Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110204

Statistical Analysis of Factors Associated with Early Marriage Among Women in Ethiopia: Application of Multilevel Logistic Regression Model

Lata Fekadu Tolera[1,] and Aboma Temesgen Sebu[2]*

[1]Department of Statistics, Jimma University, Jimma,Ethiopia
[2]Department of Statistics, Haramaya University, DireDawa, Ethiopia

Received: 14 Sep. 2020, Revised: 9 Nov. 2020, Accepted: 11 Nov. 2020
Published online: 1 May 2022

Abstract: Early marriage is still widely practiced in many parts of the world mainly in developing countries and dominant in Sub-Saharan countries. Ethiopia is also among theses Sub-Saharan countries which practices early marriage with the highest prevalence rate. The main purpose of this study was also to analyze associated factors of early marriage of women in Ethiopia with the consideration of regional variability of the practices using multilevel model. The study utilized secondary data collected on 9825 married women nested within 9 regions and two city adminstration of Ethiopia obtained from 2016 Ethiopia demographic and health survey(EDHS). Multilevel logistic regression model was applied to identify determinants factors of early marriage and its variation across regions. The results of the study revealed that among total women considered in the study, 60.8% women were married at their early age while 39.2% were married at legal age group. Among the candidate models, random intercept with fixed effect of the covariates was selected based on model information criteria among the candidate multilevel logistic regression models. Regarding the associate factors of early marriage of women, the selected model revealed that place of residence, religion of women, women education level, wealth index, husband education level, husband occupation status and total number of sibling were factors that significantly affects the early marriage of women at 5% significance level. Furthermore, the study revealed women who resides in rural area were 1.239 times more likely to experiences early married than those resides in urban areas and there were regional variability of proportion in women early marriage that the study takes into consideration as unobservable random effects. This study also recommends to implement more effective planning policies that target particular units at regional level particularly, education and wealth index have positive effect in reducing early marriage. As a result special attention needs to be paid to all regions in order to access education and improve the economic status for young women that helps in reducing early marriage.

Keywords: Multilevel model, Intra-class correlation coefficient, Early marriage, Logistic regression model

1 Introduction

Early marriage, also known as child marriage, which is defined as any marriage carried out below the age of 18 years, before the girl is physically, physiologically, and psychologically ready to shoulder the responsibilities of marriage and childbearing [1]. This practice is now and for awhile understood as harmful practice on the health, psychological, physiological and socioeconomic wellbeing of young girls as well as for the newborns [2]. It was estimated that more than 12 million women worldwide first cohabited with a partner before the age of 18 without their will and consent, with most of them living in developing countries [3]. Thus, it is argued that women early marriage is one of the most traditional practices in the globe that play a great role in lowering the status of women and children [4]. This is predominantly practiced in the rural and poor communities of the developing countries where young girls are regarded as economic burden and quickly married off to alleviate household expenses [5].

The practice of early marriage is most common in South Asia, Sub-Saharan Africa and Latin America and the Caribbean, where 48%, 42% and 29% of women aged 15-24 marry before the age of 18 respectively. It is very delicate among the developing countries such as Ethiopia as a result of the intact and deepening tradition, religion and economic

* Corresponding author e-mail: latafekadu@gmail.com

motives which are the major reasons for the persistence of early marriage practice in the country [6]. The trouble was perpetuated by poverty, a lack of education and economic opportunities and social customs that limit the rights of women and girls [7]. Early marriage also threatens the achievement of MDGs such as eradicating extreme poverty and hunger, achieving universal primarily education, promoting gender equality and empowering women, reducing child mortality, improving maternal health and combating HIV/AIDs, malaria and other diseases [8].

According to UNICEF, Ethiopia has the 15^{th} highest prevalence rate of child marriage in the world and 40% of girls in Ethiopia are married off before 18 years old [3]. According to CSA and ICF [9] report, the median age at first marriage among women age 25-49 has increased slightly since 2011 to 2016 from 16.5 years to 17.1 years. During the same period, the percentage of women marrying before age 18 has declined from 63% to 58%. However, this proportion is not the same throughout the country showing Amhara and Tigray region are much higher than the national average (82% in Amhara, 79% in Tigray, 64% in Benshangul, 64% in Gambella and 46% in Afar region. The result of this report also showed that variation in proportion of early marriage across the regions and there was no appropriate methodology applied on to identify factors associated with women early marriage.

Multilevel regression models is used to model within and between source of variation at the same time are increasingly applied in many areas of social and biomedical science data sets containing identifiable units or clusters of observations [10, 11]. The general concept is that individual interacts with the social contexts to which they belong and the properties of those groups are in turn influenced by the individuals who make up that group. Generally, the individuals and the social groups are conceptualized as a hierarchical system of individuals and groups and such systems can be observed at different hierarchical levels and variables may be defined at each level. This leads to social research in to the interaction between variable characterizing individuals and groups respectively [12]. The model is applied to the hierarchically structured data in which the units at one level are clustered with the units of the next higher level by allowing the simultaneous examination of the effects of group level and individual level variation dependence of observations within and between groups variations [13]

The nature of the data considered in this study was also based on two-stage stratified cluster sampling in which the individual observations(marriage) are correlated within groups(regions) having levels. The units at lower level are married women which were nested within units at higher level(regions). Therefore, the prevailing consequence of women early marriage and its variation due to nature of the data calls for intervention in view of identifying the determinants of early marriage and quantifying its variation among women in Ethiopia using multilevel model. This is because of most early marriage study in Ethiopia has been small-scale research, focusing on some part of communities, usually small-sized rural or urban communities. Their geographic scope limits the applicability of their result on a large scale, particularly considering the complex multi-regional and multi-ethnic setting of Ethiopia. In addition, some researcher used logistic regression model and Cox proportional hazard model to estimate the effect of covariates on women early marriage which restricts consideration of regional variability of women early marriage when data are clustered type. However, this study applied multilevel logistic regression model that permit analyzing the loss of independence observations turn out from clustering individual married women in to higher level(regions) which allows to make valid inferences when examining the effect of both individual characteristics and cluster characteristics of women early marriage in Ethiopian.

2 Methodology

Ethiopia is located in the Horn of Africa. It 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 a high central plateau that varies from 1,290 to 3,000 m (4,232 to 9,843 ft) above sea level, with the highest mountain reaching 4,533 m (14,872 ft). Administratively, Ethiopia has nine federal regions and two administrative cities.

2.1 Source of Data

Data considered for this study was 2016 Demographic and Health Survey data, which was among the comprehensive survey conducted as part of the worldwide Demographic and Health Surveys project. The survey was implemented by the Central Statistical Agency (CSA) of Ethiopia and the partner organization under the auspices of the Ministry of Health from January 18, 2016, to June 27, 2016. The data provide in-depth information on marriage, fertility, family planning, infant, child, adult and maternal mortality, maternal and child health, gender, nutrition, malaria, knowledge of HIV/AIDS and other sexually transmitted diseases.

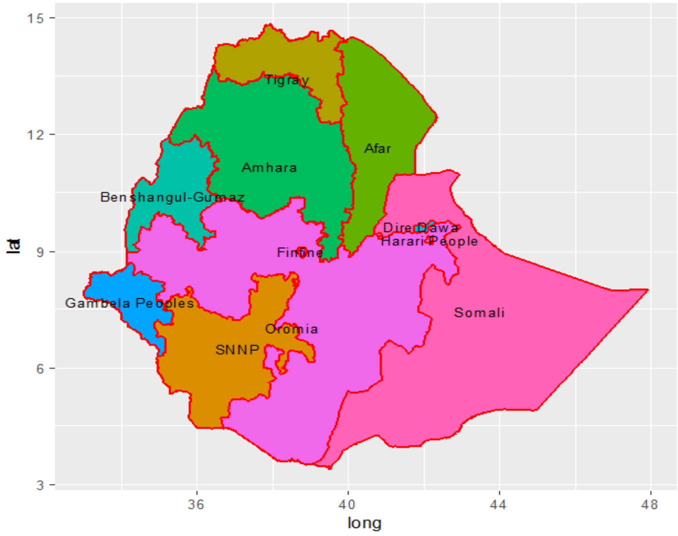

Fig. 1: Map of Study area

2.1.1 Sampling Design and Data Collection Methods

The sample was designed to provide population and health indicators at the national (urban and rural) and regional levels. The sampling frame for to select the samples were obtained from 2007 Population and Housing Census which was conducted by the Ethiopian Central Statistical Agency. Stratified two stage sampling in which each region was stratified into urban and rural areas, giving 21 sampling strata in the fist stage and samples of EAs were selected independently in each stratum in the second stage. Accordingly, total of 645 EAs including 202 EAs in urban areas and 443 EAs in rural areas were selected with probability proportional to the EA size based on the 2007 Census. In the second stage of selection, a fixed number of 28 households per cluster were selected with an equal probability systematic selection. This two-stage sampling data set has hierarchical structure and its hierarchy follows individuals/married women as level-1, and regions as level-2 (i.e. individualS women are nested into regions). All women within the reproductive age group (15-49) years that were ever married or ever lived with a man as a wife, living in Ethiopia during the study of the EDHS, 2016 was considered as target population.

Based on the selected samples all women were asked a series of questions regarding their marital status and whether they had ever married or lived with a man or not. All those who reported that they were ever married or ever-lived with a man, were asked to indicate how old they were at the time when they started, for the first time ever, living with a man as a wife, irrespective of the legality or otherwise of their union. The response to this question constitutes the woman?s age at first marriage. All the women who indicated that they had never been in a union or lived with a man were considered single and as a result they were not asked the question about the age at first marriage. This is the standard way in which age at first marriage is being measured in the worldwide DHS program [14].

2.2 Variables of the Study

2.2.1 Response Variable

The response (outcome) variable of the study was early marriage which was identified based on age at marriage using questionnaires. This age at marriage was categorized as either the age of women at first marriage which is either less than 18 years (early marriage) or 18 years and above (legal marriage) which was binary in nature. Therefore, the response variable for the ith married women was represented by a random variable Y_i with two possible values 1 or 0 and coded as follows:

$$Yi = \begin{cases} 1 & \text{if if age at first marriage is less than 18 years} \\ 0 & \text{if age at first marriage is 18 and above years} \end{cases}$$

2.2.2 Independent Variables

The primary choice of explanatory variables for this study was based on literature reviews and theoretical justification of source of data on factors influencing women early marriage at the global level and in the country. Therefore, those variables, which are reviewed from literature as determinant factors of women early marriage, were displayed on Table 1.

Table 1: The independent variables and values of the category

Variable	Category	Variable	Category
Region	Tigray, SNNP, Afar	Residence	Urban,Rural
	Gambella, Amahara,Harari	No.siblings	less than 5, greater than 5
	Oromia,Addis Ababa,Somali	Religion	Orthodox,Muslim, Protestant, Others
	Dire Dawa,Benishagule	Woman employment	Employed, unemployed
Husband occupation	Agriculture,Professional	Wealth index	Richest,Richer
	Business, Labourers, Others		Middle, Poorer,Poorest
Women's education	No education, Primary	Husband's education	No education, Primary
	Secondary, Higher		Secondary, Higher
Media Exposure	Yes, No		

2.3 Method of Data Analysis

The study explored the EDHS data conducted in 2016 which was related to women early marriage based on techniques developed for analyzing data with categorical response variables. Logistic regression and multilevel logistic regression model were also employed to identify factors related to women early marriage in Ethiopia.

2.3.1 Logistic Regression Model

Logistic regression model is apart of generalized linear models which is extension of ordinary regression model to encompass non-normal response distributions to model functions of the mean. The three components used generalized linear model are random component which identifies the distribution of the response variable; systematic component which used to specify explanatory variables used in linear predictor function and the link function which specify the function of mean of the response that equates to the systematic component [15]. Among these GLM logistic regression is the one used to model categorical binary response variable with one or more explanatory variables based on logit link functions [16]. The assumptions required for statistical tests in logistic regression are far less restrictive than those for ordinary least squares regression. This study was also used the logistic regression to model women early marriage which is binary outcome in nature. Accordingly, let **Y** be a dichotomous outcome representing status of women early marriage with categories 1 (for early married women) and 0(for married women at legal age) and it has binomial distribution which is given by:

$P(Y=y_i)=\binom{n}{y_i}\pi_i^{y_i}(1-\pi_i)^{n-y_i}$ for $i=0,1,2,...,n$

Where, π_i is the probability success (early married women) for i th women from n total women under consideration. Based on this given probability distribution of the response variable let **X** be an n x (k+1) matrix that contains the collection of k-predictor variables the conditional probability that the ith married women experiences early marriage given married women is characterized by X$_i$ in the logistic regression model is defined as:

$Logit(\pi_i)=Log(\frac{\pi_i}{1-\pi_i})=\mathbf{X}^T\beta$

Where β is by k+1 by 1 vector of independent covariate coefficients. The probability of is also derived as: $\pi_i=\frac{e^{\mathbf{X}^T\beta}}{1+e^{\mathbf{X}^T\beta}}$ and the odds of early marriage is defined by:

$\frac{\pi_i}{1-\pi_i}=e^{\mathbf{X}^T\beta}$

Furthermore, the estimated odds ratio value is greater than one indicates that the odds of the outcome in group one is larger than in group two. Thus, married women in a given group is more likely to experience early marriage than the reference group women. Once the model is estimated any factor that affects the probability of success will alter both mean and variance of the observations simultaneously [17].

The estimation of this logistic regression parameters is base on maximum likelihood (ML) estimation method which is the most used technique for model estimation. This method estimates the values for the unknown parameters by maximize the probability distribution of obtaining the observed set of data. Accordingly, let n samples of independent observations

(Y_i, X_i) for i= 1,2,...,n. Then, the joint likelihood function based on Bernoulli distribution of the outcome variable is given by:
$L(\beta|\mathbf{Y}) = \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i}$
The required parameters are estimated by maximizing the log of this likelihood function respect to each β's which is given by:
$ln(L(\beta|Y)) = \sum_{i=1}^{n} y_i ln(\frac{e^{\mathbf{X}^T\beta}}{1+e^{\mathbf{X}^T\beta}}) + (1-y_i)ln(\frac{1}{1+e^{\mathbf{X}^T\beta}})$

2.3.1.2 Model adequacy Checking and variable selection

The number of variables to be included in the model should be the minimum possible that is parsimonious and deliver optimum information. This study also begins variable selection process with univariate analysis of each independent variable on the response variable separately. Based on this fitted model any predictor variable with p-value ≤ 0.25 was considered in model [18] followed by stepwise automatic followed by including all predictor variables in the model to select significant predictors for the model based on Wald statistic.

After the section of the significant predictors for the model, the assessment for the goodness of fit which investigates how close values were predicted by the model compared with that of observed values was done based on likelihood ration test and Hosmer-Lemeshow test evaluation criteria [19]. The procedures are used to evaluate over all model including the significance test each explanatory variable in the model.

Likelihood ratio test

The likelihood ratio test (LRT) was used to assess of overall adequacy of the fitted logistic regression model. This test uses the ratio of the maximized value of the likelihood function for the full model (L_1) over the maximized value of the likelihood function for the reduced model (L_0) and its test statistic is given by:
$G^2 = -2ln(\frac{L_0}{L_1}) = -2(lnL_0 - lnL_1)$
Where, G^2 has a chi-square with degrees of freedom equal to the difference between the numbers of parameters estimated in the two models [20]. The test used to test the null hypothesis that all population logistic regressions coefficients are not significantly different from zero versus the alternative one.

Hosmer-Lemeshow test

The Hosmer-Lemeshow goodness of fit test is used to assess whether the number of predicted events from the logistic regression model reflect the number of observed events in the data. The data are ranked according to the predicted probability of the outcome from the model that is being evaluated. This test based on grouping cases in deciles in the sense that it is obtained by applying a chi-square test on contingency table by partitioning the predicted probabilities using the percentiles of the predicted event probability. Based on the test the model is not good fit, if most of subject with success are classified in the higher deciles of risk and those with failure in the lower deciles of risk and the significance of the test is less than 0.05. It is used to test the model prediction does not significantly differ from the observed versus the alternative based on Hosmer and Lemeshow test statistic is given by:
$\hat{C} = \sum_{j=1}^{g} \frac{(O_j - E_j)^2}{V_j})$
Where, $E_j = n\pi_j, V_j = n\pi_j(1-\pi_j)$, g is the number of group, O_j is the observed number of events in the j^{th} group, E_j is the expected number of events in the j^{th} group, V_j is the variance correction factors for the j^{th} group. If the observed number of events differs from what is expected by the model, the statistic will be large and there will be evidence against the null hypothesis that the model is adequate to fit the data. This statistic has an approximate chi-square distribution with degree of freedom $g-2$ [21].

Statistical tests of individual predictors

The Wald test statistic which is commonly used to test the significance of logistic regression coefficients for each independent variable. It is used to test whether the parameter associated with an explanatory variable is zero or not. For a particular explanatory variable, or group of explanatory variables, if the Wald test is significant, then it implied that parameters associated with these variables are non-zero, so that the variables should be included in the model which is

performed based on Wald test statistic given by:
$W = Z^2 = (\frac{\hat{\beta}}{SE(\hat{\beta})})^2 \sim \chi_1^2$
The statistics W is approximately distributed as chi-square with one degree of freedom [22].

2.3.2 Multilevel logistic regression model

Multilevel logistic regression model is used to model nested data by allowing analysis at several levels simultaneously, rather than having to choose at which level to carry out a single level analysis. This enable the extent of variation in the outcome of interest (women early marriage particular to this study case) to be measured at each level assumed in the model both before and after the inclusion of explanatory variables in the model. The previous logistic regression assumed that all subjects are independent in the sense that any variables affecting the dependent variable have the same effect in all regions. But, the multilevel modeling relaxes this assumption and allows these variables effects to vary across regions in hierarchical nature of the data [13]. This study also considered two-level binary logistic regression model accounting the married women at first level and regional at second level effects. Within each j^{th} region there are n_j married women. Let Y_{ij} be the binary response for i^{th} married woman in j^{th} region; and X_{ij} the associated factors of women early marriage. We define the probability of women early marriage as π_{ij} which is modeled using a logit link function can be described as:
$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \mathbf{X}_j^T\beta + \beta_{0j}$
Where, $U_j \sim IID(O, \sigma_u^2)$, U_j is the random effect at level two, without consideration of random effect U_j the model is similar to standard logistic regression model. Therefore, conditional on U_j the Y_{ij} can be assumed to be independently distributed. In order to know the variation at each level, this model can be decomposed into two components which is for individual level and the other is at regional level given by:
$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \mathbf{X}_j^T\beta$ Model for individual level and
$\beta_{oj} = \beta_0 + U_j$ Model for the regional level
The intercept β_{0j} consists of the fixed effect component(β_0) and the random effect component(U_j).

2.3.2.1 Heterogeneous Proportions

Heterogeneous proportion is the basic data structure of two-level logistic regression which is a collection of N groups, units at level-two which is within j^{th} region there are a random sample of n_j married women units. The outcome variable is dichotomous denoted by Y_{ij} for i^{th} married women nested in j^{th} region and the total sample size is $W = \sum_{j=1}^{N} n_j$. If one does not take explanatory variables into account, the probability of success is assumed constant in each group [13]. Let the early married women in group (region) j be denoted by π_j. The dichotomous outcome variable for the i^{th} married women in j^{th} region Y_{ij} can be expressed as the probability of women early married in group (region) j, $E(Y_{ij}) = \pi_j$ plus some individual dependent residual ε_{ij}that is:
$Y_{ij} = \pi_j + \varepsilon_{ij}$
The residual term is assumed to have mean zero and variance, var(ε_{ij}) = $\pi_j(1-\pi_j)$ since the outcome variable is coded 0 and 1, the group sample average is the proportion of women being early married in group (region) j given by:
$\hat{\pi}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} Y_{ij}$, where $\hat{\pi}_j$ is an estimate for the group-dependent probability and the overall sample average is the overall proportion of early marriage $\hat{\pi}$ is given as:
$\hat{\pi} = \frac{1}{W}\sum_{j=1}^{N}\sum_{i=1}^{n_j} Y_{ij}$ where W is total sample size
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 which is defined as:
$\chi^2 = \sum_{j=1}^{N} n_j \frac{(\hat{\pi}_j-\hat{\pi})^2}{\hat{\pi}(1-\hat{\pi})}$ and the statistic follows approximately a chi-square distribution with N-1 degrees of freedom.

The between group which is given $\psi^2 = S^2_{between} - \frac{S^2_{wthin}}{\tilde{n}}$ Where,$\tilde{n} = \frac{1}{N-1}(W - \frac{\sum_{j=1}^{N} n_i^2}{W})$ is also closely related to the chi-square test statistic when its calculated based probability of success which is given by:
$S^2_{between} = \frac{\hat{\pi}(1-\hat{\pi})}{\tilde{n}(N-1)}\chi^2$ Where, χ^2 is the calculate value for the heterogeneous proportions and the within group variance which is a function of group averages is given by:
$S^2_{within} = \frac{1}{W-N}\sum_{j=1}^{N} n_j\pi_j(1-\pi_j)$
The estimation of random intercept only model or the multilevel model with no predictors is also important to specify these between and within groups source of variation. The model is specified as:

$Logit(\pi_{ij}) = \beta_0 + U_{0j}$ Where, $U_{0j} \sim IID(0, \sigma_0^2)$
Where,β_0 the population average of the transformed probability and U_{0j} is the random deviation from this average for group j. For the deviations U_{0j} it is assumed that they are independent random variables with a normal distribution with mean zero and variance σ_0^2. This model does not include a separate parameter for the individual level variance because of the individual level residual variance of the Y_{ij} follows Bernoulli distribution directly from the probability of having women early married (π_j) which is given by: var(ε_{ij})= $\pi_j(1-\pi_j)$ denoted by π_0.

Another important reason of fitting multilevel logistic regression model with only random intercept is to identify the existence of intra-class (intra-regional) correlation arising from similarity of early marriage for women 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 and its values are computed as follows based on random intercept only model.

$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2}$

Where,σ_e^2 is variance of individual (lower) level units. In multilevel logit model level one residual is $\sigma_e^2 = \frac{\pi}{3} = 3.29$ which is reformulated as:

$ICC = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + 3.29}$

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 differ between groups. We assume that there are variables which potentially explain the observed success and failure. These variables are denoted by X_h (h=1,2,k) with their values indicated by X_{hij}. Since some or all of these variables could be factors of level one (married women), the probability of women early marriage is not necessarily the same for all individual in a given group [13]. Therefore, the success probability depends on the married women as well as the group (region), and is denoted by π_{ij}. The outcome variable is split into an expected value and residual as:

$Y_{ij} = \pi_{ij} + R_{ij}$

Where Y_{ij} is the dichotomous outcome variable for the ith married women in jth region,π_{ij} is the probability of i^{th} women early marriage in region j and R_{ij} is the residuals. The random intercept model expresses the log-odds, i.e. the logit of π_{ij} as a sum of a linear function of the explanatory variables for the random intercept with fixed predictors is defined as:

$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + X_h^T\beta_h + U_{0j}$

Where, logit(π_{ij}) does not include a married woman residual because it is an equation for the probability of having early married women π_{ij} rather than for the outcome Y_{ij}. $\beta_{0j} = \beta_0 + U_{0j}$ is assumed to vary randomly and probability of success is given by:

$\pi_{ij} = \frac{e^{\beta_0 + X_h^T\beta_h + U_{0j}}}{1+e^{\beta_0 + X_h^T\beta_h + U_{0j}}}$ Where, β_h is a unit difference between the X_h values of two individuals in the same group which associated with a difference of β_h in their log-odds, or equivalently, a ratio of exp(β_h) in their odds.

Furthermore, there are a cases of modifying this assumption by allowing the difference between the effects of explanatory variables to vary across regions which needs the consideration of random coefficients for the predictors. This model represents heterogeneity in relationship between the explanatory variables and the observed outcomes. Suppose that there are k factors of women early marriage $X_{1ij}, X_{2ij}, ..., X_{kij}$, and consider the model where all predictor variables have varying slopes and random intercept. That is

$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_{0j} + X_h^T\beta_h + X_h^T U_h$ by substituting the values of β_{0j} this equation is as:

$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + U_{0j} + X_h^T\beta_h + X_h^T U_h$ by decomposing into two component this model can be specified as:

$Log(\frac{\pi_{ij}}{1-\pi_{ij}}) = \beta_0 + X_h^T\beta_h + U_{0j} + X_h^T U_h$

The two components of this model represents the first component $(\beta_0 + X_h^T\beta_h)$the fixed effect parts and the second component $(U_{0j} + X_h^T U_h)$the random effect part. The coefficients of the random effect $U_{0j}, U_{1j}, U_{2j}, ..., U_{kj}$ are assumed to be independent between groups but may be correlated within groups and has multivariate normal distribution with zero mean vector and variances and co-variances matrix.

The most common methods to estimate parameters of these multilevel logistic regression models were based on maximum likelihood (ML) method. This method has two prevailing approximation, marginal quasi-likelihood (MQL) [23,24] and penalized quasi-likelihood (PQL)[25,26]. After applying these quasi likelihood methods, the model will be estimated using iterative generalized least squares (IGLS) or reweighted IGLS (RIGLS)[23]. Both MQL and PQL are based on Taylor series expansion to achieve the approximation of parameter estimation. Based on the first and second term of Taylor series expansion, MQL and PQL are often known as first order and second order of MQL and PQL, respectively. However the maximum likelihood method parameter estimation using penalized quasi-likelihood doesn?t provide the model comparison statistics (deviance and information criteria) while marginal quasi-likelihood (MQL) using approximation of the integrand Laplace approximation and Gausse-Hermite quadrature using R software can provide. In addition to that it is more complicated due to the need to perform numerical integration to obtain a marginal likelihood to maximize. So that the study used approaches based on Laplace approximation and Gausse-Hermite quadrature of glmer R package to specify number of Gauss-Hermite quadrature points only when there is one random effect.

2.3.2.2 Adequacy test and comparison of the fitted models

Likelihood ratio based on deviance of the fitted were employed to test adequacy of the fitted models. The likelihood ratio test compares the deviance (-2LL) of two models by subtracting the smaller deviance (model with more parameters) from the larger deviance (model with reduced number of parameter). The difference is a chi-square test with the number of degrees of freedom equal to the difference number of different parameters in the two models. In the case where the empty model is compared to a full model, the likelihood ratio test provides information about whether the predictors in the model together account for a significant amount of variance in the dependent variable [27]. After assessing the adequacy of the fitted models AIC and BIC were used to make comparison among the candidate models. The AIC and BIC fit indices are based on the deviance statistic, but they incorporate penalties for a greater number of parameters. The smaller the AIC or BIC value, the better is the model which are computed as follows:

$AIC = -2log(Likelihood) + 2k$

$BIC = -2log(Likelihood) + K * log(N)$

Where k is the model degrees of freedom calculated as the rank of variance c covariance matrix of the parameters and N is the number of observations used in estimation [28].

3 Results and Discussions

3.1 Description of Women Early Marriage

The study considered a total of 9825 married women from nine regional states and two administrative cities of Ethiopia. Among these women about 5976 (60.8%) of the women married at early age while 3849 (39.2%) of them were not married at this age. Among the total women who had experienced early marriage, 4838 (80.96%) of them were from rural areas, and 1138 (19.04%) were from urban areas of Ethiopia. Regarding the regional distribution of early marriage, the large proportion of women (835(13.97%)) and 833(13.94%) of them were in Amhara and Oromia respectively followed by SNNP of whom 695(11.63%) of women had experienced early marriage whereas the low proportion of early marriage in Addis Ababa accounting 216(3.61%) and followed by Dire Dawa accounting 314(5.25%) of women among the total women considered for the study. The result in 4.1 also depicted the variation early marriage was varied by religion, showing highest percentage 2771(46.37%) of them were Muslim religion followers and 2100 (35.14%) of the women were orthodox religion followers whereas small (1035 (17.32%)) proportion of early marriage were observed in protestant religion followers and only about 70 (1.17%) of the women who had experienced the early marriages were the followers of other religion. Regarding educational level of women who had experienced early marriage, from the total about 3751 (62.77%), 1579 (26.42%), 434(7.26%),and 212 (3.55%) of the women were non educated and attended primary, secondary and higher education respectively. Furthermore, the descriptive result shows most (4039 (67.59%)) of early married women were unemployed whereas only 1937 (32.51%) of these early married women had employment. Similarly, regarding the wealth index distribution of early married women about 1933 (32.35%) of the women were from the poorest family whereas about 858 (14.36%) and 907 (15.18%) of early married women were from the rich and middle income families respectively(Table 2).

Furthermore, Table 2 depicted that the percentage of early marriage for women not exposed to any mass media messages via Radio, TV or newspapers were 4691(78.50%) as compared to exposed group 1285(21.50%). With regard to husband?s education level, women whose their husband's who has no educated were highly experienced early marriage 2985(49.95%) than women with primary 1940(32.46%), secondary 634 (10.61%) and it was lower (417 (6.98%)) for women that their husband attended higher education. Similarly, the highest proportion was observed among women who?s their husband(partners) occupation was agriculture 2780 (46.52%) while the lowest proportion of women who experiences early marriage was recorded in women that their husband occupation was laborers 32 (0.54%).In relation to total number of siblings 3045 (50.95%) early marriage were observed among women came from family size greater than five while, 2931 (49.05%) were occurred among women came from parents with family size less than five, which was almost equal.

The Chi-square test result in last column of Table 2, which was used to know the association between these factor and women early marriage, showed that region place of residence, religion of respondent, women education level, wealth index, women work status, media exposure, husband education level, husband occupation status and total number of sibling were significantly associated at 5% level of significant. This chi-square test of association was also considered as preliminary result for the logistic regression model which is used to know the effects of these factors on women early marriage.

Table 2: Distribution of socioeconomic and demographic related determinant factors of women early marriage in Ethiopia

Variable	Categories	Age at First Marriage			
		Legal Marriage n(%)	Early marriage n(%)	Total n(%)	Chi-square (P-value)
Region	Tigray	341 (8.86)	620(10.37)	961 (9.78)	436.468(1.6e-87)
	Affar	221(5.74)	637(10.66)	858(8.73)	
	Amhara	294(7.64)	835(13.97)	1129(11.49)	
	Oromia	488(12.68)	833(13.94)	1321(13.45)	
	Sumale	419(10.89)	556 (9.30)	975(9.92)	
	Benishangul Gumes	292(7.59)	511(8.55)	803(8.17)	
	SNNPR	515(13.38)	695(11.63)	1210(12.32)	
	Gambella	270(7.01)	443 (7.41)	713(7.26)	
	Harari	259(6.73)	316(5.29)	575(5.85)	
	Addis Ababa	463(12.03)	216(3.61)	679(6.91)	
	DireDawa	287(7.46)	314(5.25)	601(6.12)	
Residence Place	Urban	1370(35.59)	1138(19.04)	2508(25.53)	337.3414(3.7e-75)
	Rural	2479(64.41)	4838(80.96)	7317(74.47)	
Religion	Orthodox	1488(38.66)	2100(35.14)	3588(36.52)	53.1194(1.7e-11)
	Muslim	1596(41.47)	2771(46.37)	4367(44.45)	
	Protestant	756(19.64)	1035(17.32)	1791(18.23)	
	Others	9(0.23)	70(1.17)	79(0.80)	
Women Education	None	1961(50.95)	3751 (62.77)	5712(58.14)	188.736(1.1e-40)
	Primary	1164(30.24)	1579(26.42)	2743(27.92)	
	Secondary	420(10.91)	434(7.26)	854(8.69)	
	Higher	304(7.90)	212(3.55)	516 (5.25)	
Wealth Index	Richest	1455(37.80)	1284(21.49)	2739(27.88)	311.926(2.8e-66)
	Richer	459(11.93)	858(14.36)	1317(13.40)	
	Middle	460(11.95)	907(15.18)	1367(13.91)	
	Poorer	476(12.37)	993(16.62)	1469(14.95)	
	Poorest	1000(25.98)	1933(32.35)	2933(29.85)	
Women Employment	Employed	1435(37.28)	1937(32.51)	3372 (34.32)	24.6262(7.7e-07)
	Unemployed	2414(62.72)	4039(67.59)	6453(65.68)	
Media Exposure	Yes	1204(31.28)	1285(21.50)	2489(25.33)	117.3375(1.8e-27)
	No	2645(68.72)	4691(78.50)	7336(74.67)	
Husband Education	None	1518(39.44)	2985(49.95)	4503(45.83)	282.007(7.7e-61)
	Primary	1128(29.31)	1940(32.46)	3068(31.23)	
	Secondary	596(15.48)	634(10.61)	1230(12.52)	
	Higher	607(15.77)	417(6.98)	1024(10.42)	
Husband Occupation	Agriculture	1359(35.31)	2780(46.52)	4139 (42.13)	139.8685(3.0e-29)
	Professional	1003(26.06)	1374(22.99)	2377(24.19	
	Business	606(15.74)	639(10.69)	1245(12.67)	
	Labourers	36(0.94)	32(0.54)	68(0.69)	
No. of sibling	≤ 5	2019(52.46)	2931(49.05)	4950(50.38)	10.8834(0.001)
	>5	1830(47.54)	3045(50.95)	4875(49.62)	

3.2 Associated Factors of Women Early Marriage in Ethiopia

Based on the associated factors according to the chi-square test result of Table 2 results of logistic regression analysis were obtained by using stepwise inclusion of variables that has significant association with women early marriage. The overall model evaluation, statistical tests of individual predictors and goodness-of-fit statistic were employed for each fitted model. While fitting the model initial log likelihood(the null model likelihood value)value taken into consideration to know the contribution of each predictor to the model and the result of fitted model with the predictor variable showing decrease in log likelihood value shows the contribution of considered variable to model as wells as its significant association with women early marriage. Accordingly, the initial -2LL value was 13156.2 before inclusion of the predictor variables to the model and when the model was refitted with inclusion of proposed factors associated with women early marriage in stepwise manner there was the reduction of initial log likelihood value indicating that these proposed factors have significant association with early marriage of women. The significance of each regression coefficient was also tested based Wald and score chi-square test. The test result also showed that region, place of residence, religion of respondent, women education level, wealth index, husband education level, husband occupation status and total number of sibling have significant effects at 5% level of significance as resulted in Table 7 of appendixes.

The goodness of the fitted logistic regression model was also tested based on likelihood-ratio and Hosmer and Lemeshow test. The result this test as depicted in Table 3 showed that the full model has minimum AIC and BIC values compared to the null model(fitted model with no predictor variables) showing the full model was an appropriate fit of the data compared with null model. Similarly, the likely hood ratio test which is based on the difference of log likely hood value of the null model and the full model resulting the value of 690.75 with p-value <0.0001 showed that the full model was the good fit of the data compared with null model. The Hosmer-Lemeshow test result with the calculated statistic 11.03 on 8 degree freedom with estimated p-value of 0.2 also showed that there was no difference between the observed and the model predicted values since the calculated p-value was larger than 5% level of significant. The test result of both tests showed that the model was good fit of the data and factors included in the model has significant effect on women early marriage.

Table 3: Summary of intercept only(null model)and full models statistics for overall model test

Model Information Criteria	**Null Model**	**Full Model**	**χ^2 Value(p-value)**
-2LL	13156.2	12465.45	690.75(< 0.0001)
AIC	13158.2	12529.62	
BIC	13165.4	12759.62	
	χ^2 Value	**DF**	**P-value**
Hosmer-Lemeshow Test	11.030	8	0.200

3.2.1 Results of multilevel logistic regression

With identification of important factors which has significant effects on women early marriage based on the logistic regression without the consideration hierarchical effect of the data and the multilevel logistic regression was followed which considers the hierarchical nature of the data for the improvement of the model. While fitting multilevel logistic model married women were considered as lower level (level-one) units and region was considered as the higher level (level-two) units in which the married women were nested within 11 regions of Ethiopia.

Before directly proceeding to multilevel analysis heterogeneity early marriage proportions of women among the regions of Ethiopia were tested. This test result which was based chi-square showed presence of heterogeneity based on the calculated test statistic value 436.468 and the result showed the evidence for heterogeneity of women early marriage among regional state of Ethiopia which leads to apply multilevel logistic regression model was an appropriate by considering the regional effect variations.

To select good multilevel logistic model, the three fitted candidate multilevel logistic regression were compared based on LRT (deviance based likelihood ratio test) and information criteria or model diagnostic statistic (AIC and BIC). As depicted in Table 4 of model comparison the candidate model were compared based on the model information criteria. When we compare the null model of logistic regression with null multilevel logistic regression model, the null multilevel logistic regression has minimum log likelihood value (12771) compared to likelihood of logistic regression 13156.2 and the resulting likelihood ratio test value of 383.94 with estimated p-value <0.0001 showed that null multilevel logistic regression improves model.

Furthermore, the null multilevel logistic regression likelihood value of 12771 compared with fitted model of multilevel logistic model random intercept and fixed coefficient factors likelihood value of 12511.39 was greater and their likely hood ration test statistic value of 260.86 showed that fitting the multilevel logistic regression model with fixed coefficient of factors improves the model. With similar comparison based log likelihood value and likelihood ratio test, when this multilevel logistic regression with random intercept and fixed coefficient of the covariates compared with random coefficient multilevel logistic regression model the multilevel logistic regression with random intercept with fixed coefficient of the factors has minimum log likelihood value and making the coefficient of covarietes does not improve the model. Therefore, based on the performed likelihood ratio test and information criteria multilevel model with random intercept with fixed coefficient of the factors which also has minimum AIC nad BIC values compared to other candidate model was considered as the better fit of the data.

Table 4: Comparison of multilevel logistics models with information criteria and Likelihood Ration(LLR)

Model	-2LL	LLR value(P-value)	DF	AIC	BIC
Null with random intercept	12771	283.946(0.000)	2	12775	12789
Random intercept with fixed coefficient	12511.39	260.8586(0.000)	23	12557.39	12,722.82
Random coefficients	12505.65	5.734(0.3319)	28	12561.66	12763.05

To predicts the probability of women being early married random intercept model with no predictor variable was fitted because of the model is simplest non-trivial specification of the hierarchical linear model which only the intercept varies between level two units. As shown in Table 5, the log-odds of women early marriage given in all regions under investigation was estimated as $\hat{\beta}_0$ = 0.4013 or the probability of women early marriage in Ethiopia was estimate 59.9% without the effect of other factors and the estimated variance for the random effect at regional level,σ^2_{U0}= 0.2163. The random effects of intercept only implied that the between region variance of women early marriage is 0.2163 and tells that there is a significant difference in proportion of women early marriage across regions of Ethiopia.

Table 5: Results of Parameter Estimate of Intercept-Only Model with Random Effect

Factors	Estimate(S.E)	Z	P-value	Odds Ratio
Fixed Effects				
Intercept($\hat{\beta}_0$)	0.4013(0.142)	2.826	0.0047	1.493
Random-effects				
var(U_{0j})(σ^2_{U0})	0 .2163(0.09488)	2.28	0.01131	0.09161

The significance of random effect at 5% level of significance also indicated that adding region as a random effect or between-region variance in the model was necessary in order to assess variation of women early marriage among regions.This confirming that there was statistically significance difference in proportion of women early marriage across regions providing that multilevel model is the best option to account the regional variation of women early marriage.The result of this finding was also in line with findings of Assefa et al.[29] and Workineh et al. [30]. Assefa et al.[29] suggested that women with the same characteristics in two different regions have different age at first marriage which might be because of the fact that differences in lifestyle, culture, ethnic or environmental determinants between different regions that confirm early marriage differs significantly by region of residence.

Another importance of this null model with random intercept was to calculate the between region variations based on intra-class correlation coefficient (ICC) which is the measure of the correlation between two individuals who are in the same higher level unit (region). As indicated in the methodology section on how to compute ICC low values ICC indicates that relatively small between region variations. Accordingly, from estimated random effect value of 0.2163 and given constant 3.29 the intra-class correlation coefficient was estimated 0.0616 based on the ICC formula. This estimated value depicted that about 6.16% of the variation in women early marriage can be explained by grouping the married women into regions and the remaining (100 - 6.16% = 93.84%) of the variation in experiencing early marriage was explained within region-lower level units.

After examining estimates of the region effects or residuals random effects obtained from the null model. The calculated the residuals to produce a caterpillar plot with the region effects shown in rank order together with 95% confidence intervals was depicted in Figure 2 for all 11 regions in the sample. The residuals represent regional departures from the overall mean, so a region whose confidence interval does not overlap the line at zero (representing the mean value of early marriage practice across all regions) is said to differ significantly from the average at the 5% significance level. At the lower side of the plot, there is a cluster of regions whose mean practice of early marriage was lower than average and vice versa. To assess the effect of independent variables on women early marriage, the study considered random intercept model. That is the probability of women early marriage is vary across regions, but we assumed that the effects of each explanatory variables are the same for each region. As shown in Table 5 the variance component of random intercept only multilevel model is 0.2163, whereas the variance for the random intercept with fixed coefficient model was estimated 0.1045 as depicted in Table 6. The variance of random intercept logistic model decreased compared to that of the random intercept only model. The reduction of the random effect of the intercept variance is due to the inclusion of fixed explanatory variables. That is, taking in to account the fixed independent variables can provide extra predictive value on women early marriage among regions.

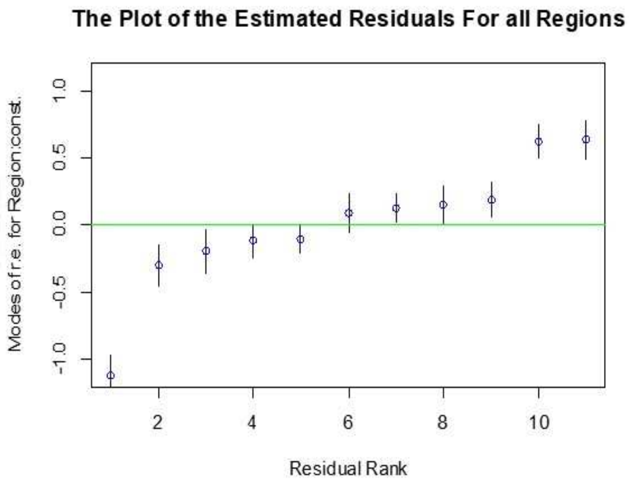

Fig. 2: Plot of estimated residuals (random effects) for all regions of women early marriage.

Table 6: Result of Parameter Estimation for random intercept with fixed coefficient multilevel logistic regression model

Factors	Categories	Estimate(SE)	P-value	OR	[95%C.I for OR]
Fixed Effects					
	Intercept	0.28702(0.13692)	0.03614**	1.332	[1.018,1.742]
Residence	Urban(ref.)				
	Rural	0.2148(0.090)	0.017*	1.239	[1.038,1.479]
Religion	Orthodox(ref.)				
	Muslim	0.1629(0.065)	0.012 **	1.176	[1.035,1.337]
	Protestant	0.0343(0.077)	0.658	1.034	[0.888,1.205]
	Others	1.6933(0.361)	<0.001***	5.437	[2.679,11.036]
Women Education	Not educated(ref.)				
	Primary	-0.1383(0.053)	0.009*	0.870	[0.784,0.966]
	Secondary	-0.0728(0.087)	0.40	0.929	[0.783,1.103]
	Higher	-0.236(0.112)	0.035*	0.789	[0.633,0.983]
Wealth Index	Richest(ref.)				
	Richer	0.1964(0.098)	0.046*	1.217	[1.002,1.477]
	Middle	0.2192(0.101)	0.030*	1.245	[1.020,1.518]
	Poorer	0.2816(0.101)	0.005*	1.325	[1.086,1.616]
	Poorest	0 .1140(0.097)	0.243	1.120	[0.925,1.357]
Women Employment	Working (ref.)				
	Not working	-0.0180(0.047)	0.704	0.982	[0.894,1.078]
Media Exposure	Yes(ref.)				
	No	-0.0941(0.062)	0.130	0.910	[0.805,1.028]
Husband education	Not educated(ref.)				
	Primary	0.0617(0.053)	0.246	1.063	[0.958,1.180]
	Secondary	-0.1647(0.074)	0.027 **	0.848	[0.732,0.981]
	Higher	-0.5411(-0.086)	<0.001***	0.582	[0.491,0.689]
Husband Occupation	Agriculture(ref.)				
	Professional	-0.0650(0.062)	0.295	0.937	[0.829,1.058]
	Business	-0.2246(0.073)	0.002***	0.798	[0.691,0.923]
	Labourers	-0.3037(0.255)	0.234	0.738	[0.447,1.217]
	Others	-0.1579(0.061)	0.010 **	0.853	[0.756,0.963]
No. of sibling	< = 5 (ref.)				
	>5	0.1020(0.043)	0.019*	1.107	[1.016,1.205]
Random Effects					
$\hat{\sigma}^2_{U0}$	0.1045166(0.047)	0.0132*	1.11	–	[0.0427, 0.2557]

In the results of the random intercept with fixed slope model, the fixed part showed that place of residence, religion of respondent, women education attainment, wealth index, husband education attainment, husband occupation status and total number of sibling were found to be statistically significant factors of women early marriage in Ethiopia at 5% level of significance.

The estimated model revealed that the odds of early marriage, in women who resides in rural area were 1.239 time morosely than women resides in urban. The study conducted by Erulkar et al.[31] using logistic regression and Workineh et al.[30] also found that significant effect of women residence on their early marriage.Similarly, the finding of Adebowale et al.[32] using Chi-square and Cox proportional hazard models showed that in Nigeria women who reside in rural area married earlier than their counterpart in urban area. This could be rural areas tend to have institutional and normative structures such as the kinship and extended family that promote early marriage and it might be the fact that women in urban areas might highly participate or attain education when compared with women in rural areas that resulted them to develop skills, gain resources and achieve maturity to manage an independent household and thus they might be delay marriage.

Regarding religion effects, the odds or early marriage in muslim religion follower women were 17.6% more experiences early marriage and it was 3.4% more in protestant religion followers in comparison with orthodox followers. Similarly, the odds of early marriage in other religion followers were 5.437% more likely than the orthodox religion followers and this result showed significant effect of religion on women early marriage.The result of this study confirmed with study conducted by Maswikwa et al.[33] that showed among 12 Sub-Saharan African countries certain religious affiliations were positively associated with child marriage where its prevalence was higher among women who believed in Islamic religion, traditional religions or no religion than those women who were Christians. Furthermore, the estimated model depicted that women educational level has significant effects on their earl marriage and women who attended primary education were 13.00% less likely to experiences early marriage than women who had no education.In the same way women who attended higher education were 21.10% less likely to experience early marriage than women with no education level controlling the effect of other variables constant. Study conducted by Peninah et al.[34] in western Uganda using Cox?s proportional hazard model also found that the risk of early marriage for women who attended primary and secondary education were lower as compared to non-educated women and a lower risk of getting early married among educated women may be due to waiting time for schooling and understanding the side effect of early marriage. It is also in line with the study by CSA of Ethiopia [35] found in Ethiopia age at first marriage greatly increases with education; women with more than secondary education get married almost eight years later than those with no education.

Another significant factors on women early marriage was wealth index and the estimated odds depicted that women who were richer and middle in economic status were 21.7% and 24.5% more likely to experience early married than women with richest economic status family respectively holding the effects of other factor constant.Similarly, the odds of women early marriage those came from poorer family were 1.325 times more likely to be early married as compared to those from the richest family. This result was also consistent with the study conducted by Workineh et al., 2015 in Ethiopian case and study conducted by Hotchkiss et al.[36] in Serbia which showed that women living in the poorest quintile of households were married earlier compared to middle wealth and the richest wealth group.

Another significant factor on women early marriage was husband background on education level and occupation. Regarding husband educational level, odds of women being early married were 15.2% and 41.8% for women husband educational level was secondary and higher education level as compared to women that their husband do not have any educated level.The study done in Bangladesh by Zahangir and Kamal [37] based on multiple logistic regression model found that no(less) educated women, marry with no(less) educated husbands and it is argued that higher educational attainment was the main force underlying the delay age at first marriage among females.Another significant factor related to Husband was occupation status of the women husband and the result showed that the odds of women being early married were reduced by 20.2% and 14.7% for women that their husbands business worker and other work category (i.e. excluding agriculture, professional, business and laborers) respectively as compared to women that their husbands were farmer husband keeping the random effect and other variables in the model constant. The effect of this husband occupation was also confirmed with the study conducted by Peninah et al.[34] and Sishaw et al.[38].

The number of sibling was also another significant factors associated with women early marriage in Ethiopia as the result shown in the fitted model. This result also depicted that chance of women early marriage increase as the number of family size increased. Thus, the odds ratio of women early marriage was increased by 10.7% for women came from total number of sibling greater than five as compared to less family size. The effect of this number of siblings for the married women also agreed with finding of Sishaw et al.[38] that showed that family size wa significantly associated with women early marriage.

3.2.2 The selected model diagnosis

The diagnostic plot for residuals like the normality for Pearson and standardized residuals of the multilevel model were also presented under Figure 3. Therefore Q-Q plot for normality of residual and scatter plots of diagnostic checking model was performed. The Q-Q plot from the following Figure in first panel verifies that the residuals are close to normally distributed. Q-Q plots for normality of random effects at regional levels are also given in the figure at panel two, and illustrates that the intercept (regional) random effects are normally distributed with mean zero and variance. Furthermore, the scatter plot from the Figure in third and forth panel verifies that the residual of the model versus index or residual versus the estimated probability did not show any systematic pattern. This suggested that the residuals are symmetric around zero (i.e. positive and negative residuals are almost equal). For acceptable fit one would expect that locally the residual average zero, the smooth line helps in detecting a deviation from this expectation. Therefore, from these residual plots the model fit is very well. Thus, the fitted multilevel logistic regression model is good fit for the given data.

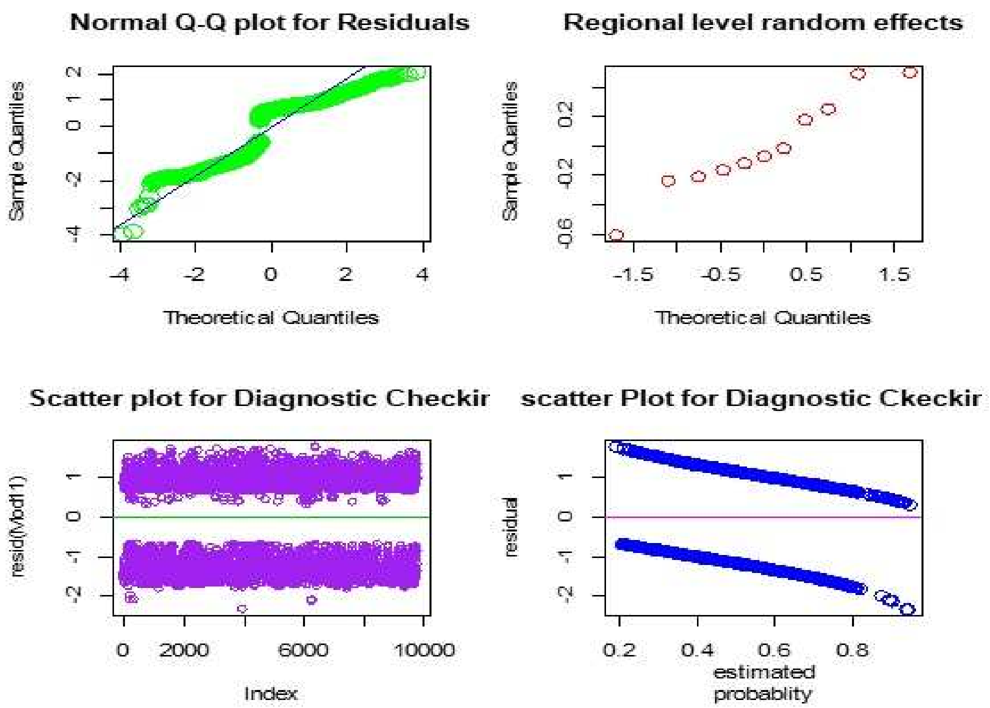

Fig. 3: Diagnosis plots for the multilevel logistic regression model

4 Conclusions

Among the candidate models considered for the study random intercept with fixed effects multilevel logistic regression model an good fit of women early marriage in Ethiopian and indicated that there was significant variation of women early marriage across the regions and city administrations of Ethiopia. The intra-cluster correlation(ICC) result revealed the correlation between women who were in the same region experiencing early marriage within region was higher than that of between regions.

The model result showed that residence, religion, women education level, wealth index, husband education attainment, husband occupation level and total number of sibling were the significant factors that affects the women early marriage in Ethiopia. Women who resides in rural area were more likely to experience early marriage than those lived in urban.

The reason for this disparity among urban rural residences might be access to education in urban areas compared to rural areas. Furthermore, women from no or low educational, living in poor family and large family size were more likely to experiences early marriage than educated or attended higher educational level, living in rich family and from less than five family sizes respectively.

This study will also forwards the consideration within and between cluster variation of pregnancy termination in Ethiopia for upcoming researchers on the same area. Additionally, special attentions to uneducated, rural areas resident,poor family background and uneducated husband of women by governmental or nongovernmental organizations to reduce the risk of women early marriage in Ethiopia as recommendations based the study results.

Acknowledgement

We acknowledge the Ethiopian central statistical agency for providing us the data utilized for the study.

Conflicts of Interests
The authors declare that they have no conflicts of interests

References

[1] UNIFPA, 2006. Ending child marriage, A guide for global policy action International Planned Parenthood Federation and the Forum on Marriage and the Rights of Women and Girls. Available at: $https://www.refworld.org/docid/48abd594c.html$
[2] ICF International, 2011. Ethiopia Demographic and Health Survey. Central Statistical Agency, Addis Ababa, Ethiopia.
[3] UNICEF, 2017. Ending Child Marriage Report. New York: UNICEF
[4] ICRW, 2010. Too young to Wed: Education and Action towards Ending Child Marriage. ICRW, Washington DC, USA.
[5] Kyari, G.V. and Ayodele, J., 2014. The socio-economic effect of early marriage in North Western Nigeria. Mediterranean Journal of Social Sciences, 5(14), p.582.
[6] Kanta, B., 2013. Child marriage in Ethiopia- A matter of concern. International journal of Management and Social Sciences, 2(5), p.74.
[7] Mengistu, M.M., 2017. Early marriage in Ethiopia: so little done but so much to do. Bus Manag, 9(3), p.1.
[8] UN., 2007. Arab millennium development goals report 2011: an inclusive approach to development in a time of transition. UN.
[9] CSA[Ethiopia]and ICF, 2016. Ethiopia Demographic and Health Survey 2016. Addis Ababa, Ethiopia, and Rockville, Maryland, USA: CSA and ICF.
[10] Goldstein, H., 2003. Multilevel statistical models (Vol. 922). John Wiley & Sons.
[11] Gelman, A. and Hill, J., 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press.
[12] Hox, J., 2002. Quantitative methodology series. Multilevel analysis techniques and applications. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
[13] Snijders, T.A. and Bosker, R.J., 2011. Multilevel analysis: An introduction to basic and advanced multilevel modeling. Sage.
[14] Ikamari, L.D., 2005. The effect of education on the timing of marriage in Kenya. Demographic Research, 12, pp.1-28.
[15] Agresti, A., 2003. Categorical data analysis (Vol. 482). John Wiley & Sons.
[16] Hosmer, D. and Lemeshow, S.,2011. Applied Logistic Regression (5th Edition). New York: John Wiley & Sons.
[17] Rodriguez,G., 2007. Logit models for binary data. Online available at: $http://data.princeton.edu/wws509/notes/c3.pdf, zuletztgepruftam$, 12, p.2015.
[18] Bursac, Z., Gauss, C.H., Williams, D.K. and Hosmer, D., 2007. A purposeful selection of variables macro for logistic regression. In SAS Global Forum Proceedings (Vol. 173, pp. 1-5).
[19] Bewick, V., Cheek, L. and Ball, J., 2004. Statistics review 13: receiver operating characteristic curves. Critical care, 8(6), p.508.
[20] Menard, S., 2002. Applied logistic regression analysis (Vol. 106). Sage.
[21] Hosmer, D.W. and Lemeshow, S., 2000. Applied Logistic Regression 2nd ed. New York John Willey & Sons.
[22] Gourieroux, C., Holly, A. and Monfort, A., 1982. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. Econometrica: journal of the Econometric Society, pp.63-80.
[23] Goldstein, H., 2011. Multilevel statistical models (Vol. 922). John Wiley & Sons.
[24] Goldstein, H. and Rasbash, J., 1996. Improved approximations for multilevel models with binary responses. Journal of the Royal Statistical Society: Series A (Statistics in Society), 159(3), pp.505-513.
[25] Laird, N., 1978. Nonparametric maximum likelihood estimation of a mixing distribution. Journal of the American Statistical Association, 73(364), pp.805-811.
[26] Breslow, N.E. and Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. Journal of the American statistical Association, 88(421), pp.9-25.
[27] Agresti, A., 2002. Categorical data analysis (Vol. 482). John Wiley & Sons.
[28] Akaike, H., 1974. A new look at the statistical model identification. IEEE transactions on automatic control, 19(6), pp.716-723.

[29] Assefa, D., Wassie, E., Getahun, M., Berhaneselassie, M. and Melaku, A., 2005. Harmful traditional practices for the Ethiopian health center team. Hawassa, College, pp.33-40.
[30] Workineh, S., Kibretb, G.D. and Degu, G., 2015. Determinants of early marriage among female children in Sinan District, Northwest Ethiopia. Health Science Journal, 9(6), p.1.
[31] Erulkar, A., 2013. Early marriage, marital relations and intimate partner violence in Ethiopia. International perspectives on sexual and reproductive health, pp.6-13.
[32] Adebowale, S.A., Fagbamigbe, F.A., Okareh, T.O. and Lawal, G.O., 2012. Survival analysis of timing of first marriage among women of reproductive age in Nigeria: regional differences. African journal of reproductive health, 16(4), pp.95-107.
[33] Maswikwa, B., Richter, L., Kaufman, J. and Nandi, A., 2015. Minimum marriage age laws and the prevalence of child marriage and adolescent birth: evidence from sub-Saharan Africa. International Perspectives on Sexual and Reproductive Health, 41(2), pp.58-68.
[34] Peninah, A., Agaba, P., Atuhaire, L.K. and Rutaremwa, G., 2010. Determinants of age at first marriage among women in Western Uganda. In European population conference.
[35] CSA, 2012. Ethiopia 2011 Demographic and Heath Survey. Federal Democratic Republic of Ethiopia, Addis Ababa, Ethiopia.
[36] Hotchkiss, D.R., Godha, D., Gage, A.J. and Cappa, C., 2016. Risk factors associated with the practice of child marriage among Roma girls in Serbia. BMC international health and human rights, 16(1), p.6.
[37] Zahangir, M.S. and Kamal, M.M., 2011. Several attributes linked with child marriage of females in Bangladesh. International Journal of Statistics and Systems, 6(1), pp.107-117.
[38] Sishaw, T., Abdumalik Katabe, M. and Akmel, A., 2018. Spatiotemporal Variation of Early Marriage in Babile Woreda, Eastern Hararghe Zone, Oromiya National Regional State, Ethiopia (Doctoral dissertation, Haramaya University).

Appendixes

Table 7: Results of binary logistic regression model

Variable	Categories	Coeff.(SE)	p-value	OR = exp(Coeff.)	95% C. I for OR
	Intercept	0.2969(0.070)	2.6e-5*	1.636	[1.268,2.112]
Region	Tigray(ref.)				
	Affar	0.332(0.124)	0.008**	1.394	[1.091,1.780]
	Amhara	0.2046(0.059)	0.000**	1.409	[1.161,1.710]
	Oromia	-0.159(0.061)	0.009*	0.770	[0.633,0.938]
	Somale	-0.263(0.071)	0.000***	0.652	[0.519,0.818]
	BenishagulGumes	-0.121(0.066)	0.067	0.819	[0.662,1.012]
	SNNPR	-0.243(0.065)	0.000***	0.670	[0.544,0.827]
	Gambella	0.051(0.073)	0.482	1.085	[0.857,1.372]
	Harari	-0.198(0.075)	0.008*	0.723	[0.568,0.920]
	Addis Ababa	-0.523(0.073)	<0.001***	0.429	[0.339,0.527]
	Dire Dawa	-0.229(0.075)	0.002**	0.689	[0.542,0.875]
Residence Place	Urban(ref.)				
	Rural	0.127(0.055)	0.022*	1.225	[1.026,1.462]
Religion	Orthodox(ref.)				
	Muslim	0.103(0.040)	0.010*	1.185	[1.041,1.350]
	Protestant	0.022(0.048)	0.639	1.041	[0.892,1.215]
	Others	0.971(0.190)	<0.001**	5.444	[2.682,11.051]
Women Education	None(ref.)				
	Primary	-0.085(0.032)	0.009*	0.871	[0.784,0.966]
	Secondary	-0.044(0.053)	0.413	0.930	[0.783,1.104]
	Higher	-0.142(0.069)	0.038*	0.790	[0.634,0.985]
Wealth Index	Richest(ref.)				
	Richer	0.11(0.060)	0.057	1.209	[0.996,1.468]
	Middle	0.130(0.062)	0.036*	1.238	[1.015, 1.511]
	Poorer	0.169(0.062)	0.006**	1.320	[1.082, 1.610]
	Poorest	0.065(0.060)	0.282	1.111	[0.916, 1.346]
Women Employment	Employed (ref.)				
Unemployed	-0.010(0.029)	0.708	0.980	[0.892, 1.076]	
Media Exposure	Yes(ref.)				
	No	-0.059(0.038)	0.118	0.90	[0.801, 1.023]
Husband Educational	Not Educated(ref.)				
	Primary	0.039(0.032)	0.213	1.065	[0.962, 1.186]
	Secondary	-0.098(0.046)	0.032*	0.852	[0.736, 0.986]
	Higher	-0.335(0.053)	<0.001*	0.584	[0.493, 0.691]
Husband Occupation	Agriculture(ref.)				
	Professional	-0.038(0.038)	0.308	0.935	[0.828, 1.057]
	Business	-0.136(0.045)	0.002**	0.800	[0.692, 0.925]
	Labourers	-0.185(0.158)	0.240	0.740	[0.448, 1.221]
	Others	-0.096(0.038)	0.011**	0.854	[0.756, 0.964]
No. of sibling	<= 5 (ref.)				
	>5	0.064(0.026)	0.015*	1.109	[1.018, 1.208]