

---

# Multilevel Modelling of Determinants of Youth Unemployment in Urban Ethiopia: Bayesian Estimation Approach

**Teshita Uke Chikako**

Department of Statistics, College of Natural and Computational Science, Bule Hora University, Bule Hora, Ethiopia

**Email address:**

tasheman09@gmail.com

**To cite this article:**

Teshita Uke Chikako. Multilevel Modelling of Determinants of Youth Unemployment in Urban Ethiopia: Bayesian Estimation Approach. *International Journal on Data Science and Technology*. Vol. 4, No. 2, 2018, pp. 67-78. doi: 10.11648/j.ijdst.20180402.15

**Received:** May 20, 2018; **Accepted:** June 5, 2018; **Published:** July 4, 2018

---

**Abstract:** The main objective of this study was to identify and explain the effects of the Demographic and Socio-economic determinant factors of Youth unemployment in urban of Ethiopia. The data used for this study is the 2016 Ethiopian Urban Employment Unemployment Survey (UEUS) which was conducted by Central Statistical Agency (CSA) of Ethiopia. The statistical methods of data analysis are multilevel logistic regression models and Bayesian multilevel models and the parameters are estimated by using maximum likelihood estimation method and Bayesian estimation method by Stata and WinBUGS software. The analysis result revealed that Out of the 3870 youth considered in the analysis, 1,757 (45.4%) youth were unemployed, while 2113 (54.6%) youth were employed at the time of data collection. Region, Sex of youth, Age of youth, Literacy status, marital status, Type of Training, Steps taken to search work, Household size and Educational level are found to be the significant determinants of youth unemployment in urban Ethiopia. The multilevel logistic model revealed that the random intercept is better fit than null and random coefficient multilevel models. The intra correlation coefficient suggests that there is clear variation of youth unemployment status across the region of urban Ethiopia. The result of classical and Bayesian multilevel shows high prevalence of unemployment among youth and the probability of being unemployed for youth was found to decline with increasing age, literacy level, training, educational level and household size.

**Keywords:** Youth Unemployment, Regional Variations, Multilevel Logistic Regression, Bayesian Multilevel

---

## 1. Introduction

Unemployment is one of the several socio-economic problems prevailing in virtually all countries of the world. It is a serious social and economic problem for developed as well as developing countries. These problems have a serious effect not only on the living standards of people and the socio-economic status of a nation, but also elevate the magnitude of corruption, poverty, crime and suicidal rates in a society [1]. A person is considered as unemployed if he or she had actively looked for work and was not employed during the last seven days although he or she was available.

Youth unemployment and underemployment reflect the failure to make use of an important factor of production, labor and for fostering economic growth [2]. As stated in an economic report [3], a spatial perspective of Africa's labor market outcomes witnesses' higher rates of unemployment in

urban areas than in rural ones. Consequently, the report added, shortages of basic social infrastructure and facilities such as housing, schooling, water, and health are aggravated in the urban areas.

Youths are among the most important resources countries need to have in order to bring about prosperity, Energetic; courageous and qualified youth can make changes to the social economic development if they are well utilized and managed, [4]. However, unemployment among young people has becoming a major policy challenge for all nations in the world. Youth unemployment is a new concept of global problems, threatening whole humanity including educated and illiterate mass especially youth to cope up the magnitude of the problem in their respective locality. In most developing countries in general, and in sub-Sahara Africa in particular, the worst affected groups in Africa's job crisis are young people, women, the disabled and the elderly, [2]. It results

substantial crises in psychological, social and economic perspectives, some of them are: increasing crime rates and violence, dependence on family, low self-esteem, poor social adaptation, depression and loss of confidence. [5] In the same manner showed that unemployment affects the socio-economic status of the family, leads to poor mental health, dependency and increases the magnitude of corruption, drug addiction, crimes and suicide in a society. The level of youth unemployment, currently, is three times higher adults.

In Africa, in the same way, unemployment continues to be serious social problems despite some improvements in recent years. In the Sub-Saharan Africa youth unemployment rate is significantly higher than the adult unemployment rate. It is as twice as adult unemployment rate, [2]. According to ILO figures, and the Sub-Saharan Africa region has the highest rate of youth unemployment after the Middle East and North Africa. The youth unemployment rate serves as an indicator of the success of the strategies applied for job creation.

Developing countries are more victimized than others and Ethiopia has its own long history of unemployment than any other countries. Despite such improvements, youth unemployment (ages 15-29 years according to Ethiopian youth policy) [6] is high and is one of the socio economic problems in the country. This shows that the economy cannot provide adequate jobs to the growing population in both rural and urban areas. A higher rate of unemployment implies the loss of individual contributions to the growth of Growth Domestic Product (GDP) as well as the national economy, and hence a waste of human resource. Unemployed youths can be considered as opportunity for future investment and development.

They can be trained for various skills and job markets and own businesses. So studying the youth unemployment is interesting.

#### *Statement of the Problem*

Due to population pressure, the number of youth looking for work is expected to increase from year to year in Ethiopia. Failure to address youth employment issues will have serious consequences for the economy and society. While these general facts are clear, the specific factors affecting youth employment in Ethiopia have received little research attention. There is, therefore, limited empirical basis for formulating policies and programs promoting youth employment. So, this study will attempt to examine the demographic and socio economics determinants of Youth unemployment in urban of Ethiopia and to provide empirical information for policy makers.

#### *Objectives of the study*

General objectives: The general objective of this study was to identify and explain the effects of the demographic and socio-economic determinant factors of youth unemployment in urban of Ethiopia.

Specific objectives: The specific objectives of this study are:

- i. To identify the factors associated with youth unemployment in urban areas of Ethiopia.
- ii. To examine the extent of the variation in youth

unemployment status within and between regions of urban Ethiopia.

- iii. To identify the factors that may explain the variation in youth unemployment between and within regions of Ethiopia.
- iv. To provide relevant recommendations for policy makers and program managers that helps them in the process of job creation mechanism.

## 2. Methodology

### 2.1. Data

The data used to analysis the Youth unemployment status of urban Ethiopia was based on the secondary data obtained from the 2016 Urban Employment and Unemployment Survey (UEUS) which was conducted by Central Statistical Agency (CSA) annual survey.

### 2.2. Study Variables

Dependent Variable: The response variable or dependent variable in this study is youth unemployment status. It indicates whether one is employed or unemployed.

Independent Variable: The explanatory variables to be studied as determinants of youth unemployment are:

Region	Marital status
Gender	Steps taken to search work
Age	Household size
Literacy status	Youth education level
Types of training (field of study)	

### 2.3. Multilevel Logistic Regression Model

Multilevel statistical approach was used to model the relation between youth unemployment status and the explanatory variables. Two levels of data hierarchy were stated (for instance individual youth and region) in a multilevel logistic regression model. Units at one level are nested within units at the next higher level. In this study the basic data structure of the two-level logistic regression is a collection of  $J$  groups (regions) and within-group  $j$  ( $j=1,2,\dots,J$ ), a random sample  $n_j$  of level-one units (individual youth). The response variable is denoted by;

$$Y_{ij} = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ youth in the } j^{\text{th}} \text{ region is unemployed} \\ 0 & \text{if the } i^{\text{th}} \text{ youth in the } j^{\text{th}} \text{ region is employed} \end{cases}$$

With probabilities,  $P_{ij} = P(Y_{ij} = 1 | X_{ij}, u_{ij})$  is the probability of being unemployed for the  $i^{\text{th}}$  youth in the  $j^{\text{th}}$  region and  $1 - P_{ij} = P(Y_{ij} = 0 | X_{ij})$  is the probability of being employed for the  $i^{\text{th}}$  youth in the  $j^{\text{th}}$  region.

#### 2.3.1. The 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

differ between groups [7].

The random intercept model expresses the log-odds, i.e. the logit of  $P_{ij}$ , as a sum of a linear function of the explanatory variables. That is,

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

Where the intercept term  $\beta_{0j}$  is assumed to vary randomly and is given by the sum of an average intercept  $\beta_0$  and group-dependent random errors  $U_{0j}$ , that is  $\beta_{0j} = \beta_0 + U_{0j}$ . As a result we have

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

Where  $\beta_0 + \sum_{h=1}^k \beta_h x_{hij}$  is the fixed part of the model. The remaining  $U_{0j}$  is called the random part of the model. It is assumed that the residual  $U_{0j}$  is mutually independent and normally distributed with mean zero and variance  $\delta_0^2$

**2.3.2. The Random Coefficient Model**

The random coefficients builds up on 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 groupings. In the random coefficient model both the intercepts and slopes are allowed to differ across the region. It is given by;

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

For this study parameter estimation can be done maximum likelihood estimation and Bayesian estimation.

**2.4. Bayesian Multilevel Logistic Regression Model**

The classical multilevel logistic regression treats the unknown parameters as fixed constants for a fixed effect and treats as random for random effect without any distribution, while the Bayesian approach treats them as random variables, which means that the parameters can vary according to a probability distribution (prior distribution). This variation can be regarded as purely stochastic for a data driven model, but it can also be interpreted as beliefs of uncertainty under the Bayesian approach. In a Bayesian formulation the uncertainty about the value of each parameter can be represented by a probability distribution, if prior knowledge can be quantified, [8]. The key ingredients to a Bayesian analysis are the likelihood function, which reflects information about the parameters contained in the data, and the prior distribution, which quantifies what is known about the parameters before observing data. The prior distribution and likelihood can be easily combined to form the posterior distribution, which represents total knowledge about the parameters after the data have been observed [9].

**2.4.1. The 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. Since individual subjects in the group are assumed independent from each other, the likelihood function over a data set of n subjects in the J=11 regions is then:

$$L(y | \beta_i, \sigma_u^2) = \prod_{i=1}^n \prod_{j=1}^J \left[ \left( \frac{e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} + U_{0j}}}{1 + e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} + U_{0j}}} \right)^{y_{ij}} \left( 1 - \frac{e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} + U_{0j}}}{1 + e^{\beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \dots + \beta_k x_{kij} + U_{0j}}} \right)^{1 - y_{ij}} \right]$$

In these models, both of  $\beta$  and  $\sigma_u^2$  are unknown and independent. A typical prior specification can be normal distribution for  $\beta$  and inverse gamma prior for  $\sigma_u^2$ , that is:  $\beta \sim N(0, \sigma^2), \sigma_u^2 \sim IG(a, b)$ .

**2.4.2. 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 for parameter 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 distribution is 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 case, 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 of before the data is collected. They are non-informative in the sense that no value is favored over any other and are also described as diffuse or at prior due to this reason and their shape. In this study the researcher uses normal distribution prior for the fixed effect ( $\beta$ ) and Inverse gamma prior for random effect ( $\sigma_u^2$ ).

#### 2.4.3. Posterior Distribution

It is obtained by multiplying the prior distribution over all parameters by the full likelihood function. All Bayesian inferential conclusions 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, [10]. 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 can be solved easily by using MCMC simulations. It's given by;

$$f(\beta, \sigma_u^2 | y) = L(y | \beta, \sigma_u^2) \times \pi(\beta) \times \pi(\sigma_u^2)$$

In the Bayesian framework, inference follows from the full posterior distribution [11]. Bayesian multilevel model inference is then based on samples drawn from the posterior distribution using an MCMC algorithm such as the Gibbs sampler. For example, the posterior means and variances of the parameters can be estimated based on these samples, and Bayesian inference can then be based on these estimated posterior means and variances. This sampling can be done using Win BUGS software. We selected very vague prior distributions in our Win BUGS analysis. That is, we chose priors and hyper parameter values in such a way that, the priors will have minimal impact relative to the data.

The empirical results from a given MCMC analysis are not deemed reliable until the chain has reached its stationary distribution. On account of this, the term convergence of an MCMC algorithm refers to whether the algorithm has reached its equilibrium (target) distribution. If this is true, then the generated sample comes from the correct target distribution. Hence, monitoring the convergence of the algorithm is essential for producing results from the posterior distribution of interest. Convergence diagnosis was adopted to answer the question of how to determine whether the sample has reached its stationary distribution, [12]. The Gibbs sampling algorithm is one of the simplest Markov chain Monte Carlo algorithms converges to the target density as the number of iterations become large [13]. There are several convergence assessment methods. Among several ways the most popular and straight forward convergence assessment methods are; Time series plot, Autocorrelation plot, Density plot and Gelman-Rubin statistic.

## 3. Results

### 3.1. Results of Multilevel Logistic Regression Analysis

In the multilevel analysis, a two-level structure is used with regions as the second-level units and youth as the first level units. This analysis is mainly aimed at a comparison among regions and within regions variation of youth unemployment in urban Ethiopia. The hierarchical structure of the data is formed such that 3870 individuals (youth) were nested in eleven geographical regions based on the 2016 Urban employment and unemployment survey.

#### 3.1.1. Result of Empty (Null) Multilevel Logistic Regression Model

The empty model contains no explanatory variables and it can be considered as a parametric version of assessing heterogeneity among regions with respect to youth unemployment status.

**Table 1.** Results of Multilevel Logistic Regression Model without Explanatory Variables.

Fixed part	Estimate	S. error	z-value	p-value
$\beta_0 =$ intercept	-0.3086	0.1062	-2.906	0.00367 **
Random effect	Estimate	S. error	z-value	p-value
$\hat{\sigma}_u^2$	0.1073	0.0535	2.025	0.0142
ICC ( $\rho$ )	0.0316	.015266	0.000	

Here the null hypothesis tested is  $\delta_0^2 = 0$ . i.e., there is no regional variation in youth unemployment status in urban areas of Ethiopia. Based on the results Wald  $\chi^2 = 4.100 = (Z\text{-value})^2 = (0.1073/0.0535)^2$ ,  $df=1$  with the corresponding p-value= 0.0285 less than 0.05, the null hypothesis has to be rejected, indicating strong evidence that the between region variance is non zero. The variance of the random factor is significant which indicates that there are regional differences in youth unemployment and thus, multilevel analysis can be considered as an appropriate approach for further analysis.

In order to get an idea of how much of variation in unemployment among youth aged between 15-29 years was attributable to the region level factors, it is useful to see the intra-region correlation coefficient (ICC)=  $0.1073/(0.1073 + 3.29) = 0.0316$ , which measures the proportion of variance of the youth unemployment that is between regions, not within regions. The intra-region correlation coefficient (ICC) in intercept only model is 0.0316 which is significant at 5% level of significance. This means that around 3.16% of the variance in youth unemployment is due to variation across (between) regions. Whereas the remaining 96.84% attributable to individual level, i.e., within region differences.

#### 3.1.2. Result of Random Intercept Multilevel Logistic Regression Model

The intercept estimation is random at the regional level,  $\text{var}(u_{0j})$ . Thus, the value of  $\text{var}(u_{0j}) = \sigma_u^2 = 0.1069$  is the estimated variance component of the intercept. The multilevel logistic regression analysis result displayed in

table 3.2 confirmed the significance of regional difference in youth unemployment in urban areas of Ethiopia. The deviance based chi-square=37.81, d.f=1, p-value=0.000<0.05 for random effects in random intercept model, suggesting that youth with the same characteristics in different regions have different unemployment status in urban areas of Ethiopia: that is, there is a clear regional effect.

The results displayed in Table 8 showed that the intra-region correlation coefficient (ICC) is estimated as  $0.1069/(0.1069 + 3.29) = 0.0315$ , which is statistically significant at the 5% level of significance. This means that about 3.15% of the total variability in youth unemployment is due to differences across regions, with the remaining unexplained 96.85% attributable to individual differences.

The multilevel analysis also revealed that Sex of youth, Age of youth, Literacy status, Marital status, Type of Training, Steps taken to search work, Household size and

Educational level were also found to be significant determinants and also contributing factors for variation in youth unemployment among the regions of urban Ethiopia.

The multilevel logistic analyses indicate that sex of individual youth is significantly associated with unemployment status of youth at 1% level of significance. Particularly, individual female youth is 1.994 times more likely (OR=1.994, CI=(1.709, 2.327)) to be unemployed than male. In addition to this youth aged between 20-24 years were 0.582(OR=0.582, CI=(0.442, 0.768)) times less likely to be unemployed than youth aged between 15-19 and youth aged between 25-29 years were 0.263(OR=0.263, (CI=0.198, 0.2349) times less likely to be unemployed than the youth aged between 15-19 years. This implies that the youth aged between 15-19 have the highest chance of unemployed when compared to youth aged between 20-24 years and 25-29 years.

**Table 2.** Results of Random Intercept and Fixed Coefficient Logistic Regression Model.

Fixed effect	categories	Esti-mate	S. E	Z-value	p-value	odds	95% CI	
							lower	Upper
Sex	Male(ref)	--	--	--	--	--	--	--
	Female	.692	.078	8.81	.000	1.998	1.713	2.331
Age	From 15-19(ref)	--	--	--	--	--	--	--
	From 20-24	-.533	.140	-3.79	.000	.587	.445	.773
	From 25-29	-1.32	.144	-9.16	.000	.266	.200	.353
Literacy	Literate(ref)	--	--	--	--	--	--	--
	Illiterate	.387	.144	2.68	.007	1.473	1.109	1.957
Marital status	Never married(ref)	--	--	--	--	--	--	--
	Married	.308	.086	3.58	.000	1.361	1.150	1.611
	Divorced/separated	.201	.151	1.33	.185	1.22	.909	1.644
Training	No training(ref)	--	--	--	--	--	--	--
	Social science	-.602	.235	-2.56	.011	.547	.345	.869
	Natural science	-.237	.236	-.99	.321	.791	.500	1.257
	Engineering	-.304	.241	-1.26	.207	.738	.460	1.183
	Health science	-2.02	.331	-6.09	.000	.133	.070	.255
	Agriculture	-.707	.254	-2.79	.005	.493	.300	.811
Steps	Other social service	-1.09	.233	-4.68	.000	.336	.213	.530
	Through advertising boards(ref)	--	--	--	--	--	--	--
	Newspaper, radio & TV	-1.41	.116	-12.17	.000	.244	.195	.307
	Assistance of friends, relatives, etc.	-1.74	.115	-15.23	.000	.175	.140	.220
Household size	Other methods	-1.13	.130	-8.77	.000	.321	.249	.413
	1-3(ref)	--	--	--	--	--	--	--
	4-6	.075	.085	.88	.377	1.078	.912	1.274
Educational level	Above 6	-.399	.138	-2.89	.005	.671	.512	.879
	No education(ref)	--	--	--	--	--	--	--
	Primary	-.283	.148	-1.91	.056	.753	.563	1.006
Constant	Secondary and above	-.400	.413	-2.79	.005	.671	.507	.888
		2.297	.310	7.41	.000		1.689	2.904
Random effect								
var( $u_{0j}$ )		.1069	.055				.0385	.296
ICC( $\rho$ )		.0315						

Another finding of this study indicates that Literacy level of youth has a significant contribution on unemployment of youth in urban areas. The odds of unemployment for youth who cannot able to read and write (illiterate) was about 49.7% (OR=1.497, CI=(1.126, 1.990)) higher than the odds of unemployment of youth who can read or write (literate). From this we conclude that the illiterate youth have high chance of unemployed when compared to literate youth. The study also showed that youth marital status has a significant

contribution on unemployment of youth in urban areas. The odds of unemployment of married youth was 1.373 times (OR=1.373, CI = (1.160, 1.626)) higher than the odds of unemployment of never married youth and the unemployment of divorced or separated youth was not statistically different from never married youth (reference category).

This study also revealed that Types of training they have taken is significantly associated with unemployment status of

youth at 5% level of significance. The odds of unemployment of youth who has trained social science was 0.541 times (OR=0.541, CI= (0.340, 0.861)) lower than the youth who has no training. That is reduction of unemployment by 45.9%. The unemployment of youth who train Health science were about 86.9% (OR=0.131, CI= (0.068, 0.251)) less likely to be unemployed than the youth who has no training. The odds of unemployment of youth who has trained Agriculture was 0.5492 times (OR=0.492, CI=(0.299, 0.811)) lower than the youth who has no training and The odds of unemployment of youth who has trained other social service was 0.331 times (OR=0.331, CI=(0.209, 0.524)) lower than the youth who has no training. This implies that the youth who has no training have the highest chance of unemployed when compared to the youth who has other type of training. This indicates type of training determines the unemployment of youth.

Another finding of this study also indicates that the Steps the youth take to search (seek) work is significantly associated with unemployment status of youth at 5% level of significance. The youth who can search work through newspaper, Radio and TV were 24.6% (OR=0.246, CI=(0.196, 0.308)) less likely to be unemployed than the youth who can search work through advertising boards (reference category). The odds of unemployment of youth who have can Seeking work through assistance of friends, relatives, etc was 0.176 times (OR=0.176, CI=(0.140, 0.220)) lower than the odds of unemployment of youth who can search work through advertising boards (reference category) and the youth who can search work through other methods were 32.1% (OR=0.321, CI=(0.249, 0.414)) less likely to be unemployed than the youth who can search work through advertising boards (reference category). From this we conclude Steps the youth take to search (seek) work determines the unemployment of youth.

A youth from large number household size (family size above 6 members) were 33.4% (OR=0.666, CI=(0.508, 0.874)) less likely to be unemployed as compared to youth from small number household size ((family size between 1-3 member)) while the odds of unemployment youth from medium number household size (family size between 4-6 member) were not significantly different from youth from small number household size ((family size between 1-3

member)).

Another finding of this study from Table 6 above indicate that youth educational level is significantly associated with unemployment status of youth at 5% level of significance. The odds of youth unemployment who had secondary and above education were significantly differing from those who had no education (reference category). But the odds of youth unemployment who had primary education were not significantly differ from those who had no education. A youth who had secondary and above education is 33.4% (OR: 0.666, 95% CI: (0.503, 0.882)) less likely to be unemployed than a youth who no education. This implies that youth with the no or primary education have the lowest chance of being employed compared to youth who have secondary and above education and the youth with the secondary and above education have the lowest chance of being unemployed compared to youth who have no or primary education.

### 3.1.3. Result of Random Coefficient Multilevel Logistic Regression Model

In random intercept model we allowed the intercept only to vary across regions by fixing explanatory covariates, but the relation between explanatory and dependent variables can differ between groups (regions in our case). This model contains a random slope for Literacy status of youth, which means that it allows the effect of the coefficient of this variable to vary from region to region. This model is more appropriate than the previous model for the variables being used since it is intuitive to assume that Literacy status of youth varies from region to region.

From table 3.3 we investigated the value of intra-region correlation coefficient, in each model. By adding level-1 predictors, the ICC increased and estimated as  $(0.0008 + 0.1276) / (0.1276 + 0.0008 + 3.29) = 0.0375$ , meaning that roughly 3.75% of the total variability in youth unemployment is attributable to the random factor and region in random coefficient multilevel binary logistic model. From the above Table again the random coefficient estimates for intercepts and the slopes vary significantly at 5% significance level or the confidence interval does not include zero, which implies that there is a considerable variation in the effects of literacy status of youth and this variables differ significantly across the regions.

Table 3. Results of Random Coefficient Logistic Regression Model.

Fixed effect	categories	Esti-mate	S. E	Z-value	p-value	odds	95% CI	
							lower	upper
Sex	Male(ref)	--	--	--	--	--	--	--
	Female	.692	.079	8.81	.000	1.998	1.713	2.331
Age	From 15-19(ref)	--	--	--	--	--	--	--
	From 20-24	-.533	.140	-3.79	.000	.587	.445	.773
	From 25-29	-1.33	.145	-9.16	.000	.265	.200	.352
Literacy	Literate(ref)	--	--	--	--	--	--	--
	Illiterate	.390	.146	2.68	.007	1.478	1.110	1.967
Marital status	Never married(ref)	--	--	--	--	--	--	--
	Married	.308	.086	3.58	.000	1.361	1.150	1.611
	Divorced/separated	.201	.151	1.33	.183	1.222	.909	1.645
Training	No training(ref)	--	--	--	--	--	--	--

Fixed effect	categories	Esti-mate	S. E	Z-value	p-value	odds	95% CI	
							lower	upper
Steps	Social science	-.602	.236	-2.55	.011	.548	.345	.869
	Natural science	-.235	.236	-.99	.321	.791	.500	1.257
	Engineering	-.305	.241	-1.26	.206	.737	.460	1.183
	Health science	-2.02	.331	-6.09	.000	.133	.070	.255
	Agriculture	-.707	.254	-2.78	.005	.493	.300	.811
	Other social service	-1.09	.233	-4.68	.000	.336	.213	.530
	Through advertising boards(ref)	--	--	--	--	--	--	--
	Newspaper, radio & Tv	-1.41	.116	-12.16	.000	.244	.195	.307
	Assistance of friends, relatives, etc	-1.74	.115	-15.23	.000	.175	.140	.219
	Other methods	-1.14	.130	-8.77	.000	.321	.249	.413
Household size	1-3(ref)	--	--	--	---	--	--	--
	4-6	.075	.085	.88	.378	1.078	.912	1.274
	Above 6	-.399	.138	-2.89	.005	.671	.512	.879
Educational level	No education(ref)	--	--	--	--	--	--	--
	Primary	-.284	.148	-1.92	.56	.753	.563	1.006
Constant	Secondary and above	-.400	.413	-2.79	.005	.671	.507	.888
		2.297	.310	7.40	.000		1.688	2.905
Random effect								
	var( $u_{0j}$ )	.1069	.138					
	var1( $u_{1j}$ )	.0008	.009				.0385	.296
	corr( $u_{1j}, u_{0j}$ )	-.010	.060					
	ICC( $\rho$ )	.0315						

**3.1.4. Model Comparison**

The choice of relevant multilevel model is an important step, and it should be based on the necessity of parsimony in the model. This means that models should be as simple as possible [14].

*Table 4. Model Comparison.*

Fitted model	Multilevel null model	Multilevel random intercept model	Multilevel random coefficient model
-log likelihood	-2639.83	-2216.64	-2216.63
Deviance	5279.9	4433.3	4433.3
AIC	5283.67	4475.3	4479.25
BIC	5296.19	4606.8	4623.3

Based on the, Deviance, AIC and BIC values, random intercept model has the smallest value among the model considered. Therefore the random intercept model best fits the data.

**3.2. Results of Bayesian Multilevel Logistic Regression Analysis**

Bayesian multilevel logistic analysis procedure was used to make inference about the parameters of a multilevel logistic model. The Gibbs sampler algorithm was implemented with 25000 iterations in three different chains, 10000 burn-in terms discarded, as to obtain 45003 samples from the full posterior distribution for the multilevel model. The Gibbs sampler with more than one chain simultaneously provide autocorrelation and time series plots of each chain in different colors that help us to check convergence. If all the chains appear to be overlapping, we are confident that convergence has been attained. The researcher use non-informative normal prior distribution with mean =0 and precision =0.001 for the fixed effect and Inverse gamma distribution with *scale* =0.1, *shape* =0.1 for random effect. This implies that, the parameters of the covariates were

estimated by 45003 Markov chain sample values, simply using the Markov Chain samples after the burn-in state. After this simulation study the covariates that are significant in classical multilevel model are also statistically significant in the Bayesian inference.

**3.2.1. Assessment of Model Convergence**

There are a lot of commonly used methods to assess the convergence of MCMC output, but in this study only some of them are used.

Time series plot: Time series plots (iteration number on x-axis and parameter value on y-axis) are commonly used to assess convergence. If the plot looks like a horizontal band, with no long upward or downward trends, then we have evidence that the chain has converged. The three independently generated chains demonstrated good "chain mixture", an indication of convergence (see Figure below).

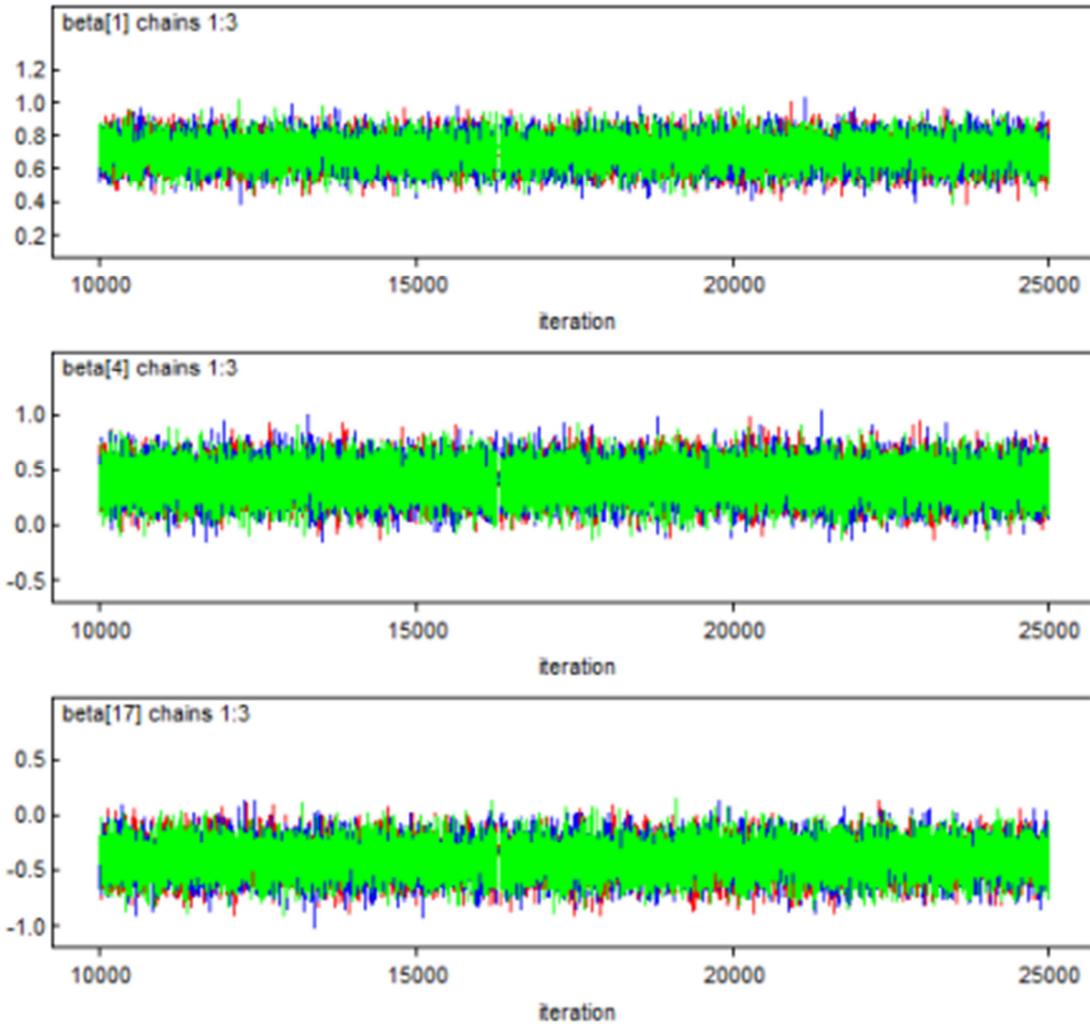


Figure 1. Convergences of Time Series Plots for the Coefficients of Sex, literacy and Household size.

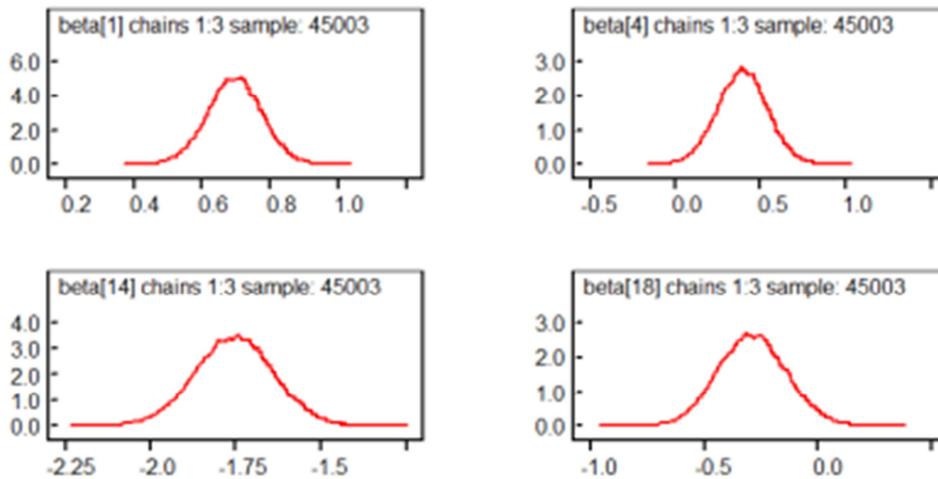


Figure 2. Convergences using kernel density plot for the Coefficients of Sex, literacy, Steps taken to search work and Educational level.

Gelman-Rubin Statistic: For a given parameter this statistic assesses the variability within parallel chains as compared to variability between parallel chains. The model is judged to have converged if the ratio of between variability to within variability is close to 1. The green line represents

the between variability, the blue line represent the within variance and the red line represents the ratio. Evidence for convergence comes from the red line being close to 1 on the y-axis. Since in the plot the red line seems exactly on 1, providing evidence for convergence.

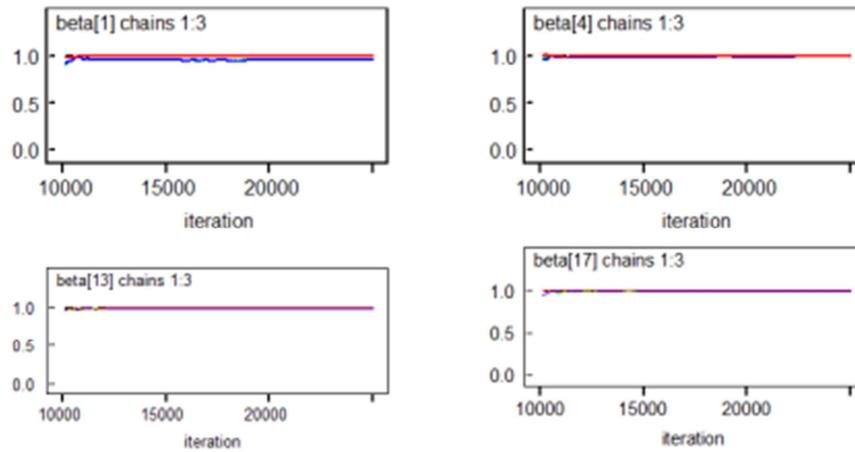


Figure 3. Convergences using Gelman-Rubin Statistics plot for the Coefficients of Sex, literacy, Steps taken to search work and Household size.

Autocorrelation: This option produces lag-autocorrelations the monitored parameters within each chain. High autocorrelation indicates slow mixing within a chain and usually slow convergence to the posterior distribution. The

plots show that the three independent chains were mixed or overlapped to each other indicating convergence. The plots displayed in Figure below indicate low autocorrelation and efficient sampling.

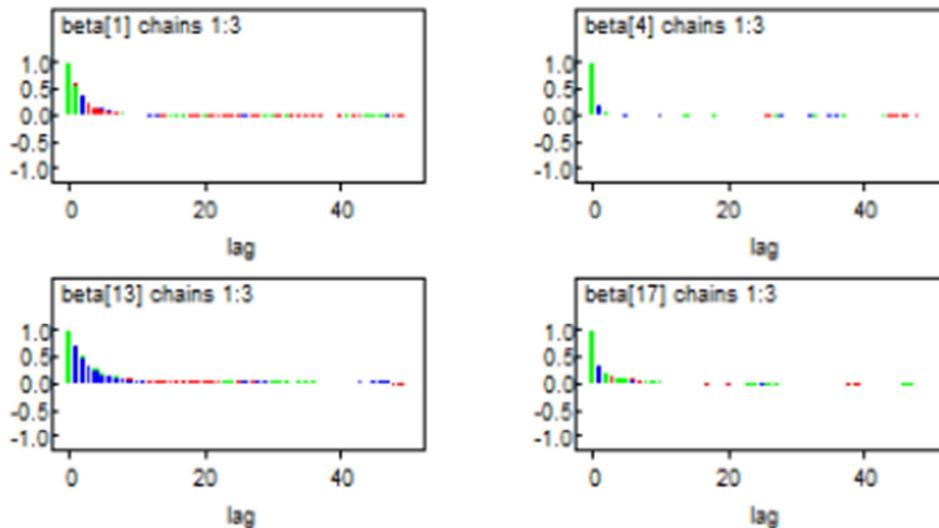


Figure 4. Convergences of autocorrelation plot for the Coefficients of Sex, literacy, Steps taken to search work and Household size.

Once model convergence is achieved, we can talk about the variables which have significant contribution for the prediction of the response variable.

**3.2.2. Interpretation of Bayesian Random Intercept Result**

To identify the effect of explanatory variables a multilevel binary logistic regression model with random intercept and fixed explanatory variables was estimated by Bayesian approach by using WinBUGS software. Note that there is little change in the estimate of the between-region variance, suggesting that the distribution of fixed explanatory variables is somewhat similar across regions of the country.

From the table below, the sample obtained from posterior distribution, summary statistics of all parameters for posterior distribution are present and the predictor variables like; Sex of youth, Age of youth, Literacy status, Marital status, Type of Training, Steps taken to search work, Household size and

Educational level were also found to be significant determinants of youth unemployment at 5% level of significance (Since the credible intervals of these variables does not contain zero (at least one category)). This shows significant variables are more determining the unemployment status of youth.

The Bayesian multilevel logistic regression analysis result displayed in Table above, also estimates the variance of random effect at the regional level,  $var(u_{0j})$ . Thus, the value of  $var(u_{0j}) = \sigma_u^2 = 0.145$  is significant at 5% level of significance (which means the 95% credible Confident intervals does not contain zero). This confirmed the significance of regional difference in youth unemployment in urban areas of Ethiopia. That is, there is a clear regional effect.

The results displayed in table above also showed that the intra-region correlation coefficient (ICC) is estimated as

0.145/(0.145 + 3.29) = 0.042. This means that about 4.2% of the total variability in youth unemployment is due to differences across regions, with the remaining unexplained 95.8% attributable to individual differences.

**Table 5.** Results of Bayesian Random Intercept Logistic Regression Model.

Fixed part	categories (level)	node	Estimate	Sd	MC error	2.5%	media	97.5%
Intercept		alpha	2.307	0.3257	0.01225	1.656	2.306	2.936
SEX	male (ref)							
	Female	beta[1]	0.6971	0.078	7.108E-4	0.5429	0.6972	0.8515
AGE	15-19 (ref)							
	20-24	beta[2]	-0.5398	0.143	0.002846	-0.8185	-0.5411	-0.2553
	24-29	beta[3]	-1.337	0.146	0.003134	-1.626	-1.338	-1.042
LITERACY	literate (ref)							
	illiterate	beta[4]	0.3936	0.145	8.189E-4	0.1077	0.394	0.6792
	no training (ref)							
	social science	beta[5]	-0.6036	0.237	0.007319	-1.071	-0.6015	-0.1397
	Natural science	beta[6]	-0.2333	0.237	0.007202	-0.7009	-0.2326	0.2347
TRAINING	Engineering	beta[7]	-0.3048	0.240	0.007347	-0.7797	-0.3037	0.1692
	Health science	beta[8]	-2.037	0.334	0.007798	-2.702	-2.034	-1.392
	Agriculture	beta[9]	-0.7067	0.256	0.007596	-1.211	-0.7059	-0.2024
	Other social service	beta[10]	-1.095	0.234	0.007466	-1.555	-1.094	-0.635
MARITL STATUS	Never married (ref)							
	Married	beta[11]	0.3111	0.086	7.557E-4	0.1418	0.3111	0.481
	Divorced/Separated	beta[12]	0.202	0.152	0.001127	-0.09836	0.2028	0.4994
STEPS	Advertising board(ref)							
	Newspaper, TV, Radio	beta[13]	-1.418	0.1158	0.001489	-1.646	-1.418	-1.192
	Assistance of friend	beta[14]	-1.755	0.115	0.001512	-1.982	-1.754	-1.532
	other method	beta[15]	-1.146	0.1288	0.001399	-1.4	-1.145	-0.8958
HOUSE HOLD SIZE	1-3 (ref)							
	4-6	beta[16]	0.07537	0.085	8.419E-4	-0.09008	0.07475	0.2456
	Above 6	beta[17]	-0.4027	0.138	0.001167	-0.6749	-0.4023	-0.1302
EDUCATIO NAL LEVEL	No education(ref)							
	Primary	beta[18]	-0.286	0.147	0.00295	-0.5721	-0.2873	0.01044
	Secondary and above	beta[19]	-0.402	0.141	0.002914	-0.6782	-0.4039	-0.1179
	Random effect							
	$\sigma_u^2$		0.145	0.104	0.001016	0.05974	0.1526	0.4706
	ICC( $\rho$ )		0.042					

The fixed effect parametric multilevel random intercept models in classical approach and in Bayesian approach were fitted. Both method give almost consistent results, but most of the parameters in Bayesian analysis had smaller standard error than the corresponding classical multilevel random intercept model and the Bayesian analysis additional solutions as posterior distribution of the parameters.

Therefore, Bayesian multilevel random intercept model gives better fit than the classical multilevel random intercept models. In the estimation of random effect there is a wide difference between the estimation of classical approach and Bayesian approaches (that is 0.1069 and 0.145 respectively), and the Bayesian multilevel random intercept model is more appropriate than the classical multilevel random intercept in explaining the variation of youth unemployment across the region of urban Ethiopia.

### 3.3. Discussion

This study is an attempt to identify some socio-economic and demographic determinants of unemployment status of youth in urban Ethiopia based on urban employment unemployment survey UEUS (2016) data.

The analysis showed that the prevalence of unemployment among youth is 45.4% and the probability of being unemployed was found to decline with increasing age,

educational level, literacy status of youth and household size. As the above analysis showed, Age is one of the determinants of unemployment status of youth. All age groups are negatively associated with unemployment. This is inconsistent with the findings by [15]. This study also revealed that youth unemployment rate is highest for the age group 15-19 followed by those aged 20-24. This indicates that, youth who are age group 25-29 years are the less likely to be unemployed than the other age group youth.

Marital status of youth is significant determinants of unemployment of youth in urban areas of Ethiopia. Married youth were more likely to be unemployed than never married youth. This result is similar with the results found by [16]; [17] and [1] and not similar with [15].

The study showed that, education level was an important determinant for unemployment status of youth and had negative effects. This result is similar with the results by other researchers [18], but it is not similar with the result of [16] and [19] that education has a positive effect on unemployment status. Youth with higher education were less likely to be unemployed compared to youth with no education at all controlling for other variables in the model. However, unemployment status of youth with primary education was not significantly differing from youth who had no education. This finding is not similar to the findings by

[17] where they concluded that the higher the level of education, the more likely to participate in the market and the more likely to be employed.

The study also indicates that training for the youth has negatively related with the unemployment status of youth. This indicates the youth who have no training are more likely to be unemployed than the youth who have different type of training. This result coincides with findings by [4]

Household size (Family size) is another predictor and negatively related with unemployment status of youth. Youth from large household size (family size) (more than 6) were less likely to be unemployed compared with youth from below 6 household sizes controlling for other variables in the model. This result contradicts with the findings of [15].

The analysis based on multilevel logistic regression provided estimates for variances of the random effects and interclass correlations. The estimates for each level were different, suggesting that the variance composition of unemployment status was different at individual and regional levels. This means that the sources of variations are individuals and regions. The result of multilevel logistic regression model comparison indicates that, the random intercept multilevel logistic regression model best fits the model than the null model and random coefficient model of multilevel logistic regression model. This study confirms with the finding of the study by [17]

From the result of intra correlation coefficient of null model we conclude that within region (individual level) variation of youth unemployment status is larger than that of between regional variations. This indicate that Region or location of individuals have significant effects on unemployment status of youth. This study confirms with the finding of the study by [17] and [15].

## 4. Conclusions

The purpose of this study is to identify major factors that determine variation of youth unemployment in urban Ethiopia. The study was based on the 2016 Urban Employment and Unemployment Survey (UEUS) which was conducted by the Central Statistical Agency (CSA) of Ethiopia in 2016.

This study revealed that youth unemployment is quite prevalent in urban Ethiopia with rate of 45.4%. Age of respondents and unemployment rate are related inversely. The majority of the unemployed youth were found in the age group 15-19 years. Relatively speaking youth in the age group 15-19 are from the most affected by unemployment.

Specifically the study revealed that unemployment status of youth is less likely for youth from secondary and above educational level as compared to those from uneducated youth. The predictors, type of training and steps taken to search work is negatively associated with unemployment status of youth and the predictors' literacy status of youth and marital status of youth are positively associated with unemployment status of youth.

From the methodological aspect, it was found out that

multilevel random intercept model is better compared to empty (null) model and random coefficient model in fitting the data and in explaining the variations of youth unemployment status across regional levels of urban Ethiopia. In addition from the random intercept model the overall variance of constant term was found to be statistically significant, implying the existence of difference in unemployment status of youth among regions of urban Ethiopia. This suggests that the youth with the same characteristics in two different regions have different unemployment status that is there is a clear regional effect.

Bayesian multilevel model also used to analyze unemployment status of youth. From classical and Bayesian multilevel models, the Bayesian multilevel analysis is preferable since it had smaller standard error than that of the classical multilevel analysis and it also provide the posterior distributions of the parameters.

## 5. Recommendations

The results obtained from this study are of great concern to policy makers because of the negative effects of unemployment on the loss of output, on the society and on the psychological wellbeing of the unemployed and immediate family members. In order to formulate policies to control the rising problem of unemployment in urban Ethiopia, it is important not only to understand the effect of reforms on the incidence of unemployment among the youth, but also on the duration of unemployment and on the probability of exiting unemployment and how it differs with demographic and economic characteristics. In response to this challenge, this paper suggests the following possible solution to reduce youth unemployment problem in urban Ethiopia.

- i. Policies that promote education and create more job opportunities should be implemented. For example, re-schooling or training of the less educated youth, increasing vocational training and labor market information. This would encourage more youth to go to work and thus generate the income required that would enable more youth in the regions to be able to increase their living standards.
- ii. The government should take a measure of action to support the very poor and to bring about rapid economic growth at the national level. To this effect, it is important to develop community based interventions giving priority to very poor households to participate in the labor market, education, health facility and areas of job access.
- iii. The policies that focus on reduction of poverty should also considered to reduce youth unemployment

---

## References

- [1] Asif M, Arshad I. A and Ali, N. (2015). A Statistical Analysis of Factors Affecting the Women Employment in Pakistan. Islamabad, Pakistan, 27(1)791-794.

- [2] ILO (2010). Growth-Employment-Poverty Reduction linkages: a Framework for Recovery and Accelerated Progress towards the Millennium Development Goals, Economic Report on Africa.
- [3] Economic Commission for Africa (ECA 2005). Economic Report on Africa 2005 Meeting the Challenges of Unemployment and Poverty in Africa. ECA Publication Cluster.
- [4] Therese F. Azeng Thierry U. Yogo (2013) Youth Unemployment and Political Instability in Selected Developing Countries.
- [5] F. Nazir, M. A. Cheema, M. I. Zafar and Z. Batool (2009). Socio-Economic Impacts of Unemployment in Urban Faisalabad, Pakistan Journal of Social Science, 18(3): 183-188. Department of Rural Sociology, University of Agriculture, Faisalabad 38040, Pakistan.
- [6] Ministry of Youth, Sport and Culture (2004). The Federal Democratic Republic of Ethiopia National Youth Policy, Addis Ababa, Ethiopia.
- [7] Snijders, T. A. B. and Roel J. Bosker (1999). An Introduction to Basic and Advanced Multilevel Modeling, First ed. Department of Statistics, University of Poone.
- [8] Kynn, M. (2005). Eliciting Expert Knowledge for Bayesian Logistic Regression in Species Habitat Modelling.
- [9] Howson, C. and Urbach, P. (1993), Scientific Reasoning: the Bayesian Approach, 2. ed. Open Court, Chicago.
- [10] Dezfuli, H, Smith, C, Galyean, W, (2009). Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis.
- [11] Gelman, A., Carlin, J. C., Stern, H. and Rubin, D. B, (1984). Bayesian Data Analysis. Chapman and Hall, New York.
- [12] Merkle, E., Shev, C. and Trisha, G. (2005). Simulation Based Bayesian Inference Using Winbugs. Winbugs Tutorial Outline.
- [13] Walsh, B. (2004). Markov Chain Monte Carlo and Gibbs Sampling.
- [14] Hox, J.(2002). Multilevel Analysis: Techniques and Applications. Mahwah, N. J: Lawrence Erlbaum.
- [15] Abebe Fikre Kasa (2011). Unemployment in Urban Ethiopia: Determinant and Impacts on Household Welfare. School of business, economic and law, University of Gothenburg.
- [16] Hongyu Yang, (1992). Female Labor Force Participation and Wages: A Case Study of Panama. In Case Studies on Women's Employment and Pay in Latin America.
- [17] Mesfin Mulu (2012). Determinants of Women Unemployment in Ethiopia: A Multilevel Model Approach. Master Thesis. Addis Ababa University, Addis Ababa, Ethiopia.
- [18] Bhorat, H. (2007). Unemployment in South Africa: Descriptors and Determinants. Paper Presented To The Commission On Growth And Development, World Bank, Washington Dc.
- [19] Pieter Sermeels (2007) The Nature of Unemployment Among Young Men in Urban Ethiopia.