

Count regression models for under-five deaths in rural Ethiopia

Yenew Alemu mihret (✉ yenewalemu@gmail.com)

Research article

Keywords: mortality (death), under-five mortality, Poisson, NB, ZIP and ZINB

Posted Date: July 25th, 2019

DOI: <https://doi.org/10.21203/rs.2.11962/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Under-five mortality is defined as the likelihood for a child born alive to die between birth and fifth birthday. Mortality of under the age of five has been the main target of public health policies and is a common indicator of mortality levels, especially in developing countries. It is also viewed as an indicator of the level of development, health and socioeconomic status of the population. The objective of this study was to identify determinants of under-five mortality in Ethiopia using the 2011 EDHS data. To achieve the objective of this study descriptive statistics and count regression models were used for data analysis using socio-economic, demographic and environmental related variables as explanatory variables and the number of under-five deaths per mother as the response variables. According to Ethiopian Demography health Survey, 2011 report the level of under-five mortality in rural parts of Ethiopia is 114 deaths per 1000 live births. Factors influencing the number of under-five deaths have been identified. The study revealed that mother's age at the first birth, breastfeeding status, wealth index, current mother working, region and mother's level of education had statistically significant on the number of under-five deaths in rural parts of Ethiopia.

Background Of The Study

Under-five mortality is defined as the likelihood for a child born alive to die between birth and fifth birthday. Mortality of under the age of five has been the main target of public health policies and is a common indicator of mortality levels, especially in developing countries[8]. It is also viewed as an indicator of the level of development, health and socioeconomic status of the population.

Sub-Saharan Africa still has the world's highest rate of child mortality, despite reducing the number of under-five deaths by 48% since 1990. Sub-Saharan Africa is the region most affected and accounts for more than one-third of deaths of children under the age of five. In 2013, about half of global under-five deaths occurred in sub-Saharan Africa and 32 percent in South Asia. Around half of under-five deaths occur in just five countries: India (21%), Nigeria (13%), Pakistan (6%), Democratic Republic of the Congo (5%) and China (4%) [20].

Under-five mortality rate was measured as 166 per 1000 live births in the year 2000 survey, 123 deaths per 1000 live births in the year 2005 survey and reduced to 88 deaths per 1000 live births in the period leading up to 2011, a 47% reduction between 2000 and 2011 (EDHS 2000, 2005 and 2011 Final Reports). Under five child mortality is one of the main problems in Ethiopia. In recent years, Ethiopia health and family planning (EHFP) has successfully implemented in a wide array of fertility and mortality reduction interventions. Besides, the growth and transformation plan (GTP) has been developed and under implementation starting from 2011 to improve access and quality of health services. However, despite all of these efforts, health care facilities in Ethiopia are limited and inadequate. Moreover, lack of health personnel, medicines and other facilities are not uniformly available. To expand our understanding about the most common and consistent factors on the risk of under-five child mortality, we have considered possible determinants of under-five child mortality using count regression model. Therefore, this study explores the Socio-demographic, environmental and socio-economic characteristics of under-five mortality in rural Ethiopia.

Despite the fact that a number of researchers have studied the risk factors of under-five mortality in Ethiopia, the majority of them studied the risk factors at national level. Such studies have overlooked an important point for policy makers as the findings at national level may not demonstrate the exact situation at the rural areas. Under-five death is showing a declining trend over the last decades. Numbers of under-five death rates were 166 per 1000 live births in 2000 and 88 per 1000 live births in 2011. According to [6] report the level of under-five mortality in rural parts of Ethiopia is 114 deaths per 1000 live births. However, the rate is still very high and requires intervention to lower the death. Mostly in the rural level, there are limited studies that focused on the determinant factors of under-five mortality using count regression models.

Logistic regression has been widely used in analyzing child death data. However, since its dependent variable is dichotomized to be either "1" (death) or "0" (alive), logistic regression undercounts the total number of under-five mortality since multiple child deaths are collapsed into a single unit to fulfill the requirements of logistic regression. Logistic regression cannot provide sufficient information for studying the pattern of multiple child deaths. Therefore, we would develop and compare count regression models for the number of under-five deaths and discuss how they can enhance our understanding of the risk factors of under-five deaths.

The general objective of the study is to identify socioeconomic, demographic and environmental factors of under-five mortality in rural Ethiopia using count regression model. The specific objectives are to examine the determinants of under-five mortality in rural parts of Ethiopia using 2011 EDHS data and to fit an appropriate count regression model for rural parts of Ethiopia.

Methods

2.1 Source of the data

The source of the data in this study is the 2011 Ethiopia Demographic and Health Survey (EDHS). The 2011 EDHS was conducted by the Central Statistics Agency (CSA) with support from the ministry of Health. The analysis presented in this study on under-five mortality was based on the 10,475 women aged 15–49 years. During the analysis stage, Statistical Package for Social Science (SPSS) version 16, South Texas Art Therapy Association (STATA) version 12 and Microsoft-Excel were used as tools of analysis.

2.2 Variables in the study

The dependent variable for this study is the number of deaths of under-five deaths per mother. Based on the [15] determinants of childhood morbidity and mortality framework for developing countries, experiences from the available similar studies and available data on the subject, the main predictors explored for under-five mortality have been grouped into demographic, socioeconomic and environmental factors. The demographic factors for this studies are

mother's age at the first birth, mother's currently breastfeeding and mother's marital status. The socioeconomic factors are mother's level of education, mother currently working, region and household's wealth index. The environmental factors are Source of drinking water and Toilet facility.

2.3 Methods of data analysis

In this study, the variable of interest is a count variable. When the response or dependent variable is a count (which can take on non-negative integer values (0, 1, 2, ...), it is appropriate to use non-linear models based on non-normal distribution to describe the relationship between the response variable and a set of predictor variables. For count data, the standard framework for explaining the relationship between the outcome variable and a set of explanatory variables includes the Poisson and negative binomial regression models. Unlike linear regression, count data regression models have counts as the response variable that can take only nonnegative integer values. The two most popular models for count data are the Poisson model and the negative binomial model.

2.3.1 Poisson regression model

This regression model is a popular and simple regression model for count data. It assumes a Poisson distribution, characterized by a positive skewed and a variance equals the mean. Poisson regression analysis is a technique which allows to model dependent variables that describe count data [3]. According to [19], the apparent simplicity of Poisson comes with two restrictive assumptions. First, the variance and mean of the count variable are assumed to be equal. The other restrictive assumption of Poisson models is that occurrences of the event are assumed to be independent of each other.

Let Y_i represent counts of events occurring in a given time or exposure periods with rate μ_i . Y_i are Poisson random variables which the p.m.f. is characterized by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, y_i denotes the value of an event count outcome variable occurring in a given time or exposure periods with mean parameter μ_i .

The likelihood function of the Poisson model based on a sample of n independent observations is given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

The log-likelihood function is

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

The likelihood equation for estimating the parameter is obtained by taking the partial derivations of the log-likelihood function and setting them equal to zero.

There are two basic criteria commonly used to check the presence of overdispersion:

1. Deviance, $D(y, \mu)$, is given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, y is the number of events, n is the number of observations and $\hat{\mu}_i$ is the fitted Poisson mean.

2. Pearson chi-square test, χ^2 is also given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

Another way of checking the presence of over-dispersion is a statistical test of the hypothesis:

$H_0: \alpha = 0$ vs $H_1: \alpha > 0$.

If P-value of $LRT \alpha < \alpha$ (level of significant), which is an indicated of over-dispersion is present; negative binomial is preferred. The negative binomial regression model is more appropriate for over-dispersed data because it relaxes the constraints of equal mean and variance.

In the general Poisson regression model, we think of μ_i as the expected number of under five-child death from the i^{th} mother and the total number children ever born from the i^{th} mother is N_i . This means parameter will depend on the population size and the total number of children ever born from the individual mother. Thus the distribution of Y_i can be written as:

$Y_i \sim \text{poisson}(N_i \mu_i)$

Where, N_i is the total fertility rate of i^{th} mother and $\mu_i = (\mathbf{X}_i^T \boldsymbol{\beta})$

The logarithm of the children ever born is introduced in the regression model as an offset variable. By including $\ln(\text{children ever born})$ as offset in the equation, it is differentiated from other coefficients in the regression model by being carried through as a constant and forced to have a coefficient of one [9].

2.3.2 Negative binomial regression model

This model is used when count data are overdispersed (i.e when the variance exceeds the mean). Overdispersion, caused by heterogeneity or an excess number of zeros (or both) to some degree is inherent to most Poisson data. By introducing a random component into the conditional mean, the negative binomial regression model addresses the issue of over-dispersion. However, it equally models both zero and nonzero counts, which might result in a poor fit for data with excessive number of zeros. Therefore, it is always necessary to check the proportion of zero counts before developing a negative binomial regression model. We used the likelihood ratio test to determine the more appropriate model between the Poisson regression and negative binomial regression. [11] used negative binomial regression to model over dispersed Poisson data.. The NB regression model is

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

with mean and variance are given by

$$E(Y_i) = \mu_i = \exp(x_i^T \beta) \text{ and } \text{Var}(Y_i) = \mu_i(1+\alpha\mu_i)$$

where, α shows the level of overdispersion and $\Gamma(\cdot)$ is the gamma function.

The likelihood function of the NB model based on a sample of n independent observations is given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

The log-likelihood function l of NB regression model is

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

For estimating regression coefficients β and dispersion parameter α , the Newton-Raphson iteration procedure is applied like Poisson model.

2.3.3 Zero- inflated model

In some cases, excess zeros exist in count data and considered as a result of over-dispersion. In such a case, the NB model cannot be used to handle the over-dispersion which is due to the high amount of zeros. To do this, zero-inflation (ZI) can be alternatively used.

2.3.3.1 Zero- inflated Poisson regression model

The Zero-inflated Poisson regression study the relationship between dependent and independent variable(s) when there are many zeros value in the dependent variable, where the relationship is the mixture between Poisson model and Logistic model. Zero-inflated Poisson Regression also provides a flexible way of modeling zero counts and an attractive interpretation.

The ZIP regression model is [14],

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, $Y_i \sim \text{ZIP}(\mu_i, \omega_i)$. The mean and variance of ZIP are given by

$$E(Y_i) = (1-\omega_i)\mu_i \text{ and } \text{Var}(Y_i) = E(Y_i)(1+\omega_i\mu_i)$$

The parameters μ_i and ω_i can be obtained by using the link functions,

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, x_i^T and Z_i^T are covariate matrices, β and γ are the $(p+1) \times 1$ and $(q+1) \times 1$ unknown parameter vectors, respectively. The log-likelihood function of ZIP model is given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, $I(\cdot)$ is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise.

To obtain the parameter estimates of ZIP regression models, β and $\hat{\gamma}$, the Newton-Raphson method can be used.

2.3.3.2 Zero-inflated negative binomial regression model

Zero-Inflated Negative Binomial (ZINB) regression is one of the methods used in troubleshooting overdispersion due to excessive zero values in the response variable (excess zeros). This model provides a way of modeling the excess number of zeros (with respect to a Poisson distribution or negative binomial distribution) in addition to allow for count data that are skewed and overdispersed. We used the vuong test, likelihood ratio based test, to compare the zero inflated negative binomial model with negative binomial regression model. A significant z-test indicates that the zero inflated models are preferred. We consider Y_i as a ZINB distribution. Specifically, we consider the distribution. [10] used the zero-inflated negative binomial (ZINB) regression to model overdispersed data with an excess of zeros. This regression model was given by

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, μ_i is the mean of the underlying negative binomial distribution, $\alpha > 0$ is the over dispersion parameter and is assumed not to depend on covariates and $0 \leq \omega_i \leq 1$. Also the parameters μ_i and ω_i depend on vectors of covariates x_i and z_i , respectively.

The log-likelihood function $l = l(\alpha, \mu_i, \omega_i; y)$, for the ZINB model is given below.

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

Furthermore, l can be written as

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

Newton-Raphson iteration procedure can be used for estimating the parameter of ZINB regression models.

2.4 Goodness of fit tests

2.4.1 Likelihood Ratio test

The Likelihood ratio test is a test of a null hypothesis H_0 against an alternative H_1 based on the ratio of two log-likelihood functions. The likelihood ratio test is a test of the overall model. The overall test statistic for likelihood ratio test is given as:

[Due to technical limitations, this equation is only available as a download in the supplemental files section.]

This statistic is called the likelihood-ratio test statistic.

Where: l_{null} is the log-likelihood of the null model and l_k is the log-likelihood of the model comprising k predictors, p is number of parameters and χ^2_{p-1} is a chi-square distribution with $p-1$ degree of freedom. If the test statistics exceeds the critical value, the null hypothesis is rejected. That means the overall model is significant. In this study, to compare Poisson and NB regression models and also ZIP with ZINB regression models, we used significance of dispersion parameter and likelihood ratio (LR) test as criterions. The statistic of likelihood ratio test for α is given by the following equation: $LRT\alpha = -2(LL_1 - LL_2)$

This statistic has a Chi-squared distribution with 1 degrees of freedom and LL is log-likelihood. If the statistic is greater than the critical value then, the model 2 is better than the model 1.

2.4.2 Vuong Test

The Vuong test is a non-nested test that is based on a comparison of the predicted probabilities of two models that do not nest [21]. That means vuong test statistics are needed to provide the appropriateness of zero-inflated models against the standard count models. For testing the relevance of using zero-inflated models versus Poisson and NB regression models, the Vuong statistic is used. Let's define [Due to technical limitations, this equation is only available as a download in the supplemental files section.]

where, $P_1(Y_i/X_i)$ and $P_2(Y_i/X_i)$ are probability mass functions of zero-inflated and Poisson or NB models, respectively. In general, $P_N(Y_i/X_i)$ is the predicted probability of observed count for case i from model N , then the Vuong test statistic is simply the average log-likelihood ratio suitably normalized. The test statistic is [Due to technical limitations, this equation is only available as a download in the supplemental files section.]

Where, \bar{m} is mean of m_i , s_m standard deviation and n sample size.

The hypotheses of the Vuong test are:

$$H_0: E[m_i] = 0$$

$$H_1: E[m_i] \neq 0$$

The null hypothesis of the test is that the two models are equivalent. Vuong showed that asymptotically, V has a standard normal distribution. As Vuong notes, the test is directional [21].

- If $V > Z\alpha/2$, the first model is preferred.
- If $V < -Z\alpha/2$, the second model is preferred.
- If $|V| < Z\alpha/2$, none of the models are preferred.

2.4.3 AIC and BIC

AIC and BIC are goodness of criteria used for model selection. The likelihood ratio test was used to compare the Poisson model and NB model. Many Monte-Carlo simulations indicate that the BIC and AIC selection criteria need to be used together [4] and [23]. The model with smallest value of AIC or of BIC is preferable. Selecting an appropriate model can be used a standard likelihood information criteria, for example, Akaike information criteria [2] or Baysians information criteria [18] abbreviated by AIC and BIC, respectively, Where

$$AIC = -2 \log \text{likelihood} + 2k$$

$$\text{BIC} = -2 \log \text{likelihood} + k \ln(n)$$

where, k = number of parameters and n = number of observations.

Result

3.1 Descriptive statistics

Information on the number of deaths of under-five children obtained from a total of 8,668 women in the rural parts of Ethiopia was studied. Table 1 showed the frequency and percentage distribution of the number of under-5 deaths in rural parts of Ethiopia based on information from 8,668 women. Summary statistics for the dependent variable used in the present study are presented in Table 2. The 8,668 observation values corresponding to each variable were used in the study. Among the 8,668 women, considered, 5,180 children died before the age of five (see Table 1). Table 2 showed that the sample mean of the response variable, the number of under-five deaths was 0.5976 while the sample variance was 1.052. The fact that the mean is smaller than the variance, suggested a case of over-dispersion. Moreover, the data has excess zeros and thus one might expect that the Poisson model would not be appropriate to predict the number of under-five deaths. As shown in Figure 1, the distribution of the number of under-five deaths has a rapidly decreasing tail and is highly skewed to right with excess zeros.

Figure 1: Histogram of the number of under-5 deaths

3.2 Comparison of Count Data Models in rural parts of Ethiopia

As shown in table 3, the test of over-dispersion, Deviance statistics and Pearson Chi-square Statistic divided by their corresponding degrees of freedom, are greater than one indicating over-dispersion. The likelihood-ratio test statistic is 348.32 with $p < 0.0001$, we reject H_0 that there is no over-dispersion, and conclude that there is significant over-dispersion in the data and the negative binomial regression model is favored over the regression model. As shown in the summary Table 4, the likelihood-ratio chi-square values for all models were found to be significant. Thus, all regression models are significant. The model with the smallest AIC and BIC and the largest log likelihood is preferred. Since ZINB model has the smallest AIC and BIC and maximum log-likelihood, ZINB model is the most appropriate and preferred model among the four models. That is, the zero-inflated negative binomial regression model with the lowest value of AIC and the maximum value of log-likelihood is the most appropriate model for describing the number of under-five children deaths. Moreover, the fact that the LRT of a statistic, $LRT_{\alpha} = -2(LL_P - LL_{NB}) = 348.32$, is greater than $\chi^2(01)$, implies that the NB is better than the Poisson model and $LRT_{\alpha} = -2(LL_{ZIP} - LL_{ZINB}) = 50.25$, greater than $\chi^2(01)$, indicates that ZINB is better than ZIP model. We used Vuong test to compare Zero-inflated regression models with other non-nested models, including Poisson regression and Negative Binomial regression models. The results indicated that Zero-inflated negative binomial (ZINB) regression model was the most appropriate count data model for this data. Although it is hard to distinguish Negative Binomial and Zero-inflated Poisson (ZIP) regression models, they performed better than the Poisson regression model. The Vuong statistic values are presented in Table 5.

Hypothesis of likelihood ratio test for nested models are H_0 : The simpler model is preferred

and H_1 : The more complex model is preferred. Based on table 6, If $P\text{-value} < 0.05$, we reject H_0 , and conclude H_1 . In addition, the following figure also confirms that the ZINB model is the most appropriate model among the four models considered.

Figure 2: residual plots for estimated models

Moreover, the residual plot in Figure 2 showed that ZINB model fits well because almost all ZINB points pass through 0 and makes a straight line.

3.3 Interpretation of data

3.3.1 Interpretation of ZINB regression model for positive counts in rural parts of Ethiopia

The results in Table 7 showed that region has a significant factor on the number of under-five deaths in the non-zero group. The expected number of under-five deaths for women from Amhara region was 1.15 times the expected number of under-five deaths for women in the reference group (Tigray) holding all other variables in the model constant. Similarly, the expected number of under-five deaths increased by a factor of 1.17 for women living in Oromiya compared to women in Tigray controlling for other variables in the model. In addition, the expected number of under-five mortality for women from Benshangul-gumuz, SNNP and Gambella had increased by a factor of 1.45, 1.36 and 1.33 as compared to the expected number of under five mortality in Tigray region, respectively, while holding all other variables in the model constant.

We can interpret mother's age at the first birth obtained in Table 7 using the reference category. The expected number of under-five deaths for women were decreased by a factor of 0.64 in the age group 15-19 compared to those in the age group less than 15 controlling for other variables in the model. Similarly, the expected number of under-five deaths for women decreased by a factor of 0.50 in the age group 20 and above as compared to those in the age group less than 15 controlling other variables in the model.

The finding of this study also revealed that mother's level of education had a significant factor on the number of under-five mortality. The expected number of under-five mortality for mothers with primary education was decreased by a factor of 0.64 as compared to those with no education (reference group) controlling other variables in the model. In addition, the expected number of the under-five mortality for mothers with secondary and higher level of education were decreased by a factor of 0.38 and 0.19 as compared to those with no education, respectively, controlling other variables in the model.

In this study, currently breastfeeding has a significant effect on the number of under-five deaths. The estimated number of under-five mortality for mothers who were breastfeeding is about 0.78 times lower than mothers who were not breastfeeding. In addition, mothers work status had a significant factor on the number of under-five deaths. The expected number of under-five deaths increased by a factor of 1.09 for working mothers as compared to that for non-working mothers while holding all other variables in the model constant.

According to the findings of this study, wealth index of the household has a significant influence on the number of under-five mortality. The expected number of under-five deaths for women in the medium and rich households was 0.92 and 0.91 times the expected number of under-five deaths for women in the poor households, respectively, while holding all other variables in the model constant.

3.3.2 Interpretation of ZINB regression model for covariates of zero counts in rural parts of Ethiopia

As shown in Table 7, mother's marital status has a significant effect on the probability of being an excess zero. The odds of being in the zero groups are increased by a factor of 2.82 for unmarried mothers as compared to married mothers controlling for other variables in the model.

3.4 Discussion

This study was carried out to identify the risk factors of under-five mortality in rural areas based on EDHS 2011 data. The total number of women from rural areas included in the present study was 8,668 among which 35.5% experienced under-five deaths due to different factors. The most appropriate model was selected from four possible count models and the ZINB regression model was selected as the most appropriate model in rural parts of Ethiopia.

As the results show, the number of under-five mortality in rural Ethiopia has a wide variation among regions. This result is consistent with the findings by [7].

In this study, mothers' education was found to be an important socio-economic predictor of the number of under-five mortality in rural Ethiopia. Under-five mortality decreased with increased level of mother's education. This result is consistent with the findings by [24], [5], [1], [13], [12] and [17].

The current study revealed that those children whose mother's age at first birth was below 15 years had the higher risk of dying relative to children whose mothers' age was 15 and above. A similar study in Kenya by [7] also found that mothers' age at first birth has a significant effect on infant and child mortality showing that a child born to a younger mother experienced the highest risk of dying. These findings are consistent with the findings of [1] and [17].

According to the results, mother's current breastfeeding status was a significant determinant of under-five mortality showing that children born to non-breastfeeding mother experience higher risk of mortality than children born to breastfeeding mothers. This result is consistent with the findings by [16] and [17].

The findings suggested that under-five child mortality risk is higher for children of poor mothers compared to children of medium and rich mothers. This finding is consistent with [16], [7], [22] and [5].

The study indicated that children born from working mothers have higher risk of mortality than non-working mothers. This finding is consistent with [5].

The results of ZINB indicated that marital status has a significant factor on the odds of being in the always zero group. The odds of being in the always zero group for unmarried mothers were higher than unmarried mothers.

Conclusion

The main objective of the study was to identify some of the factors that influence the number of infant and child mortality not only at national level but also separately at rural and urban levels. The study was based on secondary data obtained from the central statistical agency of Ethiopia. Among the four models considered for analyzing the data from women in rural areas

Factors influencing the number of under-five deaths have been identified. The study revealed that mother's age at the first birth, currently breastfeeding, wealth index, current mother working, region and mother's level of education had statistically significant effect on the number of under-five deaths in rural parts of Ethiopia.

Based on our findings we recommend that

- There is a need for comprehensive prevention strategies that will help to further reduce child mortality.
- The government/ministry of health should give greater attention to improve immunization services and concentrate on health education campaigns for mothers and for the community.
- Early marriages should be discouraged and awareness about the danger of giving birth at early ages should be created through education.
- Health interventions should particularly be targeted towards women who are suffering from illness and weakness to allow them to continue breastfeeding.
- Effort should be made for providing better access to education and health facilities for mothers so that the gap in under-five mortality is bridged.

Abbreviations

CSA: Central Statistical Agency

EDHS: Ethiopia Demographic and Health Survey

U5M: Under-five Mortality

UNICEF: United Nations Children Fund

WHO: World Health Organization

BIC: Bayesian Information Criterion

AIC: Akaike Information Criterion

PRM: Poisson Regression Model

NBRM: Negative Binomial Regression Model

ZIP: Zero Inflated Poisson

ZINB: Zero Inflated Negative Binomial

NB: Negative Binomial

LRT: Likelihood Ratio Test

UN: United Nations

Declarations

Ethics approval and consent to participate

Ethical clearance was obtained from the department of statistics of the University of Addis Ababa. The data is important for finding factors of child death.

Consent to publish

Not applicable

Availability of data and materials

All data supporting the findings and conclusion are presented in the manuscript. The datasets during the current study is accessible to the publisher.

Competing interest

The authors declare no competing interests.

Funding

I received no specific funding for this work.

Author's contributions

Conceptualized, designed, analyzed, interpreted the results and wrote the manuscript are done by myself. Dejen Tesfaw(Phd) is assisted designing the research and revising the manuscript.

Acknowledgements

I thankful to the EDHS coordinators for their assistant to give data. I also thankful to Mokennon Tadesse (associate professor) for the guidance given during the drafting of the document.

References

1. abimbola o adepoju ao akanni o falusi ao determinants child mortality rural Nigeria world rural observations 2012 4 2 38 45 issn 1944 6543 print issn 194
2. Bozdogan H. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. Psychometrika. 1987 Sep 1;52(3):345–70.

3. Cameron AC, Trivedi PK. Regression analysis of count data. Cambridge university press; 2013 May 27.
4. Dalrymple ML, Hudson IL, Ford RP. Finite mixture, zero-inflated Poisson and hurdle models with application to SIDS. Computational Statistics & Data Analysis. 2003 Jan 28;41(3-4):491-504.
5. Ermias dessie buli 2013 determinants child survival chances rural ethiopia adama science technology university adama ethiopia proceedings 59th isi worl
6. Demographic E. Health Survey Central Statistical Agency Addis Ababa. Ethiopia ICF International Calverton, Maryland, USA. 2011:180-6.
7. Ettarh RR, Kimani J. Determinants of under-five mortality in rural and urban Kenya. Rural & Remote Health. 2012 Jan 1;12(1).
8. Garenne M, Gakusi E. Health transitions in sub-Saharan Africa: overview of mortality trends in children under 5 years old (1950-2000). Bulletin of the World Health Organization. 2006;84:470-8.
9. Ko MK, Sawangdee Y, Gray R, Hunchangsith P. Ecological analysis of community-level socioeconomic determinants of infant and under-five mortality in Myanmar: an analysis of the 2014 Myanmar population and housing census. Journal of Health Research. 2017;31(1):57-68.
10. Gurmu S, Trivedi PK. Excess zeros in count models for recreational trips. Journal of Business & Economic Statistics. 1996 Oct 1;14(4):469-77.
11. Hilbe JM. Negative binomial regression. Cambridge University Press; 2011 Mar 17.
12. Wang L, Jacoby H. Environmental Determinants of Child Mortality in Rural China: A Competing Risks Approach. The World Bank; 2004 Mar 12.
13. Van der Klaauw B, Wang L. Child mortality in rural India. World Bank Policy Research Working Paper. 2004 Apr 21(3281
14. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics. 1992 Feb 1;34(1):1-4.
15. Mosley WH, Chen LC. An analytical framework for the study of child survival in developing countries. Population and development review. 1984 Jan 1;10(0):25-45.
16. Mustafa HE, Odimegwu C. Socioeconomic determinants of infant mortality in Kenya: analysis of Kenya DHS 2003. J Humanit Soc Sci. 2008;2(8):1934-722.
17. Mani K, Dwivedi SN, Pandey RM. Determinants of under-five mortality in Rural Empowered Action Group States in India: An application of Cox frailty model. International Journal of MCH and AIDS. 2012;1(1):60.
18. Raftery AE. Choosing models for cross-classifications. American sociological review. 1986 Feb 1;51(1):145-6.
19. Sturman MC. Multiple approaches to analyzing count data in studies of individual differences: The propensity for type I errors, illustrated with the case of absenteeism prediction. Educational and Psychological Measurement. 1999 Jun;59(3):414-30.
20. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, Gerland P, New JR, Alkema L. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. The Lancet. 2015 Dec 5;386(10010):2275-86.
21. Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society. 1989 Mar 1:307-33.
22. Silva P. Environmental factors and children's malnutrition in Ethiopia. The World Bank; 2005 Jan 12.
23. Wang P, Puterman ML, Cockburn I, Le N. Mixed Poisson regression models with covariate dependent rates. Biometrics. 1996 Jun 1:381-400.
24. Seyoum Y, Sharma MK. Survival Analysis of Under Five Mortality in Rural Parts of Ethiopia. International Journal of Statistics in Medical Research. 2014 Aug 5;3(3):266-81.

Tables

Table 1: Frequency distribution of number of under-5 deaths in rural parts of Ethiopia

Number of deaths per mother	Frequency	percent
0	5591	64.50
1	1802	20.79
2	772	8.91
3	303	3.50
4	131	1.51
5	42	0.48
6	15	0.17
7	5	0.06
8	7	0.08
Total	8668	100.0

Table 2: Summary data for the dependent variable, number of under-5 deaths

	Minimum	Maximum	Mean	Variance	Skewness
Number of under-5 deaths	0.00	15.00	0.5976	1.052	2.537

Table 3: The results of over-dispersion test after fitting a Poisson regression

Statistics	Value	Degree of freedom	Value ----- degree of freedom	P-value
Deviance	9620.636	8648	1.11	0.000
Pearson chi-square	10027.21	8648	1.16	0.000

Table 4: Fit statistic of count regression models

criteria	Poisson	NB	ZIP	ZINB
Log-likelihood	-8412.6821	-8238.5233	-8253.059	-8227.931
-2LL	16,825.3648	16,477.0466	16,506.118	16,455.862
AIC	16865.36	16519.05	16548.12	16499.86
BIC	17006.71	16667.46	16696.53	16655.35
Likelihood ratio test	561.03 (0.000)	405.93 (0.000)	424.52 (0.000)	403.19 (0.000)

Table 5: Model comparisons by Vuong test for non-nested models

Model	Vuong Statistic(V)	Preferred model
ZIP VS Poisson	8.16	ZIP
ZINB VS NB	2.46	ZINB

Table 6: Model comparisons by likelihood ratio test for nested models

Model	Likelihood ratio test (p-value)	Preferred model
NB VS Poisson	0.000	NB
ZINB VS ZIP	0.000	ZINB

Table 7: Estimated coefficients of ZINB regression model

variables	Coef.	Std. Err.	z	P> z	[95% CI]
Region,Tigray(ref.)					
Affar	0.0416065	0.0724009	0.57	0.566	-.1002966 .1835095
Amhara	0.1411311	0.0690014	2.05	0.041	.0058908 .2763715
Oromiya	0.1569805	0.0657261	2.39	0.017	.0281598 .2858012
Somali	0.0403798	0.0779589	0.52	0.604	-.1124168 .1931763
Benshangul- Gumuz	0.3706548	0.0706173	5.25	0.000	.2322474 .5090622
SNNP	0.3081286	0.065627	4.7	0.000	.179502 .4367552
Gambela	0.2633426	0.0806256	3.27	0.001	.1053194 .4213659
Harari	-0.0685959	0.0986595	-0.7	0.487	-.261965 .1247732
Dire Dawa	0.1177358	0.092601	1.27	0.204	-.0637588 .2992305
Mother level of education, no education(ref.)					
primary	-0.4637706	0.0458781	-10.11	0.000	-.55369 -.3738512
secondary	-0.9691338	0.3312863	-2.93	0.003	-1.618443 -.3198247
higher	-1.6755	0.7211637	-2.32	0.02	-3.088955 -.2620451
Wealth index, poor(ref.)					
medium	-0.0869593	0.0434646	-2.00	0.045	-.1721482 -.0017703
rich	-0.0910793	0.0423884	-2.15	0.032	-.174159 -.0079995
Mother age at the first birth,<15(ref.)					
15-19	-0.4489802	0.0492739	-9.11	0.000	-.5455553 -.352405
20 and above	-0.6945656	0.0542082	-12.81	0.000	-.8008118 -.5883194
Currently breastfeeding, no(ref.)					
yes	-0.2534029	0.0347195	-7.3	0.000	-.3214518 -.185354
Current mother working, no(ref.)					
yes	0.0755776	0.0364916	2.07	0.038	.0040554 .1470999
_cons	-0.4335085	0.0812158	-5.34	0.000	-.5926885 -.2743285
log(children ever born)	1.000				
inflate					
Marital status, married(ref.)					
Unmarried	1.03762	0.2837085	3.66	0.000	.4815612 1.593678
_cons	-1.959543	0.3535019	-5.54	0.000	-2.652394 -1.266692
/lnalpha	-1.274886	0.2225431	-5.73	0.000	-1.711062 -.8387092
alpha	0.2794629	0.0621925			.1806738 .4322681

ref. = reference category of the variable.

Figures

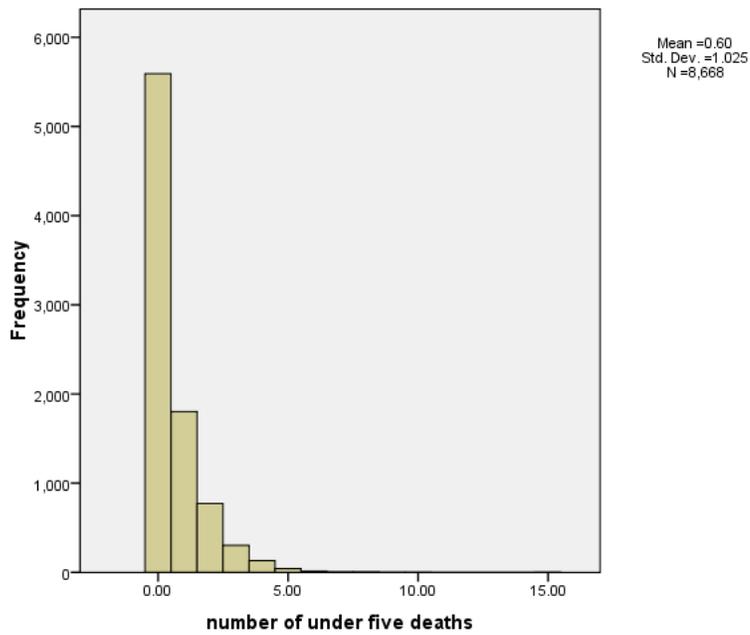


Figure 1

Histogram of the number of under-5 deaths

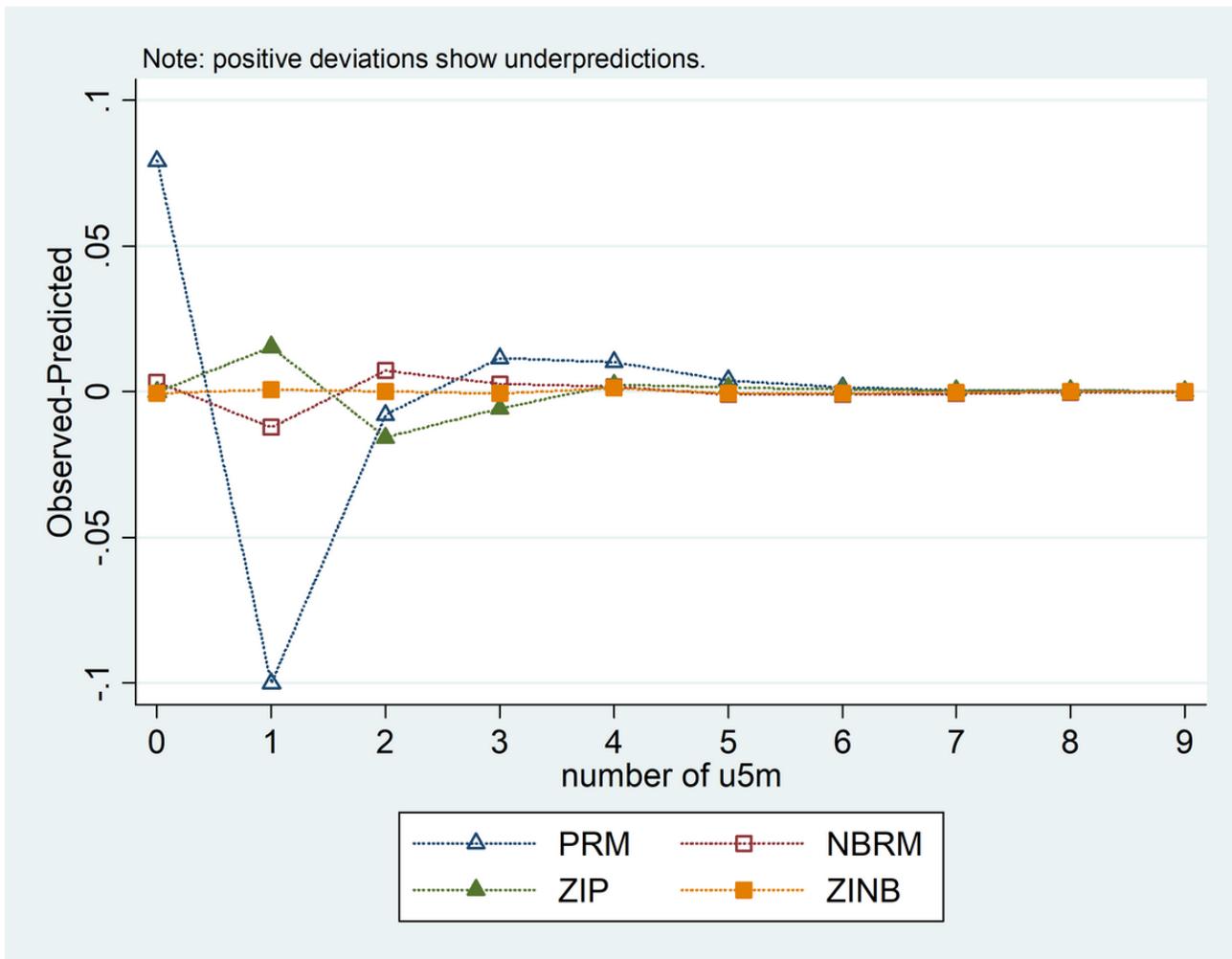


Figure 2

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [eq13.jpg](#)
- [eq8.jpg](#)
- [eq12.jpg](#)
- [eq14.jpg](#)
- [eq10.jpg](#)
- [eq1.jpg](#)
- [eq3.jpg](#)
- [eq2.jpg](#)
- [eq11.jpg](#)
- [eq7.jpg](#)
- [eq9.jpg](#)
- [rurall.xls](#)
- [eq5.jpg](#)
- [eq4.jpg](#)
- [eq16.jpg](#)
- [eq15.jpg](#)
- [eq6.jpg](#)