

Innovations

Bayesian Multilevel Analysis of Youth Unemployment in Ethiopia

Tesfaye Denano^{1*}, Eyasu Anjulo¹, Netsanet asfaw¹, and Sintayehu Sibera¹

¹Department of statistics, Wolaita Sodo University, Postal Address 138, Ethiopia

Corresponding Author: **Tesfaye Denano**

Abstract

Youth unemployment is currently a critical macroeconomic problem in Ethiopia. The main objective of this study was to examine the extent of the variation in youth unemployment status within and between woredas and youths by using Bayesian multilevel logistic regression analysis. A sample of 379 youth is randomly selected from the Wolaita zone by stratified sampling techniques. Alcohol abuse, educational level, family size, parents' follow-up, job searching habits, access to credit, youth training, availability of jobs in the area, parents' employment status and urban migration are found to be significantly affecting the unemployment status of the youth. The effects of youth level variables: Alcohol abuse, educational level, family size, parents follow up, job searching habits, and parent's employment status are significantly varying from woreda to woreda. The result indicates that a large proportion of unemployment status variation is accounted for within between woredas. The government or concerned bodies should take necessary measures to eliminate young unemployment by expanding the Small Enterprises training program and assisting them with job placement, land, financing, and training for those who are jobless. To reduce rural-urban youth mobility, it is preferable to enhance rural livelihoods by using modern agricultural technologies.

Keywords: 1.Unemployment, 2.Bayesian multilevel, 3.Wolaita Zone, 4.Youth, 5.Variation, 6.Ethiopia

1. Introduction

Unemployment is one of Africa's macroeconomic issues, particularly in Ethiopia. Unemployment is used, to describe people who are employable and actively looking for work but are unable to find any work. Those in the labor force who are employed but do not have an acceptable job are included in this category(Ajufo 2013, Awad and Hussain 2022).

The capacity of youth to participate in labor force activities is a social and economic issue for a country's economy. Youth unemployment is frequently greater than adult unemployment, reflecting the challenges that many nations confront in easing the transition from school to work. Youth in developing nations confront the difficulty of not just finding a productive job but also finding safe and acceptable labor(Economic and Affairs 2004, Dagume and Gyekye 2016).

This article identified the variation between youth and woreda levels in the Wolaita zone, Ethiopia. We identify key policy inputs for constructing government labor force initiatives and economic growth approaches that might assist Ethiopian adolescents to improve their labor market results. In the Bayesian multilevel analysis, a two-level model is used with woredas as the second-level and youth as the first-level. This is basically with the expectation that there

would be a difference in the unemployment status of the youth among the woreda. In this part, first, we have to test; if there are differences in the mean unemployment status of youth among woreda before proceeding to Bayesian multilevel model analysis was done. From multilevel analysis, (Tesfaw and Derebew 2014) found that variation of academic achievement of Commerce second-year undergraduate regular students was higher among departments, while there was similarity within departments. (Roberts 2004, Gentry, Kuhnert et al. 2007) are used in Unemployment rates region as level-two and district as level-one in the multilevel model. In this study, we used youth as the first level and woredas as the second level to examine the extent of the variation in youth unemployment status between woredas in the study area.

2. Research Methodology

2.1. Research Design

The study adopted a cross-sectional survey research design to examine the extent of the variation in youth unemployment status within and between woredas-level and youths-level.

2.1.1. Data Sources and Method of Data Collection

We used seconder and primary data sources. To get the background information of respondents in the study area, secondary data were reviewed from various sources such as reports and other records. Other published and unpublished documents, the internet, and previous findings which are found to be relevant to the study will be also used. The other related information and primary data were collected from unemployed youth in the study area by using questionnaire interviews. The target population consists of professional unemployed and employed youths at the time of the survey. To prove the validity of the instrument pilot testing for the questionnaire was done before the main data collection.

According to the sample size determination formula, that was adopted in this study

$$n = \frac{(z_{\alpha/2})^2 P(1 - P)}{d^2} \left[1 + \frac{1}{N} \left[\frac{z_{\alpha/2}^2 P(1 - P)}{d^2} - 1 \right] \right] \quad (1)$$

Where d is some margin of error to tolerate in estimation; P is the proportion of unemployed youth; N is the total number of youth; n is the total sample size; Z is the value of standard normal distribution for a given level of significance (α). In fixing this sample size will be, $d= 0.05$ to increase precision, and $\alpha=0.05$ were used.

2.2. Multilevel Modeling

A multilevel model allows simultaneous estimation of parameters from various levels and quantifies between and within-cluster variations. Hierarchical models take account of the variability associated with each level of the hierarchy. These models have also been referred to as mixed models, random coefficient models, and covariance component models (Breslow and Clayton 1993, Wolfinger and O'connell 1993, Clayton 1996, Neuhaus and Segal 1997, Dean and Nielsen 2007).

2.2.1. Variance Components Model(null model)

The variance components model is given by:

$$Y_{mk} = \eta_0 + \gamma_{0k} + \varepsilon_{0mk} \tag{2.1}$$

Where Y_{mk} is the unemployment status of m^{th} youth in k^{th} woreda; γ_{0k} 's are woreda level random effects that are independently identically normally distributed with zero means and constant variances δ_0^2 ; ε_{0mk} 's are errors that are independently identically normally distributed with mean zero and constant variance δ_ε^2 , η_0 is the overall average unemployment status in the study area.

The interclass correlation measures the correlation between observations within the cluster

$$\rho \text{ or } ICC = \frac{\delta_{0u}^2}{\delta_{0u}^2 + \delta_\varepsilon^2} \tag{2.2}$$

2.2.2. Random Intercept Model

The random intercept model written in the form of:

$$Y_{mk} = \underbrace{\eta_0}_{\text{overall mean}} + \underbrace{\eta_1 x_{1mk} + \dots + \eta_n x_{nmk} + \gamma_{0k}}_{\text{woreda level}} + \underbrace{\eta_{n+1} x_{n+1,mk} + \dots + \eta_p x_{p,mk}}_{\text{youth level}} + \varepsilon_{mk} \tag{2.3}$$

Where Y_{mk} is the unemployment status of m^{th} youth in k^{th} woreda; x_{nmk} is are covariates from the two levels, γ_{0k} 's are woreda level random effects that are independently identically normally distributed with zero means and constant variances δ_0^2 ; ε_{0mk} 's are errors that are independently identically normally distributed with mean zero and constant variance δ_ε^2 , η_0 is overall average unemployment status in the study area.

2.2.3. Random Coefficients Model

The random coefficient model written in the form of:

$$Y_{mk} = \underbrace{\eta_0}_{\text{overall mean}} + \underbrace{\eta_1 x_{1mk} + \dots + \eta_n x_{nmk} + \gamma_{0k}}_{\text{woreda level}} + \underbrace{\eta_{n+1,k} x_{n+1,mk} + \dots + \eta_{pk} x_{p,mk} + \gamma_{1k} x_{n+1,mk} + \dots + \gamma_{pk} x_{p,mk}}_{\text{youth level}} + \varepsilon_{mk} \tag{2.4}$$

Woreda level indicators are kept fixed in equation (2.4). The random effects $\gamma_{0k}, \gamma_{1k}, \dots, \gamma_{pk}$ are independently normally distributed with zero means and constant variances. ε_{0mk} are errors independently normally distributed with zero means and constant variances.

The posterior distribution of the parameters for Bayesian inference is given in generic form as:

$$\prod(\eta, \delta_{0u}^2, \delta_{0\varepsilon}^2 | y, x) \propto L(y, x | \eta, \delta_{0u}^2, \delta_{0\varepsilon}^2) \prod(\eta) \prod(\delta_{0u}^2, \delta_{0\varepsilon}^2)$$

2.3. The goodness of Fit Test

It is useful to be able to judge whether a model is a good fit for the data. For this study, the test of goodness of fit was employed using deviance. The maximum likelihood procedure produces a statistic called the deviance, which indicates how well the model fits the data. The test compares the deviance (-2 log-likelihood) of two models by

subtracting the smaller deviance (model with more parameters) from the larger deviance (model with lower parameters). The overall model evaluation is also examined using Akaike Information Criteria (AIC), deviance information criteria(DIC) and Schwartz Information Criteria (BIC). The smaller the value, the better the model will be(White-Newsome, Sánchez, et al. 2012, Naimoli, Frymus, et al. 2014, Sullivan, Pearce-Higgins, et al. 2017).

3. Results and discussion

3.1. Descriptive analysis

When the data collection was undergone the respondents were specifically requested about their employment status earlier to the survey time. During the survey time the maximum number of the respondents was unemployed from the total sample size of 379 interviewed respondents 296(78.1 percent) were unemployed and 83(21.9 percent) of the respondents were employed at the time of data collection period as shown in table-1 below

Table-1: Unemployment Status of Respondent (n=379)

sex of respondents * employment status Cross tabulation				$\chi^2 - value$	p-values
Count	employment status			1.476	0.024
		Employed	Unemployed		
Sex	Female	27	118	145	
	male	56	178	234	
Total		83(21.9)	296(78.1%)	379	

The overall average variation between woredas is estimated to be 58.2%, with a standard error of 3.545, and the between and within woreda variations are also statistically significant at a 5% significant level. The results from the analysis of the variance components model (2.1) are displayed in Table 2. The results of ICC was tells us that about 74.3% of the total variation in unemployment can be accounted for by which kebeles each youth is in.

Table-2: Bayesian variance components model (n=379)

Fixed effect	Estimates	Std.error	Confidence interval (95%)		p-value
			Lower	Upper	
Intercept(η_0)	58.2	3.545	53.32	63.51	0.000*
Variance components					
Woreda variance (η_0)	86.8	24.843	38.99	127.2	
Residual variance (ϵ_{mk})	30.05	1.165	17.9	22.33	
ICC	0.743				
DIC	163.751				

*, indicates that statistically significant at 1%

Note: The interclass correlation (ICC) for this model is equal to:

$$\rho = \frac{\delta_{0k}^2}{\delta_{0k}^2 + \delta_e^2} = \frac{86.8}{86.8 + 30.05} = 0.7428$$

Results of Random Intercept Model with Covariates

Results of the analysis of the random intercept model (2.3) are displayed in Table 3. Intercept and several covariates are significant. There is much reduction of DIC to 1,814.98 in the model (2.3) and the between and within woreda variations are statistically significant. About 43.02% of the variation in the unemployment status of the youths is accounted for between woredas, while 53.4% is accounted to the youth’s level. Among covariates at the youth level, the following are found significant at 5% significant level: sex, age, alcohol abuse, marital status, work experience, parent’s employment status, family size, and job searching habits. Unemployment of youth in the study area can decrease by decreasing youth drinking alcohol and increasing work experiences, job searching habits, educational level, parent’s employment status, and parent's follow-up. Alcohol abuse, work experience, and family size are negatively associated with employment status. For instance, a youth who is addicted to alcohol for a unit of time is likely to reduce his average employment status by 1.583%. An increase in work experience by a unit of year increases employment by 0.21%.

For the woreda level, the significant predictors are: access to credit, youth training, availability of new jobs, and social network as well as positively associated with unemployment status. Increasing access to credit, youth training, availability of new jobs and social network increases the employment status of youth by 1.08%, 1.5%, 4.792%, and 1.735%, respectively.

Table 3. Random intercept Bayesian model

Fixed effect	Estimates	Std. error	Confidence interval 95%)		p-values
			Lower	Upper	
Intercept(η_0)	9.8475	0.729	8.511	11.320	0.002
Youth level					
Age	0.324	0.176	-0.037	6.550	0.014
Sex	3.249	0.401	2.397	3.970	0.151
Educational level	0.326	0.082	0.167	0.482	0.002
Marital status	0.828	0.389	0.039	1.523	0.044
Alcohol abuse	-1.583	0.335	0.913	2.232	0.003
Job search habit	3.760	0.317	3.145	4.403	0.001
Work experience	-0.210	0.097	-0.006	0.372	0.038
Family size	-0.211	0.105	-0.0015	0.412	0.026
Family employment status	0.671	0.278	0.176	1.455	0.026
Work demand at home	0.214	0.092	0.022	0.376	0.130
Parents follow up	1.044	0.315	0.437	1.654	0.000
Woreda level covariates					
Access to credit	1.068	0.384	0.343	1.801	0.005
youth training	1.501	0.847	-0.092	3.135	0.016
Availability of new job	4.792	0.922	3.093	6.546	0.000
Social network	1.735	0.845	0.149	3.407	0.022
Random part					
Woreda level	6.485	4.777	5.689	21.78	

variance(η_0)					
Residual variance(ε_{mk})	8.590	1.026	15.04	18.94	
ICC	0.4302				
DIC	1,814.98				

$$\rho = \frac{\delta_{0k}^2}{\delta_{0k}^2 + \delta_e^2} = \frac{6.485}{6.485 + 8.59} = 0.4302$$

Random coefficients Bayesian model

In table-4, the results show that the random effects models are significant. The intercept, as well as a number of variables are significant. This model has the lowest DIC=162.985 in equation (2.4), indicating that the Bayesian random coefficients model is the best to fit the data. The variation between and within woredas are statistically significant. Around 45.3 percent of the variations in youth unemployment is explained by woredas level, while 54.36 percent is explained at the youth level.

Among independent variables at the youth level, the following are found significant at 5% significant level: Age, sex, educational level, marital status, alcohol abuse, job search habits, work experience, family size, parent’s employment status, and parents' follow-up. Work demand at home is not significant. Unemployment of youth in the study area decreased by increasing educational level, work experience, parent's employment status, parents follow up, and by decreasing alcohol abuse, and family size. Alcohol abuse and parents' follow-up are negatively related to employment status. For instance, a youth who is alcohol addicted a unit of time is likely to reduce his employment status by 4.749%. An increase in parent’s education by a unit decrease unemployment by 2.013%. Parents need to reduce the time the youth is to spend on alcohol and rather motivate the youth to give more time to for searching jobs and increasing their educational level. Parent’s follow-up of the youth is so important that the youth always goes to increase his educational level and gets proper mind maturity.

For the woreda level, the significant independent variables are access to credit availability, youth training, availability of new jobs in the woreda, and social network. Youth training access to credit is negatively related to the youth's employment status. Increasing youth training, access to credit in the area, availability of new jobs, and social network by a unit decrease the unemployment status of youth by nearly 3.204%, 4.506%, 14.377%, and 5.206%, respectively.

Table-4: Random coefficients Bayesian model (n=379)

Fixed effect	Estimates	Std.error	Confidence interval 95%)		p-values
			Lower	Upper	
Intercept (η_0)	29.544	2.187	25.533	33.96	0.000
Youth level					
Age	0.972	0.528	-0.112	19.650	0.213
Sex	9.750	1.204	7.192	11.92	0..123
Educational level	0.978	0.246	0.503	1.447	0.042
Marital status	2.485	1.167	0.120	4.569	0.205
Alcohol abuse	4.749	1.010	2.739	6.696	0.000*
Job search habit	3.760	0.317	3.145	4.403	0.036**

Work experience	0.630	0.292	-0.018	1.116	0.049**
Family size	-0.633	0.315	-0.0045	1.236	0.063***
parents employment status	2.013	0.834	0.528	4.365	0.056***
Work demand at home	0.643	0.276	0.067	1.128	0.130
Parents follow up	3.132	0.945	1.311	4.963	0.010**
Woreda level variables					
Access to credit	3.204	1.152	1.029	5.404	0.000*
youth training	4.506	2.541	-2.79	9.411	0.000*
Availability of new job	14.377	2.766	9.279	19.638	0.000*
Social network	5.206	2.535	0.447	10.221	0.000*
Random part					
Woreda level variance (η_0)	20.460	14.331	17.067	65.340	
Residual variance (ε_{mk})	24.800	3.078	45.120	56.820	
Parents employment status	12.015	19.189			
Work experience	0.394	0.169			
Parent follow up	0.450	4.989			
Work demand at home	0.345	3.891			
ICC	0.452				
DIC	162.985				

*, **, and *** indicate that statistically significant at 1%, 5%, and 10% respectively

Note: DIC is a deviance information criterion

$$\rho = \frac{\delta_{0k}^2}{\delta_{0k}^2 + \delta_e^2} = \frac{20.460}{20.46 + 24.80} = 0.45$$

4. Conclusions and recommendations

The main objective of this study was to examine the extent of the variation in youth unemployment status within and between woredas and youths by using the Bayesian approach. A sample of 379 youth is randomly selected from 5 kebeles of the Wolaita zone based on a stratified sampling. The unemployment status of youth is considered an outcome variable. Between and within woreda variations are studied.

In comparison to likelihood approaches, the Bayesian method has offered more exact estimates of model parameters. The data is best fit by the Bayesian random coefficients model. The Bayesian random intercept model performs similarly and is interpretable.

The factors identified to be significantly affecting the unemployment status of youth are: (i) youth level: alcohol abuse, family size, parents' employment status, educational level, job searching habit, family size, work experience, and parents' follow up. (ii) For woreda level: the significant independent variables are access to credit, youth training, social network, and availability of a new job. Sex, age, and marital status are not significant. There are unemployment variations between woredas between woreda and within woreda in the study area. Variation of 45.2 % is accounted for between woredas and 54.36% to the youths.

The government or concerned bodies should take necessary measures to eliminate young unemployment by expanding the Small Enterprises training program and assisting them with job placement, land, financing, and training for those who are jobless. To reduce rural-urban youth mobility, it is preferable to enhance rural livelihoods by using modern agricultural technologies.

Ethics Approval and Consent to Participate

Ethical permission was granted for this study by the Wolaita Sodo University (WSU) of Research Ethics Committee. Before gathering samples from Wolaita zone, a verbal agreement was obtained from WSU. The best guidelines and regulations were followed, and the study's aim was explained to the Wolaita Sodo University of Research Ethics and Review Committee accepted the oral informed consent method described in the paper.

Consent for publication

-Not Applicable

Availability of data and materials

The data sets generated and/or analyzed during the current study are publically available upon reasonable request from corresponding author.

Funding source: there was no specific fund for this work

Competing interests: The authors declare that they have no competing interests.

Author Contributions

All authors contributed to the data collection, data analysis, and writing of the paper. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank to Wolaita Sodo University Research and Community Service for their follow-up this project work.

References

1. Ajufo, B. I. (2013). "Challenges of youth unemployment in Nigeria: Effective career guidance as a panacea." *African research review*7(1): 307-321.
2. Awad, A. and M. A. Hussain (2022). "Youth in the Labour Market: Does Inequality of Opportunity Matter? The Experience of the Democratic Republic of Congo and Madagascar." *Journal of the Knowledge Economy*: 1-27.
3. Breslow, N. E. and D. G. Clayton (1993). "Approximate inference in generalized linear mixed models." *Journal of the American Statistical Association*88(421): 9-25.
4. Clayton, D. G. (1996). "Generalized linear mixed models." *Markov chain Monte Carlo in practice*1: 275-302.

5. Dagume, M. A. and A. Gyekye (2016). "Determinants of youth unemployment in South Africa: Evidence from the Vhembe district of Limpopo province." Environmental economics(7, Iss. 4): 59-67.
6. Dean, C. and J. D. Nielsen (2007). "Generalized linear mixed models: a review and some extensions." Lifetime data analysis**13**(4): 497-512.
7. Economic, U. N. D. o. and S. Affairs (2004). World youth report 2003: the global situation of young people, UN Department of Economic and Social Affairs New York, NY, USA.
8. Gentry, W. A., K. W. Kuhnert, S. P. Mondore and E. E. Page (2007). "The influence of supervisory-support climate and unemployment rate on part-time employee retention: A multilevel analysis." Journal of Management Development.
9. Naimoli, J. F., D. E. Frymus, T. Wuliji, L. M. Franco and M. H. Newsome (2014). "A community health worker "logic model": towards a theory of enhanced performance in low-and middle-income countries." Human resources for health**12**(1): 1-16.
10. Neuhaus, J. M. and M. R. Segal (1997). An assessment of approximate maximum likelihood estimators in generalized linear mixed models. Modelling Longitudinal and Spatially Correlated Data, Springer: 11-22.
11. Roberts, J. M. (2004). "Monetary policy and inflation dynamics." Available at SSRN 633222.
12. Sullivan, M. J., J. W. Pearce-Higgins, S. E. Newson, P. Scholefield, T. Brereton, and T. H. Oliver (2017). "A national-scale model of linear features improves predictions of farmland biodiversity." Journal of Applied Ecology**54**(6): 1776-1784.
13. Tesfaw, D. and M. Derebew (2014). "Multilevel analysis on determinants of academic achievement of second-year regular students: The case of Addis Ababa university school of commerce." Journal of Research & Method in Education: 2320-2737.
14. White-Newsome, J. L., B. N. Sánchez, O. Jolliet, Z. Zhang, E. A. Parker, J. T. Dvonch and M. S. O'neill (2012). "Climate change and health: indoor heat exposure in vulnerable populations." Environmental Research**112**: 20-27.
15. Wolfinger, R. and M. O'connell (1993). "Generalized linear mixed models a pseudo-likelihood approach." Journal of Statistical Computation and Simulation**48**(3-4): 233-243.

***Corresponding Author email: tesfustat@gmail.com**