**College of Natural Sciences**

**Department of Statistics**

# An applications of Bayesian multilevel logistic regression model on age of initiation for sexual intercourse among women in Ethiopia

---

By:

Tariku Irana

A Thesis Submitted to Jimma University, College of Natural Sciences, Department of Statistics in Partial Fulfillment for the Requirements of the Degree of Master of Science in Biostatistics

September 6, 2021

Jimma, Ethiopia

# An applications of Bayesian multilevel logistic regression model on age of initiation for sexual intercourse among women in Ethiopia

**MSc. Thesis**

**By:**

**Tariku Irana**

**Main Advisor: Akalu Banbeta (Ass't Prof.)**
**Co-Advisor: Reta Habtamu (MSc.)**

**August, 2021**
**Jimma, Ethiopia**

SCHOOL OF GRADUATE STUDIES

JIMMA UNIVERSITY

COLLEGE OF NATURAL SCIENCES

**Approval Sheet**

As thesis research advisors, we here by certify that we have read the thesis prepared by Tariku Irana under our guidance, which is entitled **"An applications of Bayesian multilevel logistic regression model on age of initiation for sexual intercourse among women in Ethiopia"**, in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and ful-fill university and department style requirements; (2) its illustrative materials including tables.

Approved by:

| | | |
|---|---|---|
| Akalu Banbeta (Ass't Prof.)<br>Name of Major Advisor | Signature | 01/09/2021<br>Date |
| Reta Habtamu (MSc)<br>Name of Co-Advisor | Signature | 01/09/2021<br>Date |
| Samuel Fekadu (MSc.)<br>Name of the Internal Examiner | Signature | 01/09/2021<br>Date |
| Shibru Temesgen(PhD)<br>Name of the External Examiner | Signature | 01/09/2021<br>Date |
| Kibrealem Sisay(MSc.)<br>Name of the Chairperson | Signature | 06/09/2021<br>Date |

Jimma University

School of Graduate Studies

Department of Statistics

As thesis research advisors, we here by certify that we have read the thesis research prepared by **Tariku Irana Dinagde** under our guidance, which is entitled "An applications of Bayesian multilevel logistic regression model on age of initiation for sexual intercourse among women in Ethiopia, in its final format it is consistent and acceptable. Hence we recommend that the thesis accepted as it fulfills the university and department style requirements for the degree of Master of Science in Biostatistics.

| | | |
|---|---|---|
| Akalu Banbeta (Ass't Prof.)<br>Main Advisor | ...........<br>Signature | ...............<br>Date |
| Reta Habtamu (MSc.)<br>Co-advisor | ...........<br>Signature | ...............<br>Date |

As the members of the board of examiners of MSc. thesis open defense examination of **Tariku Irana Dinagde**, we certify that we have read and evaluated the thesis and examined the candidate. We recommend that the thesis has been accepted as it fulfills the requirements for the degree of Master of Science in Biostatistics.

| | | |
|---|---|---|
| ...........<br>Name of External Examiner | ...........<br>Signature | ...............<br>Date |
| ...........<br>Name of Internal Examiner | ...........<br>Signature | ...............<br>Date |
| ...........<br>Name of Chairman | ...........<br>Signature | ...............<br>Date |

**DECLARATION**

I declare that this thesis is my original work and that all source materials used for this thesis have been properly cited and acknowledged. This thesis has been submitted in partial fulfillment of the requirements for MSc. degree of Master of Science in Biostatistics at Jimma University. I earnestly declare that this thesis is not submitted to any other institution anywhere for the award of any academic degree, diploma, or certificate.

| | | |
|---|---|---|
| Tariku Irana Dinagde | ........... | ............... |
| Name | Signature | Date |

This thesis has been submitted for examination with my approval as a University advisor and Co- advisor.

| | | |
|---|---|---|
| Akalu Banbeta (Ass't Prof.) | ........... | ............... |
| Main Advisor | Signature | Date |
| Reta Habtamu (MSc.) | ........... | ............... |
| Co-advisor | Signature | Date |

# Contents

# List of Figures

# List of Tables

**Acknowledgment**

Above all, I would like to thank the almighty God for the gift of health, wisdom and strength throughout any steps for the achievement of my thesis and more of my entire life.

My deep gratitude goes to my advisor Mr. Akalu Banbata (Assistant Professor) for his professional and deep comments, advice and guidance at any stage of this thesis. I appreciated his friendly contact and his punctuality for comments beginning from the proposal development up to the final work of this thesis. Simply, without his frequent supervision, this thesis may not reach at the stage it is now.

I sincerely appreciate my co-advisor Mr. Reta Habtamu (MS.c) for his valuable suggestions and comments for the successful realization of this thesis. It is a great pleasure to express his commitment to share his experience, knowledge and his smooth contacts is also appreciable.

Also I would like to thank gratefully and sincerely instructor Asefu Nigussie (Madda Walabu University) and instructor Gudata Aga (Ambo university) who help me during various working time of this thesis work and also I would like to thank Ambo University, my employer for the financial support, motivation, and giving chance to pursue this master program and to all Jimma University staff members of the Statistics program.

My sincere gratitude goes to my family for their tenacious love and encouragement. I do not believe words can express my feelings to Amanu Fayisa Irana for supporting me in all direction and all my best friends for your unconditional love and supports. Wishes you long life .

**ABSTRACT**

**Background:** Globally, 333 million new cases of curable sexual transmitted infections estimated in each year. Approximately 11% of all births worldwide and nearly 95% of these birth occur in developing countries. In Africa region all female in Liberia and Kongo had first sexual intercourse at age of 18 has 78% and 77% respectively. In Ethiopia, 29% had the first sexual intercourse before age 15 years, 62% before age 18 years 58% marry before the age of 18, which can increase the vulnerability of early sexual debut.

**Methods:** The 2016 Ethiopian Demographic and Health Survey (EDHS) dataset were used and a total of 11962 participants reproductive age were included. Bayesian multilevel logistic regression model was applied to investigate the risk factors and regional variations in early sexual debut among women in Ethiopia. The prevalence of early sexual debut have been analyzed with factors like age, education, place of residence, wealth index, occupation, region, religion and marital status.

**Results:** Out of 11962 (75.2%) had their first sexual intercourse before the age of 18 years was early sexual debut among women in Ethiopia respectively. About 69.7% women to having experience of sexual debut where living in rural area and 30% living in urban area. The odd of women who come from urban area 42.1% less likely (OR=0.579 ) to having experienced early sexual debut compared to women who come from rural area. Poorer was 20.3% (OR=1.203) times more likely to experience early sexual debut when compared to poorest women. The odd of women working is increased by 13.6%(OR=1.136) compared to women who have not working. Bayesian multilevel random coefficient model is the most significant model and best fit to the data.

**Conclusions:** Age, education, place of residence, wealth index, occupation and marital status was significantly affect the early sexual debut. Educational coverage and community-level of wealth status are important intervention area to delay the age of early sexual debut.

**Keywords:** Early sexual debut, Bayesian multilevel, Ethiopia

## ACRONYMS

| | |
|---|---|
| ACF: | Autocorrelation Function |
| AIC: | Akaike Information Criterion |
| ANC: | Antenatal care |
| BDIC: | Bayesian Deviance Information Criterion |
| BML: | Bayesian Multilevel |
| CI: | Credible interval |
| CSA: | Central Statistical Agency |
| DIC: | Deviance Information Criterion |
| EAs: | Enumeration areas |
| EDA: | Exploratory data analysis |
| EDHS: | Ethiopian Demographic and Health Survey |
| ESS: | Effective of sample size |
| FMoH: | Federal Ministry of Health |
| HIV: | Human Immunodeficiency Virus |
| ICC: | Intra-class correlation coefficient |
| IGLS: | Iterative generalized least squares procedure |
| LMIC: | low and middle income countries |
| MCMC: | Markov chain Monte-Carlo |
| MCSE: | Monte Carlo Standard Error |
| MLE: | Maximum Likelihood Method |
| ORs: | Odd Ratios |
| PACF: | Partial Autocorrelation Function |
| STI: | Sexual Transmitted infection |
| WHO: | World Health Organization |

# 1 Introduction

## 1.1 Background of the study

Early sexual initiation was taken as an experience of first sexual intercourse before the age of 18 years [1, 2]. According to World Health Organization early sexual initiation is associated with adolescents increased risks of unwanted pregnancy, unsafe abortion, developing fistula and sexual transmitted infections including human immunodeficiency virus (HIV) [3]. Around 16 million adolescent girls aged 15-19 give birth in each year, approximately 11% of all births worldwide and nearly 95% of these birth occur in developing countries [3].

Globally, 333 million new cases of curable STIs estimated in each year occur, connected to the highest rates among aged 20–24 years , followed by aged 15–19 years [4]. Around 218 million women of reproductive age (15–49) and adolescent women (15-19) an estimated around 21 million are pregnancy are living in low and middle income countries(LMICs) [5]. Approximately 111 million women of reproductive age (15–49) and around 10.5 million adolescent are unintended pregnancy [5]. All maternal deaths in worldwide,approximately 86% is accounting for sub-Saharan Africa and South Asia,the highest proportion (32%) of young people aged 10 – 24 years Sub-Saharan Africa [6, 7].

According to world health organization on Africa region all female in Liberia and Kongo had first sexual intercourse at age of 18 has 78% and 77% respectively [8]. Approximately 31% of females between the ages of 20 and 24 in the African region were married before age 18 years. The region for early marriage were Niger (76%), Central African Republic (68%), Chad (67%), Mali (54%), and South Sudan (52%) are the top five countries [8].

In Ethiopia, around 6 million women who are within the age group of 15–19 accounting for 12% of the total female population with extremely limited media access (26% report at least weekly exposure to radio, 18% of television and 9% to newspapers); consequently, 39% had first sex before age 18 and 28% of recent births to women younger than 20 were unplanned [9].

The 2016 Ethiopian Demographic and Health Survey (EDHS) data used for this study is based on two stage stratified cluster sampling. The appropriate approach to analyzing age of sexual initiation among women in Ethiopia from this survey is therefore based on nested sources of variability. Here the units at lower level (level-1) are individuals (early sexual debut ) who are nested within units at higher level (region). Beside the nested source of variability; the response variable in this study is early sexual debut which is binary response. Therefore, for analyzing the data, Bayesian multilevel model is used in this study. The reason why the Bayesian approach is preferred over the usual frequentest technique is that the power of information obtained from the approach is much better as it is the combination of likelihood data and prior information about the distribution of the parameter.

## 1.2 Statement of the Problem

Ethiopia maternal mortality ratio for 2017 was 401 deaths per 100,000 live births. According to the report of EDHS in 2011 the maternal mortality ratio were 676 per 100,000 live births which was higher as compared to the global average and also the report was revealed that, 29% first had sexual intercourse before age 15 years, 62% before age 18 years 58% marry before the age of 18, which can increase the vulnerability of early sexual debut [10]. This is indicate the burden on age of initiation for sexual intercourse among women in Ethiopia.

A study conducted in Ethiopia by correlates of sexual initiation among adolescent and youth in Addis Ababa the proportion of sexual initiation aged between 15-24 years was 49.3% [11]. Youth who begin early sexual activity are more likely to practice risky sexual behaviors, such as multiple sexual partners and incorrect or inconsistent condom use. As a result, they increase the risk of sexually transmitted infection (STIs), including HIV/AIDS, unwanted pregnancy, unsafe abortion, early childbirth, vaginal fistula, substantial socio-economic and psychosocial problems. These problems are the greatest threats to health and wellbeing of the youth [12].

Studies conducted in Ethiopia used logistic regression or multivariate logistic regression and spatial analysis to identify important risk factors of the age initiation of sexual intercourse among female, youth and women in Ethiopia [13, 14]. It is found out a number of socioeconomic and demographic characteristics of women greatly influence age of initiation of sexual intercourse of women such as age of women, education, respondent occupation , place of residence, religion , wealth index, radio, television and region. However, their study did not investigate the relative contribution at the individual and regional level of factors influencing women's age of initiation of sexual intercourse of women using the Bayesian multilevel logistic regression.

To fill this gap, Bayesian multilevel logistic regression model has been employed for identifying factors affecting age of initiation of sexual intercourse of among women in Ethiopia at individual and regional levels.

In general, this study were addressed the following research questions:

1) What are the most influential factors that have significant contribution on early sexual debut among women in Ethiopia?.

2) Do regions differ in the occurrence of age of initiation of sexual intercourse of among women in Ethiopia?

3) From the study variables which predictors have variation across regions?

4) Which model is a good fit to the data well?

### 1.2.1 General Objectives of the study

The general objective of this study was to examine the extent of regional disparity on age of initiation for sexual intercourse and to examine factors associated with early sexual debut.

### 1.2.2 Specific objectives of the study

1) To examine the most influential factors that has significantly associated with early sexual debut among women in Ethiopia .

2) To investigate between and within regional variations of age of initiation of sexual intercourse of among women in Ethiopia.

3) To determine from the study variables, the variation of predictors across regions.

4) To compare Bayesian multilevel empty model, the intercept model and random coefficient model.

### 1.2.3 Significance of the study

The results of this study could provide information to government and other concerned organizations in setting policies, society, strategies and further investigation for understanding the effect of age of initiation of sexual intercourse.

The other basic significance of the study is that it may also further assist other researchers interested in this area and they may use it as a benchmark for their future works.Its also expected to give some knowledge about the determinants or risk factors of early sexually transmitted disease(infection)

# 2 Literature Review

## 2.1 Overview of age of initiation of sexual intercourse

Early sexual debut is defined as having sexual intercourse before the age of 18 year. Sexual initiation usually happens during adolescence and early onset may result in unwanted consequences in several aspects and also early initiation of sexual activity affects the sexual and reproductive health of the young population [2]. Furthermore, girls before 15 years are five times at higher risk of death and those 15–19 years are twice more likely to die than women aged 20–24 years in pregnancy or childbirth [15, 16]. Complications from Pregnancy and childbirth are the primary cause of disease (1 out of 7 girls) among before aged 19 years in third world countries [17].

In 2016, an estimated 21 million girls aged 15–19 years in developing countries became pregnant, approximately 12 million of whom gave birth and nearly half of pregnancies to girls aged 15–19 years in developing countries are unintended and the estimated 5.6 million abortions that occur each year among adolescent girls aged 15–19 years, 3.9 million are unsafe, contributing to maternal mortality, morbidity and lasting health problems [18]. The health consequences, pregnancy and childbirth complications are the leading cause of death among girls aged 15–19 years globally, with 99% of global maternal deaths of women aged 15–49 years accounting for low- and middle-income countries [19].

In African countries , an early sexual debut is high which ranging from 26.8% in Nigeria to 55% in Ghana [20, 21]. Among Nigerian adolescents aged 15-19 years, a fifth of them were found to have initiated sexual activities [22]. A study in Ghana indicated that age at first sexual experience increased from 16 years in 1988 to 18 years in 2014 [23]. The studies conducted in Nigeria based on geographical variations and contextual effects on age of initiation of sexual intercourse among women in Nigeria using multilevel and spatial study of age at first sexual intercourse found strong evidence that women's odds of starting sexual intercourse early was significantly associated with respondents' current age, education attainment, religion affiliation, ethnicity, geopolitical region, and community median age of marriage. The log odds of early sexual debut increased with increasing age and decreased with increasing level of education [24].

### 2.1.1 Sexual intercourse in Ethiopia

Like other African countries, experiences of early sexual initiation among Ethiopian female adolescents are high, ranging from 19% in Shire-Endasillasie to 20.4% in Ambo university, through to 27.6% in Legehida district of Amhara region [25, 26]. According to few local studies in Ethiopia, being a female increases the likelihood of early sexual initiation [27, 28] as well as, educational status, place of residence, income and religion are found to be factors associated with an early sexual initiation in female adolescents [25, 29].

The study based on geographical variations of early age sexual initiation among reproductive-age women in Ethiopia: evidence from EDHS 2016 using the independent variable age, religion, wealth index, type of residence, respondent occupation, region, marital status and education. Out of these variable place of residence, educational attainment, wealth index, and marital status were identified as statistically significant factors on early sexual debut and Poorer [AOR = 1.28, 95% CI: (1.12–1.46)] and richest [AOR = 1.46, 95% CI: (1.23, 1.74)] were 1.28 and 1.46 times more likely to experience early sexual initiation as compared to the poorest women respectively [30].

### 2.1.2 Determinants of sexual debut among women in Ethiopia

A number of socio-demographic factors have been identified to be contributing to early sexual debut. Risk factors or early sexual debut can be categorized in two levels (**individual level**: current age , education, religion, respondent occupation and marital status) and (**community level** is residence, wealth index and region).

### Age

Age at first sexual activity in many areas tends to begin at a younger age than in the past. The mean age of marriage has gradually been increasing while the age of puberty in both sexes appears to be falling [31]. The study done in Nigeria using multilevel and spatial distribution revealed that age of women is significant association on early sexual debut. This variable are three groups(15-24,25-34,and 34+). The median age at first sexual intercourse for all women included in the study was 15 years [24].

### Education

Education has been found to be a protective factor to early sexual experience because it tends to delay sexual intercourse to a later stage in life compared to those that a less educated [23]. Adolescent with no education or little education are more likely to initiate early sexual intercourse [32]. This is because education support adolescent with relevant information about sexual and reproductive health needs and also support knowledge about their sexual needs and rights that may affect their future careers and opportunity [32]

The study conducted in Nigeria based on A Cox Proportional Hazard Model revealed that education significant association with Age at First Sexual Intercourse with (p value $0.00 < 0.05$). This variable had four categories(no education, primary, secondary and higher). It was found that the hazard of having an early sexual intercourse for those with primary, secondary and higher education decreases by 14.3% (1-0.857) x 100, 24.3% and 47.6% respectively when compared with those with no education [33].

The study conducted to early initiation of sexual intercourse among adolescent females in Ethiopia revealed that primary, secondary, and higher education level decreases the odds of early sexual debut compared to those with no formal education (primary AOR = 0.44, 95% CI: 0.35, 0.56), (secondary AOR = 0.19, 95% CI: 0.13, 0.28), and (higher AOR= 0.31, 95% CI: 0.15, 0.63) [34]. The other study is depicted that female youth who were no attending school were 14 times more likely initiate sex at or before age

18 than attending higher education(AOR = 14.1, 95% CI = (8.1, 24.7) [35].

Similar the study conducted in Ethiopia depicted that as the education level of teenage women increase the percentage of teenage pregnancy decrease.The percentage of teenage pregnancy based the education level is (34.0%)for uneducated, (11.5%) primary and (6.7%) secondary and above education level [36].

### Wealth Index

The protective effects of wealth against early sexual experience has been established [32, 37]. Belonging to a rich household wealth index reduces the risk of early sexual experience, For example, the study of in Nepal found that belonging to a rich wealth index is associated with lower likelihood of early sexual activities [37]. A study carried out in Amhara region to establish the odds of being married below the age of 18 in the poorer and poorest level 1.38 and 2.37 more likely to be early married compared to those women in the richest level (AOR = 1:38 and CI: 1.16,4.83; AOR = 2:37 and CI: 2.19, 7.83), respectively [38].

### Respondent Occupation

The study conducted in Nigeria using cox proportional hazard regression model and logistic regression model revealed that respondent occupation where associated with early sexual debut for female youth and female youth who were working also had higher odds (OR 1.28, C.I. 1.03, 1.58) of using a condom at last sex compared to those who were not working [39].
The other study show that percentage of teenage pregnancy based on current working status of women shows that 17.0% of teenage women those who did not work were pregnant where as 13.8% of those who were workings are pregnant [36].

### Place of Residence

The study in Nigeria found that being a female youth from a rural area increased the risk of early sexual activities. Another study also found that women staying in a rural area is more at risk of early sexual experience compared to those staying in urban area

[40]. The other study on the correlates of early sexual debut among sexually active youth in Ghana found that early sexual experience is significantly higher among youth residing in urban areas compared to youth in rural areas [32] Youth in the urban areas maybe more exposed to urban lifestyles including media and Internet activities which may influence their sexual behavior [32].

### Marital Status

The study conducted on determinants of teenage pregnancy in rural Ethiopia based on marital status of teenage, the percentage of teenage pregnancy and motherhood was the highest for those who were married (59.0%) and the lowest for those who were single (1%), while 30% for those who were divorced or widowed [41].

### Religion

According to the national census conducted in 2007, over 32 million people or (43.5%) Ethiopian orthodox Christians, more than 25 million or (33.9% ) Muslim, 13.7 million, or (18.6%), were Protestants, less than two million or (2.6%) adhered to traditional beliefs,0.7% roman catholicism and 0.7% reported to the others [42]. The other research done determinants of age at first sexual intercourse among women in rural Ethiopia religion is statistical significant association age at first sexual intercourse among women and it show that the total women, 35.1% were Orthodox, 42.3% Muslim, 19.6% Protestant, and 3% of them were from other religion followers at the time of the survey [14].

### Regional

The regional variation in sexual behavior underlines the powerful role of environmental factors in shaping behavior and its consequences for sexual health. As of 2008 ,in Nigeria the study conducted on geographical and contextual effect on age of initiation of sexual intercourse among women in Nigeria by using multilevel modeling and spatial analysis region is statistical significant of early sexual debut and north west and north east had the highest proportion of women who had reported early sexual debut (61% to 78%). The proportion of women who had reported early sexual intercourse was lowest in South west, south south, south east, and Kwara State (8% to 25%) [24].

### 2.1.3 Over view of Bayesian multilevel model

There are a number of reasons for using multilevel models. The classical logistic regression analysis treats the units of analysis as independent observations. One consequence of not considering hierarchical structures is that standard errors of regression coefficients will be underestimated, leading to an overstatement of statistical significance and wrong conclusion. Standard errors for the coefficients of higher-level predictor variables will be the most affected by ignoring grouping [43]. According to the study conducted the temperature is depending on where and when it is measured. The study was aimed to test the spatial modeling of annual minimum and maximum temperature in Iceland, the observed data, consisting of measurements of temperature, exhibit a latent dependence structure in the sense that the temperature was 14 dependent on where and when it is measured. It also indicated that including the random effects were significant [44].

In Nigeria the study based on geographical and contextual effect on age of initiation of sexual intercourse among women in Nigeria by using multilevel modeling and spatial analysis, the purpose of this study was to examine the extent of regional and state disparities in age of initiation of sexual intercourse and to examine individual- and community-level predictors of early sexual debut. The variable age, education attainment, ethnicity, region and community median age of marriage are statistical significant association of sexual debut. Multilevel logistic regression models were applied to data on 5531 ever or currently married women who had participated in 2003 Nigeria Demographic and Health Survey and the result is showed that the spatial distribution of age of initiation of sexual intercourse was nonrandom and clustered with a Moran's I = 0.635 (p = .001). There was significant positive spatial relationship between median age of marriage and spatial lag of median age of sexual debut (Bivariate Moran's I = 0.646, (p = .001) [24].

The study based on Bayesian multilevel model on maternal mortality in Ethiopia revealed that educational attainment, wealth index, an age of mother, status and number of living children was a significant factor of maternal mortality and Bayesian multilevel random coefficient was found to be appropriate to fit the data [45]. The association between water pipe dependence and chronic obstructive pulmonary disease, by comparing frequentest and Bayesian methods results show as Bayesian approach have advantages over the frequentest one, particularly in case of a low power of the frequents analysis [46]. The data collection procedure is the hierarchical level or structures that means the levels are nested one another; Kinds of the literature indicated that the Bayesian models are given preference because the technique is more robust and precise than the traditional (classical) statistics since it is usually criticized based on the priors and information from the likelihood. Thus, this collective information has been strengthening the better determination of the parameter [47].

Study conducted in Ethiopia established on Bayesian Multilevel Analysis of Utilization of Antenatal Care Services in Ethiopia revealed that Bayesian multilevel binary logistic regression of random coefficient model factors, place of residence, religion, educational attainment of women, husband educational level, employment status of husband, beat, household wealth index, and birth order were found to be the significant factors for usage of ANC and Bayesian random coefficient model was found better in fitting the data appropriately based upon their deviance information criteria compared to the empty and random intercept model [48].
The study developed on anemia prevalence among children aged 6-59 months in Ethiopia by comparing classical and Bayesian approaches. As the author comparison between Bayesian approach and classical approach results indicated a reduction of standard errors is associated with the coefficients obtained from the Bayesian approach and Bayesian produces precise estimates and more robust compared to the classical [36].

According to the study conducted spatial pattern of perinatal mortality and its determinants in Ethiopia by using multilevel logistic regression models and spatial analysis. The data from 2016 Ethiopia Demographic and Health Survey (EDHS) out of 7230 women who at delivered at seven or more months gestational age nested within 622 enumeration areas (EAs) were used . The result is spatial distribution of perinatal mortality in Ethiopia revealed a clustering pattern. The global Moran's I value was

0.047 with p-value 0.001. Perinatal mortality was positively associated with the maternal age, being from rural residence, history of terminating a pregnancy, and place of delivery, while negatively associated with partners educational level,higher wealth index, longer birth interval, female being head of household and the number of antenatal care (ANC) follow up [49].

A number of efficient algorithms are available for obtaining maximum likelihood (ML) estimates of a multilevel model, for example the iterative generalized least squares procedure (IGLS) or restricted maximum likelihood estimates. Nevertheless, Bayesian methods can implement multilevel models without statistical limitations. Bayesian MCMC methods yield inferences based upon samples from the full posterior distribution and allow exact inference in cases where, as mentioned above, the likelihood based methods yield approximations. Here, we apply a fully Bayesian approach as suggested in [50] which is based on Markov priors and uses Markov Chain Monte Carlo (MCMC) techniques for inference and model checking. There is no established method for determining an appropriate number of iterations and burn-in size. Rather, the researcher use a trial-and-error process in which the ultimate goal is to obtain stable parameter estimates that minimize simulation error. As with the computational intensity this steps require more time on the part of the researcher. However, for choice of the model, we routinely used the DIC developed in [51], as a measure of fit and model complexity.

# 3 Methods of Data Analysis

## 3.1 Description of the Study Area

The study conducted in Ethiopia ($3^0 - 14^0 N$ and $33^0$–$48^0 E$), located at the horn of Africa. Governmentally, the country is divided into nine regional states and two city administrations [52]. It is the second-most populous nation in Africa next to Nigeria. Ethiopia is bordered by Eritrea to the North, Djibouti, and Somalia to the East, Sudan and South Sudan to the West, and Kenya to the South. Ethiopia has eleven geographic or administrative regions: nine regional states (Tigray, Afar, Amhara, Oromia, Somali, Benishangul-Gumuz, SNNPR, and Harari) and two city administrations (Addis Ababa and Dire Dawa) with a capital city of Addis Ababa.

### Source of data

This study was conducted using the data from the 2016 Ethiopia demographic and health survey (EDHS)which was collected from January 18 to June 27, 2016. The datasets used for this study are publicly available on the DHS Program repository https://www.dhsprogram/to all registered users, downloaded with permission. In Ethiopia, this national and subnational representative household survey is conducted every five years.

The sampling frame used for the 2016 EDHS is the Ethiopia Population and Housing Census (PHC), which was conducted in 2007 by the Ethiopia CSA.

The 2016 EDHS is the fourth Demographic and Health Survey conducted in Ethiopia. It was implemented by the CSA at the request of the Ministry of Health (MoH).

### Sampling Technique and Sample Size

The 2016 EDHS was the fourth survey conducted in Ethiopia as part of the worldwide Demographic and Health Surveys project. It was implemented by the Central Statistical Agency (CSA) at the request of the Federal Ministry of Health (FMoH). Data collection took place from January 18, 2016, to June 27, 2016 with national representative of 18,008 households were selected based on a nationally representative sample that provides estimates at the national and regional levels and for urban and rural areas.

The data provide in-depth information on family planning, fertility, marriage, infant, child,adult and maternal mortality, maternal and child health, gender, nutrition, malaria, knowledge of HIV/AIDS and other sexually transmitted diseases.
The 2016 EDHS sample was selected by considering two-stage cluster design and census enumeration areas (EAs) were the sampling units for the first stage. A typical two-level stratification involves first stratifying the population by region at the first level and then by urban-rural within each region. The sample included 645 EAs (202 in urban areas and 443 in rural areas).

### Inclusion and exclusion Criteria

For this study women of reproductive age groups (15-49 years) women who didn't have history of sexual intercourse(virgin) and never married where excluding and women who have at least early sexual intercourse included. To handle missing values we used list wise deletion which is a common approach and easy to perform by deleting all incomplete observations from the analysis. The result was unbiased when data are MCAR [53, 54] missing citation even so,the disadvantage of this method is reduction of sample size. The sample for this study would be consisted of 15683 age at first sex(sexual debut , from which only 11962 of them would be considered in this study.

## 3.2 Methods of Data Analysis

Firstly, descriptive statistics was performed and also assess the patterns in prevalence early sexual debut among women in Ethiopia. The descriptive analysis was performed using count,percentages and Chi-squared test to do early sexual debut status. Furthermore, the statistical significance of the discrete variables are tested Using Chi-square test by considering the research objective by using SPSS 20 software.

Secondly, appropriate Bayesian multilevel logistic regression models were fitted for early sexual debut by using 2016 EDHS data. All models that used Bayesian multilevel estimation approach was implemented within Bayesian framework using in MCMC approaches and using MLwiN 2.02 software and R version 4.0 software. The statistical model that used to analysis the data was Bayesian multilevel Logistic Regression Model. The reason for proposing multilevel approach is that during sample selection in the data collection procedure there is hierarchical level or structures that mean individual women were nested under the Region.

## Response variable

The dependent variable was early sexual debut , categorized dichotomously as "Yes/No" variable. Respondents who were engaged in sexual intercourse before the age of 18 were categorized as "Yes" and those who didn't as "No." Therefore the $i^{th}$ women who are early sexual debut in the $j^{th}$ region is represented by a random variable $Y_{ij}$ ,with two possible values coded as 1 and 0. Hence, the response variable for the $i^{th}$ women in early sexual debut in the $j^{th}$ region is measured as a dichotomous variable.
$Y_{ij}$= 1 if women early sexual debut $j$ region
0 other wise
With $i = 1, 2, 3, ....n$ and $j = 1, 2, 3, ...., k$. Where: $n$- is the number of women who are early sexual debut in each region $j$ and $k-$ is the number of regions.
A two-level model with a binary response ($y$, reported as having started sexual intercourse at an early age) for women who are early sexual debut $i$ living in region $j$ of the form:
$P(Y_{ij} = 1) = \pi_{ij}, P(Y_{_{ij}} = 0) = 1 - \pi_{ij}Y_{ij} \sim$ bernoulli $(1, \pi_{ij})$

## Explanatory variables

The explanatory variable are women's age in years, place of residence, region, religion, marital status, education level, wealth index and working status. See the code of explanatory variable from appendix **A**.

## 3.3 Multilevel Logistic Regression Model

When the data is collected in hierarchical structure multilevel model is appropriate to analyze the data and to see variation between individual and between classes. Multilevel models are used to account for the correlation of observations within a given group by incorporating group specific random effects. These random effects can be nested (women nested within regions, with random effects at the women and region levels) [55].
The main statistical model of multilevel analysis is the hierarchical generalized linear model, an extension of the generalized linear model that includes nested random coefficients. Multilevel/hierarchical modeling explicitly accounts for the clustering of the units of analysis, individuals nested within groups. Such data structures are viewed as a multistage sample from a hierarchical population. The best way to the analysis of multilevel data is an approach that represents within group as well as between groups relations within a single analysis, where group refers to the units at the higher levels of the nesting hierarchy in this case it is region.

Multilevel models makes sense to use probability models to represent the variability within and between groups, in other words, to consider the unexplained variation within groups and the explained variation between groups as random variability. In this study not only unexplained variation of early sexual debut but also unexplained variation between regions was regarded as a random variable. Such variation can be analyzed through statistical models known as random coefficients models. The multilevel logistic regression analysis considers the variations due to hierarchy structure in the data. Hence, the study will help for examination of the effects of regional level and individual level variation of observations [56]. Conventional logistic regression assumes that all experimental units are independent in the sense that any variable which related with early sexual debut has the same effect in all regions, but multilevel models are used to assess whether the effect of predictors vary from region to region.

### 3.3.1 Two level model

In this study, the clustering of the data points within geographical regions offers a natural 2-level hierarchical structure of the data, i.e. women were nested within regions.Let $Y_{ij}$ be the age at first sex status for individual in region , coded 1 or 0 associated with level-one unit, women nested within level two unit region . Also let $\pi_{ij}$ be the probability that the response variable for individual in region j equals 1, and $\pi_{ij} = Pr(Y_{ij} = 1)$. Here, $Y_{ij}$ follows a Bernoulli distribution. Like the logistic regression $\pi_{ij}$ is modeled using the link function,logit. The binary response $Y_{ij}$ is which equals 1 if women i in region j was sexual debut and 0 other wise. Similarly, a j subscript is added to the proportion so that $\pi_{ij} = Pr(Y_{ij} = 1)$. If we have a single explanatory variable, $X_{ij}$, measured at the individual level, then single level is extended to a two-level model as follows:

$$\eta_{ij} = logit(\pi_{ij}) = \beta_{oj} + \beta_1 X_{ij} \tag{1}$$

Where $\beta_{oj} = \beta_o + U_{ij}$

where $i = 1, 2, 3, ...11962$ and $j = 1, 2, 3, ...11$ $U_{ij}$ Is the random effect at regional level. Therefore, conditional on $U_{ij}$ , the $Y_{ij}$s can be assumed to be independently distributed. Here, $U_{ij}$ is a random quantity and follows $N(0, \delta_u^2)$. The outcome variable is dichotomous which is women at early sexual debut and denoted by $i = 1, 2, 3, ...11962$ and $j = 1, 2, 3, ...11$ for level-one unit in group j.

The model (1) is often described as follows.

$logit(\pi_{ij}) = \beta_{oj} + \beta_1 X_{ij}$ ........ [Level 1 model]

. $\beta_{oj} = \beta_o + U_{ij}$ ............ [Level 2 model]

Where: $u_{uj}$ follows normal distribution with mean zero and variance $\delta_o^2$. For models with multiple variables at level-1 or level-2, the above level-1 and level-2 sub models are generalized in an obvious way.

### 3.3.2 Testing heterogeneous proportions

For the proper application of multilevel analysis, the first logical step is to test heterogeneity of proportions between groups. To test whether there are indeed systematic differences between the groups the well-known chi-square test for contingency table can be used. In this case the chi-square test statistics is given by:-

$$\chi^2 = \sum_{j=1}^{N} n_j \frac{(\bar{y_{.j}} - p)^2}{p.(1-p.)} \sim \chi^2(N-1) \tag{2}$$

- Where $\bar{y_{.j}}$ :- is group average, obtained as $\bar{y_{.j}} = \frac{1}{n_j}\sum_{i=1}^{nj} y_{ij}$ is the proportion of successes in region j

which is an estimate for the group-dependent probability$p_j$ and $n_j$is the number of observations,N is the number of groups.

- p :- Is the overall average $p_. = \bar{y_{..}} = \frac{1}{n_j}\sum_{j=1}^{g}\frac{1}{n_j}\sum_{i=1}^{nj} y_{ij}$ is the overall proportion of successes.
- The decision is based on chi-square distribution with N-1 degrees of freedom.

**Empty Model**

The empty two-level model for a binary outcome variable refers to a population of groups (level-two units (regions)) and specifies the probability distribution for group dependent probabilities Pj in $Y_{ij} = P_j + \epsilon_{ij}$ without taking further explanatory variables into account. Here, the logit transformed model, $logit(\pi_{ij})$ can have the normal distribution. Consequently, the empty model can possibly be expressed in the form of the following formula:

$$logit(\pi_{ij}) = \beta_0 + U_{oj} \tag{3}$$

In the equation above, $\beta_0$ indicates the population average of the transformed probability and $U_{oj}$ is the random deviations from this average for region j. The residual term that is associated with the group dependent deviations, $U_{oj}$ has a unique effect of region j on the response variable; and it is assumed to be normally and independently distributed with mean zero and variance ,$\delta_o^2$ that is $U_{oj} \sim N(o, \delta^2)$ . In this situations, the level 2 residual can possibly capture the variations across regional means. In this model, the amount of variance regarding early sexual debut that is attributable within group characteristics (women) and between group difference (region) can be investigated.

The above equation is does not include a separate parameter for the level one variance. The reason is the level one residual variance of binary outcome variable follows directly the success probability indicated as follow: $var(\epsilon_{ij}) = \pi_{ij}(1 - \pi_{ij})$ are early sexual debut dependent residuals. The other reason of applying multilevel analysis is the existence of intra-class (intraregional) correlation arising from similarity of early sexual debut in the same region compared to those of different regions. The intra-class correlation coefficient (ICC) measures the proportion of variance in the outcome explained by the grouping structure. ICC can be calculated using an intercept-only model or an empty model. The ICC can be calculated as:

$$ICC = \frac{\delta_u^2}{\delta_u^2 + \delta_\epsilon^2} \tag{4}$$

Where $\delta_u^2$ is the variance between the group which can be estimated by $U_{oj}$ and $\delta_\epsilon^2$ is within-group variance [57] and follows a logistic distribution with variance $\frac{\pi^2}{3} \approx 3.29$ [58]. Denote $\pi_o$ the probability corresponding to the average values $\beta_o$ as defined by $P(\pi_o) = \beta_o$ for the logit function, the so-called logistic transformation of $\beta_o$ , is defined as:

$$\pi_o = logit(\beta_o) = \frac{e^{\beta_o}}{1 + e^{\beta_o}} \tag{5}$$

Note that due to the non-linear nature of the logit link function, there is no simple relation between the variance of the deviations$U_{oj}$. However, there is an approximate formula which is valid when the variances are small and is given by:
$var(\pi_j) = (\pi_o(1 - \pi_o))^2\delta_o^2$
Note that an estimate of population variance $var(\pi_j)$ can be obtained by replacing sample estimates of $\pi_o$ and $\delta_o^2$ The resulting approximation can be compared with the nonparametric estimate:

$$T^2 = S_{between}^2 - \frac{S_{within}^2}{n} \tag{6}$$

HYPOTHESIS

$H_O$ = There is no regional variation in early sexual debut among women in Ethiopia.
$H_1$ = There is regional variation in early sexual debut among women in Ethiopia

### 3.3.3 Analysis of Random Intercept Model

The random intercept model is used to model unobserved heterogeneity in the overall response by introducing random effects. In the random intercept model the intercept is the only random effect meaning that the groups differ with respect to the average value of the response variable, but the relation between explanatory and response variables cannot be differ between groups. The random intercept model expresses the log odds, i.e the logit of$\pi_{ij}$ , as a sum of linear functions of the explanatory variables. That is,

$$logit(\pi_{ij}) = log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} \tag{7}$$

i=1,2,.....,n, j=1,2,....,11 Where the intercept term $\beta_{oj}$ is assumed to vary randomly and is given by the sum of an average intercept $\beta_o$ and group-dependent deviations; $\beta_{oj} + U_{oj}$ . As a result we have:

$$logit(\pi_{ij}) = \beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj} \tag{8}$$

solving for

$$\pi_{ij} = \frac{e^{\beta_{oj}+\sum_{h=1}^{k}\beta_h X_{hij}+U_{oj}}}{1+e^{\beta_{oj}+\sum_{h=1}^{k}\beta_h X_{hij}+U_{oj}}} \tag{9}$$

The above equation is does not include a level one residual because it is an equation for the probability$\pi_{ij}$ rather than for the outcome $Y_{ij}$, where $\beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj}$ is the fixed part of the model. The remaining $U_{oj}$ is called the random or the stochastic part of the model. It is assumed that the residual $U_{oj}$ is mutually independent and normally distributed with mean zero and variance $\delta_o^2$ [59].

### 3.3.4 Random Coefficients Model (full model)

Random coefficient logistic regression is based on linear models for the log-odds that include random effects for the groups or other higher level units. The random coefficients build upon the random intercept model by allowing the effects of individual predictors to vary randomly across level 2, that is, level 1 slope coefficients are allowed to take on different values in different aggregate groups. In the random coefficient model both the intercepts and slopes are allowed to differ across the regions. Consider a model with group-specific regression of logit of the success probability $logit(\pi_{ij})$ on a single level -one explanatory variable X:

$$logit(\pi_{ij}) = \beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj} + \sum_{h=1}^{k} U_{hj} X_{hij} \tag{10}$$

The term $\sum_{h=1}^{k} U_{hj} X_{hij}$ can be regarded as a random interaction between group and the explanatory variables. This model implies that the groups are characterized by two random effects: their intercepts and their slopes. It assumes that for different groups, the pairs of random effects $(U_o, U_h, h = 1, 2, 3, ..., k)$ are independent and identical distributed. The random intercept variance, $var(U_{oj}\delta_o^2)$ , the random slope variance, $var(U_{1j}\delta_1^2)$ and the covariance between the random effects $(cov(U_{oj})U_{1j} = \delta_{oj})$ are called variance components [59].

## Estimation of between and within-group variance

Consider a population having two-levels, the basic data structure of two-level logistic regression analysis is a collection of N groups (units at level-two (region)) and within group $j (j = 1, 2, 3, ..., N)$ a random sample of nj level-one units (individual females).The outcome variable is dichotomous and denoted by $y_{ij}, (i = 1, 2, 3, ....11962, j = 1, 2, 3, ..., 11)$ for women in region j. The total sample size is $M = \sum_{j=1}^{N} n_j$ Then, the theoretical variance between the groups (region) dependent probabilities, i.e., the population value of $var(p_j)$, can be estimated by:

$$\hat{T}^2 = s^2 between - \frac{s^2 within}{\hat{n}} \tag{11}$$

Where $\hat{n} = \frac{1}{n-1}(M - \frac{\sum_{j=1}^{n} n_j^2}{M}) = n - \frac{\hat{s}^2(n_j)}{Nn}$
For dichotomous dependent variable, the observed between- groups variance is closely related to the Chi squared test statistic [56]. They are given by the formula:

$$s^2 between = \frac{p(1-\hat{p})}{\hat{n}N - 1}\chi^2 \tag{12}$$

Where, $\chi^2$is as given by equation (2), and the within- group variance in dichotomous case is a function of the group:

$$s^2 within = \frac{1}{m-1} \sum n_j p_j (1 - p_j) \tag{13}$$

where $p_j$ is the proportion of successes in group j and $p$ is the overall proportion of successes, $n_j$ is the sampled observation in group j, $\bar{n}$ is the average of the sampled observation in groups, M is the total sampled observations, and N is the number of groups (in this case 9 regions and two administrative). Multilevel logistic regression analysis can be employed in the simplest case without explanatory variables,(usually called the empty model) and also with explanatory variables by allowing only the intercept term or both the intercept and slopes (regression coefficients) to vary randomly. In this study, multilevel logistic regression model is considering the data to be analyzed on the case of two-levels, individual women is considered as level 1 and region is considered as level 2 [56].

## 3.4 Bayesian multilevel logistic regression model

Statistical model that used for this data to analysis was the Bayesian multilevel logistic model. The Markov chain Monte-Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest [60]. Sampling process of MCMC approaches is pretty heavy but has no bias and, so, these methods are preferred when accurate results are expected, without regards to the time it takes [61]. In this study a Bayesian multilevel logistic regression approach for binary outcomes was preferred, which takes into account the hierarchical structure of data and properly estimates the parameters and accuracy intervals [62].

Bayesian multilevel logistic analysis procedure was used to make inference about the parameters of a multilevel logistic model. The Bayesian method gives estimates of parameters by sampling them from their posterior distributions through an MCMC method. This approach was employed to model early sexual intercourse among women in Ethiopia. The metropolis Hasting algorithm were implemented using non-informative uniform prior distribution with scale parameter (0, 1) for the fixed effects and inverse gamma distribution with a scale of 0.001 and shape 0.001 for random effect and inverse wishart for random part [57]. The data used in this study have a hierarchical structure. Here, the level-1 units are women and the level-2 units are the regions that constitute the groups into which the women are clustered or nested.The multilevel logistic regression model is appropriate for research designs where data for respondents are organized more than one level i.e, nested data.

A multilevel logistic regression model can account for lack of independence across levels of nested data (i.e., individuals nested within regions).For simplicity of presentation two-level models for this study, i.e., models accounting for women-level and regional -level effects. In this data structure,level-1 is the women level and level-2 is the regional level. Within each level-2 unit, there is nj in the $j^{th}$ region. Burnoulli distribution is used. Then burnoulli distribution is given by
$\pi_{ijk} : Y_{ijk} \sim burnoulli(1, \pi_{ijk})$
The probability was related to a set of categorical predictors, X; and a random effect for each level, by a logit-link function as.

$$logit(\pi_{ijk}) = (log\frac{\pi_{ijk}}{1-\pi_{ijk}}) = \beta_{oj} + \sum_{i=1}^{h} \beta_{hi} X_{ijk} \tag{14}$$

where $X_{ijk=} = X_{1jk}, X_{2jk}, ..., X_{ijkp}$ represents the first and the second level of covariates for variable h, The linear predictor on the right-hand side of the equation consisted of a fixed part $\beta_o + \beta X_{ijk}$ estimating the conditional coefficients for the explanatory, and two random intercepts attributable to communities $U_{ojk}$ with each assumed to have an independent and identical distribution and variance estimated at each level.

The analysis was done in three steps. In Model 1 (empty model), no explanatory variable was included. In Model 2, only individual-level factors were included. In Model 3, community contextual factors were added to Model 2. The results of fixed effects (measures of association) were shown as odds ratios (ORs) with 95% confidence intervals (CIs). The results of random effects (measures of variation) were presented as variance partition coefficient and percentage change in variance.

### 3.4.1 Likelihood Function

Statistical inferences are usually based on maximum likelihood estimation (MLE). MLE chooses the parameters that maximize the likelihood of the data, and is intuitively appealing. In MLE, parameters are assumed to be unknown but fixed and are estimated with some confidence. In Bayesian statistics, the uncertainty about the unknown parameters is quantified using probability. So that, the unknown parameters are regarded as random variables.

The likelihood function used in the Bayesian approach is equivalent to that of the classical inference. The joint distribution of n independent Bernoulli trials is the product of each Bernoulli densities, where the sum of independent and identically distributed Bernoulli trials has a Binomial distribution. Specifically, let $Y_{1j}, Y_{2j}, ..., Y_{ij}$ be independent Bernoulli trials with success probabilities$\pi_{1j}, \pi_{2j}, ..., \pi_{ij}$ that is $Y_{ij} = 1$ (the probability of having sexual intercourse for $i^{th}$ women in the $j^{th}$ th region) and also $1 - \pi_{ij}$ (failure probabilities) is the probability of i th women not having sexual intercourse in the $j^{th}$ region,for i = 1, 2, ..., n and j = 1, 2, ...11. Since, the trials are independent, the joint distribution of $Y_{1j}, Y_{2j}, ..., Y_{ij}$ is the product of n Bernoulli probabilities.

The probability of success in logistic regression varies from one subject to another, depending on their covariates. Thus, the likelihood function is illustrated below as product of n Bernoulli trials:

$$L(\frac{\pi_{ij}}{Y_{ij}}) = \Pi_{ij}(\pi_{ij})^{Yij}(1 - \pi_{ij})^{1-Yij} \tag{15}$$

and the linear predictor or the logit functions is:

$$logit(\pi_{ij}) = log(\frac{\pi_{ij}}{1 - \pi_{ij}}) = \beta_{oj} + \sum_{h=1}^{k} \beta_{hj} X_{hij} + U_{oj} + \sum_{h=1}^{k} U_{hj} X_{hij} \tag{16}$$

Where :

$$\pi_{ij} = \frac{e^{\beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj} + \sum_{h=1}^{k} U_{hj} X_{hij}}}{1 + e^{\beta_{oj} + \sum_{h=1}^{k} \beta_h X_{hij} + U_{oj} + \sum_{h=1}^{k} U_{hj} X_{hij}}} \tag{17}$$

$\pi_{ij}$ represents the probability of the event for subject ij who has covariate vector $X_{ij}, Y_{ij} = 1$, indicates the presence (women having sexual intercourse before 18 years ) and $Y_{ij} = 0$ the absence (women having sexual intercourse after 18 years) of the event for the given subject.

### 3.4.2 Prior distribution

The prior distribution is a probability distribution that represents the prior information associated with the parameters of interest. It is a key aspect of a Bayesian analysis. There are two types of prior distribution: **Informative priors** and **Non-informative priors**.

An informative prior is a prior distribution that is used when information about the parameter of interest is available before the data is collected. Typically, informative prior distributions are created from historical studies, pure expert knowledge (experience) and a combination of both. Even if there is prior knowledge about what we are examining, in some cases we might prefer not to use this and let the data speak for themselves.

In this study, we wish to express our prior ignorance in to the Bayesian system. This leads to non-informative priors. A non-informative prior distribution that is used to express complete ignorance of the value before the data is collected. In this study, the prior distributions for fixed effect parameter was $P(\beta) \sim$ Uniform distributions and for random effect terms was, $P(\frac{1}{\delta_u^2}) \sim gamma(\alpha, \beta)$ where $\alpha$ and $\beta$ are scale (0.001) and shape (0.001) parameters which are fixed constant or we can write $P(\beta_0) \propto 1$ and $P(\delta_{uo}^2) \propto$ inverse Gamma $(\alpha, \beta)$ where $\alpha$ and $\beta$ are shape (0.001) and scale (0.001) parameter which is fixed constant.

Let us denote the parameters $\beta_o, \beta_1, ..., \beta_k$ and $\Omega_u$ as prior distributions would be given as follows; $P(\beta_o) \propto 1, P(\beta_1) \propto 1, ..., P(\beta_k) \propto 1$, and $p(\Omega_u) \propto$ inverse wishart $(m * S_u, v)$ distribution.

The parameter $\Omega_u$ is the variance-covariance matrices and Su is an estimate for the true value of $\Omega_u$ and V is the number of row in the variance-covariance matrix.
The Wishart distribution is the multivariate extension of the gamma distribution; although most statisticians use the Wishart distribution in the special case of integer degrees of freedom, in which case it simplifies to a multivariate generalization of the $X^2$ distribution [63]. As the distribution $X^2$ describes the sums of squares of n draws from a univariate normal distribution, the Wishart distribution represents the sums of squares (and cross-products) of n draws from a multivariate normal distribution.

### 3.4.3 Posterior Distribution

It is obtained by multiplying the prior distribution over all parameters by the full likelihood function. All Bayesian inferential are based on the posterior distribution of the model generated. The inference is performed by sampling from posterior distribution until the convergence to the posterior distribution is achieved [64]. The major problem in the Bayesian approach is that in most cases the full form of the posterior distribution cannot be obtained in closed form, that is, the posterior density may not belong to standard distribution. Such problem cannot be solved easily. In order to solve such problems the researcher used MCMC simulations. The Markov chain Monte-Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest [60].
The full conditional distribution for parameter $\beta_0$ is:

$$P(\beta_0/\delta_{u0}^2, Y_{ij}) \propto \Pi_j \pi_{ij}^{Yij}(1-\pi_{ij})^{1-y_{ij}} \propto \Pi_j[(\frac{e^{\beta_0+U_{oj}}}{1+e^{\beta_0+U_{oj}}})^{Y_{ij}}(\frac{1}{1+e^{\beta_0+U_{oj}}})^{(1-Yij)}]^1 \quad (18)$$

For parameter $\delta_{uo}^2$ the full conditional distribution is:
$P(\delta_{uo}^2/\beta_0, Y_{ij}) \propto \Pi_j(\pi_{ij})^{Yij}(1-\pi_{ij})^{1-Yij}*$ inverse gamma $(\frac{n}{2}+n(\alpha-1), n\beta)$

$$P(\delta_{uo}^2/\beta_0, Y_{ij}) \propto \Pi_j(\pi_{ij})^{Yij}(1-\pi_{ij})^{1-Yij} * \frac{\beta^\alpha}{\Gamma(\alpha)} X^{-\alpha-1} e^{-\beta/x} I(x>0) \quad (19)$$

where **n** total numbers
Estimating $\beta$ of the posterior distribution may be difficult, for this reason, we need to use the non-analytic method such as simulation techniques. The most popular method

of simulation technique is Markov Chain Monte Carlo (MCMC) methods.

### 3.4.4 MCMC Methods

Bayesian estimation was performed using the software MLwiN a specialized program for performing multilevel analysis. MLwiN can carry out iterative generalized least squares or restricted iterative generalized least squares estimation as initial value for MCMC estimations [65]. The Markov chain Monte Carlo (MCMC) method is a general simulation method for sampling from posterior distributions and computing posterior quantities of interest. With the MCMC method, it is possible to generate samples from an arbitrary posterior density p ( ./y) and use these samples to approximate expectations of quantities of interest. In Bayesian statistics, there are two MCMC algorithms that the researcher can use: the Gibbs sampler and Metropolis hasting algorithm. While for comparing them, Metropolis hasting estimation methods used for both fixed and random effects [66].

The Metropolis–Hastings algorithm works by generating a sequence of sample values Bayesian inference is solved by randomly drawing a very large sample from the posterior distribution. The idea of drawing a large sample from the posterior distribution is called Markov Chain Monte Carlo using MCMC techniques. The Bayesian approach applies probability theory to a model derived from substantive knowledge and theory, deal with realistically complex situations; the approach can also be termed full probability modeling. There has recently been huge progress in methods for Bayesian computation, generally exploiting modern computer power to carry out simulations known as Markov Chain Monte Carlo (MCMC) methods. The MCMC simulation is used to do the integration numerically rather than analytically by sampling from the posterior distribution of interest even when the form of that posterior has no known algebraic form [67].

#### Metropolis-Hastings Algorithm

Metropolis-Hasting algorithm is an iterative algorithm that produces a Markov chain and permits empirical estimation of posterior distributions. Therefore, in this study metropolis-hasting algorithm were used to estimate the fixed and the random effects parameters for prevalence of sexual intercourse among women. The Metropolis-Hasting algorithm (MH) generates samples from a probability distribution using full joint density function. Metropolis-Hastings algorithm correctly applied for non-Gaussian data and if the posterior distribution doesn't follow some known distribution [60].

The metropolis-Hasting Algorithm follows the following steps:

1) Establish starting values S for the parameter:$\theta^{j=0} = S$ set j = 1

The starting values can be obtained via maximum likelihood estimation.

2) Draw a candidate parameter,$\theta^c$ from a proposal density $\alpha(.)$.

The simulated value is considered a candidate because is not automatically accepted as a draw from the distribution of interest. It must be evaluated for acceptance.

3) Compute the ratio $\frac{f(\theta^c)\alpha(\theta^{j-1}|\theta^c)}{f(\theta^{j-1})\alpha(\theta^c|\theta^{j-1})}$

4) Compute R with a $U(0,1)$ random draw u. If R less than u, then set $\theta^j = \theta^c$ Otherwise, set $\theta^j = \theta^{j-1}$

5) Set j = j + 1 and return to step 2 until enough draws are obtained. Once convergence is reached, all simulation values are from the target posterior distribution and a sufficient number will be drawn so that all areas of the posterior will be also explored.

## 3.5 Model selection and comparison

Model selection is to select the best model among several choices based on an evaluation of the performance of the models. A widely used statistic for comparing models in a Bayesian framework is the Deviance Information Criterion. In Bayesian, the lowest expected deviance has the highest posterior probability. Assessing goodness of fit involves investigating how close the values are predicted by the model with that of observed values [68]. The deviance information criterion (DIC) is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion, also known as the Schwarzcriterion).

It is particularly useful in Bayesian model selection problems where the posterior distributions of the models have been obtained by Markov chain Monte Carlo (MCMC) simulation. Like AIC and BIC it is an asymptotic approximation as the sample size becomes large. It is only valid when the posterior distribution is approximately multivariate normal. Define the deviance as $D(\theta) = -2log(P(Y|\theta)) + c$, where y are the data, $\theta$ are the unknown parameters of the model and $P(Y|\theta)$ is the likelihood function. C is a constant that cancels out in all calculations that compare different models, and which therefore does not need to be known. The expectations $\bar{D} = E[D(\theta)]$ is a measure of how well the model fits the data; the larger this is, the worse the fit.The effective number of parameters of the model is computed as $P_D = \bar{D} - D(\theta)$ ,where. $\theta$ is the expectations of $\theta$. The larger this is, the better it is for the model to fit the data.The deviance information criterion can be described as:

$$DIC = D + pD \tag{20}$$

The idea is that models with smaller DIC should be preferred than models with larger DIC. Models are penalized both by the value of $\bar{D}$ ,which favors a good fit, but also (in common with AIC and BIC) by the effective number of parameters pD. Since $\bar{D}$ will decrease as the number of parameters in a model increases, the pD term compensates for this effect by favoring models with a smaller number of parameters. The advantage of DIC over other criteria, for Bayesian model selection, is that the DIC is easily calculated from the samples generated by a Markov chain Monte Carlo simulation. AIC and BIC require calculating the likelihood at its maximum over $\theta$, which is not readily available from the MCMC simulation. But to calculate DIC, simply compute $\bar{D}$ as the average of $D(\theta)$ over a sample values of $\theta$, and $D(\bar{\theta})$ as the value of D evaluated at the average of the samples of $\theta$. Then the DIC follows directly from these approximations.

## 3.6 Tests for Convergence(Model diagnostic)

It is important to establish whether a sequence of Markov chain Monte Carlo iterations has converged, that is, reached its stationary distribution. To examine the convergence of MCMC, considering a different method will be useful for detecting poorly sampled Markov Chains. Among several ways of a test of convergence, the most popular and straight forward convergence assessment methods was used for this study. The most common ways of checking goodness of fit are: diagnosis for convergence and mixing and posterior-predictive check. We have used the following in our study for convergence tests for the variables are Time series plots, kernel density plot, Monte Carlo Standard Error (MCSE), the efective of sample size (ESS) and Partial Autocorrelation Function (PACF). The following methods are more likely to be considered for this study:

**Monte Carlo Standard Error:** Small values of the MC error indicate that the quantity of interest has been calculated with precision.
**Autocorrelation:** High correlation between the parameters of a chain tends to give slow convergence, where as high autocorrelation within a single parameter chain leads to slow mixing and possibly individual non convergence to the limiting distribution because the chain will tend to explore less space in finite time. That is, low or high values indicate fast or slow convergence, respectively. In analyzing Markov chain autocorrelation, it is helpful to identify lags in the series in order to calculate the longer-run trends in correlation, and in particular whether they decrease with increasing lags.

**Time series plots or trace plots:** Time series plots (iteration number on X axis and parameter value on y-axis) are commonly used to assess convergence [69]. If the plot looks like a horizontal band, with no long upward or downward trends, then we have evidence that the chain has converged.
**Density plot:** The plots of all statistically significant covariates indicated that none of the coefficients have bimodal density and hence the simulated parameter values have converged.
**The Effective Sample Size:** The Effective Sample Size is a measure of efficiency that provides an estimate of the equivalent number of independent observations that are contained in the chain; this will of course be directly related to the degree of autocorrelation or dependence in the sequence for that parameter. A related concept to the MCMC convergence would be the inefficiency factor which is useful to measure the efficiency of the MCMC sampling algorithm. It is given as:-
Inefficiency factor $= 1 + 2\sum_{k=1}^{\infty} \rho(k)$ ,where $\rho(k)$ is the sample autocorrelation at lag

k calculated from sample draws. A large value of efficiency factor indicates that we need large MCMC iteration. The effective sample size, the number of MCMC output divided by the inefficiency factor. Convergence of posterior estimate has been checked using an effective sample size that is all the effective sample sizes of the estimates are greater than 200. So, the more samples saved, the more accurate would be in posterior estimates. This is an indication of efficient posterior estimate. Closely related measure of mixing is effective sample size [70].

Let the output of MCMC denoted by N, then it can be given as:

Effective sample size

$$ESS = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho(k)} \tag{21}$$

**Assessing Model Accuracy:** After model convergence would be achieved, we need to run the simulation for a further number of iterations to obtain samples that can be used for posterior inference. The more samples we save, the more accurate would be our posterior estimates. One way of assess the accuracy of the posterior estimates is by calculating the Monte Carlo standard error for each parameter. This is an estimate of the difference between the mean of the sampled values (which we are using as our estimate of the posterior mean for each parameter) and the true posterior mean. As a rule of thumb, the simulation should be run until the Monte Carlo error for each parameter of interest is less than about 5% of the sample standard deviation.

# 4 Results and Discussion

## 4.1 Descriptive Summary

From the total of 11962 women 8997(75.2%) were below the age of 18 and the rest 2665(24.8%) were above 18 years was early sexual debut among women in Ethiopia respectively in (Figure 4.1).

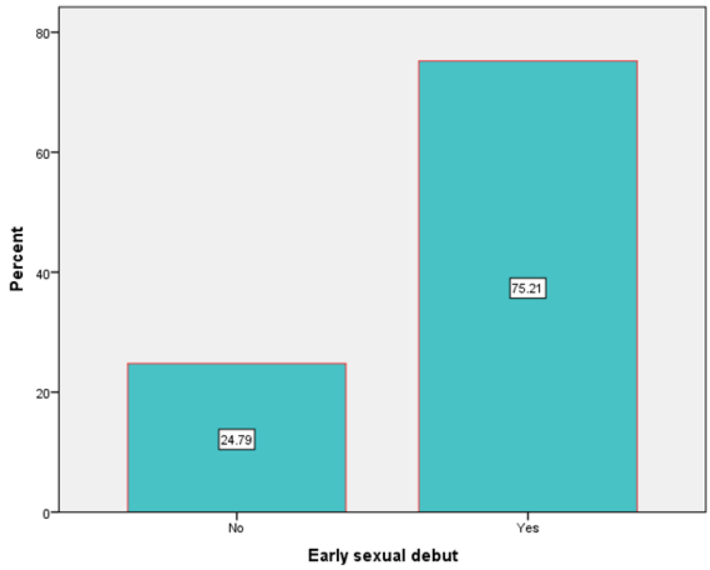



Figure 4.1: Bar char for proportion of early sexual debut in Ethiopia

As test of association states women age in years is important characteristic of the population under study and the descriptive result depicts that from a total of 3076 women age group 15-24, about 81.8% where experienced early sexual debut, from a total of 4752 women age group 25-34, about 70.8% where experienced early sexual debut and from a total of 4134 women age group 34+, about 75.4% where experienced early sexual debut. Education level another important characteristic of the population under study and the descriptive result depicts that from a total of 6547 no education level, about 81.6% where experienced early sexual debut, from total of 3388 primary education level, about 78.1% where experienced early sexual debut, from total of 1238 secondary education level, about 58.1% where experienced early sexual debut and from total of 789 higher education level,about 36.6% where experienced early sexual debut.

Place of residence of women is also another important characteristic of the population under study and the descriptive result depicts that from a total of 3621 women living in urban area, about 61.4% where experienced early sexual debut and from a total of 8341 women living in rural area, about 81.2% where experienced early sexual debut. Region is important characteristic of the population under study and the descriptive result depicts that from a total of 1285 Tigray region, about 80.2% where experienced early sexual debut, from a total of 973 about 86.2% where experienced early sexual debut, from a total of 1389 Amhara region about 87.2% where experienced early sexual debut, from a total of 1494 Oromia region about 78.1% where experienced early sexual debut, from a total of 1090 Somali region about 69.4% where experienced early sexual debut, from a total of 905 Benishangul region about 81.5% where experienced early sexual debut, from a total of 1320 SNNPR region about 70.2% where experienced early sexual debut, from a total of 882 Gambela region about 79.7% where experienced early sexual debut, from a total of 704 Harari region about 71.7% where experienced early sexual debut, from a total of 1114 Addis Ababa city administration about 50.5% where experienced early sexual debut and from a total of 806 Dire Dawa city administrative about 69.2% where experienced early sexual debut.

Wealth index is also another important characteristic of the population under study and the descriptive result depicts that from a total of 3392 poorest about 80.1% where experienced early sexual debut, from a total of 1678 poorer about 82.9% where experienced early sexual debut, from a total of 1527 middle about 81.8% where experienced early sexual debut, from a total of 1502 richer about 79.2% where experienced have early sexual debut and from a total of 3861 richest about 63.3% where experienced early sexual debut.Respondent occupation is important characteristic of the population under study and the descriptive result depicts that from a total of 5998 had didn't have a job about 76.6% where experienced early sexual debut and from a total of 5964 had have a job about 73.8% where experienced early sexual debut.

Marital status is also another important characteristic of the population under study and the descriptive result depicts that from a total of 566 single about 48.9% where experienced early sexual debut, from a total of 9599 married about 76.5% where experienced early sexual debut, from a total of 219 living with partner about 63.9% where experienced early sexual debut, from a total of 451 widowed about 78.5% where experienced early sexual debut, from a total of 875 divorced about 79.9% where experienced early sexual debut and from a total of 252 separated about 71.0% where experienced early sexual debut. The study explored percentages of early age of initiation for sexual intercourse across regions of Ethiopia and accordingly, early sexual debut prevalence is highest in Amhara region (87.2%) and followed by Afar region (86.2%) and Benishangul region (81.5%) whereas those Addis Ababa city administrative shares smallest proportion (50.5%).

Table 4.1: Individual and community characterestics of respondents EDHS(n=11962)

| Variables | Early sexual debut | | |
|---|---|---|---|
| | No(%) | Yes(%) | Total (%) |
| Women Age in Years | | | |
| 15-24 | 561(18.2) | 2515(81.8) | 3076(25.7) |
| 25-34 | 1389(29.2) | 3363(70.8) | 4752(39.7) |
| 34+ | 1015(24.6) | 3119(75.4) | 4134(34.6) |
| Education | | | |
| No education | 1205(18.4) | 5342(81.6) | 6547(54.7) |
| Primary | 741(21.9) | 2647(78.1) | 3388(28.3) |
| Secondary | 519(41.9) | 719(58.1) | 1238(10.3) |
| Higher | 500(63.4) | 289(36.6) | 789(6.6) |
| Religion | | | |
| Orthodox | 1275(26.7) | 3495(73.3) | 4769(39.9) |
| Catholic | 20(29.4) | 48(70.6) | 68(0.6) |
| Protestant | 572(27.4) | 1514(72.6) | 2086(17.4) |
| Muslim | 1061(21.6) | 3840(78.4) | 4901(41.0) |
| Traditional | 18(24.0) | 57(76) | 75(0.6) |
| Other | 19(30.2) | 44(69.8) | 63(0.5) |
| Type of place of residence | | | |
| Urban | 1396(38.6) | 2225(61.4) | 3621(30.3) |
| Rural | 1569(18.8) | 6772(81.2) | 8341(69.7) |
| Region | | | |
| Tigray | 255(19.8) | 1030(80.2) | 1285(10.7) |
| Afar | 134(13.8) | 839(86.2) | 973(8.1) |
| Amhara | 178(12.8) | 1211(87.2) | 1389(11.6) |
| Oromia | 327(21.9) | 1167(78.1) | 1494(12.5) |
| Somali | 334(30.6) | 756(69.4) | 1090(9.1) |
| Benishangul | 167(18.5) | 738(81.5) | 905(7.6) |
| SNNPR | 393(29.8) | 927(70.2) | 1320(11.0) |
| Gambela | 179(20.3) | 703(79.7) | 882(7.4) |
| Harari | 199(28.3) | 505(717) | 704(5.9) |
| Addis Ababa | 551(49.5) | 563(50.5) | 1114(9.3) |
| Dire Dawa | 248(30.8) | 558(69.2) | 806( 6.7) |
| Wealth Index | | | |
| Poorest | 679(19.9) | 2719(80.1) | 3394(28.4) |
| Poorer | 287(17.1) | 1391(82.9) | 1678(14.0) |
| Middle | 278(18.2) | 1249(81.8) | 1527(12.8) |
| Richer | 307(20.4) | 1195(79.6) | 1502(12.6) |
| Richest | 1418(36.7) | 2443(63.3) | 3861(32.3) |
| Respondents occupation | | | |
| Not Working | 1402(23.4) | 4596(77.6) | 5998(50.1) |
| Working | 1563(26.2) | 4401(73.8) | 5964(49.9) |
| Current marital Status | | | |
| Single | 289(51.1) | 277(48.9) | 566(4.7) |
| Married | 2251(23.5) | 7348(76.5) | 9599(80.2) |
| Living with partner | 79(36.1) | 140(63.9) | 219(1.8) |
| Widowed | 97(21.5) | 354(78.5) | 451(3.8) |
| Divorced | 179(20.1) | 699(79.9) | 875(7.3) |
| Separated | 73(29.0) | 179(71) | 252(2.1) |
| Early sexual Debut | | | |
| NO | | | 2965(24.8) |
| Yes | | | 8997(75.2) |

The bivariate association between early sexual debut status of women and predictors were showed in Table 4.2 indicates that early sexual debut was strongly associated with Women age in years , education, place of residence, region, wealth index, respondent occupation and marital status were found to have a significant association with to early sexual debut at the 5% significance level.

Table 4.2: Chi-square association between early sexual debut and predictor variables

| Variable names | Chi-square value | D.f | P-value |
|---|---|---|---|
| Intercept | 22.685 | 1 | 0.000* |
| Women Age | 168.177 | 2 | 0.000* |
| Education | 390.854 | 3 | 0.000* |
| Religion | 4.130 | 5 | 0.531 |
| place of residence | 11.740 | 1 | 0.001* |
| Regions | 235.112 | 10 | 0.000* |
| Wealth Index | 10.134 | 4 | 0.038* |
| Respondent occupation | 4.445 | 1 | 0.035* |
| Marital status | 61.553 | 5 | 0.000* |

*** indicates a significant variable at 5% level of significance**

## 4.2 Test of Heterogeneity

Before analyzing multilevel data, one has to test the heterogeneity of early sexual debut status among regional states of Ethiopia from which essential clues would be obtained for incorporating the random effects. Therefore, the Pearson chi-square for the proportion of sexual intercourse among women across the region has been investigated in the Table 4.3 Consequently, as it obtained by cross tabulations the Pearson Squareness ($\chi^2$) = 242.276 with 10 degrees of freedom.

The P-value is less than 0.05 level of significance, implying strong evidence of heterogeneity for the early sexual debut among women across regional states of Ethiopia. Hence we have enough evidence to reject the null hypothesis and conclude that there is heterogeneity status of sexual intercourse among women in regional states of Ethiopia.

Table 4.3: Heterogeneity of sexual intercourse among women between regional states of Ethiopia

| Chi-square test | | | |
|---|---|---|---|
| Statistics | Value | D.f | P-value |
| Pearson Chi-square | 242.276 | 10 | 0.000 |
| N of valid cases 11962 | | | |

## 4.3 Analysis based on empty model

The simplest important specification of the hierarchical linear model is a model in which only the intercept varies between level two units and no explanatory variables are entered in the model. The empty model contains no explanatory variables and it can be considered as a parametric version of assessing heterogeneity of early sexual intercourse among women in the regions. The variance of the random factor is significant which indicates that there are regional differences in early sexual debut.

From Table 4.4, the overall posterior mean of early sexual debut without incorporating the covariate is estimated to be $\beta_{oj} = 1.250$ and the between region (level two) variance of early sexual debut of women is estimated as $\delta^2_{oj} = 0.727$ which is found to be significant because the credible interval of the respective parameters was greater than zero. Indicating the variations of early sexual debut of women among regional states of Ethiopia. Here the null hypothesis tested is $\delta^2_{oj}= 0$. i.e., there is no regional variation in the incidence of early sexual debut of among women in Ethiopia. Based on the Table 4.4 result, the values are significant at 95% credible interval, which means that the interval is greater than zero, therefore, the null hypothesis had been rejected indicating strong evidence that the between region variance is greater than zero. The variance of the random factor is significant which indicates that there was regional differences in prevalence of early sexual debut among women in Ethiopia.

Table 4.4: **Posterior summaries for parameters of the empty model**

| Model | Post. Mean(Coefficient) | SD | p-value | 2.5% | 50% | 97.5% |
|---|---|---|---|---|---|---|
| Fixed intercept($\beta_{oj}$) | 1.250 | 0.044 | 0.000 | 1.162 | 1.250 | 1.335 |
| Random intercept var($\upsilon_{oj}$ ) | 0.727 | 0.071 | | 0.588 | 0.723 | 0.865 |

The variance $\delta_{\epsilon}^{2}$ and $\delta_{oj}^{2}$ of estimate the variations among individual of women and among regions of the country respectively. In the variance component model it is possible to decompose the variance in to regional level (higher level) and individual level. Individual (level-1) variance was to assess how much of the variation is due to the individual themselves and how much of the variation is due to regional level. In order to get an idea of how much of the variation in early sexual debut of among women was attributable to the region level factors, it is useful to see the intra-region correlation coefficient (ICC) as$\rho = \frac{0.727}{0.727+3.29} = 0.18098$ ,which measures the proportion of variance of the early sexual debut that is between regions, not within regions. This means that around 18.1% of the variance in early sexual debut are due to variation among regions. Whereas the remaining 81.9% attributable to individual level, i.e., within regional differences. Thus, Bayesian multilevel analysis can be considered as an appropriate approach for further analysis.

## 4.4 Model comparison of Bayesian multilevel logistic regressions

From Table 4.5 below, we see that the comparison of the fit of Bayesian multilevel logistic regression models using the summary of the fitted model. The researcher is going to compare Bayesian multilevel empty model, random intercept model and random coefficient model using DIC further strengthened the advantage of the Bayesian multilevel model. The model which has small DIC is the best model for the data set of early sexual debut among women in Ethiopia. The result, shows that Bayesian multilevel random coefficient model was an improved fit as compared to the rest models in any combination of variables in the data set.

The DIC diagnostics of random intercept Bayesian multilevel logistic regression model are reduced by 671.06 from the Bayesian multilevel logistic regression of an empty model. This show as adding covariate variables to the model indicates how the variable was determined the occurrence of early sexual debut among women.Thus; Bayesian multilevel logistic regression for random intercept was the better model as compared to Bayesian multilevel for an empty model. The DIC diagnostics of Bayesian multilevel logistic regression of random coefficient model is reduced by 24.59 from Bayesian multilevel for random intercept so, this Bayesian multilevel random coefficient model is a great improvement suggesting that this model is the appropriate model than a Bayesian multilevel empty model and Bayesian multilevel for intercept model to determine women in early sexual debut factors.

The average deviance from the complete set of iterations $\bar{D}($ also decreased from an empty model to random intercept and from random intercept to the random coefficient model. $D(\bar{\theta})$ shows that the deviance at the expected value of the unknown parameters and it also shows the decreasing trend from an empty model to random intercept and from random intercept to the random coefficient model. Also the model complexity is measured by **pD**(The effective number of parameters in the model), the larger the **pD** is easier to fit the data. Based on this fact, the third model has the largest value of this measure, it is selected again.

Table 4.5: DIC values for model comparisons.

| **BDIC for model comparison** | | | | |
|---|---|---|---|---|
| Model | $\bar{D}$ | $D(\theta)$ | pD | DIC |
| BML logistic regression of Empty model | 12123.36 | 11731.17 | 392.19 | 12515.55 |
| BML logistic regression of Random intercept model | 11515.07 | 11185.65 | 329.42 | 11844.49 |
| BML logistic regression of Random coefficient model | 11381.04 | 10942.17 | 438.87 | 11819.90 |

**Comparison of Multilevel Random Coefficient and Binary Logistic Models.**

The goodness of fit of various statistical models can be examined using various statistical measures. These measures are used to select a suitable model among the possible candidate models. In Table 4.6 below, some model comparison measures have been listed. Among these, the most commonly used statistical measures used to select a model having good predictive capacity are AIC and log likelihood. The model having the smallest AIC is considered to have a good fit and predictive performance that can minimize the difference between the actual and predicted observations. On the basis of the AIC values in Table 4.6, the multilevel random coefficient model has a good fit or better predictive performance than the binary logistic regression model. The comparison of multi-level models in table 4.5 above also confirmed that the multilevel random coefficient model has a good fit.

Table 4.6: Comparison of goodness of fit of the binary logistic and multilevel random coefficient model

| Model comparison measure | Multilevel random coefficient model | Binary logistic regression |
|---|---|---|
| -2 log*likelihood | 6030.6 | 6179.633 |
| AIC | 12173.3 | 12379.27 |

## 4.5 Random intercept model

Based on our analysis the intercept and some covariates are significant. In this Bayesian intercept model, the intercept is allowed to vary across the region after incorporating covariates of the early sexual debut. This means that, the intercept $\beta_o$ is shared by all regions, while the random effect $U_{oj}$ is specific to region j and the random effect is assumed to be a normal distribution with variance $\delta_u^2$. The result shows that the variance of the random effect is significant which indicates that there are regional differences in early sexual debut among women in the given data set. The Bayesian multilevel logistic regression analysis result of intercept model displayed in (appendix **C**; Table 5.3) also estimates the variance of random effect at the regional level var( $U_{oj}$) = 0.398 since the 95% credible interval was greater than zero under the interval, which indicates that there is a significant regional variation.

This confirmed the significance of the regional difference in prevalence of early sexual debut among women in Ethiopia. In general, the variance component for random intercept is found to be significant because the lower limit of the credible interval is greater than zero, indicating strong evidence of the variations across regions for prevalence of early sexual debut. The results showed that the intra-region correlations coefficient (ICC) is estimated as $\rho = \frac{0.398}{0.398+3.29} = 0.1079$. This means that about 10.8% of the total variability in early sexual debut of women is due to difference across regions, with the remaining unexplained 89.2% attributable to individual differences.

## 4.6 Analysis based on random coefficient model

Based on our analysis the Bayesian random coefficients model are displayed in Table 4.5 below ,we interpret the results as follows. Some of the independent variables were found to be significant on the early sexual debut among women in Ethiopia. The results revealed that women age in years, education, place of residence, wealth index of house hold, respondent occupation and marital Status were found to be significant, indicating strong effects on early sexual debut among women and also contributing to variations among regional state in Ethiopia. However, the impacts of religion was found to be insignificant.

The interpretation of the odds ratio for multilevel logistic regression model has no difference from the interpretation in logistic regression model. But the difference between them is the multilevel logistic regression model is additional information than single level regression model, which are they include additional term which is called the random part. From the output of the random coefficient Bayesian multilevel models, we interpret the results as follows: The women age in years has a significant association with early sexual debut. The odd ratio of women in age group of 25-34 years are 57.4% times less likely (OR=0.426, 95% credible interval:(-0.984,-0.726)) to experience early in sexual debut compared to women in age group 15-24 years. Likewise the odd of Women in age group of 34+ years are 53.9% times less likely (OR=0.461, 95% Credible interval:(-0.914,-0.632)) to experience women in early sexual debut compared to women in age group 15-24 years.

Table 4.7: Result of random coefficient Bayesian multilevel model

| Fixed effect Variables | category | Coefficient | SD | p-value | Exp($\beta$) | 2.5% | 50% | 97.5% |
|---|---|---|---|---|---|---|---|---|
| | Intercept | 1.557 | 0.140 | 0.000 | 4.744 | 1.287 | 1.556 | 1.834 |
| Women Age in years | 15-24(ref) | | | | | | | |
| | 25-34 | -0.853 | 0.067 | 0.000 | 0.426 | -0.984 | -0.853 | -0.726 |
| | 34+ | -0.774 | 0.073 | 0.000 | 0.461 | -0.914 | -0.773 | -0.632 |
| Education | No education(ref) | | | | | | | |
| | Primary | -0.342 | 0.064 | 0.000 | 0.710 | -0.468 | -0.342 | -0.217 |
| | Secondary | -1.177 | 0.101 | 0.000 | 0.308 | -1.371 | -1.177 | -0.978 |
| | Higher | -2.020 | 0.110 | 0.000 | 0.133 | -2.241 | -2.021 | -1.807 |
| Religion | Orthodox(ref) | | | | | | | |
| | Catholic | -0.233 | 0.322 | 0.229 | 0.792 | -0.827 | -0.243 | 0.428 |
| | Protestant | -0.190 | 0.084 | 0.010 | 0.826 | -0.360 | -0.190 | 0.027 |
| | Muslim | -0.142 | 0.072 | 0.022 | 0.867 | -0.281 | -0.142 | -0.005 |
| | Traditional | -0.420 | 0.335 | 0.104 | 0.657 | -1.055 | -0.424 | 0.260 |
| | Other | -0.287 | 0.324 | 0.188 | 0.750 | -0.917 | -0.288 | 0.341 |
| Place of residence | Rural(ref) | | | | | | | |
| | Urban | -0.546 | 0.117 | 0.000 | 0.579 | -0.775 | -0.547 | -0.315 |
| Wealth index | Poorest(ref) | | | | | | | |
| | Poorer | 0.185 | 0.089 | 0.019 | 1.203 | 0.010 | 0.185 | 0.360 |
| | Middle | 0.166 | 0.093 | 0.038 | 1.180 | 0.020 | 0.167 | 0.342 |
| | Richer | 0.191 | 0.109 | 0.037 | 1.21 | 0.015 | 0.190 | 0.413 |
| | Richest | 0.295 | 0.125 | 0.011 | 1.343 | 0.046 | 0.293 | 0.545 |
| Respondent occupation | Not Working(ref) | | | | | | | |
| | Working | 0.128 | 0.054 | 0.006 | 1.136 | 0.022 | 0.127 | 0.233 |
| Marital Status | Single(ref) | | | | | | | |
| | Married | 0.806 | 0.110 | 0.000 | 2.238 | 0.582 | 0.806 | 1.104 |
| | Living with partner | 0.408 | 0.190 | 0.016 | 1.503 | 0.037 | 0.410 | 0.782 |
| | Widowed | 1.024 | 0.168 | 0.000 | 2.784 | 0.700 | 1.024 | 1.353 |
| | Divorced | 1.014 | 0.141 | 0.000 | 2.756 | 0.726 | 1.016 | 1.285 |
| | Separated | 0.711 | 0.188 | 0.000 | 2.036 | 0.343 | 0.711 | 1.089 |
| Random Effect | | | | | | | | |
| $\delta^2_{uo}$ | - | 0.456 | 0.061 | 0.000 | | 0.336 | 0.452 | 0.575 |
| Random slope | | | | | | | | |
| $\delta^2_{u11}$ | | 0.440 | 0.159 | 0.000 | | 0.129 | 0.422 | 0.751 |
| $\delta^2_{u21}$ | | 0.473 | 0.164 | 0.000 | | 0.151 | 0.454 | 0.795 |

Based on our analysis, education level of household also showed a statistical significant association with early sexual debut. Women with primary education level was 29% (OR=0.710, 95% credible interval:(-0.468,-0.217)) times less likely odds of having experienced early sexual debut as compared to no education level of women. Like wise women with secondary education and higher education level 69.2% and 86.7% times less likely than women having experienced early sexual debut as compared to no education level women (OR=0.308, 0.133 95% Credible interval:(-1.371, -0.978, -2.241, -1.807)) respectively. This means that a women who are not educated are mostly affected by experience early sexual debut and being educated reduces the risk of having an early sexual intercourse.

Place of residence of women are also statistical association on early sexual debut among women in Ethiopia. The odd of women who come from urban area are 42.1 less likely %(OR=0.579, 95% credible interval : (-0.775,-0.315 ))to having experienced early sexual debut compared to women who come from rural area. This means women who living in urban area are more affected early sexual intercourse than women who living in rural area.
Wealth index of household also showed a statistical significant association with early sexual debut. Poorer was 20.3% (OR=1.203, 95% credible interval:(0.010, 0.360)) times more likely to experience early sexual initiation when compared to poorest women. Women in middle,richer,and richest are 18.0%, (OR=1.180, 95% credible interval: (0.020,0.342 ), 21% (OR= 1.21, 95% credible interval: (0.015, 0.413) and 34.3% (OR=1.343, 95% credible interval: (0.046, 0.545) times more likely to experience early sexual initiation when compared to poorest women respectively. This means women whose wealth index are poorer, middle, richer, and richest are most affected by early sexual debut.

Respondent occupation of women are also statistically significant association of women in early sexual debut. The odd of women whose working group are 13.6% (OR=1.136, credible interval: (0.022,0.233) times more likely to experience early sexual initiation when compared to not working group. This means age of initiation of sexual intercourse among women is most affected women whose are working.
This study also revealed that, marital status of women are also statistically significant association of women in early sexual debut. The odd of women whose marital status are married 2.238 (OR=2.238, 95% credible interval: (0.582, 1.104),living with partner are 50.3% (OR=1.503, 95% credible interval: (0.037, 0.782), widowed are 2.784 (OR=2.784, 95% credible interval: (0.700,1.353), divorced are 2.756 (OR=2.756, 95% credible interval: (0.726,1.285) and separated are 2.036 (OR=2.036, 95% credible interval: (0.343,1.089) times more likely to experience early sexual debut when compared to whose marital status are single.

The Bayesian multilevel logistic regression analysis of random coefficient model results displayed in Table 4.7 above, also estimates the variance of random effect at the regional level, Var ($U_{oj}$). Thus, the value of Var ($U_{oj}$) = 0.456 indicate there was significant variation (which means the 95% credible intervals is greater than zero). This confirmed the significance of the regional difference in women in sexual debut in the regional state of Ethiopia.The researcher tried to identify to see the level of variation;that the intra-region correlation coefficient ICC is estimated as $\rho = \frac{0.456}{0.456+3.29} = 0.1217 =$ This means that about 12.2% of the total variability of early sexual debut are due to differences across regions, with the remaining unexplained 87.8% attributable to individual differences.

In the random intercept model, we allowed the intercept only to vary across regions by fixing explanatory covariates. In this model, researcher has tested the variable that have significant impact on occurrence of early sexual debut among women the Bayesian multilevel intercept model by observing their respective region effect. Consequently, regional level variables which are supposed to varying regionally such as wealth index of household(richer) and women education level(secondary) have been examined. The region wise intercept ($U_{oj}$), education level slopes ($U_{u11}$) and wealth index slopes ($U_{u21}$) vary significantly. There was a significant variation in the effects of these explanatory variables across the regions.

The researcher revealed that the variance of women education of secondary category has slopes ($U^2_{u11}$) = 0.440 with a credible interval of (95% CI: 0.129,0.751) and the variance of wealth index of richer category has slope ($U^2_{u21}$) = 0.473 with credible interval (95% CI: 0.151,0.795) the interval was greater than zero. This indicate the random slope of women educational level and wealth index of womens are statistically significant across the region. This means women educational level and wealth index of women early sexual debut among women are varies from region to region.

Since their p-value is less than 0.05. The null hypothesis has to be rejected, indicating that the variance of the random factor is significant which means that there was a significant variation in the effects of these explanatory variables across the regions or the random slope of those variables allows the effect of the coefficient of this variable to vary from region to region. but another variables in the model did not change from region to region which means their random part variance is not statistically significant. The significance of these two variables further indicates that a model with a random coefficient is more appropriate to explain regional variation than a model with fixed coefficients.

## 4.7 Assessment of Bayesian random coefficient model Convergence

Once a model has been developed, we now would like to know how effective the model is in describing the outcome. This is referred to as goodness of fit. These are a method used to determine whether the algorithm has reached its equilibrium or target distribution. There are several ways used to monitor convergence. The most common ways of checking goodness of fit are: diagnosis for convergence of the MCMC chains was confirmed by visual inspection the trace plot, kernel density, effective sample size and Monte Carlo (MC) error. Small values of the MC error indicate that the quantity of interest has been calculated with precision. To use summary statistics of the estimated posterior distributions for inference the realized value of the parameters (the MCMC value) should converge. To check this we have to use a suitable diagnosis to evaluate mixing a convergence of a sampler. From different methods of checking convergence, trace and history plots, kernel density plot and autocorrelation are among the common [71]. Tests used for checking convergence of a Bayesian multilevel random coefficient model were as follows:

**Time Series Plots:** Are commonly used to assess convergence of the parameter estimates in Bayesian analysis. The plot with number of iterations on the x-axis and parameter values on the y-axis for each significant parameter. The plot looks like a horizontal band, with no longer upward or downward trends, then we have evidence that the chain has converged. For all simulated parameters, time series plot indicates a good convergence since the chains are mixed together (see below figure and appendix

**E**).

**Kernel Density Plot:** This is also the statistical techniques to recognize convergence in Bayesian analysis. The plot on below figure and appendix **E** shows that the coefficients of significant variables were approximately normally distributed. Thus, this indicates that the Markov chain has attained its posterior distribution.

**Autocorrelation Plot:** It is a test used for convergence of Bayesian analysis. The ACF measures how correlated the values in the chain are with their close neighbors.The lag is the distance between the two chains to be compared. High autocorrelation indicates slow mixing within a chain and usually slow convergence to the posterior distribution. So, the plots displayed below indicate low autocorrelation as we have seen from below figure and appendix **E**.

**Effective sample size:** Convergence of posterior estimate has been checked using an effective sample size that is all the effective sample sizes of the estimates are greater than 200. So, the more samples you save, the more accurate would be your posterior estimates.This is an indication of efficient posterior estimate. It has been presented in below figure and appendix **E**.

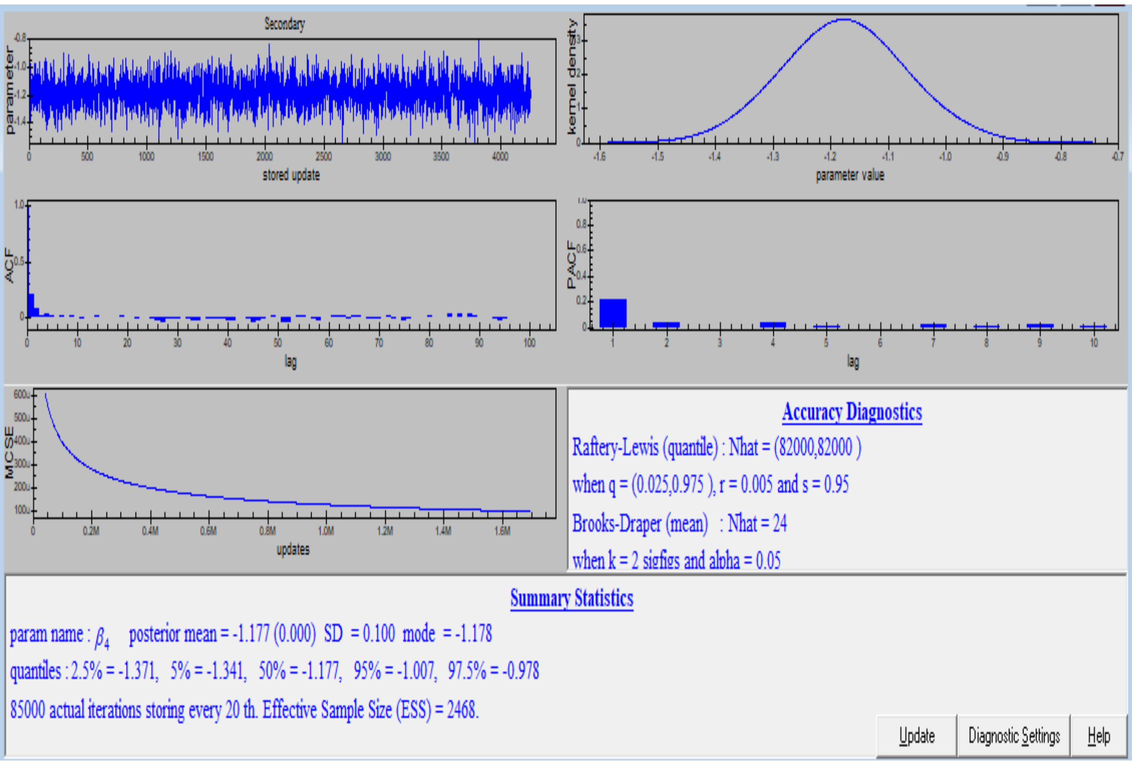



Figure 4.2: Convergence for women educational level in secondary

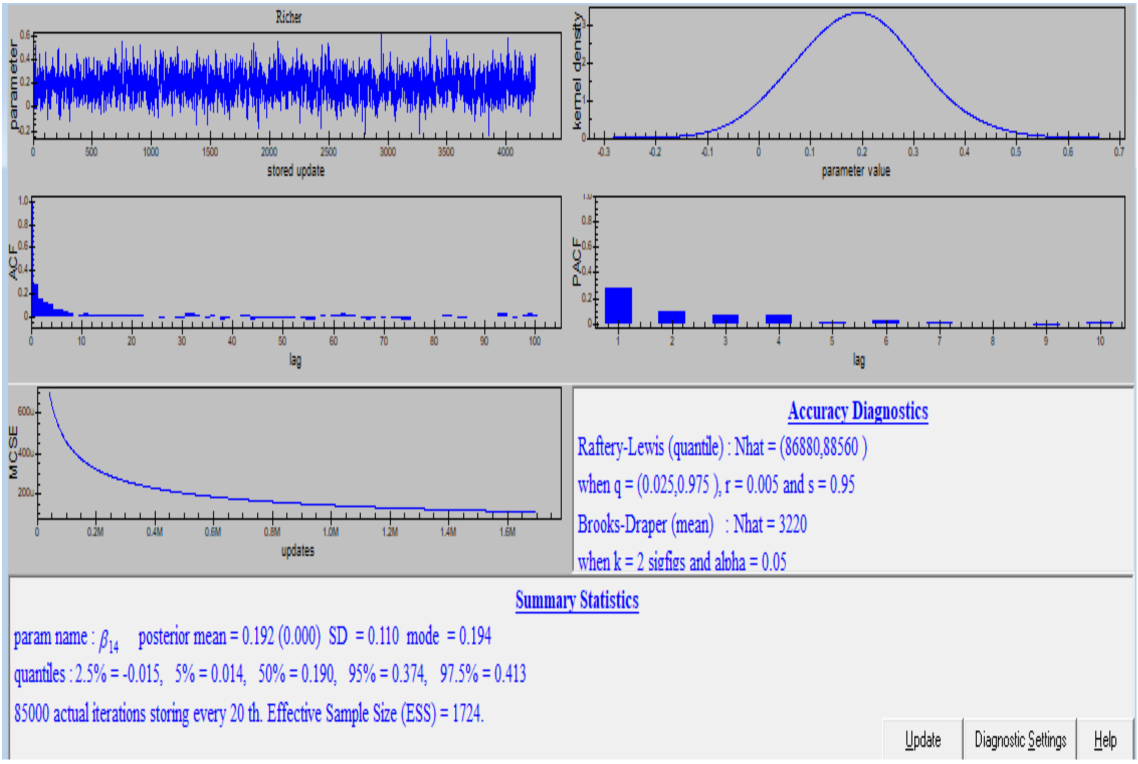



Figure 4.3: Convergence for wealth index of household in richer class

## 4.8 Discussions

Findings from this study shows that, the prevalence of early sexual initiation among reproductive-age women was 75.2%. The result of this study was higher than a study conducted in South Africa 70.8%, 26.8% in Nigeria and 55% in Ghana [20, 21]. According to the descriptive data analysis,the highest prevalence of early sexual debut was observed in Amhara (87.2%) Afar (86.2%) and Benishangul (81.5%) as opposed to the low prevalence which was recorded in Addis Ababa (50.1%). In addition, there was also higher proportion of rural residence (81.2%) respondents who participated in the study compared to Urban residence (61.4%). The result was confirmed by similar findings investigated previously study in Ethiopia [34].

We consider Bayesian multilevel logistic regression model was used. Firstly, the researcher has fitted the Bayesian multilevel empty model, intercept and coefficient models. In this regard, the Bayesian random coefficient model was found better in fitting the data appropriately based upon their deviance information criteria compared to the empty and random intercept model. The result was confirmed by Bayesian multilevel analysis of utilization of antenatal care services in Ethiopia [48, 45]. In the multilevel analysis women are nested within various region in Ethiopia. Before the analysis of data using multilevel approach, the necessity of multilevel analysis was investigated through Chi-square test statistic. The heterogeneity test and the significance of variance of random effect suggest that early sexual debut among women differs among regions. The results which are obtained from the Bayesian multilevel logistic regression of random coefficient models are discussed as follows: Based on Bayesian multilevel logistic regression of random coefficient model age of women in years, education, respondent occupation, wealth index, place of residence and marital status are statistically significant association with early sexual debut. This finding is consistent with a study done in Ethiopia [48, 30] and Nigeria [24, 33].

The above result is revealed that the odd of Women in age group of 25-34 years are 57.4% (OR=0.426, 95% Credible interval:(-0.984,-0.726)) and age group of 34+ years are 53.9% (OR=0.461, 95% Credible interval:(-0.914,-0.632)) times less likely to experience early in sexual debut compared to women in age group 15-24 years respectively. Age difference can make women more vulnerable to health risks and social isolation by creating power dynamics. These power dynamics can increase girls' vulnerability to emotional, physical, and sexual abuse. Young married girls are more likely to be illiterate and of low social status. They tend to have no access to financial resources and restricted mobility; they are therefore less likely to leave home to socialize with others, limiting their ability to obtain information on reproductive health, contraception, HIV, and other sexually transmitted infections (STIs). This power differential can also limit girls ability to negotiate contraceptive or condom use, putting them at high risk for contracting STIs and HIV. The result was confirmed by similar findings investigated Nigeria [24] and Ethiopia [30]. According to this study, education level of household also showed a statistical significant association with early sexual debut. Being educated reduces the risk of having an early sexual intercourse and might protect themselves from being engaged. Support education beyond primary school. Investments must be made to support girls education. Evidence suggests that educated girls are less likely to agree to marry at a young age. Development programs need to be creative in implementing programs that support a girl through the critical drop-out period, along with secondary and vocational opportunities that are acceptable to the girls families. The result was confirmed by similar findings investigated Nigeria [33, 24] and Ethiopia [30, 34, 35].

The odd of women who come from urban area are 42.1 less likely %(OR=0.579 95% Credible interval: (-0.775, -0.315)) to having experienced early sexual debut compared to women who come from rural area. This means women who living in urban area are more affected early sexual intercourse than women who living in rural area. Women in the urban areas maybe more exposed to urban lifestyles including media and Internet activities which may influence their sexual behavior.This finding was in line with studies conducted in Ethiopia [30], Nigeria [40] and Ghana [32].

Wealth index of household also showed a statistical significant association with early sexual debut. Women in Poorer, middle, richer, and richest are 20.3% (OR=1.203, credible interval:( 0.010,0.360)), 18.0%, (OR=1.180,95% credible interval 0.020,0.342 ), 21% (OR= 1.21, 95% Credible interval:( 0.015, 0.413)) and 34.3% (OR=1.343, 95% Credible interval:(0.046,0.545)) times more likely to experience early sexual initiation when compared to poorest women respectively. Societal expectations also put peculiar financial pressure on women, such as membership of different associations or social clubs, desire to be trendy in fashion and lifestyles there is a general consensus that a large number of women are pushed into the sex-for-money lifestyle in order to meet diverse social and familial demands. This finding was in line with studies conducted in Ethiopia [30, 38, 36], Ghana [32] and Nigeria [37].

The odd of women in early sexual debut of marital status are married 2.238 (OR=2.238, 95% Credible interval:( 0.582, 1.104)), living with partner are 50.3% (OR=1.503, 95% Credible interval: (0.037, 0.782)), widowed are 2.784 (OR=2.784, 95% Credible interval: (0.700, 1.353)), Divorced are 2.756 (OR=2.756, 95% Credible interval:(0.726, 1.285)) and separated are 2.036 (OR=2.036, 95% Credible interval:(0.343, 1.089)) times more likely to experience early sexual debut than women whose marital status are single. The possible explanation is early initiation of sexual activity and higher numbers of marital status are wide variety of negative life outcomes, including increased rates of infection with sexually transmitted diseases, increased rates of out-of-wedlock pregnancy and birth, increased single parenthood, decreased marital stability, increased maternal and child poverty, increased abortion, increased depression, and decreased happiness. This might be engaged in early marriage (before the age of 18 years) which is the potential scenario for women to be engaged in early age sexual intercourse activity. This finding was in line with studies conducted in Ethiopia [30, 36].

According to this study respondent occupation are significantly association on early sexual debut among women. The positive impact of occupational health service locally may be observed in reducing morbidity and work-related injuries. In addition, this also means fewer losses to employer and worker as there will be a reduction of wage losses and decreased compensation costs. An increasing number of workers in industrial countries complain about psychological stress and overwork. This finding was in line with studies conducted in Nigerian youth [39], Ethiopia [36] and it was inconsistent the study conducted in Ethiopia [30]. Another finding of this study showed religion was not significant association with early sexual debut. This finding was inconsistent the study conducted in Ethiopia [30]and Nigerian youth [39]. Bayesian multilevel coefficient model analysis indicated that there were regional variations of early sexual debut. From the result of the model adequacy Bayesian multilevel logistic random coefficient model is the best-fitted model [48, 45].

The model would be implemented using MLwiN 2.02 versions [72]. For each model, 85,000 MCMC iterations were run, with the initial 15,000 burn-in terms discarded, and thereafter keeping every 20th sample value to make observations independent or low autocorrelations. The 60,000 iterations left were used to assess convergence of the chain and parameter estimation. For convergence check researcher has implemented four methods such as trace plot, density plot, autocorrelation, and effective sample size. With these four methods the convergence of the posterior estimate was correctly achieved and MC error for each significant predictor was found to be less than 5% of its posterior standard error.

# 5 Conclusions and Recommendations

## 5.1 Conclusions

Accordingly, the study used 11962 from 2016 EDHS there are (75.2%) had their first sexual intercourse before the age of 18 years was early sexual debut among women in Ethiopia respectively. The variable in single logistic regression such as women age in years, place of residence, region, education, respondent occupation, wealth index and marital status are significant association with early sexual debut. In single level the P-value is less than 0.05 level of significance, implying strong evidence of heterogeneity for the early sexual debut among women across regional states of Ethiopia.

Based on Bayesian multilevel logistic regression women age in years, place of residence, region, education, respondent occupation, wealth index and marital status are significant association with early sexual debut. The intraregional correlation coefficient (ICC) of Bayesian multilevel null model implied that 18.1% of the total early sexual debut variation is due to the variation between regions, which confirmed that the variation in age of initiation of early sexual intercourse among women across region in Ethiopia. This variation of regionals is due to differences in women background characteristics such as education level and wealth index, since the random part (regional level) variance of these two variables were significant at 5%level of significance.
There is a considerable variation in the effects of education level and wealth index so, these variables differ significantly across the regions and made the variation between regions.

## 5.2 Recommendation

Based on the findings of this study, the following recommendations are announced.

- To better the government and stakeholders should give more attention and emphasis on early age of initiation sexual intercourse.
- Female education should be encouraged, also religious bodies should encourage young women to postpone first sex until they are ready.

- To better enforcing existing laws about age at marriage and implementing programs to delay marriage and greater involvement of teachers and school administrators, health officials, and other authorities is critical in helping girls resist parental and social pressures to marry early.
- Improving girls and young women access to education is important for rising the women age at first sexual intercourse, which is vital for empowering them and enhancing their participation in any sector.
- The government should design programs to improve the socioeconomic standards of the poor, there is a clear need for intervention to reduce economic inequalities and ultimately poverty among the populace.
- To better advocating for changes in social attitudes and norms through multi sectoral and integrated community-based programs such as through religious institutions and associations, health institutions, other local civic organizations, and schools are the best channels for raising awareness of the negative consequences of early sexual debut and the many economic, social, and health benefits of delaying sexual debut.
- To better more emphasis on ergonomics and occupational psycho social factors would be needed in the services industry should be increase.
- The respondent should be able to learn on the job and be allowed to continue to learn as their career progresses.

# References

[1] Meleko, A., Mitiku, K., Kebede, G., Muse, M. and Moloro, N., 2017. Magnitude of pre-marital sexual practice and its associated factors among Mizan preparatory school students in Mizan Aman town, South West Ethiopia. J Community Med Health Educ, 7(539), pp.2161-0711.

[2] Kassahun, E.A., Gelagay, A.A., Muche, A.A., Dessie, A.A. and Kassie, B.A., 2019. Factors associated with early sexual initiation among preparatory and high school youths in Woldia town, northeast Ethiopia: a cross-sectional study. BMC public health, 19(1), pp.1-8

[3] World Health Organization, 2018. WHO recommendations on adolescent sexual and reproductive health and rights.

[4] Carazzone, C. and Mazzarelli, S., 2019. The Human Rights-based Approach to Girls' and Women's Health. Advocating for Girls' and Women's Health and Human Rights, p.21.

[5] Sully, E., Biddlecom, A., Darroch, J.E., Riley, T., Ashford, L.S., Lince-Deroche, N., Firestein, L. and Murro, R., 2020. Adding it up: investing in sexual and reproductive health 2019.

[6] World Health Organization, 2016. World health statistics 2016: monitoring health for the SDGs sustainable development goals. World Health Organization.

[7] Tauschek, B., 2020. The Number of Missing Women.

[8] Nations, U., 2017. Department of Economic and Social Affairs, Population Division (2017). World population prospects: the 2017 revision, key findings and advance tables. In Working Paper No. ESA/P/WP/248.

[9] Federation IPP. Sexual and reproductive health of young women in Ethiopia; 2014

[10] Demographic, C.E., 2012. Health Survey 2011 Addis Ababa. Ethiopia Central Statistical Agency.

[11] Bizuneh, H., 2019. Correlates of sexual initiation among adolescent and youth in Addis Ababa, Ethiopia: a community based cross sectional study. MOJ Public Health, 8(3), pp.108-13.

[12] Campus, D., 2018. Assessing factors influencing early sexual initiation among adolescents (13 to 19 years) in Ghana: a qualitative study.

[13] Kassahun, E.A., Gelagay, A.A., Muche, A.A., Dessie, A.A. and Kassie, B.A., 2019. Factors associated with early sexual initiation among preparatory and high school youths in Woldia town, northeast Ethiopia: a cross-sectional study. BMC public health, 19(1), pp.1-8.

[14] Yihenew Mitiku. Determinants of Age at First Sexual Intercourse Among Women in Rural Ethiopia. Biochemistry and Molecular Biology. Vol. 4, No. 6, 2019, pp. 80-85. doi: 10.11648/j.bmb.20190406.11

[15] Ngowa, J.D.K., Kasia, J.M., Pisoh, W.D., Ngassam, A. and Noa, C., 2015. Obstetrical and perinatal outcomes of adolescent pregnancies in Cameroon: a retrospective cohort study at the Yaoundé general hospital. Open Journal of Obstetrics and Gynecology, 5(02), p.88.

[16] Sakai, T., Sugawara, Y., Watanabe, I., Watanabe, T., Tomata, Y., Nakaya, N. and Tsuji, I., 2017. Age at first birth and long-term mortality for mothers: the Ohsaki cohort study. Environmental health and preventive medicine, 22(1), pp.1-14.

[17] Vogel, J.P., Pileggi-Castro, C., Chandra-Mouli, V., Pileggi, V.N., Souza, J.P., Chou, D. and Say, L., 2015. Millennium Development Goal 5 and adolescents: looking back, moving forward. Archives of Disease in childhood, 100(Suppl 1), pp.S43-S47.

[18] Darroch, J.E., Woog, V., Bankole, A. and Ashford, L.S., 2016. Adding it up: costs and benefits of meeting the contraceptive needs of adolescents.

[19] World Health Organization, 2015. Trends in maternal mortality: 1990-2015: estimates from WHO, UNICEF, UNFPA, World Bank Group and the United Nations Population Division: executive summary (No. WHO/RHR/15.23). World Health Organization.

[20] O. Asante, K., Nketiah-Amponsah, E., Andoh-Arthur, J., Boafo, I.M. and Ampaw, S., 2018. Correlates of early sexual debut among sexually active youth in Ghana. International quarterly of community health education, 39(1), pp.9-17.

[21] Olufemi, A.T., Paulin, O.I. and Akinbode, O.O., 2018. Prevalence and predictors of early sexual debut among adolescents in Ogbomoso, Nigeria. American Journal of Public Health, 6(3), pp.148-154.

[22] Fatusi, A.O. and Blum, R.W., 2008. Predictors of early sexual initiation among a nationally representative sample of Nigerian adolescents. BMC public health, 8(1), pp.1-14.

[23] Amoateng, A.Y. and Baruwa, O., 2018. Changes in the timing of sexual intercourse in Ghana: evidence from the demographic and health survey data, 1988-2014. African Population Studies, 32(3).

[24] Uthman, O.A., 2008. Geographical variations and contextual effects on age of initiation of sexual intercourse among women in Nigeria: a multilevel and spatial analysis. International Journal of Health Geographics, 7(1), pp.1-12.

[25] Guta, D.B.D.M.G. and Yeshambel, B.M., 2016. Assessment of early sexual initiation and associated factors among Ambo University Undergraduate students, Ambo, Ethiopia. Assessment, 25.

[26] Wassachew, A.K., 2015. Early Sexual Debut and Associated Factors Among High School Adolescents Legehida District, Amahara Regional State, Ethiopia (Doctoral dissertation, Haramaya University).

[27] Girma, D., Hailu, G., Ayana, M. and Ketema, K., 2015. Factors early sexual initiation among governmental preparatory school students, Addis Ababa, Ethiopia. J Community Med Health Educ, 5(1), p.333.

[28] Kassa, G.M., Tsegay, G., Abebe, N., Bogale, W., Tadesse, T., Amare, D. and Alem, G., 2015. Early sexual initiation and associated factors among Debre Markos University students, North West Ethiopia. Sci J Clin Med, 4(5), p.80.

[29] Ayalew, A., Abreha, K., Shumey, A. and Berhane, K., 2015. Magnitude and predictors of early sexual debut among high and preparatory school students in northern Ethiopia: a school-based Crosssectional study. Journal of Health Education Research Development, pp.1-8.

[30] Nigatu, A.M., Birhanu, A.Y. and Endehabtu, B.F., 2020. Geographical variations of early age sexual initiation among reproductive-age women in Ethiopia: evidence from EDHS 2016. Archives of Public Health, 78, pp.1-12.

[31] Ethiopia, A.A., 2006. DESSALEGN WIRTU (Doctoral dissertation, ADDIS ABABA UNIVERSITY).

[32] O. Asante, K., Nketiah-Amponsah, E., Andoh-Arthur, J., Boafo, I.M. and Ampaw, S., 2018. Correlates of early sexual debut among sexually active youth in Ghana. International quarterly of community health education, 39(1), pp.9-17.

[33] Obite, C.P., Bartholomew, D.C., Duru, I.B., Ismaila-Cosmos, J. and Chukwuemeka, C., 2020. A Cox Proportional Hazard Model Approach to Age at First Sexual Intercourse in Nigeria. Open Journal of Statistics, 10(02), p.252.

[34] Turi, E. et al. (2020) why Too Soon? Early Initiation of Sexual Intercourse Among Adolescent Females in Ethiopia: Evidence from 2016 Ethiopian Demographic and Health Survey, pp. 269-275.

[35] Kim, H.W., Park, S., Ahn, H.Y. and Park, E.J., 2015. The effects of an HPV education program by gender among Korean university students. Nurse education today, 35(4), pp.562-567.

[36] Kawo, K.N., Asfaw, Z.G. and Yohannes, N., 2018. Multilevel analysis of determinants of anemia prevalence among children aged 6–59 months in Ethiopia: classical and Bayesian approaches. Anemia, 2018.

[37] Odimegwu, C. and Somefun, O.D., 2017. Ethnicity, gender and risky sexual behaviour among Nigerian youth: an alternative explanation. Reproductive health, 14(1), pp.1-15.

[38] Aychiluhm, S.B., Tesema, A.K. and Tadesse, A.W., 2021. Early Marriage and Its Determinants among Married Reproductive Age Group Women in Amhara Regional State, Ethiopia: A Multilevel Analysis. BioMed Research International, 2021.

[39] Odimegwu, C. and Somefun, O.D., 2017. Ethnicity, gender and risky sexual behavior among Nigerian youth: an alternative explanation. Reproductive health, 14(1), pp.1-15.

[40] Motsima, T. and Malela-Majika, J.C., 2016. The effects of early first sexual intercourse amongst Lesotho women: evidence from the 2009 Lesotho demographic and health survey. African journal of reproductive health, 20(2), pp.34-42.

[41] Kawo, K.N., Zeleke, A.T. and Dessie, D.B., 2019. Determinants of Teenage Pregnancy in Rural Ethiopia. J Health, Med and Nursing, 68, pp.8-16.

[42] 2007 Ethiopian census, firstdraft(https://web.archive.org/web/20120604045256/http://www. 07firstdraft.pdf), Ethiopian Central Statistical Agency (accessed 6 May 2009)

[43] Rasbash, J. and Browne, W.J., 2008. Non-hierarchical multilevel models. In Handbook of multilevel analysis (pp. 301-334). Springer, New York, NY.

[44] Geirsson, Ó.P., Hrafnkelsson, B. and Simpson, D., 2015. Computationally efficient spatial modeling of annual maximum 24-h precipitation on a fine grid. Environmetrics, 26(5), pp.339-353.

[45] Shibiru, J. and Dabala, J., 2021. Bayesian multilevel model on maternal mortality in Ethiopia. Journal of Big Data, 8(1).

[46] Salameh, P., Waked, M., Khayat, G. and Dramaix, M., 2014. Bayesian and frequentist comparison for epidemiologists: A non mathematical application on logistic regressions. The Open Epidemiology Journal, 7(1).

[47] Ogunsakin, R.E. and Siaka, L., 2017. Bayesian inference on malignant breast cancer in Nigeria: a diagnosis of MCMC convergence. Asian Pacific journal of cancer prevention: APJCP, 18(10), p.2709.

[48] Kitabo, C.A. and Damtie, E.T., 2020. Bayesian Multilevel Analysis of Utilization of Antenatal Care Services in Ethiopia. Computational and Mathematical Methods in Medicine, 2020.

[49] Yadeta, T.A., Mengistu, B., Gobena, T. and Regassa, L.D., 2020. Spatial pattern of perinatal mortality and its determinants in Ethiopia: Data from Ethiopian Demographic and Health Survey 2016. PloS one, 15(11), p.e0242499.

[50] Fahrmeir, L. and Lang, S., 2001. Bayesian inference for generalized additive mixed models based on Markov random field priors. Journal of the Royal Statistical Society: Series C (Applied Statistics), 50(2), pp.201-220.

[51] Spiegelhalter, D.J., Best, N.G., Carlin, B.P. and Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. Journal of the royal statistical society: Series b (statistical methodology), 64(4), pp.583-639.

[52] Demographic, N., 2012. Health Survey 2011. Addis Ababa, Ethiopia.

[53] Schafer, J.L. and Graham, J.W., 2002. Missing data: our view of the state of the art. Psychological methods, 7(2), p.147.

[54] Allison, P.D., 2001. Missing data. Sage publications.

[55] Benjamin R, Savile. Bayesian multilevel model and Medical applications.2008

[56] Snijders, T.A. and Bosker, R.J., 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage.

[57] Browne, W.J. and Rasbash, J., 2009. MCMC estimation in MLwiN. Bristol: Centre of Multilevel Modelling, University of Bristol.

[58] Snijders, T.A., 2005. Power and sample size in multilevel modeling. Encyclopedia of statistics in behavioral science, 3(157), p.1573.

[59] Snijders, T.A. and Bosker, R.J., 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling. sage.

[60] Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications.

[61] Larget, B. and Simon, D.L., 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Molecular biology and evolution, 16(6), pp.750-759.

[62] Bornmann, L., Mutz, R. and Daniel, H.D., 2013. Multilevel-statistical reformulation of citation-based university rankings: The Leiden ranking 2011/2012. Journal of the American Society for Information Science and Technology, 64(8), pp.1649-1658.

[63] Nydick, S.W., 2012. The wishart and inverse wishart distributions. Electronic Journal of Statistics, 6, pp.1-19.

[64] Dezfuli, H., Kelly, D., Smith, C., Vedros, K. and Galyean, W., 2009. Bayesian inference for NASA probabilistic risk and reliability analysis.

[65] Goldstein, H., 2011. Multilevel statistical models (Vol. 922). John Wiley and Sons.

[66] Zeynep O.and Mehmet A. Comparison of Markov Chain Monte Carlo Convergence Diagnostic test forBayesian logistic Random Effect Models. International Journal of Research in Medical and Health Sciences. 2016;Vol.9.

[67] Spiegelhalter, D.J., Abrams, K.R. and Myles, J.P., 2004. Bayesian approaches to clinical trials and health-care evaluation (Vol. 13). John Wiley and Sons.

[68] Milbrandt, E., Kersten, A., Watson, S., Rahim, M., Clermont, G., Angus, D. and Linde-Zwirble, W., 2005. Rising use of intensive care unit services in Medicare. Critical Care, 9(1), pp.1-1.

[69] Merkle, E. and Zandt, T.V., 2005. WinBUGS Tutorial Outline.

[70] Kass, R.E., Carlin, B.P., Gelman, A. and Neal, R.M., 1998. Markov chain Monte Carlo in practice: a roundtable discussion. The American Statistician, 52(2), pp.93-100.

[71] Ntzoufras, I., 2011. Bayesian modeling using WinBUGS (Vol. 698). John Wiley and Sons.

[72] Rasbash, J., Browne, W., Healy, M., Cameron, B. and Charlton, C., 2005. MLwiN version 2.02 [Computer software]. Bristol, England: Centre for Multilevel Modelling.

# Appendices

## 5.3 Appendix A: Variable categories and coding

Table 5.1: **Variable in the study**

| Variable | Measure | code |
|---|---|---|
| Dependent variable<br>Early sexual initiation | sexual intercourse before age 18 years or younger | Yes=1<br>No=0 |
| Individual characteristics<br>Age | Completed years | 15-24 =1<br>25-34 =2<br>35+ =3 |
| Education | | No education =0<br>Primary =1<br>Secondary =2<br>Higher =3 |
| Religion | | Orthodox =1<br>Catholic =2<br>Protestant =3<br>Muslim =4<br>Traditional =5<br>Others =6 |
| Community characteristics<br>Type of residence | Cluster classification | urban =1<br>Rural =2 |
| Region | | Tigray =1<br>Afar =2<br>Amhara =3<br>Oromia =4<br>Sumali = 5<br>Benshagul gumuz =6<br>SNNPR =7<br>Gambela =8<br>Harari =9<br>Addis Ababa = 10<br>Dire Dawa =11 |
| Wealth Index | | poor =1<br>poorest =2<br>Middle= 3<br>Rich =4<br>Richest = 5 |
| Respondent Occupation | | Not working =1<br>Working =2 |
| Marital status | | single =0<br>living with partner =1<br>Married = 2<br>widowed =3<br>Divorced =4<br>Separated = 5 |

## 5.4 Appendix B: Result of single level logistic regression model

Table 5.2: Model if term removed in single level

| Variable | | Model LogLikelihood | Change in -2LogLikelihood | df | Sig.oftheChange |
|---|---|---|---|---|---|
| Step 1 | Education | -6699.913 | 867.931 | 3 | .000 |
| Step 2 | Education | -6390.661 | 606.446 | 3 | .000 |
| | Regions | -6268.149 | 361.422 | 10 | .000 |
| Step 3 | Women Age | -6087.961 | 162.231 | 2 | .000 |
| | Education | -6339.751 | 665.813 | 3 | .000 |
| | Regions | -6169.542 | 325.394 | 10 | .000 |
| Step 4 | Women Age | -6071.832 | 191.816 | 2 | .000 |
| | Education | -6270.003 | 588.158 | 3 | .000 |
| | Regions | -6127.730 | 303.612 | 10 | .000 |
| | Marital status | -6006.903 | 61.958 | 5 | .000 |
| Step 5 | Women Age | -6059.248 | 179.611 | 2 | .000 |
| | Education | -6189.657 | 440.430 | 3 | .000 |
| | place of residence | -5975.927 | 12.969 | 1 | .000 |
| | Regions | -6106.001 | 273.117 | 10 | .000 |
| | Marital status | -5999.690 | 60.495 | 5 | .000 |
| Step 6 | Women Age | -6054.210 | 181.017 | 2 | .000 |
| | Education | -6183.385 | 439.366 | 3 | .000 |
| | place of residence | -5970.375 | 13.347 | 1 | .000 |
| | Regions | -6099.294 | 271.185 | 10 | .000 |
| | Wealth Index | -5969.446 | 11.489 | 4 | .022 |
| | Marital status | -5994.096 | 60.788 | 5 | .000 |
| Step 7 | Women Age | -6053.994 | 184.858 | 2 | .000 |
| | Education | -6183.423 | 443.715 | 3 | .000 |
| | place of residence | -5968.254 | 13.379 | 1 | .000 |
| | Regions | -6092.696 | 262.263 | 10 | .000 |
| | Wealth Index | -5966.950 | 10.770 | 4 | .029 |
| | Respondent occupation | -5963.702 | 4.274 | 1 | .039 |
| | Marital status | -5992.360 | 61.591 | 5 | .000 |

## 5.5 Appendix C: Result of Bayesian multilevel Logistic regression analysis of MLwiN output

Table 5.3: Result random Intercept Bayesian multilevel model

| Fixed effect | | | | | | | |
|---|---|---|---|---|---|---|---|
| Variable | Category | Coefficient | SD | p-value | 2.5% | 50% | 97.5% |
| | Intercept | 1.551 | 0.142 | 0.000 | 1.276 | 1.554 | 1.830 |
| Women age in years | 15-24(ref) | | | | | | |
| | 25-34 | -0.842 | 0.067 | 0.000 | -0.976 | -0.842 | -0.713 |
| | 34+ | -0.768 | 0.072 | 0.000 | -0.909 | -0.768 | -0.625 |
| Education | no education(ref) | | | | | | |
| | Primary | -0.339 | 0.065 | 0.000 | -0.464 | -0.340 | -0.210 |
| | Secondary | -1.195 | 0.091 | 0.000 | -1.376 | -1.195 | -1.015 |
| | Higher | -2.005 | 0.110 | 0.000 | -2.219 | -2.006 | -1.789 |
| Religion | Orthodox(ref) | | | | | | |
| | Catholic | -0.239 | 0.317 | 0.224 | -0.843 | -0.242 | 0.392 |
| | Protestant | -0.177 | 0.081 | 0.014 | -0.336 | -0.175 | -0.021 |
| | Muslim | -0.141 | 0.071 | 0.021 | -0.277 | -0.143 | -0.005 |
| | Traditional | -0.399 | 0.321 | 0.112 | -1.014 | -0.405 | -0.222 |
| | Other | -0.288 | 0.319 | 0.183 | -0.910 | -0.289 | 0.345 |
| Place of residence | Rural(ref) | | | | | | |
| | Urban | -0.525 | 0.116 | 0.000 | -0.748 | -0.528 | -0.304 |
| Wealth Index | Poorest(ref) | | | | | | |
| | Poorer | 0.190 | 0.089 | 0.015 | 0.015 | 0.190 | 0.368 |
| | Middle | 0.167 | 0.094 | 0.040 | -0.020 | 0.167 | 0.348 |
| | Richer | 0.157 | 0.096 | 0.050 | -0.032 | 0.156 | 0.349 |
| | Richest | 0.296 | 0.122 | 0.007 | 0.056 | 0.295 | 0.540 |
| Respondent Occupation | Not Working(ref) | | | | | | |
| | Working | 0.120 | 0.053 | 0.012 | 0.016 | 0.119 | 0.223 |
| Marital Status | Single(ref) | | | | | | |
| | Married | 0.790 | 0.113 | 0.000 | 0.580 | 0.789 | 1.015 |
| | Living with partner | 0.386 | 0.186 | 0.018 | 0.022 | 0.382 | 0.753 |
| | Widowed | 1.005 | 0.167 | 0.000 | 0.688 | 1.004 | 1.337 |
| | Divorced | 0.991 | 0.140 | 0.000 | 0.727 | 0.990 | 1.269 |
| | Separated | 0.689 | 0.187 | 0.000 | 0.334 | 0.687 | 1.063 |
| Var$(U_{oj}) = \delta^2_{uo}$ | - | 0.398 | 0.050 | 0.007 | 0.301 | 0.395 | 0.496 |

## 5.6 Appendix D: MLwiN Result for equations of Bayesian multilevel logistic regression models

**MLwiN equation result for Bayesian multilevel logistic regression-empty model**

$$\text{Early\_sexual\_debut}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$
$$\text{logit}(\pi_{ij}) = \beta_{0j}\text{constant}$$
$$\beta_{0j} = 1.250(0.044) + u_{0j}$$

$$\left[u_{0j}\right] \sim \text{N}(0,\ \Omega_u)\ :\ \Omega_u = \left[0.727(0.071)\right]$$

$$\text{var}(\text{Early\_sexual\_debut}_{ij}|\pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

PRIOR SPECIFICATIONS

$p(\beta_0) \propto 1$

$p(1/\sigma^2_{u0}) \sim \text{Gamma}(0.001,0.001)$

*Deviance(MCMC)* = 12123.359(11962 of 11962 cases in use)

## MLwiN equation result for Bayesian multilevel logistic regression-random intercept model

$$\text{Early\_sexual\_debut}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$$

$$\begin{aligned}\text{logit}(\pi_{ij}) = {} & \beta_{0ij}\text{constant} + -0.842(0.066)\text{25-34}_{ij} + -0.768(0.072)\text{34+}_{ij} + -0.340(0.065)\text{Primary}_{ij} + -1.195(0.090)\text{Secondary}_{ij} + -2.005(0.109)\text{Higher}_{ij} + \\ & -0.243(0.317)\text{Catholic}_{ij} + -0.178(0.082)\text{Protestant}_{ij} + -0.141(0.071)\text{Muslim}_{ij} + -0.403(0.325)\text{Traditional}_{ij} + -0.292(0.323)\text{Other}_{ij} + -0.526(0.116)\text{Urban}_{ij} + \\ & 0.192(0.089)\text{Poorer}_{ij} + 0.168(0.093)\text{Middle}_{ij} + 0.158(0.096)\text{Richer}_{ij} + 0.295(0.122)\text{Richest}_{ij} + 0.120(0.053)\text{Working}_{ij} + 0.790(0.113)\text{Married}_{ij} + \\ & 0.390(0.188)\text{Living with partner}_{ij} + 1.005(0.168)\text{Widowed}_{ij} + 0.992(0.141)\text{Divorced}_{ij} + 0.689(0.187)\text{Separated}_{ij}\end{aligned}$$

$$\beta_{0ij} = 1.551(0.142) + u_{0j}$$

$$\left[u_{0j}\right] \sim \text{N}(0, \Omega_u) : \Omega_u = \left[0.398(0.050)\right]$$

$$\text{var}(\text{Early\_sexual\_debut}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$$

PRIOR SPECIFICATIONS

$p(\beta_0) \alpha\ 1$
$p(\beta_1) \alpha\ 1$
$p(\beta_2) \alpha\ 1$
$p(\beta_3) \alpha\ 1$
$p(\beta_4) \alpha\ 1$
$p(\beta_5) \alpha\ 1$
$p(\beta_6) \alpha\ 1$
$p(\beta_7) \alpha\ 1$
$p(\beta_8) \alpha\ 1$
$p(\beta_9) \alpha\ 1$
$p(\beta_{10}) \alpha\ 1$
$p(\beta_{11}) \alpha\ 1$
$p(\beta_{12}) \alpha\ 1$
$p(\beta_{13}) \alpha\ 1$
$p(\beta_{14}) \alpha\ 1$
$p(\beta_{15}) \alpha\ 1$
$p(\beta_{16}) \alpha\ 1$
$p(\beta_{17}) \alpha\ 1$
$p(\beta_{18}) \alpha\ 1$
$p(\beta_{19}) \alpha\ 1$
$p(\beta_{20}) \alpha\ 1$
$p(\beta_{21}) \alpha\ 1$

$p(1/\sigma^2_{u0}) \sim \text{Gamma}(0.001, 0.001)$

*Deviance(MCMC)* = 11515.072(11962 of 11962 cases in use)

**MLwiN equation result for Bayesian multilevel logistic regression-random coefficient model.**

$\text{Early\_sexual\_debut}_{ij} \sim \text{Binomial}(\text{denom}_{ij}, \pi_{ij})$

$\text{logit}(\pi_{ij}) = \beta_{0i}\text{constant} + -0.852(0.067)\text{25-34}_{ij} + -0.774(0.072)\text{34+}_{ij} + -0.342(0.065)\text{Primary}_{ij} + \beta_{4}\text{Secondary}_{ij} + -2.021(0.110)\text{Higher}_{ij} +$
$-0.237(0.320)\text{Catholic}_{ij} + -0.189(0.085)\text{Protestant}_{ij} + -0.142(0.073)\text{Muslim}_{ij} + -0.417(0.331)\text{Traditional}_{ij} + -0.287(0.328)\text{Other}_{ij} +$
$-0.547(0.117)\text{Urban}_{ij} + 0.185(0.090)\text{Poorer}_{ij} + 0.166(0.095)\text{Middle}_{ij} + \beta_{14j}\text{Richer}_{ij} + 0.295(0.126)\text{Richest}_{ij} +$
$0.127(0.053)\text{Working}_{ij} + 0.806(0.111)\text{Married}_{ij} + 0.409(0.192)\text{Living with partner}_{ij} + 1.024(0.167)\text{Widowed}_{ij} +$
$1.012(0.140)\text{Divorced}_{ij} + 0.711(0.188)\text{Separated}_{ij}$

$\beta_{0j} = 1.556(0.140) + u_{0j}$
$\beta_{4j} = -1.177(0.100) + u_{4j}$
$\beta_{14j} = 0.192(0.110) + u_{14j}$

$$\begin{bmatrix} u_{0j} \\ u_{4j} \\ u_{14j} \end{bmatrix} \sim N(0, \Omega_u) : \Omega_u = \begin{bmatrix} 0.456(0.060) & & \\ -0.142(0.084) & 0.441(0.159) & \\ -0.212(0.088) & 0.196(0.153) & 0.473(0.164) \end{bmatrix}$$

$\text{var}(\text{Early\_sexual\_debut}_{ij} | \pi_{ij}) = \pi_{ij}(1 - \pi_{ij})/\text{denom}_{ij}$

PRIOR SPECIFICATIONS

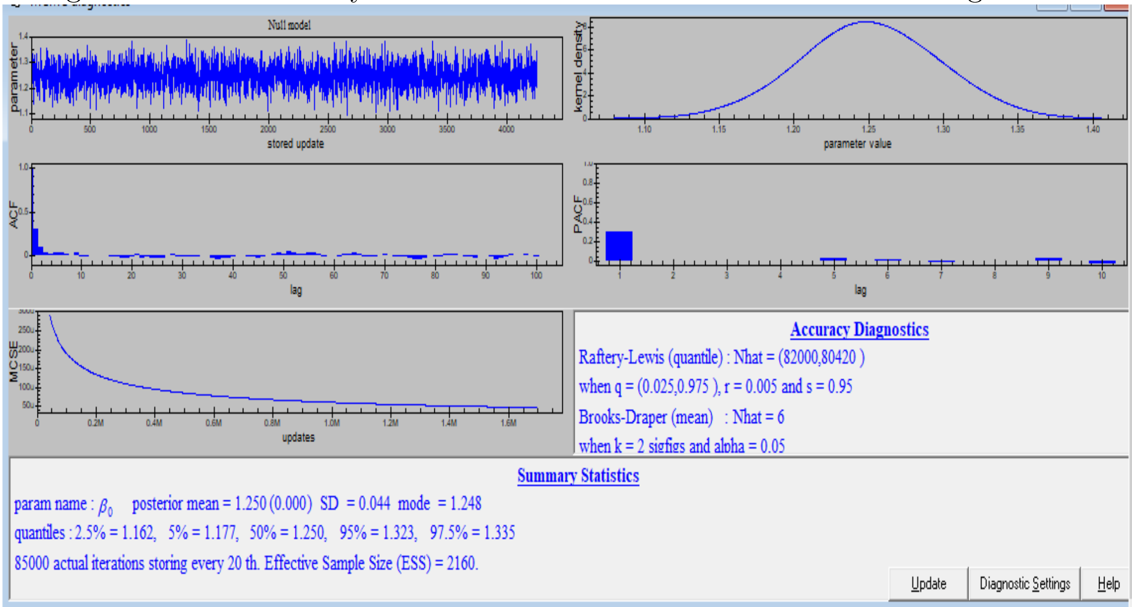
$p(\beta_0) \propto 1$
$p(\beta_1) \propto 1$
$p(\beta_2) \propto 1$
$p(\beta_3) \propto 1$
$p(\beta_4) \propto 1$
$p(\beta_5) \propto 1$
$p(\beta_6) \propto 1$
$p(\beta_7) \propto 1$
$p(\beta_8) \propto 1$
$p(\beta_9) \propto 1$
$p(\beta_{10}) \propto 1$
$p(\beta_{11}) \propto 1$
$p(\beta_{12}) \propto 1$
$p(\beta_{13}) \propto 1$
$p(\beta_{14}) \propto 1$
$p(\beta_{15}) \propto 1$
$p(\beta_{16}) \propto 1$
$p(\beta_{17}) \propto 1$
$p(\beta_{18}) \propto 1$
$p(\beta_{19}) \propto 1$
$p(\beta_{20}) \propto 1$
$p(\beta_{21}) \propto 1$

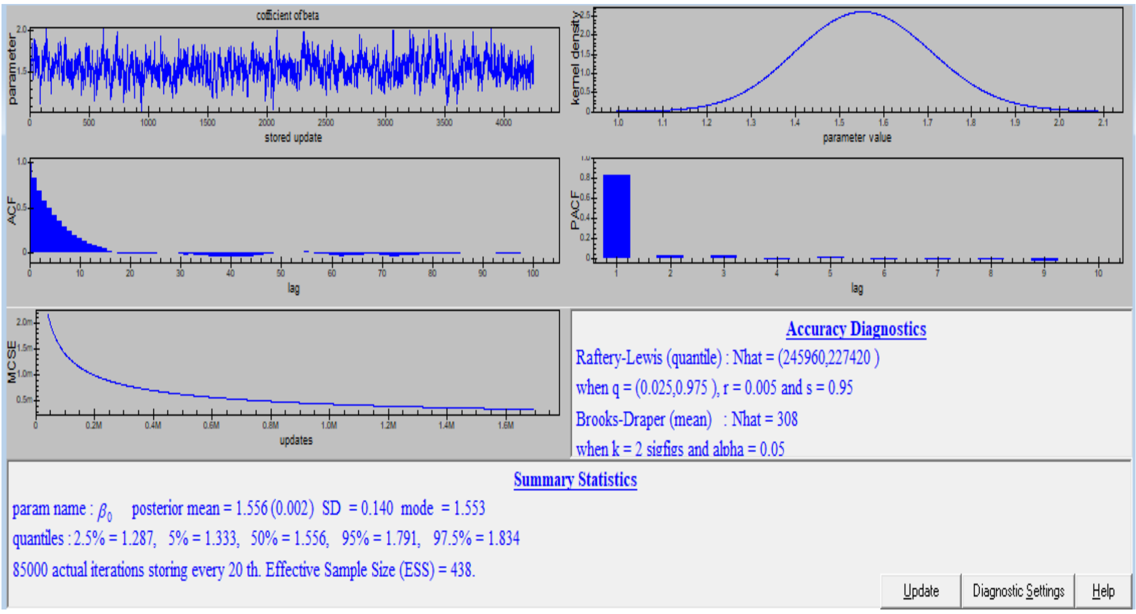
$$p(\Omega_u) \sim \text{inverse Wishart}_3[3*S_u, 3],\ S_u = \begin{bmatrix} 0.434 & & \\ -0.136 & 0.379 & \\ -0.212 & 0.178 & 0.450 \end{bmatrix}$$

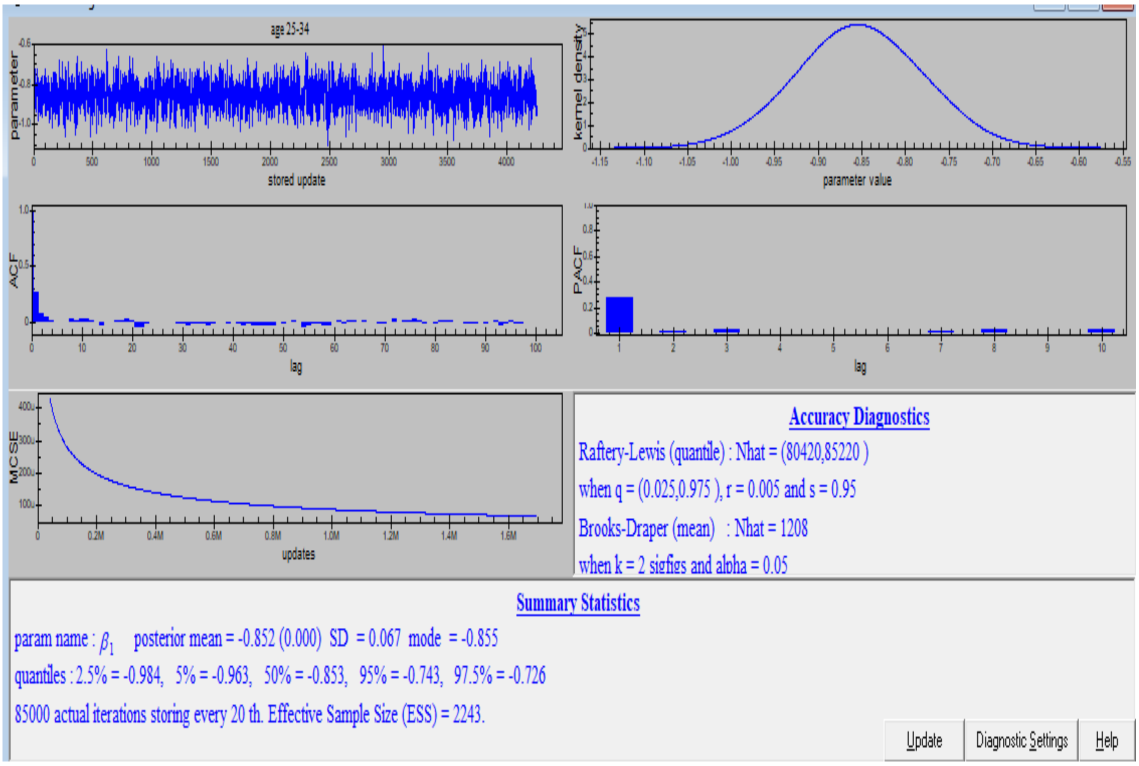
*Deviance(MCMC)* = 11381.038(11962 of 11962 cases in use)

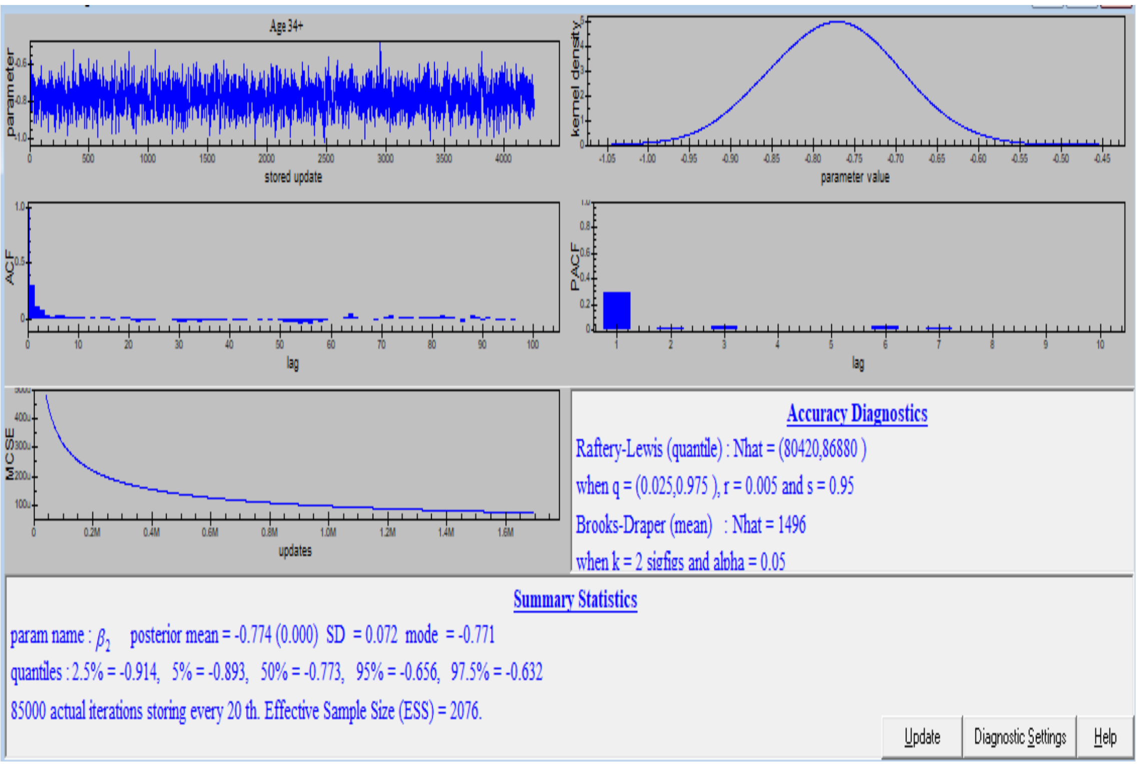
## 5.7 Appendix E: List of Figures for diagnostics

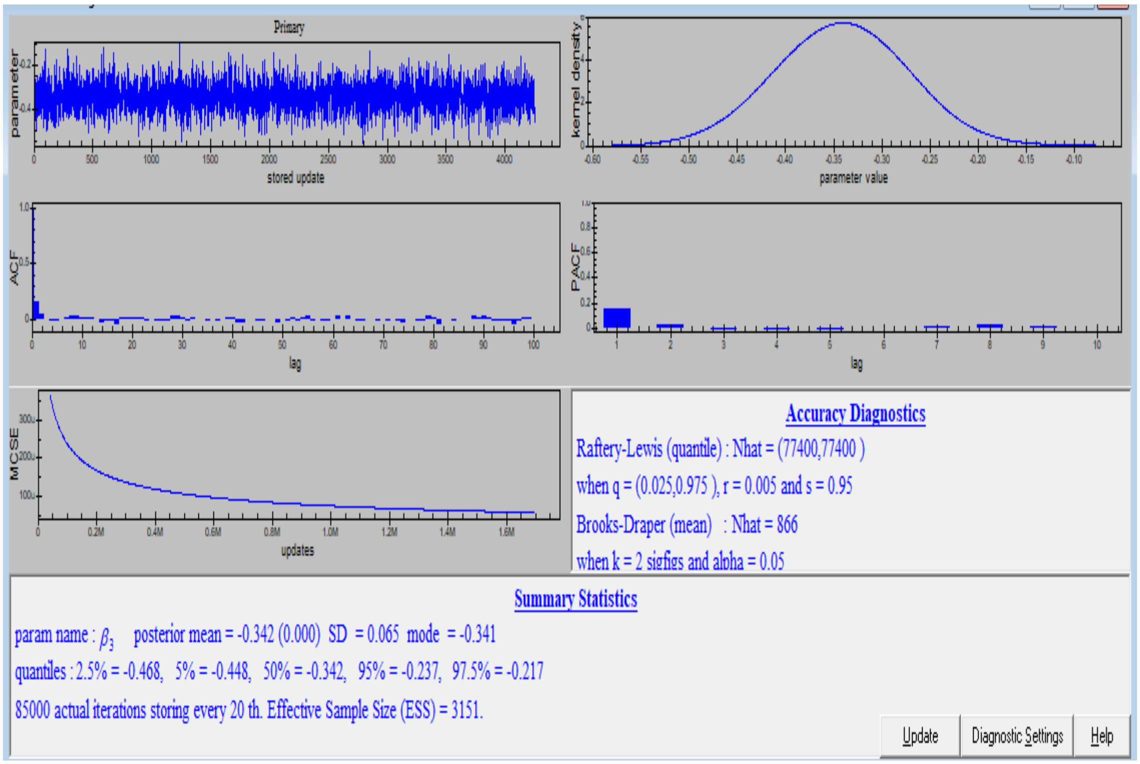
Figure 5.1: Plots of Bayesian Multilevel random Coefficients Convergence Test

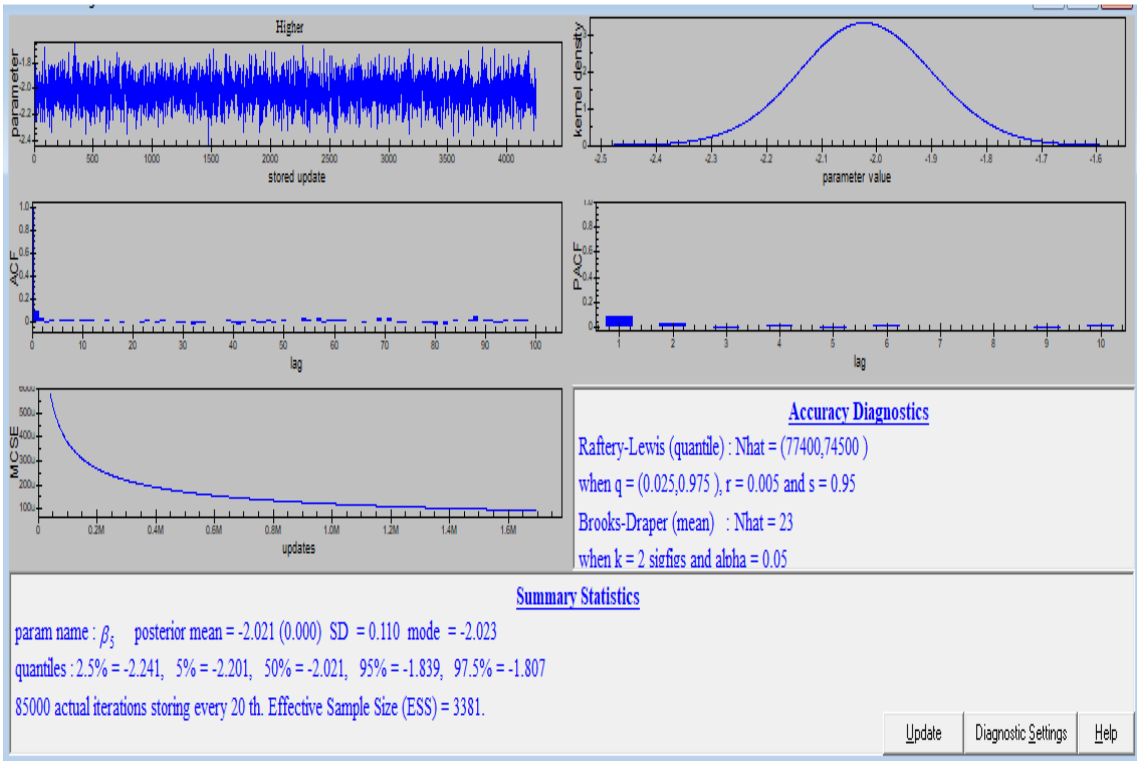


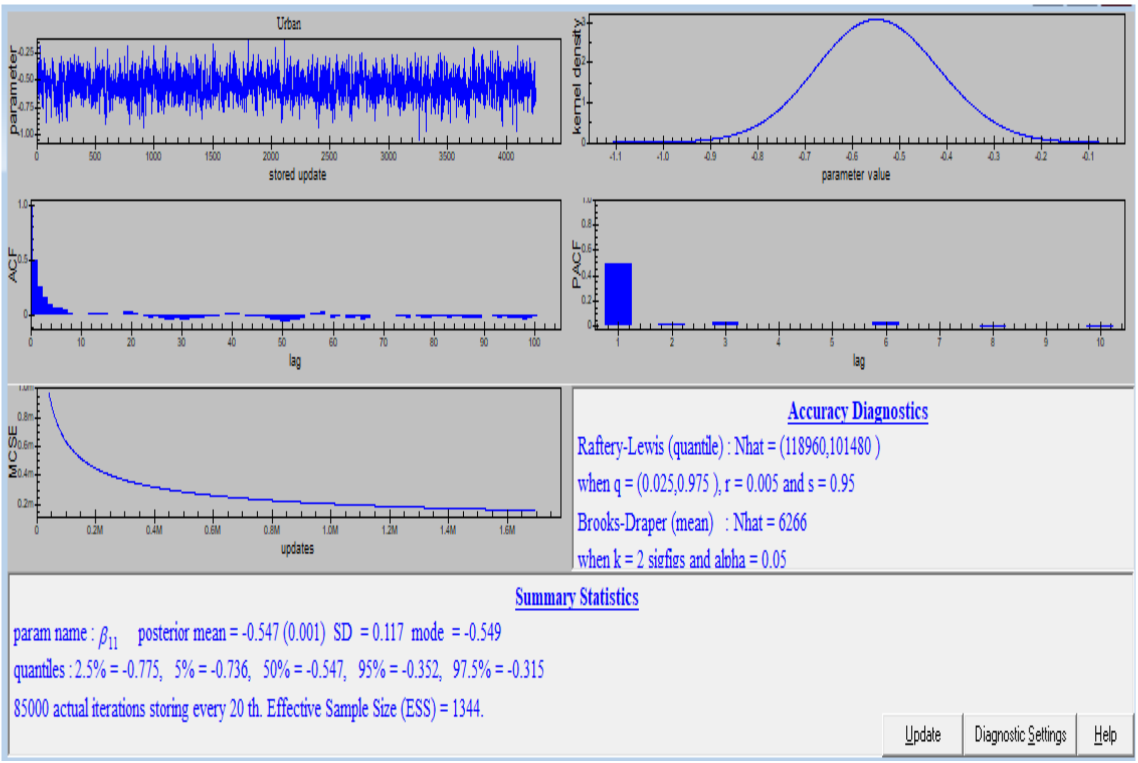



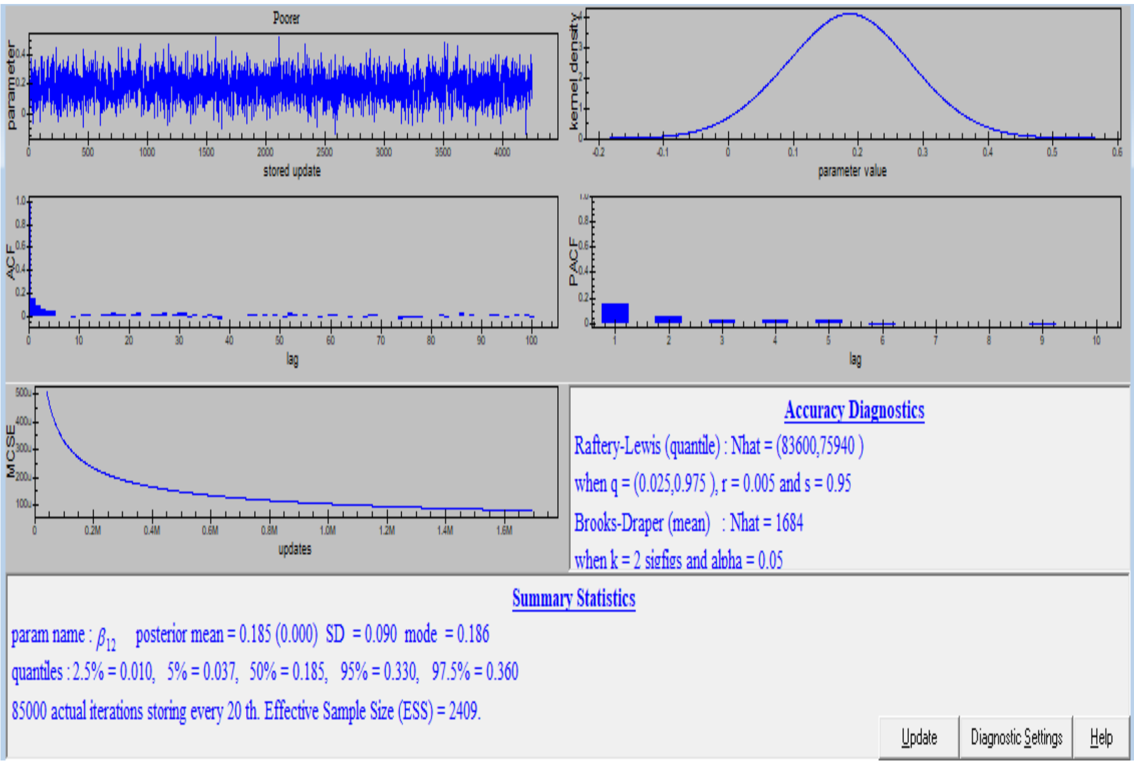

age 25-34
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (80420,85220 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 1208
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_1$ posterior mean = -0.852 (0.000) SD = 0.067 mode = -0.855
quantiles : 2.5% = -0.984, 5% = -0.963, 50% = -0.853, 95% = -0.743, 97.5% = -0.726
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 2243.
Update
Diagnostic Settings
Help

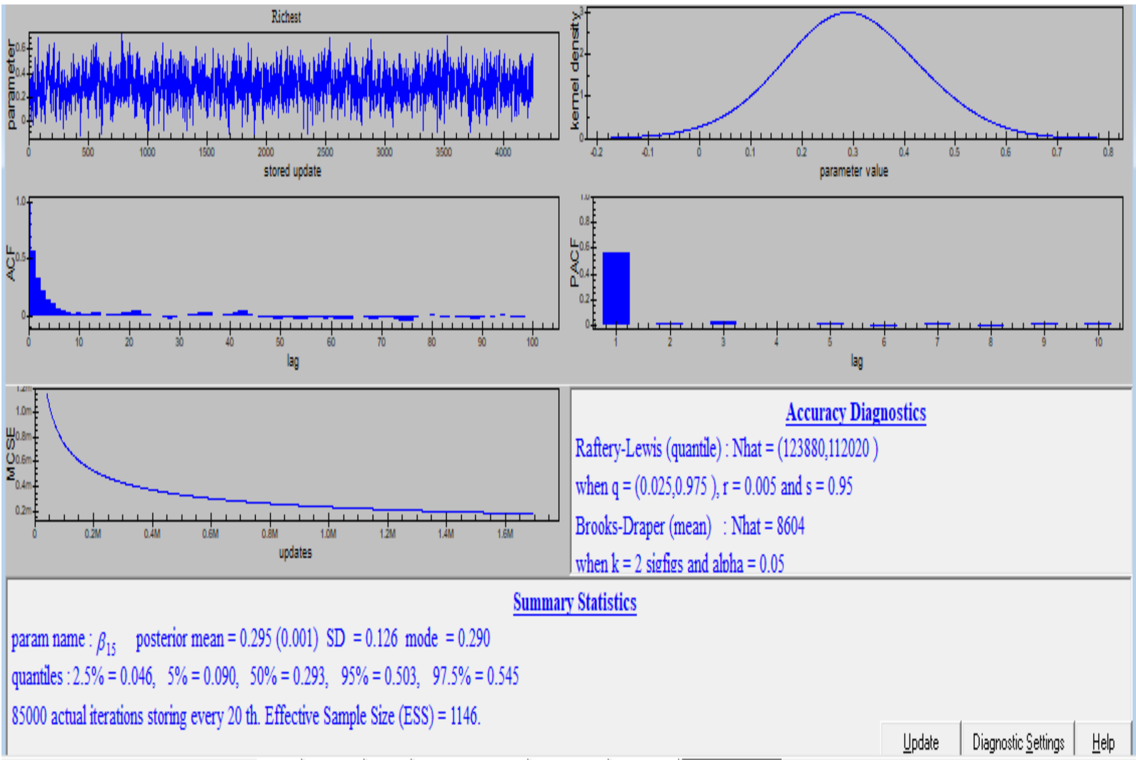

Age 34+
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (80420,86880 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 1496
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_2$ posterior mean = -0.774 (0.000) SD = 0.072 mode = -0.771
quantiles : 2.5% = -0.914, 5% = -0.893, 50% = -0.773, 95% = -0.656, 97.5% = -0.632
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 2076.
Update
Diagnostic Settings
Help

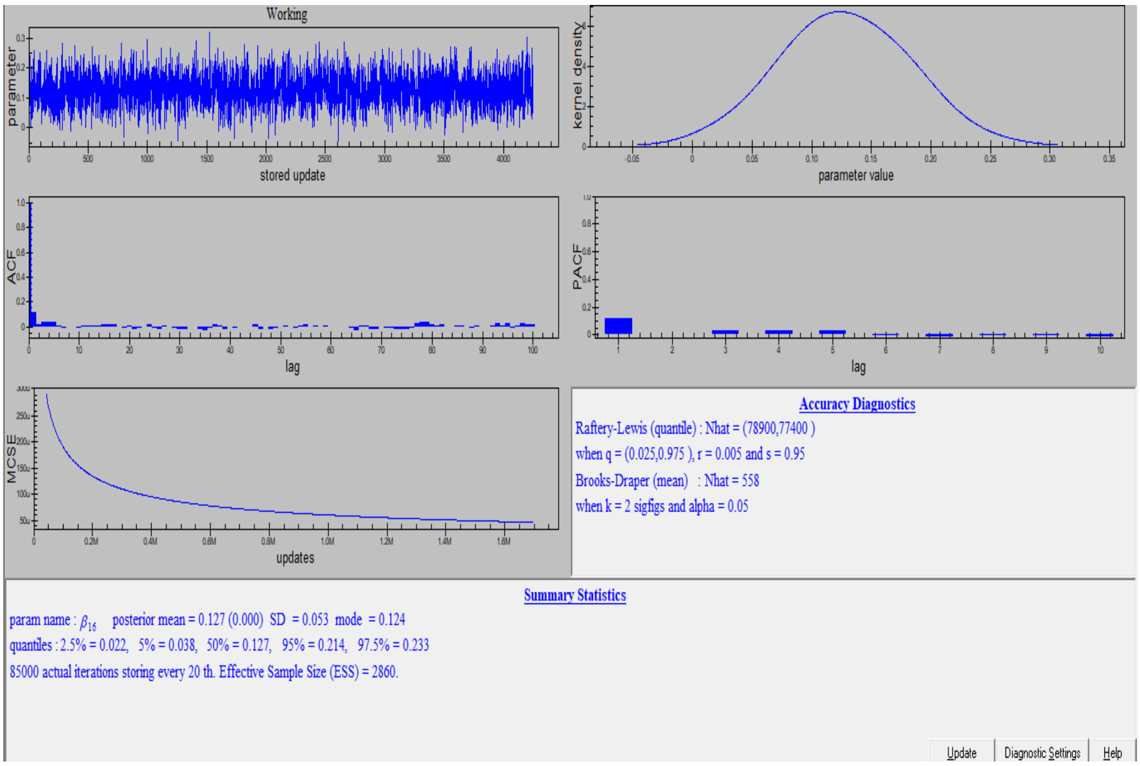

Primary
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (77400,77400 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 866
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_3$ posterior mean = -0.342 (0.000) SD = 0.065 mode = -0.341
quantiles : 2.5% = -0.468, 5% = -0.448, 50% = -0.342, 95% = -0.237, 97.5% = -0.217
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 3151.
Update
Diagnostic Settings
Help

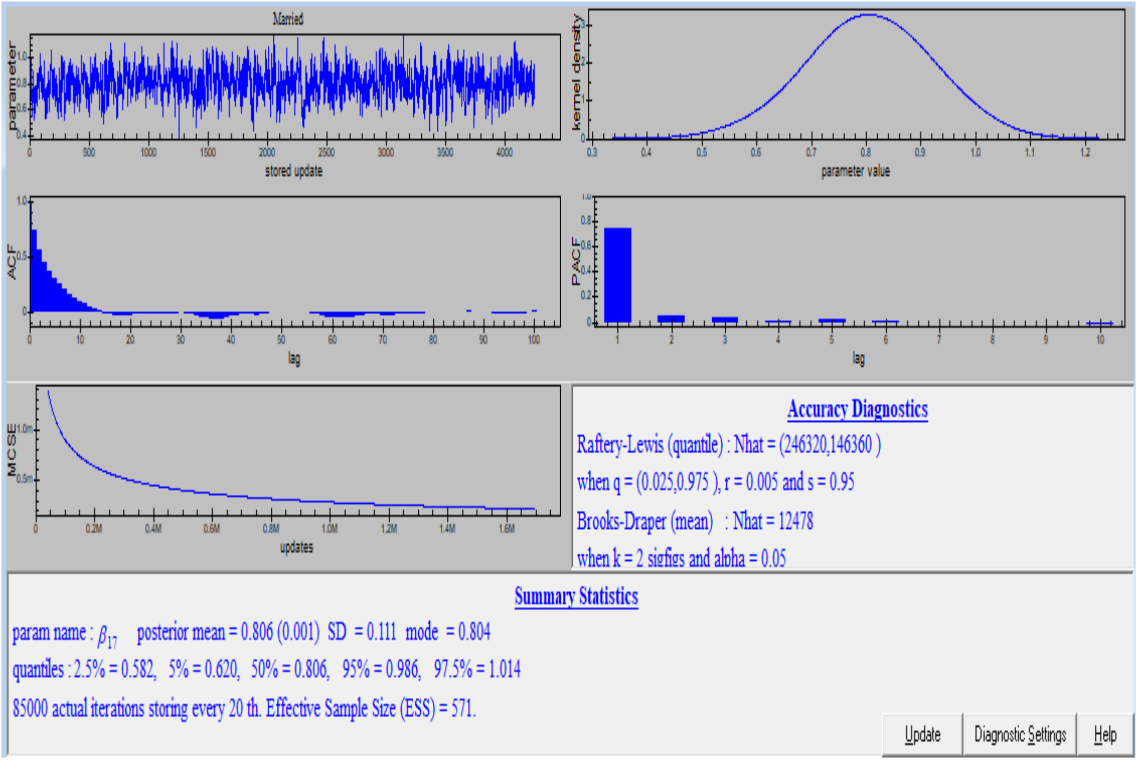

Higher
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (77400,74500 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 23
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_5$ posterior mean = -2.021 (0.000) SD = 0.110 mode = -2.023
quantiles : 2.5% = -2.241, 5% = -2.201, 50% = -2.021, 95% = -1.839, 97.5% = -1.807
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 3381.
Update
Diagnostic Settings
Help

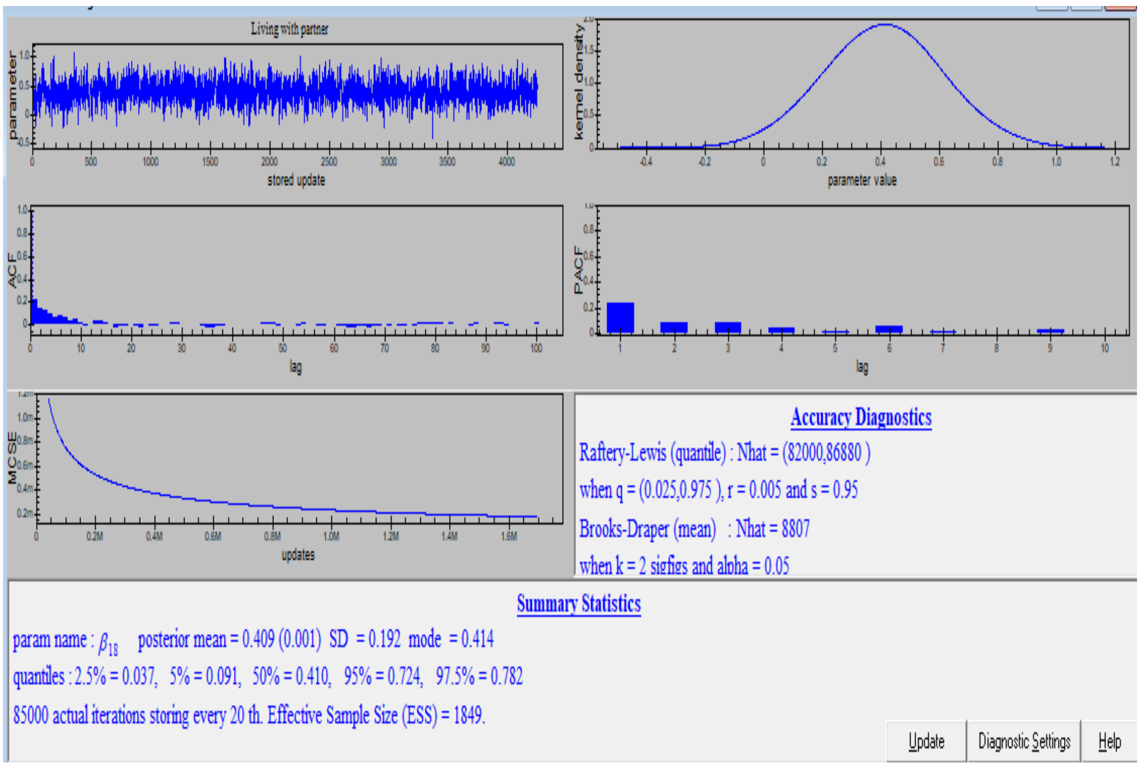

Urban
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (118960,101480 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 6266
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{11}$ posterior mean = -0.547 (0.001) SD = 0.117 mode = -0.549
quantiles : 2.5% = -0.775, 5% = -0.736, 50% = -0.547, 95% = -0.352, 97.5% = -0.315
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1344.
Update
Diagnostic Settings
Help

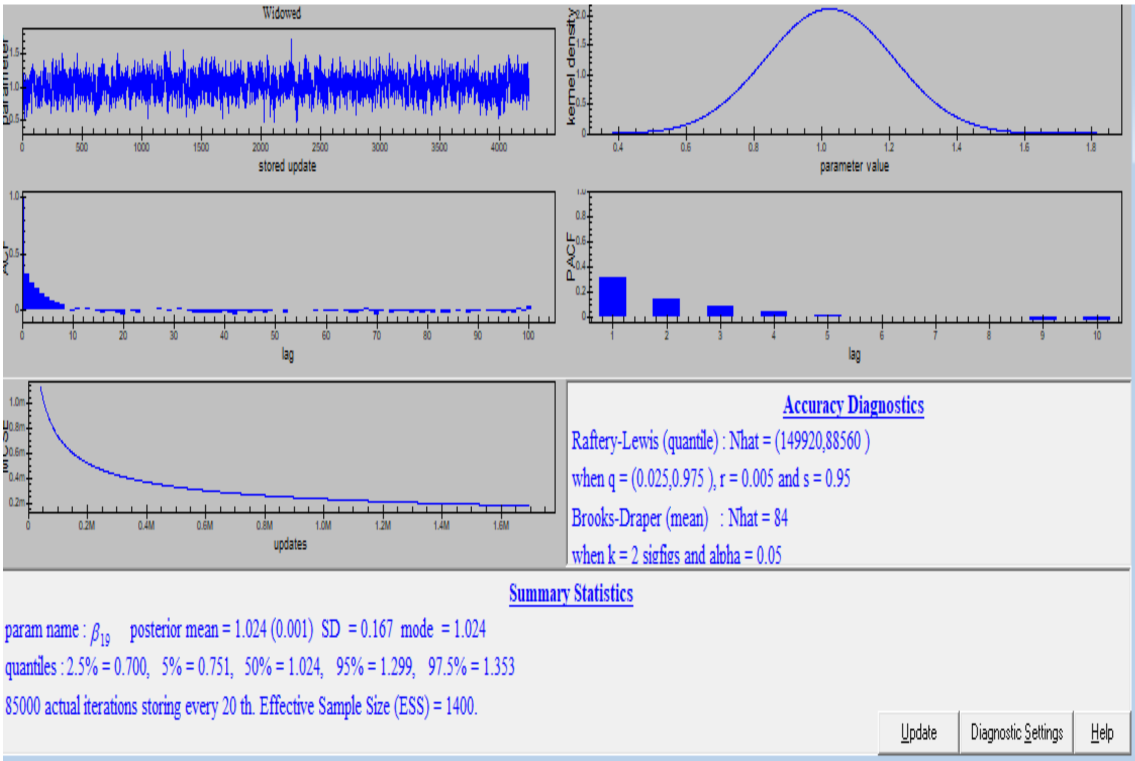

Poorer
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (83600,75940 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 1684
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{12}$ posterior mean = 0.185 (0.000) SD = 0.090 mode = 0.186
quantiles : 2.5% = 0.010, 5% = 0.037, 50% = 0.185, 95% = 0.330, 97.5% = 0.360
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 2409.
Update
Diagnostic Settings
Help

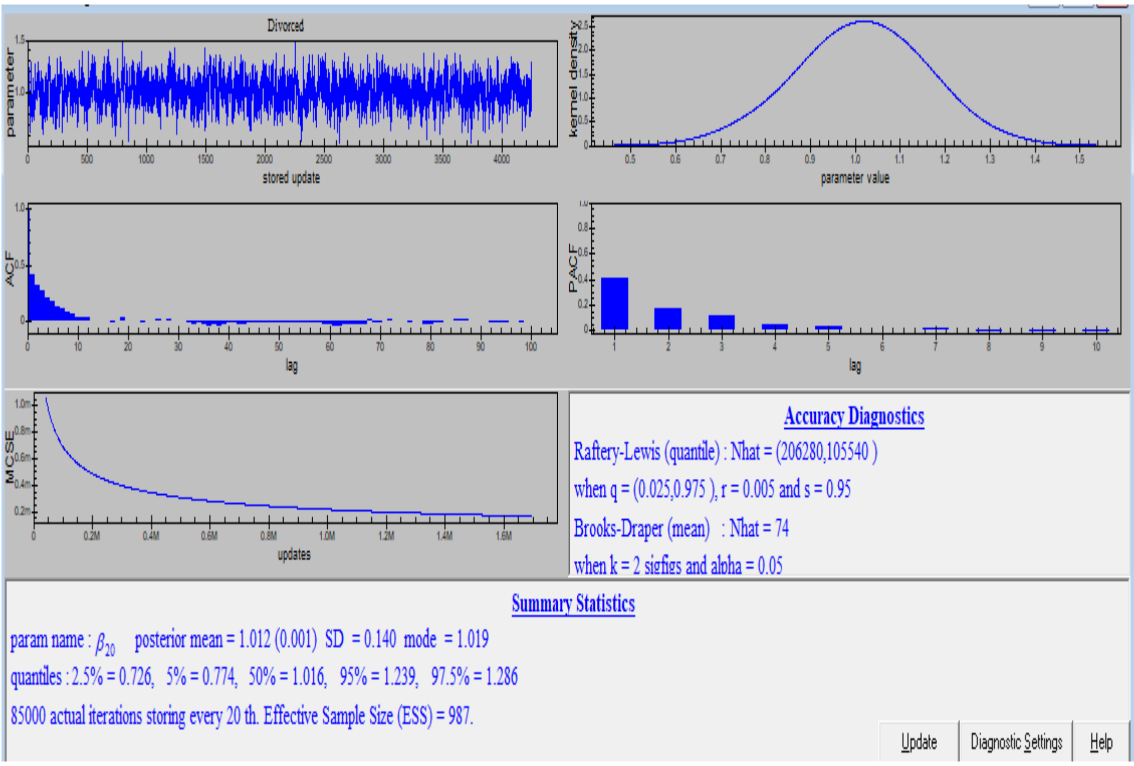

Richest
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (123880,112020 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 8604
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{15}$ posterior mean = 0.295 (0.001) SD = 0.126 mode = 0.290
quantiles : 2.5% = 0.046, 5% = 0.090, 50% = 0.293, 95% = 0.503, 97.5% = 0.545
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1146.
Update
Diagnostic Settings
Help

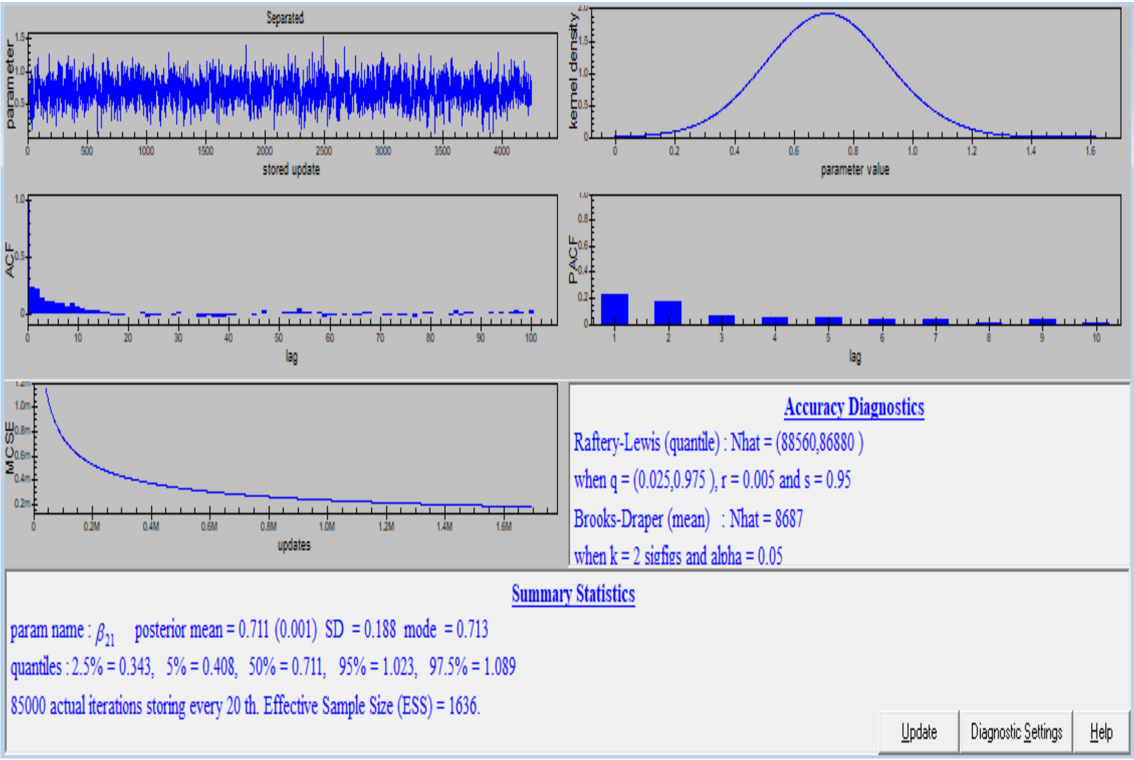

Working
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (78900,77400 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 558
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{16}$ posterior mean = 0.127 (0.000) SD = 0.053 mode = 0.124
quantiles : 2.5% = 0.022, 5% = 0.038, 50% = 0.127, 95% = 0.214, 97.5% = 0.233
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 2860.
Update
Diagnostic Settings
Help

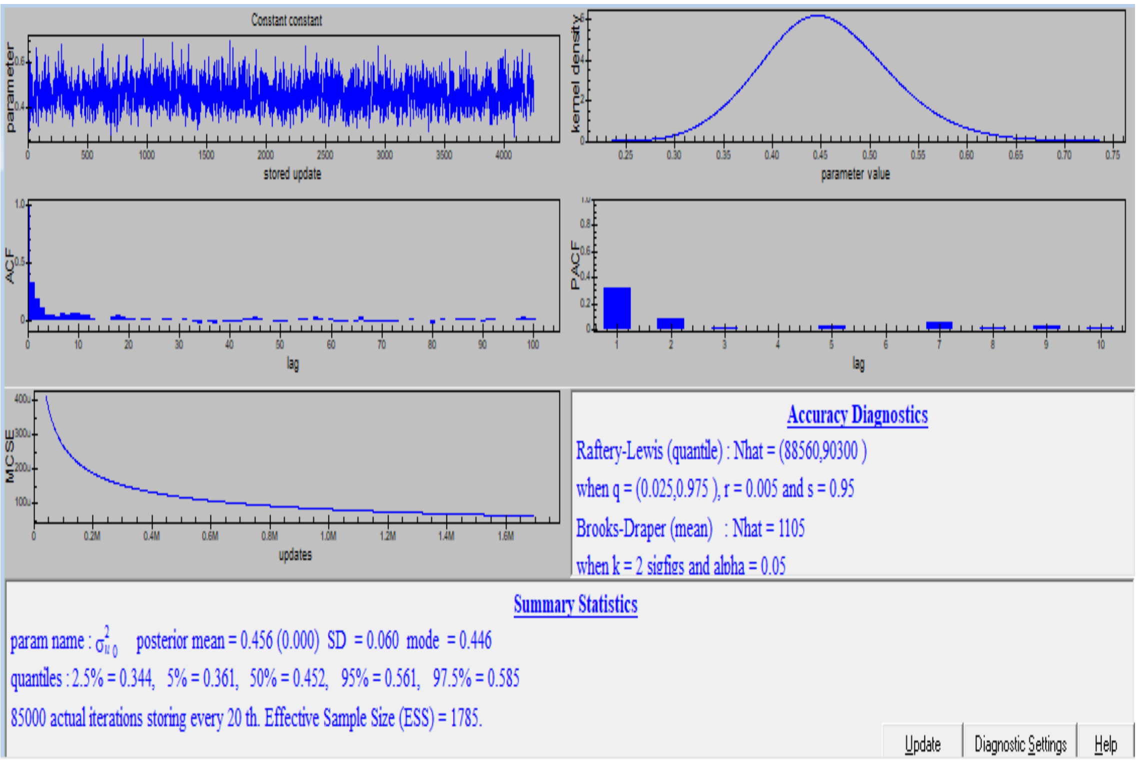

Married
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (246320,146360 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 12478
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{17}$ posterior mean = 0.806 (0.001) SD = 0.111 mode = 0.804
quantiles : 2.5% = 0.582, 5% = 0.620, 50% = 0.806, 95% = 0.986, 97.5% = 1.014
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 571.
Update
Diagnostic Settings
Help

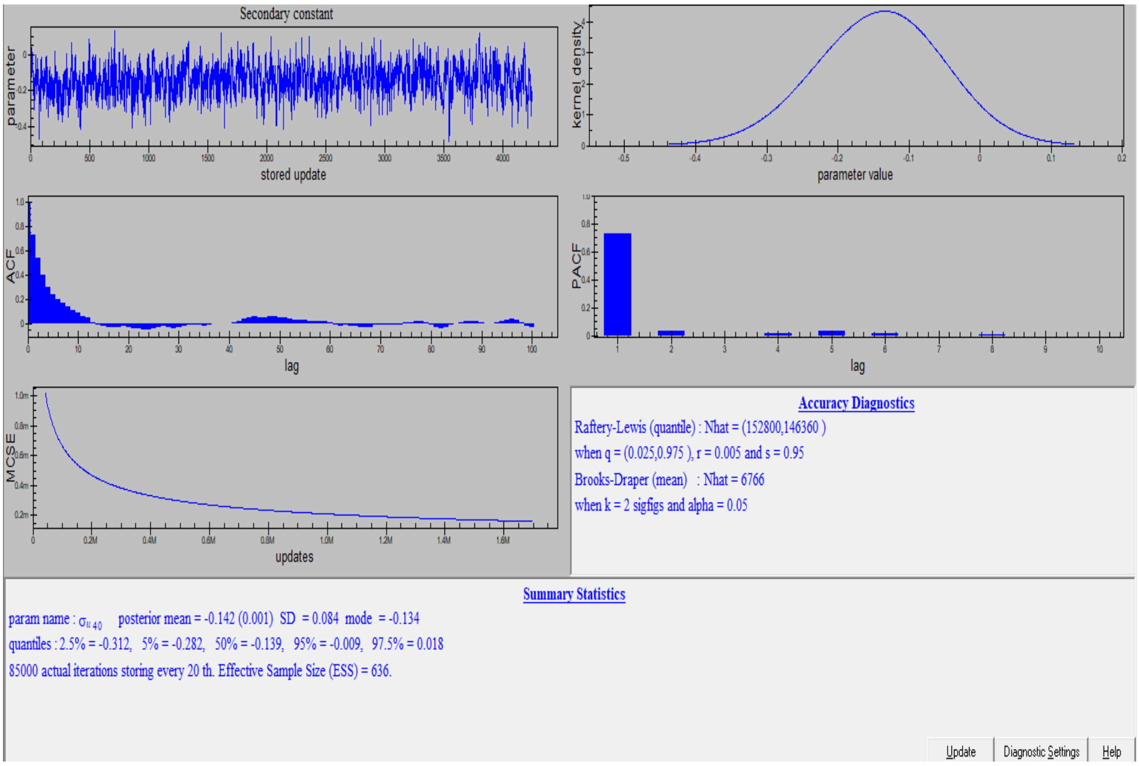

Living with partner
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (82000,86880 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 8807
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{18}$ posterior mean = 0.409 (0.001) SD = 0.192 mode = 0.414
quantiles : 2.5% = 0.037, 5% = 0.091, 50% = 0.410, 95% = 0.724, 97.5% = 0.782
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1849.
Update
Diagnostic Settings
Help

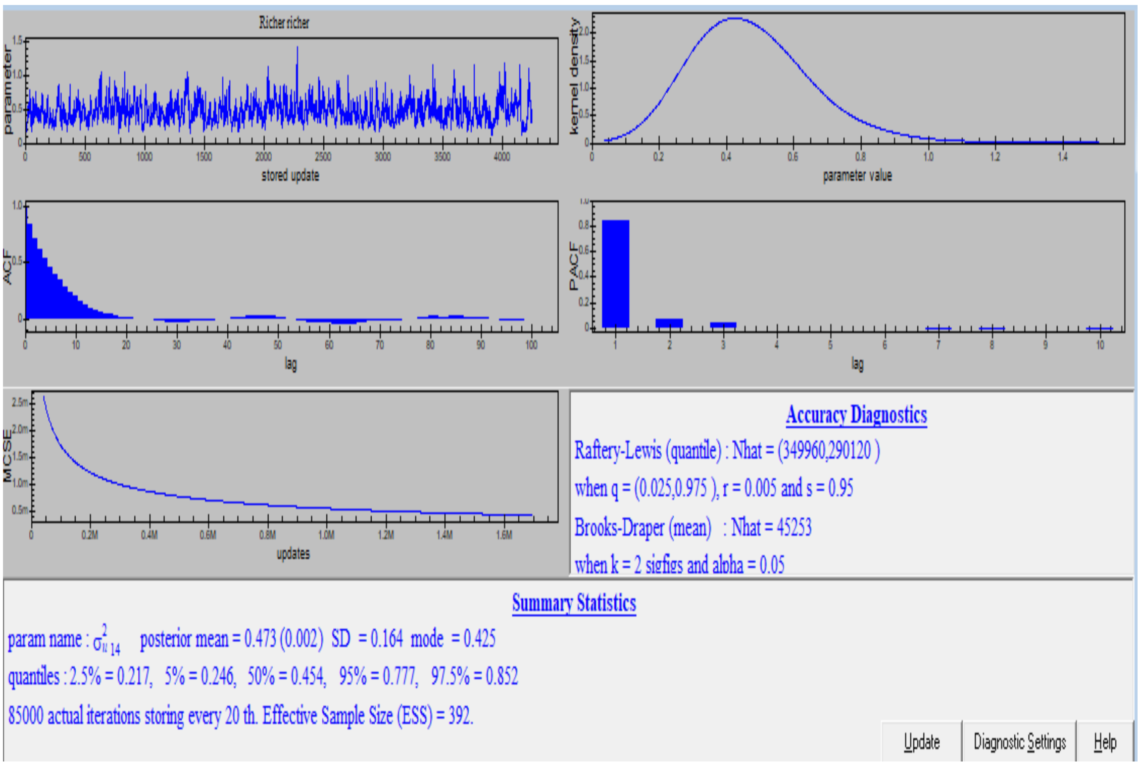

Widowed
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (149920,88560 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 84
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{19}$ posterior mean = 1.024 (0.001) SD = 0.167 mode = 1.024
quantiles : 2.5% = 0.700, 5% = 0.751, 50% = 1.024, 95% = 1.299, 97.5% = 1.353
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1400.
Update
Diagnostic Settings
Help


Divorced
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (206280,105540 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 74
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{20}$ posterior mean = 1.012 (0.001) SD = 0.140 mode = 1.019
quantiles : 2.5% = 0.726, 5% = 0.774, 50% = 1.016, 95% = 1.239, 97.5% = 1.286
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 987.
Update
Diagnostic Settings
Help


Separated
stored update
parameter value
lag
lag
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (88560,86880 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 8687
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\beta_{21}$ posterior mean = 0.711 (0.001) SD = 0.188 mode = 0.713
quantiles : 2.5% = 0.343, 5% = 0.408, 50% = 0.711, 95% = 1.023, 97.5% = 1.089
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1636.
Update
Diagnostic Settings
Help


Constant constant
stored update
parameter value
lag
lag
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (88560,90300 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 1105
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : $\sigma^2_{u0}$ posterior mean = 0.456 (0.000) SD = 0.060 mode = 0.446
quantiles : 2.5% = 0.344, 5% = 0.361, 50% = 0.452, 95% = 0.561, 97.5% = 0.585
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 1785.
Update
Diagnostic Settings
Help


Secondary constant
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (152800,146360 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 6766
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : σ_{u 40} posterior mean = -0.142 (0.001) SD = 0.084 mode = -0.134
quantiles : 2.5% = -0.312, 5% = -0.282, 50% = -0.139, 95% = -0.009, 97.5% = 0.018
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 636.
Update
Diagnostic Settings
Help


Richer richer
parameter
stored update
kernel density
parameter value
ACF
lag
PACF
lag
MCSE
updates
Accuracy Diagnostics
Raftery-Lewis (quantile) : Nhat = (349960,290120 )
when q = (0.025,0.975 ), r = 0.005 and s = 0.95
Brooks-Draper (mean) : Nhat = 45253
when k = 2 sigfigs and alpha = 0.05
Summary Statistics
param name : σ²_{u 14} posterior mean = 0.473 (0.002) SD = 0.164 mode = 0.425
quantiles : 2.5% = 0.217, 5% = 0.246, 50% = 0.454, 95% = 0.777, 97.5% = 0.852
85000 actual iterations storing every 20 th. Effective Sample Size (ESS) = 392.
Update
Diagnostic Settings
Help

# multilevel random coefficient model

```
> fit1<- glmer(tariku$Early_sexual_debut~tariku$Women_Age+tariku$Education+tariku$Religion+tariku$pleace__of_residence+tariku$Wealth_Index+tariku$Respondent_occupation+tariku$Maritul_status+(1+tariku$Education+tariku$Wealth_Index)|tariku$Regions,data = tariku,family = binomial("logit"), glmerControl(calc.derivs = FALSE))
```

| AIC | BIC | logLik | deviance | df.resid |
|---|---|---|---|---|
| 12173.3 | 12587.1 | -6030.6 | 12061.3 | 11906 |

# single level binary logistic regression model

```
> fit2<- glm(tariku$Early_sexual_debut~tariku$Women_Age+tariku$Education+tariku$Religion+tariku$pleace__of_residence+tariku$Wealth_Index+tariku$Respondent_occupation+tariku$Maritul_status++tariku$Education+tariku$Wealth_Index,family = binomial("logit"))

> AIC(fit2)
[1] 12379.27
> BIC(fit2)
[1] 12453.16
> logLik(fit2)
'log Lik.' -6179.633 (df=10)
```