+6
1.00E+6
5.00E+5
0.0
6.00E-6
7.00E-6
8.00E-6

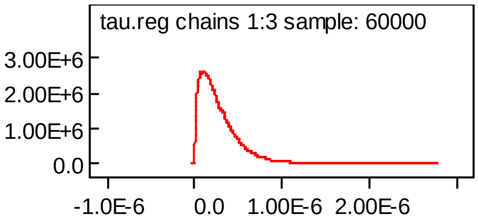
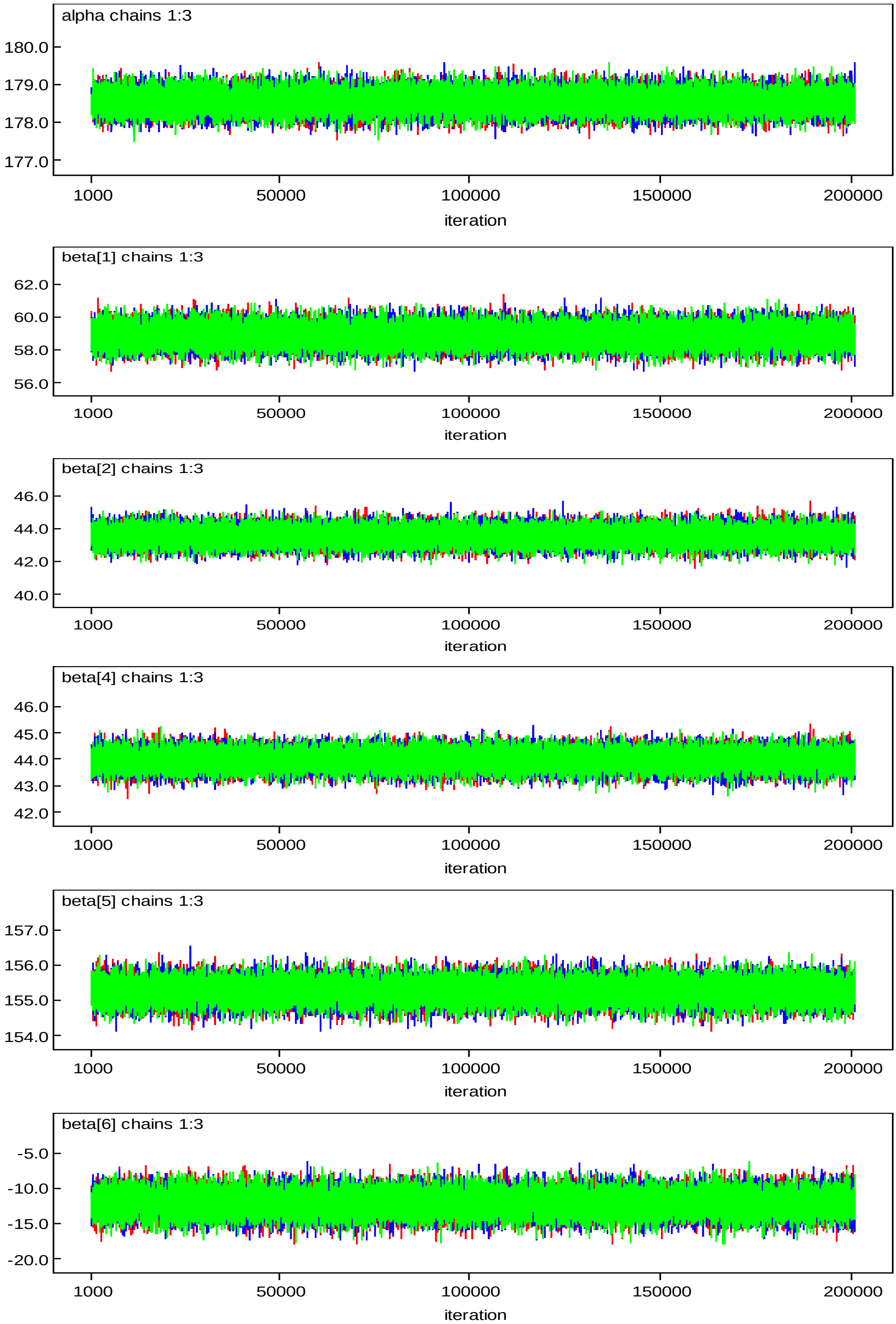
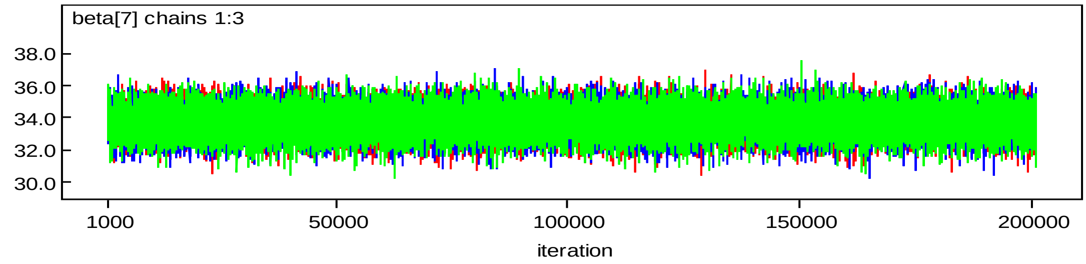
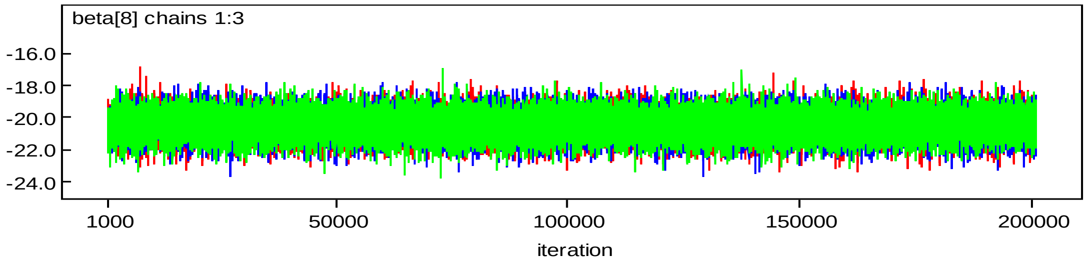
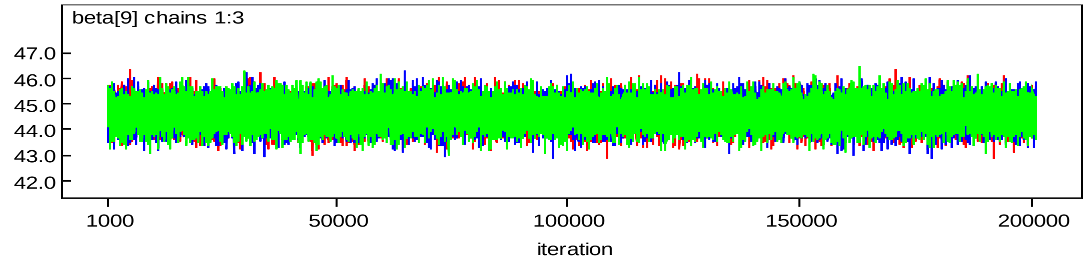

tau.reg chains 1:3 sample: 60000
3.00E+6
2.00E+6
1.00E+6
0.0
-1.0E-6
0.0
1.00E-6
2.00E-6

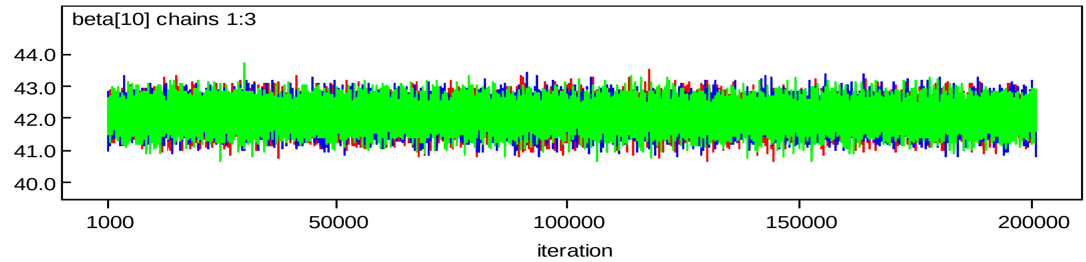
Figure 25: Time Series for Pure Hierarchical Bayesian Priors

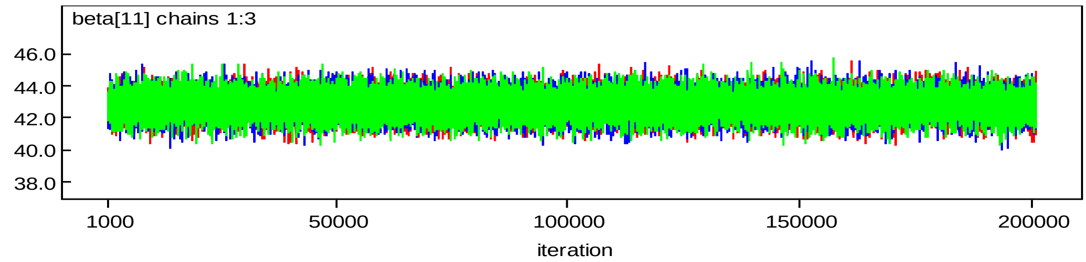


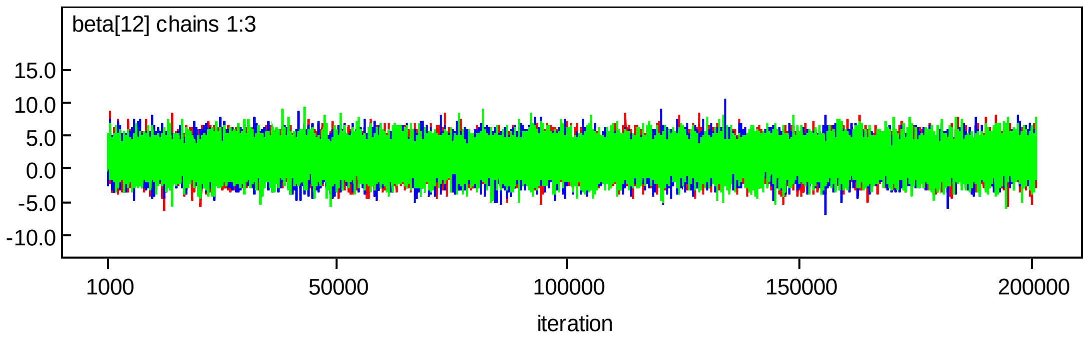

beta[7] chains 1:3
38.0
36.0
34.0
32.0
30.0
1000
50000
100000
150000
200000
iteration

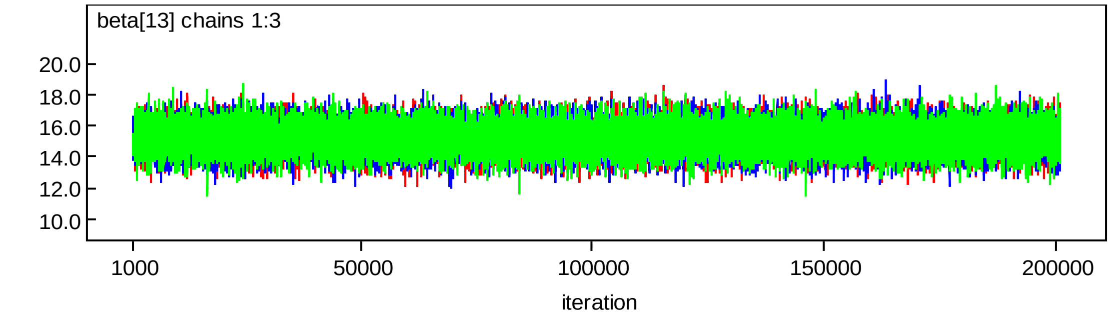

beta[8] chains 1:3
-16.0
-18.0
-20.0
-22.0
-24.0
1000
50000
100000
150000
200000
iteration

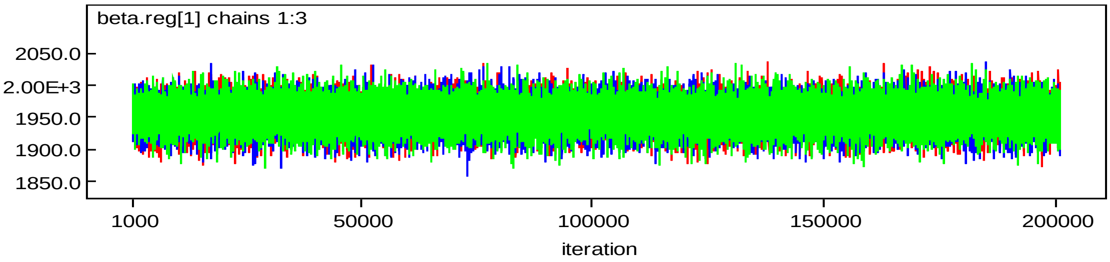

beta[9] chains 1:3
47.0
46.0
45.0
44.0
43.0
42.0
1000
50000
100000
150000
200000
iteration

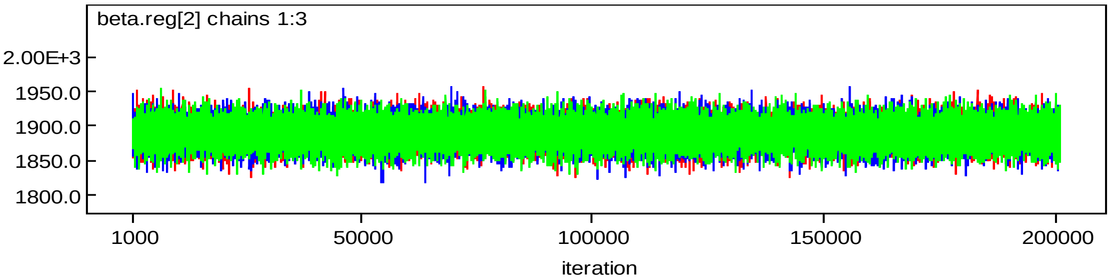

beta[10] chains 1:3
44.0
43.0
42.0
41.0
40.0
1000
50000
100000
150000
200000
iteration

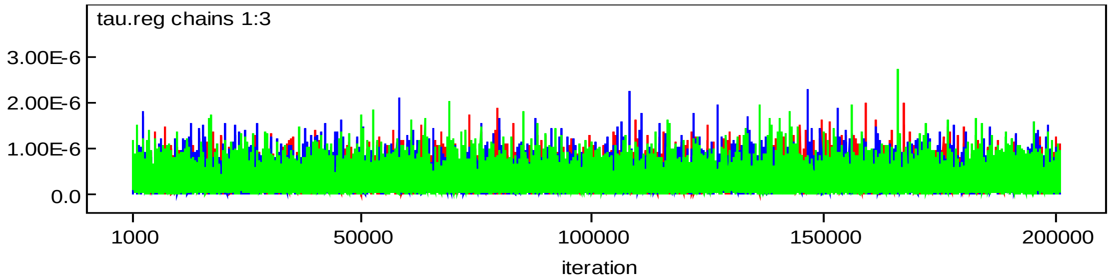

beta[11] chains 1:3
46.0
44.0
42.0
40.0
38.0
1000
50000
100000
150000
200000
iteration

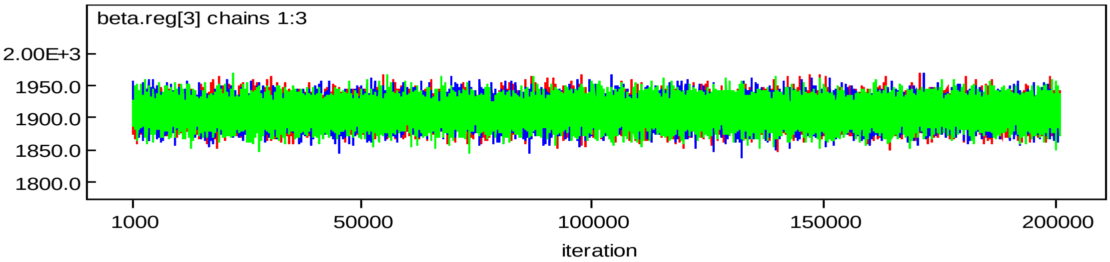

beta[12] chains 1:3
15.0
10.0
5.0
0.0
-5.0
-10.0
1000
50000
100000
150000
200000
iteration


beta[13] chains 1:3
20.0
18.0
16.0
14.0
12.0
10.0
1000
50000
100000
150000
200000
iteration


beta.reg[1] chains 1:3
2050.0
2.00E+3
1950.0
1900.0
1850.0
1000
50000
100000
150000
200000
iteration


beta.reg[2] chains 1:3
2.00E+3
1950.0
1900.0
1850.0
1800.0
1000
50000
100000
150000
200000
iteration


tau.reg chains 1:3
3.00E-6
2.00E-6
1.00E-6
0.0
1000
50000
100000
150000
200000
iteration

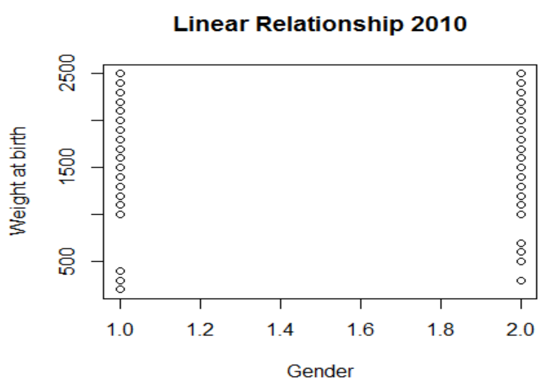

beta.reg[3] chains 1:3
2.00E+3
1950.0
1900.0
1850.0
1800.0
1000
50000
100000
150000
200000
iteration

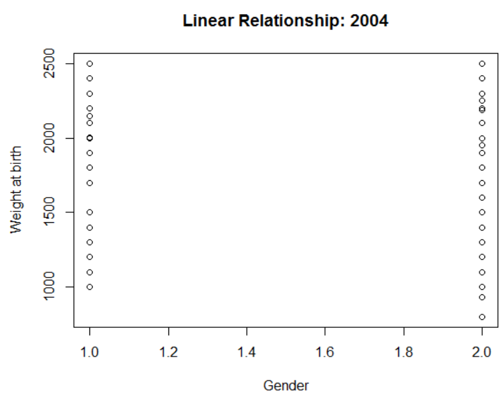
4.4 Multiple Linear Regression Assumptions

A good model should meet the assumptions mentioned and explained in page 28, that is, linearity, normality, homoscedasticity and independence of observations. This section fits these above-mentioned assumptions to the Malawi data to check whether the model.

Linear Relationship

This refers to estimation of the relationship between dependent and independent variables, i.e. if the relationship is linear in nature. If the relationship between weight at birth and the independent variables is not linear, the results of the regression analysis will under-estimate the true relationship. This result in an increased risk of Type II error for that independent variable, and in case of multiple regression, an increased risk of Type I error (over-estimation). One of the ways to detect non-linearity is the use of residual plots (Pedhazur (1997) and Cohen (1983)). Scatter plot of residuals below show the linear relationship between the dependent (weight at birth) and the independent variables (gender of a child and region):





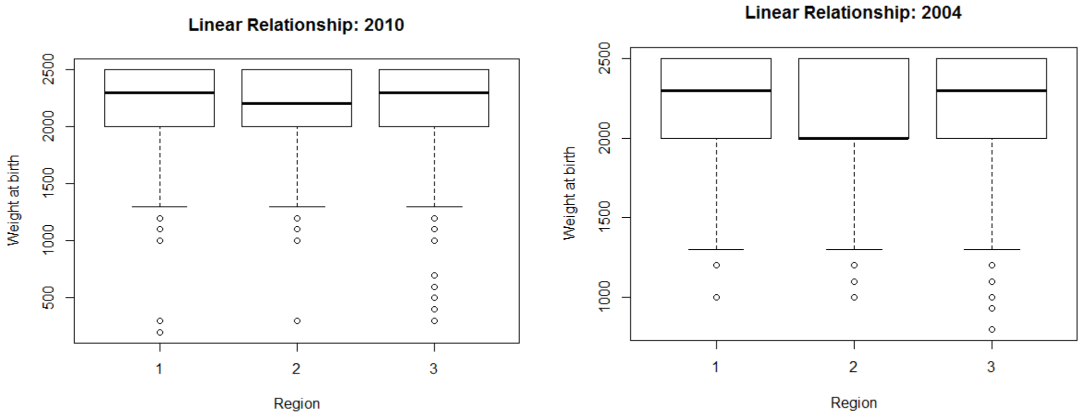


Figure 26: Linear Relationship

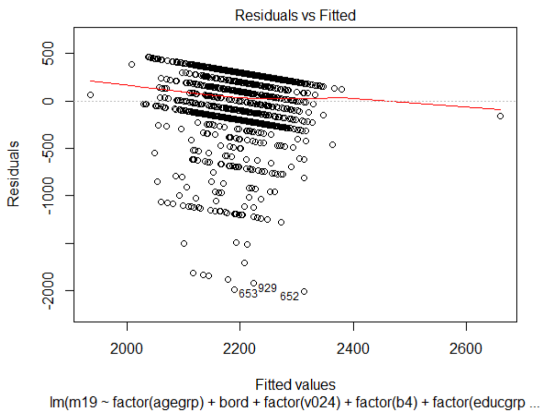
The linear relationship between weight at birth and gender is shown in Figure 26 for the years 2010 and 2004. Most female children have smaller weight at birth than female children (1.0 in the x-axis present male children). There are few children whose weight at birth is less than 1kg when compared to those who are greater than 1kg. Most children weigh close to 2.5kg which is the maximum weight for this study.

Data is based on the country of Malawi that is categorized into the following three regions, Northern, Central and the Southern region. The box-plot in figure 28 shows that we have more outliers in the Southern region, that is, the Southern region has more children who are underweight compared to other regions. According to the box-plot, the minimum child's weight at birth is approximately 1.25kg for all regions if outliers are ignored.

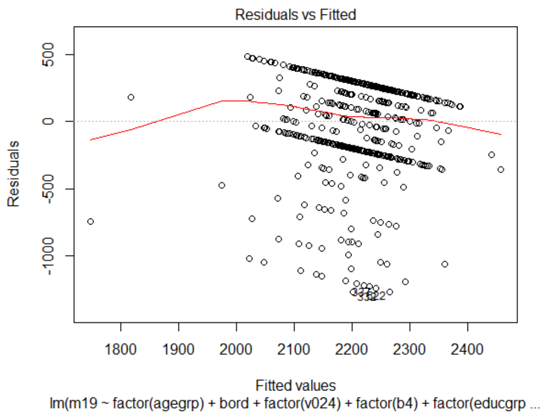
Homoscedasticity

This refers to instances when the variation of observation around the regression line (the residual standard error) is constant. To check this, scatter plots for residuals are used. For a weight at birth model, the residual versus fitted plot is given in figure 27 below.

Year 2010



Year 2004



Year 2000

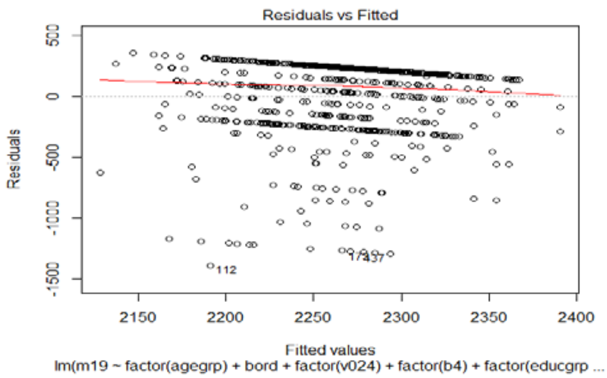


Figure 27: Test for Homoscedasticity

Above is the residual versus fit plot for years 2010 and 2004 which is the most frequently created plot. It is a scatter plot of residuals on the *y-axis* and fitted values (estimated responses) on the x-axis. The plot is used to detect non-linearity, unequal error variances and outliers. The characteristics of well-behaved residuals versus fit plots and what they are suggesting about the appropriateness of the simple linear regression model are:

- The residuals “bounce randomly” around the 0 line. This suggests that the assumption that the relationship is linear is meaningful.

- The residuals roughly form a “horizontal band” around the 0 line. This suggests that the variances of the error terms are the same.
- No single residual “stands out” from the basic pattern of residuals. This suggest that there are no outliers.

Based on the above plot of residuals, the linear assumption is meaningful for our model, since the residuals bounce randomly around the 0 line. The variances are not the same based on the results above - there does not seem to be a horizontal line in the residuals. The residual plot versus fit above shows that there are potential outliers. To check this, the year 2010 boxplot for quantitative data is used. The child’s birth order variable seems to have 4 outliers from the plot below plot.

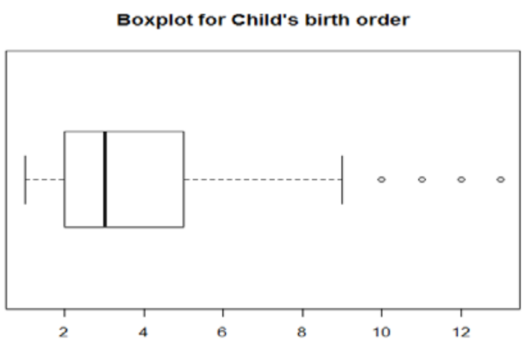


Figure 28: Boxplot for Child’s Birth Order Number

The Goldfeld-Quandt Test can test for heteroscedasticity. The parametric and nonparametric are also used to test the hypothesis that the residuals from a squares regression are homoscedastic. The parametric test uses F-statistics, whereas the nonparametric test uses number of picks in the ordered sequence of unsigned residuals.

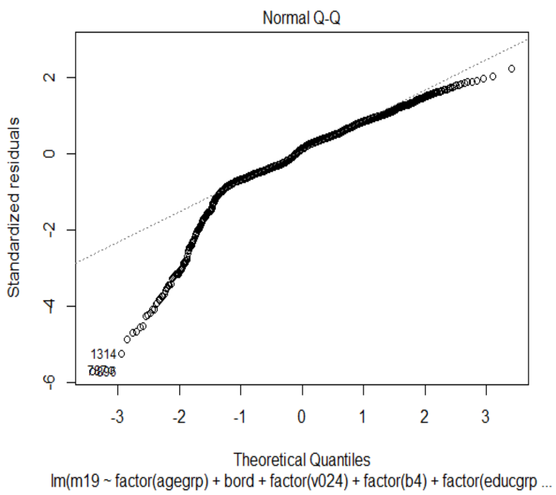
Multivariate Normality

This refers to some tests that check if a set of data is similar to the multivariate normal distribution. Non-normally distributed variables (highly skewed or kurtotic variables, or variables with substantial outliers) can distort relationships and significance tests. This results in analyses that are not exact. Several pieces of information are used to test this assumption; skewness, visual inspection of data plots, kurtosis etc.

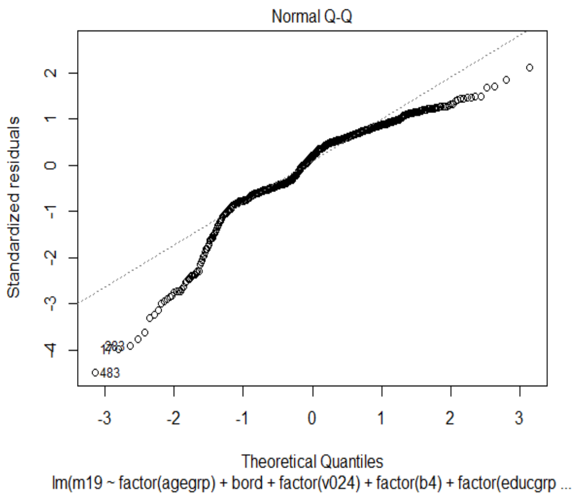
The most essential test that can be used is a Normal Quantile-Quantile (Q-Q plot). This is a graphical technique that determines if two data sets come from populations with a common distribution. A Q-Q plot is a plot of the quantiles of the first data set against the quantiles of the second data set. A 45-degrees reference line is also plotted. If the two sets come from a population with same distribution, the points should fall approximately along this reference line. The advantages of the Q-Q plot are:

- Sample sizes do not have to be equal.
- There are many other features of the distribution that can be tested simultaneously, i.e. shifts in scale, changes in symmetry, and the presence of outliers can be detected from this plot.

2010



2004



2000

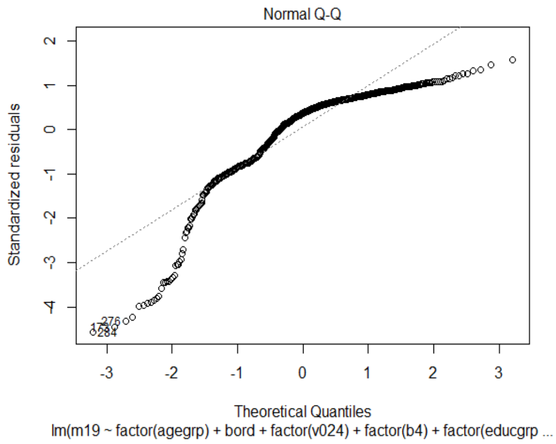


Figure 29: Normal Q-Q Plot

The model fitted seem to be negatively skewed for all the years i.e. year 2010, 2004 and 2000.

No/Little Multicollinearity

Multicollinearity occurs when the independent variables are dependent upon one another. The independence assumption that the error term of the mean is uncorrelated is also important in this instance. This means that the standard mean error of the dependent variable is independent from the independent variables. Criteria of checking multicollinearity includes; Correlation matrix, Tolerance, Variance Inflation Factor, Condition Index.

Table 18 Variance Inflation Factor

Variable	**GVIF**	**DF**	$\mathbf{GVIF}^{(\frac{1}{2*DF})}$
Mothers Age	3.599	5	1.132
Birth order	3.566	1	1.840
Region	1.055	2	1.011
Gender	1.021	1	1.005
Education level	1.332	2	1.071
Antenatal visits	1.013	1	1.004
Work	1.059	2	1.009
HIV status	1.033	1	1.009

Generalized Variance Inflation Factor (GVIF) measures how much the variance of the estimated regression coefficients are inflated compared to when the predictor variables are not linearly related. The following table has guidelines on how to interpret the VIF:

Table 24: VIF Interpretation Table

VIF	**Status of predictors**
VIF=1	Not correlated
1<VIF<5	Moderately correlated
VIF>5 to 10	Highly correlated

Most of variables in the birth weight model are moderately correlated since the 1<VIF<5. The size of the standard error is determined by taking the square root of the variance inflation factor, compared with what it would be if that variable were uncorrelated with the other predictor variables in the model. For this model, the variance inflation factor for mother's age is 3.467. ($\sqrt{3.467} = 1.861$). This means that the standard error for the coefficient of mother's age is 1.861 times as large as it would be if the mother's age variable was uncorrelated with the other predictor variables.

Table 25: Test for autocorrelation

Lag	Autocorrelation	D-W statistic	p-value
1	0.0826	1.8338	0.004

Autocorrelation refers to correlation between members of a series of numbers arranged in ti me.

The hypothesis test for autocorrelation is:

- The null hypothesis (H_0) is that there is no correlation among residuals, i.e. they are i ndependent.
- The alternative (H_0) is that the residuals are auto correlated.

Based on the results from the about table, the null hypothesis is rejected since the p-value is zero for the model.

<u>Outliers and Leverage</u>

An observation that is significantly different from all other ones can make a large difference in the results of regression analysis. Various factors might lead to outliers, such as misplaced decimal points, recording or transmission errors, exceptional phenomena such as earthquakes or strikes, or member of a different population slipping into the sample. Outliers play an important role in regression. There are two types of outliers:

- Observation of outliers that results because of the response variable, these observations represent the model failure and they are called *Outliers.*
- Outliers with respect to the predictors in the model are called *leverage points.* These points are either good (unusually large or small among X values but is not a regression outlier) or bad (points situated far from the regression around which most are centered) leverage points.

Table 26: Outlier Test

	rstudent	Unadjusted p-value	Bonferonni p
896	-5.714959	1.2396e-08	1.8697e-05
787	-5.658189	1.7202e-08	1.9823e-05
1314	-5.391927	1.6836e-07	2.7787e-04
409	-4.355373	1.3876e-06	1.1732e-03
627	-4.309751	2.2033e-06	4.7856e-03
426	-4.309751	2.2033e-06	4.7856e-03
226	-4.236285	5.7873e-06	8.1376e-03
1003	-4.125142	6.2211e-06	9.6567e-03
45	-4.025755	2.4012e-05	3.7504e-02
644	-4.843507	2.3606e-05	3.8498e-02

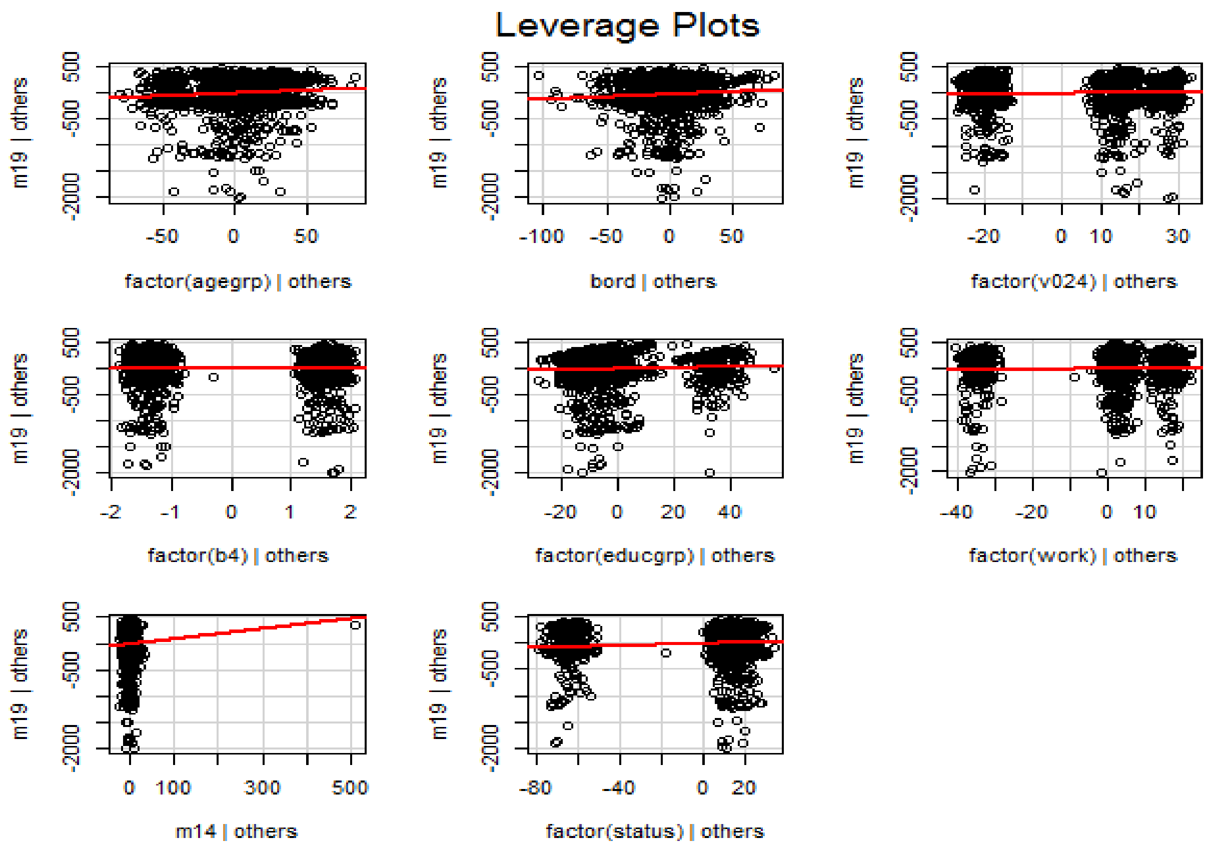


Figure 30: Leverage plot

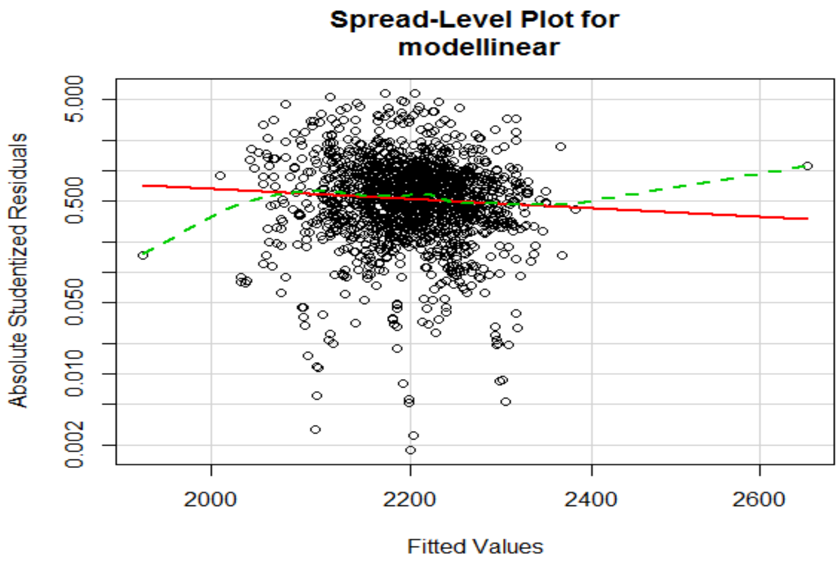


Figure 31: Spread Level Plot

The Spread-and-Level Plot (S-L plot) is a guidance plot that assist in promoting equal spread and symmetry among groups of a factor variables through an appropriate transformation. It creates plots for examining the possible dependence of spread on level, or an extension of these plots to the studentized residuals from linear models.

Final Decision:

The linearity and homoscedasticity assumption for our model is met since the residual plot has no pattern, the regression line (red line from the plot) is fairly flat and there seems to be a constant variation. This indicates that the expected residuals are normally distributed for the model. The VIF values for all the independent variables in the model are moderately correlated since the $1<VIF<5$. This implies that there is little multicollinearity in the model. The hypothesis test for outliers is:

H_0: There are no outliers.

H_1: There is at least one outlier.

At 5% level of significance the null hypothesis is rejected since all the unadjusted and Bonferroni p-values are less than 0.05. We then conclude that there is at least one outlier in the independent variables data.

Chapter 5: Discussion and Conclusion

5.1: Discussion

Findings of this study highlighted that the variables, age of a mother, birth order number of a child, mother's employment status and HIV status of a mother are significant factors influencing low birth weight. Results obtained for a non-hierarchical multiple linear regression model are almost similar to those of a multiple multilevel linear regression model. However, the hierarchical model was necessary and provided better- quality results because findings revealed that there were significant differences in causes of low birth weight between geographical regions. This result was expected in the context of low birth weight in Malawi. The analyses were fitted using Bayesian multilevel models to check for better quality results. A Bayesian linear regression analysis with non-informative and informative priors were conducted.

The non-informative Bayesian hierarchical results show that the variables, age of a mother, birth order number of a child and HIV status of a mother are significant factors influencing low birth weight of a child in this study. The non-informative Bayesian hierarchical model provided results that are almost similar to the classical model (multiple hierarchical results) approach except for mothers working status which is not significant under non-informative Bayesian hierarchical results. A non-informative Bayesian model is expected to lead to identical results with classical model. When prior information from previous surveys is included in the Bayesian multilevel model (Bayesian hierarchical model with informative priors), the results indicates that the variables, age of a mother, birth order number of a child, gender of a child, mothers' education level, working status of a mother and HIV status of a mother are significant factors influencing low birth weight of a child. Results from the same model also revealed that antenatal visit was not a significant determinant of low birth weight in Malawi.

The Bayesian multilevel method is a preferred method because it is theoretically stronger and provides more realistic results. Several authors such as OJO et al. (2017) have shown that the inclusion of scientific prior knowledge in the Bayesian model leads to better results. All significant variables in the Bayesian model with informative priors were also significant in other models (classical and Bayesian non-informative) while the opposite was not true. The inclusion of the priors resulted in the significance of key determinants of low birth weight which would have been omitted otherwise. In addition to that, the Bayesian model showed that women who are working are more likely to have a child with better weight at birth than those who are not working. This result is different from the result of multilevel classical multiple linear model. It clearly illustrates that the Bayesian model provides more realistic and comprehensive results in line with the literature. In this study, Bayesian multilevel multiple linear regression with informative priors appeared as the preferred model to consider in the study of weight at birth. Findings of this study points out that mothers aged 15 to 19, with primary education or no education at all, HIV positive, with male children and with high birth order numbers have high risk of low birth weight and deserve more attention. Bayesian informative models are in line with the literature review in terms of the determinants of underweight factors (Ngwira, A., & Stanley, C. C., 2015). page

5.2 Conclusion and Recommendations

From the analysis of the statistical methods considered in this study, it can be concluded that young mothers who are not educated are the main category to have low birth weight children. This is due to their ignorance of how to take care of themselves during pregnancy - a factor which works against child birth weight or size at birth. In addition to the above, birth order number and gender of a child can be a determinant of low birth weight. This study found that the educational level of mothers in Malawi has a relationship with low birth weight of a child. The Bayesian informative technique is then considered to provide better results than the classical technique. The availability of scientific evidence to back up choice of prior information strengthens the preference for Bayesian technique in this study. Bayesian results showed that Mother's age, birth order number, mother's level of education, antenatal visits and HIV status of a mother are most likely to influence birth weight.

It is recommended that more attention is given to women aged 15-19 in terms of nutrition during pregnancy to avoid low birth weight. Maternal nutrition status in turn may have a direct effect on child birth weight. Improvement of adolescent nutrition and maternal malnutrition for improved pregnancy outcomes should be implemented in Malawi. It is also recommended to take different measures to monitor the weight at birth according to the gender of the baby, age and education of the mother. This can be achieved through having more antenatal visits during pregnancy where a doctor is consulted. It is highly recommended for mothers who are HIV positive to take an antiretroviral drug such as AZT during pregnancy, as this reduces the risk of HIV transmission from mother to child by 67%. Improvement of adolescent nutrition and maternal malnutrition for improved pregnancy outcomes should be implemented in Malawi.

Bibliography

Badshah, S., Mason, L., McKelvie, K., Payne, R., & Lisboa, P. J. (2008). Risk factors for low birthweight in the public-hospitals at Peshawar, NWFP-Pakistan. *BMC Public Health*, *8*(1), 197.

Black, R. E., Morris, S. S., & Bryce, J. (2003). Where and why are 10 million children dying every year?. The lancet, 361(9376), 2226-2234.

Bale, J. R., Stoll, B. J., & Lucas, A. O. (2003). The Problem of Low Birth Weight.

Berger, J. O. (2013). Statistical decision theory and Bayesian analysis. Springer Science & Business Media.

Berry, W. D. (1993). Understanding regression assumptions (Vol. 92). Sage Publications.
Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7).

Black, R. E., Morris, S. S., & Bryce, J. (2003). Where and why are 10 million children dying every year?. The lancet, 361(9376), 2226-2234.

Cumming, G. (2013). Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis. Routledge.

Crawley, M. J. (2007). Mixed-Effects Models. *The R Book, Second Edition*, 681-714.
Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. (2007). Linear and nonlinear mixed effects models. *R package version*, *3*, 57.

De Cock, K. M., Mbori-Ngacha, D., & Marum, E. (2002). Shadow on the continent: public health and HIV/AIDS in Africa in the 21st century. The Lancet, 360(9326), 67-72.

El Beitune, P., & Duarte, G. (2006). Antiretroviral agents during pregnancy: consequences on hematologic parameters in HIV-exposed, uninfected newborn infant. European Journal of Obstetrics & Gynecology and Reproductive Biology, 128(1), 59-63.

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., & Heiberger, R. (2015). Package ‘car’.

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: estimation rather than hypothesis testing. Br Med J (Clin Res Ed), 292(6522), 746-750.

Gibbs, C. M., Wendt, A., Peters, S., & Hogue, C. J. (2012). The impact of early age at first childbirth on maternal and infant health. Paediatric and perinatal epidemiology, 26(s1), 259-284.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). *Bayesian data analysis* (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Geronimus, A. T. (1996). Black/white differences in the relationship of maternal age to birthweight: a population-based test of the weathering hypothesis. *Social science & medicine*, *42*(4), 589-597.

Grytten, J., Skau, I., & Sorensen, R. J. (2014) Educated mothers, healthy infants:The impact of a school reform on the birth weight of Norwegian infants 1967- 2005. Social Science & Medicine, 105. 84-92

Horta, B. L., Victora, C. G., Menezes, A. M., Halpern, R., & Barros, F. C. (1997). Low birthweight, preterm births and intrauterine growth retardation in relation to maternal smoking. Paediatric and perinatal epidemiology, 11(2), 140-151.

Jacoby, W. G. (2005). Regression iii: Advanced methods. Retrieved October, 18, 2011.

Johnson, D. (2014). The Association between Maternal Age and Low Birth Weight Offspring, NHANES 2007-2008.

Judith, P. & Kelley, L. (2000). Low birthweight. UN ACC Sub-Committee on Nutrition.

Kamman EE, Wand MP. Geoadditive Models. Journal of the Royal Statistical Society C. 2003; 52: 1–18.

Katzmarzyk, P. T., Craig, C. L., & Bouchard, C. (2001). Original article underweight, overweight and obesity: relationships with mortality in the 13-year follow-up of the Canada Fitness Survey. Journal of clinical epidemiology, 54(9), 916-920.

Kéry, M. (2010). Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press.

Kutner, M. H., Nachtsheim, C., & Neter, J. (2004). Applied linear regression models. McGraw-Hill/Irwin.

Malawi, D. H. S. (2010). Malawi Demographic and Health Survey. Zomba, Malawi, and Calverton.

McCormick, M. C. (1985). The contribution of low birth weight to infant mortality and childhood morbidity. New England journal of medicine, 312(2), 82-90.

Muula, A. S., & Maseko, F. C. (2006). How are health professionals earning their living in Malawi?. BMC health services Research, 6(1), 97.

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis. John Wiley & Sons.

Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (1996). Applied linear statistical models (Vol. 4, p. 318). Chicago: Irwin.

Nelson R. Low Birth Weights: Spatial and Socio-economic Patterns in Kamloops. Western Geography.1998; 8/9(1998/99): 31–59.

Ngwira, A., & Stanley, C. C. (2015). Determinants of low birth weight in Malawi: Bayesian geo-additive modelling. PloS one, 10(6), e0130057.

Oja, H. (2010). Multivariate linear regression. In *Multivariate Nonparametric Methods with R* (pp. 183-200). Springer New York.

Ojo OB, Lougue S, Woldegerima WA (2017) Bayesian generalized linear mixed modeling of Tuberculosis using informative priors. PLoS ONE 12(3): e0172580.

Oken, E., Kleinman, K. P., Rich-Edwards, J., & Gillman, M. W. (2003). A nearly continuous measure of birth weight for gestational age using a United States national reference. BMC pediatrics, 3(1), 6.

Okosun, I. S., Liao, Y., Rotimi, C. N., Dever, G. E. A., & Cooper, R. S. (2000). Impact of birth weight on ethnic variations in subcutaneous and central adiposity in American children aged 5–11 years. A study from the Third National Health and Nutrition Examination Survey.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. Practical assessment, research & evaluation, 8(2), 1-9.

Osei, F. B., Duker, A. A., & Stein, A. (2012). Bayesian structured additive regression modeling of epidemic data: application to cholera. *BMC medical research methodology*, *12*(1), 118.

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. Quality & Quantity, 41(5), 673-690.

O'Campo P, Xue X, Wang M, O'Brien CM. Neighbourhood Risk Factors for Low Birthweight in Baltimore: A Multilevel Analysis. AmericanJournal of Public Health. 1997b; 87: 1113–1118.

Padilla, Y. C., & Reichman, N. E. (2001). Low birthweight: Do unwed fathers help?. Children and Youth Services Review, 23(4-5), 427-452.

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte Carlo methods for item response models. Journal of educational and behavioral Statistics, 24(2), 146-178.

Raudenbush, S.W., & Bryk, A.S. Hierarchical Linear Models. Applications and Data Analysis Methods. Newbury Park, CA: Sage, 2nd ed., 2002.

Reinsel, G. C., &Velu, R. P. (1998). Multivariate Linear Regression. In *Multivariate Reduced-Rank Regression* (pp. 1-14). Springer New York.

World Health Organization. (1992). Low Birth Weight: A tabulation of available information.

Sellström, E., & Bremberg, S. (2006). Review Article: The significance of neighbourhood context to child and adolescent health and well-being: A systematic review of multilevel studies. Scandinavian journal of public health, 34(5), 544-554.

Seltzer, M. & Choi, K. 2002 Model checking and sensitivity analysis for multilevel models. In N. Duan & S. Reise (Eds.), multilevel modeling: Methodological advances, issues, and applications. Hillsdale, NJ: Lawrence Erlbaum.

Silvestrin, S., da Silva, C. H., Hirakata, V. N., Goldani, A. A., Silveira, P. P., & Goldani, M. Z. (2013). Maternal education level and low birth weight: a meta-analysis. Jornal de Pediatria (Versão em Português), 89(4), 339-345.

Simon, L. J. (2004). Detecting multicollinearity using variance inflation factors. Penn State Department of Statistics, The Pennsylvania State University.

Sobiech, N. M. (2014). State-level analysis of variation in infant mortality rate: A socio-ecological approach.

Snijders, T.A.B., and Bosker, R.J. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. London etc.: Sage Publishers, 1999.

UNAIDS, J. (2010). Global report: UNAIDS report on the global AIDS epidemic 2010. Geneva: UNAIDS.

World Health Organization. (2004). International statistical classification of diseases and related health problems (Vol. 1). World Health Organization.

APPENDIX A

Keys for Determinants

<u>Agegrp – Mother's age group</u>
Agegrp1 - (15 – 19)
Agegrp2 - (20 – 24)
Agegrp3 - (25 – 29)
Agegrp4 - (30 – 34)
Agegrp5 - (35 – 39)
Agegrp6 - (40 – 49)

bord - Child's birth order number

b4 - Child's gender

<u>Educgrp - Mother's education</u>
Educgrp 0 - No education
Educgrp 1 - Primary education
Educgrp 2 - Secondary +

<u>Work – Mother's working status</u>
Work 0 - Not working
Work 1 - Was working previous year
Work 2 - Working

M14 - Antenatal visits for the pregnancy

<u>Mother's HIV status</u>
Status 1 - HIV positive
Status 2 - HIV negative

Non-Informative Bayesian - Linear Regression codes

Model 2010

```
{
for (i in 1:Nobs) {
m19[i] ~ dnorm(mu[i], tau.e)
        mu[i] <- alpha + beta[1]*agegrp_2[i] + beta[2]*agegrp_3[i] + beta[3]*agegrp_4[i] + beta[4]*agegrp_5[i] +
beta[5]*agegrp_6[i] + beta[6]*bord[i] + beta[7]*v024_2[i]+beta[8]*v024_3[i] + beta[9]*b4_2[i]+ beta[10]*educgrp_2[i] +
beta[11]*educgrp_1[i] + beta[12]*work_1[i] + beta[13]*work_2[i] + beta[14]*m14[i] + beta[15]*status_2[i]
    }

tau.e ~ dgamma(0.01, 0.001)     # residual error variance
alpha ~ dnorm (1,0.00001) # intercept
beta[1] ~ dnorm (1,0.00001) # regression coefficients
beta[2] ~ dnorm (1,0.00001) # regression coefficients
beta[3] ~ dnorm (1,0.00001) # regression coefficients
beta[4] ~ dnorm (1,0.00001) # regression coefficients
beta[5] ~ dnorm (1,0.00001) # regression coefficients
beta[6] ~ dnorm (1,0.00001) # regression coefficients
beta[7] ~ dnorm (1,0.00001) # regression coefficients
beta[8] ~ dnorm (1,0.00001) # regression coefficients
beta[9] ~ dnorm (1,0.00001) # regression coefficients
beta[10] ~ dnorm (1,0.00001) # regression coefficients
beta[11] ~ dnorm (1,0.00001) # regression coefficients
beta[12] ~ dnorm (1,0.00001) # regression coefficients
beta[13] ~ dnorm (1,0.00001) # regression coefficients
beta[14] ~ dnorm (1,0.00001) # regression coefficients
beta[15] ~ dnorm (1,0.00001) # regression coefficients
}
list(tau.e=2,  alpha=1, beta=c(1,2,3,2,1,5,1,5,6,9,7,1,2,4,2))
list(tau.e=3,  alpha=0.1, beta=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
list(tau.e=0.2,  alpha=4, beta=c(0.1,0.2,0.3,0.2,0.1,0.4,0.8,0.1,0.7,0.3,0.4,0.7,0.2,0.3,0.4))

list(Nobs=1589)
m19[]     agegrp_2[]     agegrp_3[]     agegrp_4[]     agegrp_5[]     agegrp_6[]
    bord[]     v024_2[]  v024_3[]  b4_2[]    educgrp_2[]    educgrp_1[]    work_1[]  work_2[]  m14[]
    status_2[]
25000     1     0     0     0     3     1     0     1     0     1     0
    1     4     1
END
```

Informative Bayesian - Linear Regression codes

Mixed Bayesian Priors

```
Model 2010
{
for (i in 1:Nobs) {
m19[i] ~ dnorm(mu[i], tau.e)
    mu[i] <- alpha + beta[1]*agegrp_2[i] + beta[2]*agegrp_3[i] + beta[3]*agegrp_4[i] + beta[4]*agegrp_5[i] + beta[5]*agegrp_6[i]
+ beta[6]*bord[i] + beta[7]*v024_2[i]+beta[8]*v024_3[i] + beta[9]*b4_2[i]+ beta[10]*educgrp_2[i] + beta[11]*educgrp_1[i] +
beta[12]*work_1[i] + beta[13]*work_2[i] + beta[14]*m14[i] + beta[15]*status_2[i]
  }

tau.e ~ dgamma(936.61,0.0011)    # residual error variance
alpha ~ dnorm (0,0.00001) # intercept
beta[1] ~ dnorm (71.22,2.42) # regression coefficients
beta[2] ~ dnorm (69.66,3.34) # regression coefficients
beta[3] ~ dnorm (108.79,5.15) # regression coefficients
beta[4] ~ dnorm (93.39,7.78) # regression coefficients
beta[5] ~ dnorm (215.67,10.94) # regression coefficients
beta[6] ~ dnorm (-19.96,0.30) # regression coefficients
beta[7] ~ dnorm (40.13,2.22) # regression coefficients
beta[8] ~ dnorm (48.21,1.71) # regression coefficients
beta[9] ~ dnorm (24.40,1.10) # regression coefficients
beta[10] ~ dnorm (-23.48,1.29) # regression coefficients
beta[11] ~ dnorm (20.88,4.18) # regression coefficients
beta[12] ~ dnorm (29.55,10.43) # regression coefficients
beta[13] ~ dnorm (47.15,1.13) # regression coefficients
beta[14] ~ dnorm (-0.69,0.16) # regression coefficients
beta[15] ~ dnorm (6.73,0.89) # regression coefficients

}

list(tau.e=2,  alpha=1, beta=c(1,2,3,2,1,5,1,5,6,9,7,1,2,4,2))
list(tau.e=3,  alpha=0.1, beta=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
list(tau.e=0.2,  alpha=4, beta=c(0.1,0.2,0.3,0.2,0.1,0.4,0.8,0.1,0.7,0.3,0.4,0.7,0.2,0.3,0.4))

list(Nobs=1589)

m19[]   agegrp_2[]   agegrp_3[]   agegrp_4[]   agegrp_5[]   agegrp_6[]   bord[]
        v024_2[]  v024_3[]  b4_2[]  educgrp_2[]  educgrp_1[]  work_1[]  work_2[]  m14[]  status_2[]
2500    0   1   0   0   0   3   1   0   1   0   1   0
        1   4   1
END
```

Pure Bayesian Priors

```
Model 2010
{
for (i in 1:Nobs) {
m19[i] ~ dnorm(mu[i], tau.e)
        mu[i] <- alpha + beta[1]*agegrp_2[i] + beta[2]*agegrp_3[i] + beta[3]*agegrp_4[i] + beta[4]*agegrp_5[i] + beta[5]*agegrp_6[i] + beta[6]*bord[i] + beta[7]*v024_2[i]+beta[8]*v024_3[i] + beta[9]*b4_2[i]+ beta[10]*educgrp_1[i] + beta[11]*educgrp_2[i] + beta[12]*work_1[i] + beta[13]*work_2[i] + beta[14]*m14[i] + beta[15]*status_2[i]
      }

tau.e ~ dgamma(930.54,0.00107)      # residual error variance
alpha ~ dnorm (0,0.00001) # intercept
beta[1] ~ dnorm (84.195,2.29) # regression coefficients
beta[2] ~ dnorm (66.65,3.15) # regression coefficients
beta[3] ~ dnorm (89.38,4.89) # regression coefficients
beta[4] ~ dnorm (59.02,7.40) # regression coefficients
beta[5] ~ dnorm (168.95,10.42) # regression coefficients
beta[6] ~ dnorm (-7.37,0.18) # regression coefficients
beta[7] ~ dnorm (39.92,2.18) # regression coefficients
beta[8] ~ dnorm (58.91,1.63) # regression coefficients
beta[9] ~ dnorm (43.54,1.12) # regression coefficients
beta[10] ~ dnorm (46.04,1.56) # regression coefficients
beta[11] ~ dnorm (28.8,3.53) # regression coefficients
beta[12] ~ dnorm (50.2,6.91) # regression coefficients
beta[13] ~ dnorm (50.54,1.75) # regression coefficients
beta[14] ~ dnorm (0.88,0.55) # regression coefficients
beta[15] ~ dnorm (-7.37,0.18) # regression coefficients

}

list(tau.e=2,  alpha=1, beta=c(1,2,3,2,1,5,1,5,6,9,7,1,2,4,2))
list(tau.e=3,  alpha=0.1, beta=c(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1))
list(tau.e=0.2,  alpha=4, beta=c(0.1,0.2,0.3,0.2,0.1,0.4,0.8,0.1,0.7,0.3,0.4,0.7,0.2,0.3,0.4))

list(Nobs=1589)
m19[]    agegrp_2[]         agegrp_3[]         agegrp_4[]         agegrp_5[]         agegrp_6[]         bord[]
         v024_2[]  v024_3[]  b4_2[]    educgrp_2[]        educgrp_1[]        work_1[]  work_2[]  m14[]     status_2[]
2500     0         1         0         0         0         3         1         0         1         0         1         0
END
```

Non-Informative Bayesian - Multilevel Regression Models

```
model
{
for (i in 1:Nobs) {
        m19[i] ~ dnorm(mu[i], tau.e)
        mu[i] <- alpha + beta[1]*agegrp_2[i] + beta[2]*agegrp_3[i] + beta[3]*agegrp_4[i] + beta[4]*agegrp_5[i] + beta[5]*agegrp_6[i] + beta[6]*bord[i] + beta[7]*b4_2[i]+ beta[8]*educgrp_1[i] + beta[9]*educgrp_2[i] + beta[10]*work_1[i] + beta[11]*work_2[i] + beta[12]*m14[i] + beta[13]*status_2[i] +beta.reg[v024[i]]
                    }
     # priors on regression coefficients and variances
            tau.e ~ dgamma(0.001, 0.001)      # residual error variance
            sigma2.e <- 1/tau.e
            alpha ~ dnorm (0, 0.00001) # intercept
            beta[1] ~ dnorm (0,0.00001) # regression coefficients
            beta[2] ~ dnorm (0,0.00001) # regression coefficients
            beta[3] ~ dnorm (0,0.00001) # regression coefficients
            beta[4] ~ dnorm (0,0.00001) # regression coefficients
            beta[5] ~ dnorm (0,0.00001) # regression coefficients
            beta[6] ~ dnorm (0,0.00001) # regression coefficients
            beta[7] ~ dnorm (0,0.00001) # regression coefficients
            beta[8] ~ dnorm (0,0.00001) # regression coefficients
            beta[9] ~ dnorm (0,0.00001) # regression coefficients
            beta[10] ~ dnorm (0,0.00001) # regression coefficients
            beta[11] ~ dnorm (0,0.00001) # regression coefficients
            beta[12] ~ dnorm (0,0.00001) # regression coefficients
            beta[13] ~ dnorm (0,0.00001) # regression coefficients
            beta.reg[1] ~ dnorm (0,tau.reg) # regression coefficients
            beta.reg[2] ~ dnorm (0,tau.reg) # regression coefficients
            beta.reg[3] ~ dnorm (0,tau.reg) # regression coefficients
            tau.reg ~ dgamma (0.001,0.001) # regression coefficients
}
```

Initial Values

```
list(tau.e=2,  alpha=1, beta=c(1,2,3,2,1,5,1,5,6,9,7,1,3), beta.reg=c(4,6,6), tau.reg=1.0)

list(tau.e=3,  alpha=0.1, beta=c(1,1,1,1,1,1,1,1,1,1,1,1,1), beta.reg=c(-0.9,-0.6,0.2), tau.reg=2)

list(tau.e=0.2,  alpha=4, beta=c(0.1,0.2,0.3,0.2,0.1,0.4,0.8,0.1,0.7,0.3,0.4,0.7,0.2), beta.reg=c(4,6,3), tau.reg=3)

list(Nobs=1589)
m19[]   agegrp_2[]      agegrp_3[]      agegrp_4[]      agegrp_5[]      agegrp_6[]      bord[]
        v024[]   b4_2[]   educgrp_2[]     educgrp_1[]     work_2[]  work_3[]  m14[]   status_2[]
2500    0       1       0       0       0       3       2       1       0       1       1       0
        4       1
END
```

Informative Bayesian - Multilevel Regression Models

```
model
{
for (i in 1:Nobs) {
        m19[i] ~ dnorm(mu[i], tau.e)
        mu[i] <- alpha + beta[1]*agegrp_2[i] + beta[2]*agegrp_3[i] + beta[3]*agegrp_4[i] + beta[4]*agegrp_5[i] + beta[5]*agegrp_6[i]
+ beta[6]*bord[i] + beta[7]*b4_2[i]+ beta[8]*educgrp_1[i] + beta[9]*educgrp_2[i] + beta[10]*work_1[i] + beta[11]*work_2[i] +
beta[12]*m14[i] + beta[13]*status_2[i] +beta.reg[v024[i]]
                    }
     # priors on regression coefficients and variances
             tau.e ~ dgamma(12.081, 0.08278)      # residual error variance
             sigma2.e <- 1/tau.e
             alpha ~ dnorm (178.55, 14.88) # intercept
             beta[1] ~ dnorm (58.905, 2.90) # regression coefficients
             beta[2] ~ dnorm (43.52, 4.18) # regression coefficients
             beta[3] ~ dnorm (70.72, 5.97) # regression coefficients
             beta[4] ~ dnorm (43.96, 9.03) # regression coefficients
             beta[5] ~ dnorm (155.26, 11.78) # regression coefficients
             beta[6] ~ dnorm (-11.065, 0.37) # regression coefficients
             beta[7] ~ dnorm (33.78, 1.38) # regression coefficients
             beta[8] ~ dnorm (-20.64, 1.64) # regression coefficients
             beta[9] ~ dnorm (44.66, 4.98) # regression coefficients
             beta[10] ~ dnorm (42.08, 7.68) # regression coefficients
             beta[11] ~ dnorm (42.78, 1.93) # regression coefficients
             beta[12] ~ dnorm (-0.885, 0.18) # regression coefficients
             beta[13] ~ dnorm (15.115, 1.34) # regression coefficients
             beta.reg[1] ~ dnorm (0,tau.reg) # regression coefficients
             beta.reg[2] ~ dnorm (0,tau.reg) # regression coefficients
             beta.reg[3] ~ dnorm (0,tau.reg) # regression coefficients
             tau.reg ~ dgamma (0.001,0.001) # regression coefficients
}
```

Initial Values

```
list(tau.e=2,  alpha=1, beta=c(1,2,3,2,1,5,1,5,6,9,7,1,3), beta.reg=c(4,6,6), tau.reg=1.0)

list(tau.e=3,  alpha=0.1, beta=c(1,1,1,1,1,1,1,1,1,1,1,1,1), beta.reg=c(-0.9,-0.6,0.2), tau.reg=2)

list(tau.e=0.2,  alpha=4, beta=c(0.1,0.2,0.3,0.2,0.1,0.4,0.8,0.1,0.7,0.3,0.4,0.7,0.2), beta.reg=c(4,6,3), tau.reg=3)

list(Nobs=1589)
m19[]    agegrp_2[]          agegrp_3[]          agegrp_4[]          agegrp_5[]          agegrp_6[]          bord[]
         v024[]     b4_2[]   educgrp_2[]         educgrp_1[]         work_2[]  work_3[]  m14[]     status_2[]
2500     0          1        0          0        0          3        2         1         0         1         1         0

END
```