International Journal of Healthcare Sciences ISSN 2348-5728 (Online)
Vol. 3, Issue 1, pp: (332-355), Month: April 2015 - September 2015, Available at: www.researchpublish.com

Modeling the Number of Antenatal Care Service Visits Among Pregnant Women in Rural Ethiopia: Zero Inflated and Hurdle Model Specifications

Dawit Sekata

Department of Statistics, Wollega University, Nekemte, P.O.Box 395, Ethiopia

Abstract: **Antenatal care is a preventive obstetric health care program aimed at optimizing maternal fetal outcome through regular monitoring of pregnancy [55]. Even if WHO recommends a minimum of four ANC visits for normal pregnancy, existing evidence from developing countries including Ethiopia indicates there are few women who utilize it due to different reasons. 1127 pregnant women from EDHS of 2011 were used to analyze the determinants of the barriers in number of antenatal care service visits among pregnant women in rural Ethiopia. The data were found to have excess zeros (51.5%); the variance, 7.196, is much higher than its mean, 1.85.Thus several count models such as Poisson, NB, ZIP, ZINB, HP, and HNB regression models were fitted to select the model which best fits the data. Each of these models was compared using SAS version 9.2 by their likelihood ratio test (LR), Voung test and the information criteria's. Through the analysis, access to modern ANC visits during pregnancy was low in rural Ethiopia than the national value. Lack of awareness, absence of education, heavy workload, poverty, and shortage of health post were significantly associated with not attending ANC visits. Hurdle Poisson regression model was found to be better fitted with data which is characterized by excess zeros and high variability in the non-zero outcomes.**

Keywords: **ANC; Hurdle; NB; Poisson; Zero-Inflated.**

1. INTRODUCTION

The WHO [55] estimates that about 536,000 women of reproductive age die each year because of pregnancy related complications. Nearly all of these deaths (99%) occur in the developing world [56]. Ethiopia is one of the countries with an unacceptably highest maternal mortality and the infant mortality rate in the world [12]. In Ethiopia, the maternal mortality was estimated to be 673 deaths per 100,000 live births and infant mortality rate was 77 per 1,000 live births, which is among the highest in the world [9]. This is because 85% of the population lives in rural areas where the availability of health services, especially maternal health care services, is extremely difficult. Overall, access for maternity care is on average 26% for rural and 76% for urban areas [55]. Antenatal care is a preventive obstetric health care program aimed at optimizing maternal fetal outcome through regular monitoring of pregnancy [55]. Its main objective is to ensure a normal pregnancy with delivery of a healthy baby from a healthy mother. That is why WHO [53] recommends at least four ANC visits for women without complicated pregnancy in developing countries which would include compulsory blood pressure measurement, urine and blood tests, exclude PIH and anemia, assess maternal and fetal well-beings, review and modify birth and emergency plans, advice and counsel, give preventive measures and non-compulsory weight and height check at each visit [55]. Currently, 71% of women worldwide receive any ANC; in industrialized countries, more than 95% of pregnant women have access to ANC. In sub-Saharan Africa, 69% of pregnant women have at least one ANC visit [42]. While different studies have looked at diverse risk factors for antenatal care

(ANC) and delivery service utilization in the country, MoH of Ethiopia in 2007 reported that about 52% Ethiopian women received one or more ANC visits, less than 17% received professionally assisted delivery care and 19% received postnatal care [40]. Coverage for ANC is usually expressed as the proportion of women who have had at least one ANC visit during her pregnancy. As emphasized in the 2005 EDHS, access to modern ANC visits during pregnancy in rural areas are remained very limited by international standards. [9]. According to EDHS, 2005, only 6% of women make their first ANC visit before the fourth month of pregnancy [55]. The median duration of pregnancy for the first ANC visit was 5.6 months. The median duration of pregnancy for the first ANC visit was 4.2 months for urban women compared with 6.0 for rural women. In urban area where the health services are physically accessible and ANC at the public services are provided free of charge, only 32.4% of women seek the service before 16 weeks of gestation [9]. The report identified that, 72% of mothers with at least secondary school education received ANC compared to 45% and 21% of mothers' with primary and no education respectively. Well oriented women's about pregnancy complications are also more likely to report four or more visits. In most countries, the greatest proportionate difference occurs between women following socioeconomic, demographic, health and environmental related factors [14]. In Metekel zone, Northwest Ethiopia, 49.8% of pregnant women had received at least one antenatal care visit during the pregnancy of their last delivery [18]. According to the study report, lack of awareness, low educational status and socio-economic characteristics, place of residence, educational status, husband's educational status, possessing radio, monthly income and knowledge about antenatal care were found to have a statistically significant reasons mentioned for not attending antenatal care utilization in the zone [18]. The proportion of women who received antenatal care for their recent births in Samre Saharti District, Tigray, Ethiopia was 54% [58]. According to the study, education, parity, family education, history of obstructed labor and ANC visit were significant predictors for the selection of delivery place. Similarly, in Maichew Town, Southern Tigray Ethiopia, 80% of pregnant women had at least one antenatal visit during their pregnancy period [22]. On the contrary, a study conducted in Southwestern Ethiopia in 2009 [3] showed that 28.5% of pregnant women in Yem Special Woreda received ANC at least once but the majority 71.5% reported that they did not attend ANC up to their last pregnancy. In 2010, 86.3% in Hadiya Zone of Southern Ethiopia had received at least one antenatal visit during their last pregnancy [59]. Maternal age, husband attitude, family size, maternal education, and perceived morbidity were major predictors of antenatal care service utilization [59]. The EDHS, 2005 and community and family survey conducted in SNNPR to assess maternity care utilization, also reflected the above situation [18]. However, information coming from community-based studies related to the Health Extension Programme (HEP) in rural areas is still very limited due to different factors. That is why this study is aimed to statistically analyze the determinants of the barriers in number of antenatal care service visits among pregnant women in rural Ethiopia. Furthermore, this study will provide valuable information about count data models when the assumption of the standard Poisson regression is violated (when there is greater variability in the response counts than one would expect if the response distribution truly were Poisson). According to some study, the negative binomial and ZIP model appears to be superior when the event -stage distribution is positive and when there is moderate to moderately-high zero-inflation but not extreme zero –inflation [19, 39]. In such occasions, it is of interest to examine the applicability of the Zero Inflated models (ZIP, ZINB) and Hurdle models (Hurdle Poisson and Hurdle NB) in addition to NB and Poisson regression models and compare their performances in terms of their goodness-of-fit statistics, AIC, BIC, likelihood ratio test and theoretical soundness.

2. RESEARCH ELABORATIONS

2.1 Sources of Data and Variables in the Model:

The data used for this study was taken from the 2011 EDHS which is a nationally representative survey of women in the fertile age (15-49 years age) groups taken from the CSA, Ethiopia. This survey is the third compressive survey designed to provide estimates for the health and demographic variables of interest for the the whole urban and rural areas of Ethiopia as a domain. Women who had 9 months pregnancy during the survey interview were included in the analysis. The study includes 1127, 27.03% of fertile age group in the country, of rural women who had at least 9 months of pregnancy period during survey in 2011 DHS.

The response variable of this study is the number of antenatal care visits of pregnant women from early pregnancy to their 9 months of pregnancy period. Thus, this paper attempts to include socioeconomic, demographic and health and environmental related factors that are assumed as a potential determinants for the barriers in the number of antenatal care service visits, adopted from literature reviews and their theoretical justification. Detailed descriptions of these factors are presented in Table 1.

Table 1: Variable Description for the Analyzed ANC Visits

Dependent Variable	Description
ANC	The number of ANC service visits
Independent Variable	
MEDUC	Mother education Status: (0) she has no education (1) She can read and write.
REGION	Region of the Women: (0) Women from better progressed regions (1) Otherwise
RESID	Pregnant mother residing with husband/partner: (0) No (1) Yes
WLOAD	Workload inside and/or outside home: (0) No problem (1) a problem
WEALTH	Wealth index: (0) poor (1) middle (2) if rich
HPOST	Availability & accessibility of health post: (0) No problem (1) a problem
AWARN	Awareness about ANC and pregnancy complication: (0) No (1) yes
SIGN	Had seen sign of pregnancy complications: (0) No (1) Yes
PWANTD	Pregnancy wanted when became pregnant: (0) No (1) yes

Source: Taken from CSA of Ethiopia, EDHS 2011

2.2 Statistical Models:

An event count refers to the number of times an event occurs within a fixed interval often measured as a nonnegative, discrete, and constrained by a lower bound, which is typically zero. The lower bound constraint presents the greatest obstacle for analyzing count data when assuming a normal distribution for its skewness so that standard models, such as OLS regression are not appropriate. The Poisson regression is commonly used method to model count data formed under two principal assumptions: one is that events occur independently over given time and the other is that the conditional mean and variance are equal. However, the equality of the mean and variance rarely occurs; the variance may be either greater the mean (overdispersion) or less than the mean (underdispersion) [38].

Even though there are several statistical models, in the literature of statistical modeling for counts, some models may not be appropriate to deal with some specific types of data. That is why Cameron clarified that the use of standard OLS regression leads to significant deficiencies unless the mean of the counts is high [19]. When there is preponderance of zeros, several models have been proposed for analyzing such data. Substantively, the choice between these models should be based solely on the data generating process. However, datasets can vary as a function of both the proportions of zeros and the distribution for the non-zeros. Sometimes overdispersion of a data may not be significant if the percentage of zeros is too high (might be 80% or more) and in such case ZIP and ZINB have nearly identical estimate of the parameters [39]. But the paper suggests that ZIP does not fit the data well, if there is over-dispersion with moderate percentage of zeros. Hurdle model has a higher flexibility to fit a model with mixture of distribution for zeros and positive counts. And it performs in a competitive way with ZIP and ZINB [39].

Hence, from a number of models proposed to handle zero-inflated counts, this thesis focuses on Poisson, Negative Binomial, ZIP, ZINB, Hurdle Poisson models, and Hurdle Negative Binomial models and accessed different tests for comparing their performances.

2.2.1 Poisson Regression Model:

Given that the dependent variable (number of ANC visits) is a non-negative integer; most of the recent thinking in the field is the use Poisson regression model as a starting point. In a standard Poisson regression model, the probability of pregnant women i having y_i antenatal care service visits until her nine (9) months of pregnancy period (where y_i is a non-negative integer) is given by:

$p(y_i) = \frac{Exp(-\mu_i)\mu_i^{y_i}}{y_i!}$, y_i= 0, 1, 2, …($\mu_i > 0$) [44] ... (**Eq 1**)

Where $p(y_i)$ is the probability of 9 month pregnant women entity i having y_i antenatal care service visits in nine (9) months of pregnancy period and μ_i is the Poisson parameter for pregnant women i, which is equal to pregnant women entity i's expected number of antenatal care service visits in nine (9) months, $E(y_i)$. Poisson regression models are estimated by specifying the Poisson parameter μ_i (the expected number of antenatal care service visits) as a function of explanatory variables, the most common functional form being $\mu_i = Exp(\beta X_i)$, where X_i is a vector of explanatory variables and β is a vector of estimable parameters.

The log-likelihood function is:

$l(\mu_i) = l(\mu_i; y) = \sum_{i=1}^{n}\{y_i \ln(\mu_i) - \mu_i - \ln(y_i!)\}$…… (**Eq 2**)

Let X be a $n\times(p+1)$ matrix of explanatory variables. The relationship between y_i and i^{th} row vector of X, x_i linked by $l(\mu_i)$ is: $ln(\mu_i) = \eta_i = x_i^T\beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}$ [44]

There are two principal assumptions in the Poisson model we need to regard: one is that events occur independently over time or exposure period, the other is that the conditional mean and variance are equal [4]. The latter assumption is quite important. If it fails, the fitted model should be reconsidered.

Although the Poisson model has served as a starting point for count or frequency analysis for several decades, researchers have often found that count data exhibit characteristics that make the application of the simple Poisson regression problematic. Specifically, Poisson models cannot handle over- and under-dispersion and they can be adversely affected by low sample means and can produce biased results in small samples. Therefore, it should be performed using some appropriate modeling procedure.

2.2.2 Negative Binomial Regression Model:

The negative binomial (or Poisson-gamma) model is an extension of the Poisson model to overcome possible over-dispersion in the data. This model assumes that the Poisson parameter follows a gamma probability distribution. The negative binomial model is derived by rewriting the Poisson parameter for each observation i as $\mu_i = Exp(\beta X_i + \varepsilon_i)$ where $Exp(\varepsilon_i)$ is a gamma-distributed error term with mean 1 and variance k. The addition of this term allows the variance to differ from the mean as:

$Var[y_i] = E[y_i][1 + kE[y_i]] = E[y_i] + kE[y_i]$.

The probability mass function for the negative binomial distribution is:

$p(Y_i) = \binom{y_i + r - 1}{y_i} p^r(1-p)^{y_i}, r = 0,1,2,\ldots$ [44] …(**Eq 4**)

The parameter p is the probability of success in each trial and it is calculated as:

$p = \frac{r}{\mu_i + r}$ where, $\mu_i = E(Y)$= mean of the observations; and r = inverse of the dispersion parameter k ($i.e. r = \frac{1}{k}$). When the parameter r is extended to a real, positive number, its PMF can be rewritten using the gamma function:

$p(Y_i = y_i) = \frac{\Gamma(y_i + r)}{\Gamma(r)\Gamma(y_i+1)} p^r(1-p)^{y_i}, y_i \epsilon\{0\}UZ^+$..(**Eq 5**)

Where $\Gamma(.)$ is the gamma function. The mean and variance of the negative binomial are $E[y_i] = \mu = r\frac{1-p}{p}$ and $Var[y_i] = \frac{1-p}{p^2}$ [8] respectively. It is common to parameterize r and p in the terms of k and μ. Define $k = \frac{1}{r}$, $\mu = \frac{1-p}{kp}$, solving yields $p = \frac{1}{(1+k\mu)}$. After re-parameterization, the above model becomes

$p(Y_i = y_i) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{\Gamma\left(\frac{1}{k}\right)\Gamma(y_i+1)} \left(\frac{1}{(1+k\mu)}\right)^{\frac{1}{k}} \left(\frac{k\mu}{1+k\mu}\right)^{y_i}$….. (**Eq 6**)

The mean of this parameterization is $E[y_i] = \mu$ and $Var[y_i] = \mu + k\mu^2$. This is known as the "NB-2" model because it has a quadratic variance function. In this model $k \geq 0$ and if $k = 0$, then it reduces to a Poisson.

The negative binomial model can be estimated using maximum likelihood. The NB2 likelihood function is:

$l(\mu_i|k, y_i) = \sum_{i=1}^{n}[y_i \ln(\frac{k\mu_i}{k\mu_i+1}) - \frac{1}{k}\ln(k\mu_i + 1) + ln\,\Gamma\left(y_i + \frac{1}{k}\right) - ln\,\Gamma(y_i + 1) - ln\Gamma\left(\frac{1}{k}\right)$ ....**(Eq 7)**

In the NB regression model, μ_i is linked to the covariates: $\mu_i = Exp(x_i\beta)$. In the context of the NB GLM, the mean response for the number of antenatal care service visits is assumed to have a log-linear relationship with the covariates and is structured as:

$ln(\mu_i) = \beta_0 + \sum_{i=1}^{p} \beta_i\, x_i$ …………………….. **(Eq 8)**

Where, x_i = selected determinants of the barriers in number of ANC; $\beta' s$= regression coefficients to be estimated; and, p = total number of covariates in the model [44].

The Poisson regression model is a limiting model of the negative binomial regression model as k approaches zero, which means that the selection between these two models is dependent upon the value of k. The parameter k is often referred to as the overdispersion parameter.

The Poisson-gamma/negative binomial model has its limitations, most notably its inability to handle under-dispersed data, and dispersion-parameter estimation problems when the data are characterized by the low sample mean values and small sample sizes [3, 32, 35]. Furthermore, although the negative binomial model can solve an overdispersion problem, it may not be enough flexible to handle when there are excess zeros. In such cases, one can use the zero-inflated models as well as hurdle models to solve the problem.

2.2.3 Zero-Inflated Models:

There are situations where a major source of overdispersion is a preponderance of zero counts, and the resulting overdispersion cannot be modeled accurately with negative binomial model. In such scenarios, one can use zero-inflated Poisson or zero-inflated negative binomial model to fit the data [47].

According to Lord, Zero-inflated techniques permit the researcher to answer two questions that pertain to low base rate-dependent variables: (a) what predicts whether or not the event occurs, and (b) if the event occurs, what predicts frequency of occurrence? In other words, two regression equations are created: one predicting whether the count occurs and a second one predicting the occurrence of the count [32]. Moreover, zero-inflated models have statistical advantage to standard Poisson and negative binomial models in such a way that they model the preponderance of zeros as well as the distribution of positive counts simultaneously[41].

2.2.3.1 Zero-Inflated Poisson Regression Models:

ZIP model operates on the principle that the excess zero density that cannot be accommodated by a traditional count structure. The probability of an ANC visitation entity being in zero or non-zero states can be accounted by a splitting regime that models a woman who is not visited for ANC versus a woman who has visited for ANC during their pregnancy period determined by a binary logit or probit model [29, 54].

In one regime (R_1) the outcome is always a zero count, while in the other regime (R_2) the counts follow a standard Poisson process. Suppose that: $p[y_i \epsilon R_1] = \omega_i$; $p[y_i \epsilon R_2] = (1 - \omega_i); i = 1, 2, \ldots\ldots, n$. Where ω_i = Inflation Probability

Then, this two-state process gives a simple two-component mixture distribution with PMF [26].

$$p(Y_i = y_i) = \begin{cases} \omega_i + (1-\omega_i)e^{-\mu_i};\ when\ y_i = 0 \\ (1-\omega_i)\frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!};\ when\ y_i > 0 \end{cases} \qquad \mu_i > 0; and\ 0 \le \omega_i \le 1 \ldots\ldots\ldots\ldots\ldots\ldots \textbf{(Eq 9)}$$

As before, covariates enter the model through the conditional mean, μ_i, of the Poisson distribution: $\mu_i = Exp(x_i^T\beta)$, where x_i^T is a $(1 \times p)$ vector of the i^{th} observation on the covariates, and β is a $(p \times 1)$ vector of coefficients.

Clearly, $E(y_i) = (1 - \omega_i)\mu_i = \mu_i$ and $Var(y_i) = \mu_i + \left(\frac{\omega_i}{1-\omega_i}\right)\mu_i^2 = (1 - \omega_i)(\mu_i + \omega_i\mu_i^2)$ indicating that the marginal distribution of y_i exhibits over-dispersion of the data $(if\ \omega_i > 0)$. It is clear that this reduces to the standard Poisson model when $\omega_i = 0$. This over-dispersion does not arise from heterogeneity, as is case when the Poisson model is generalized to the Negative Binomial model. Instead, it arises from the splitting of the data into the two regimes. In practice, the presence of over-dispersion may come from one or both of these sources [17, 30].

2.2.3.2 Zero-Inflated NB Regression Models:

Zero-Inflated Negative Binomial (ZINB) regression model also assumes two distinct data generation processes. The result of a Bernoulli trial is used to determine which of the two processes is used. For mother i, with probability ω_i the only possible response of the first process is zero counts, and with probability of $(1-\omega_i)$ the response of the second process is governed by a negative binomial with mean μ_i. The zero counts are generated from both the first and second processes, where a probability is estimated for whether zero counts are from the first or the second process. The overall probability of zero counts is the combined probability of zeros from the two processes.

A ZINB model for the response y_i (the number of ANC visits during pregnancy) can be written as:

$$P(Y)=\begin{cases}\omega_i+(1-\omega_i)(1+v\mu_i)^{\frac{1}{k}}; & when\ y_i=0\\ (1-\omega_i)\frac{\Gamma(y_i+\frac{1}{k})}{\Gamma(y_i+1)}\frac{(k\mu_i)^{y_i}}{(1+k\mu_i)^{y_i+\frac{1}{k}}}; when\ y_i>0\end{cases} \quad \text{.... (Eq 12)}$$

In this case, the mean and variance of the y_i are:

$E[y_i]=(1-\omega_i)\mu_i \& Var[y_i]=(1-\omega_i)\mu_i(1+\mu_i(\omega_i+k))$ Where μ_i is the mean of the underlying negative binomial distribution, and v is the over-dispersion parameter [29]. The ZINB distribution reduces to the ZIP distribution as $k\to 0$. The parameter μ_i is modeled as a function of a linear predictor, that is, $\mu_i=Exp(x_i^T\beta)$. β is the $(p+1)\times 1$ vector of unknown parameters associated with the known covariate vector $x_i^T=(1,x_{i1},\dots,x_{ip})$, where p is the number of covariates not including the intercept. The parameter ω_i, which is often referred as the zero-inflation factor, is the probability of zero counts from the binary process. For common choice and simplicity, ω_i is characterized in terms of a logistic regression model by writing as $logit(\omega_i)=Z_j^T\gamma$. γ is the $(q+1)\times 1$ vector of zero-inflated coefficients to be estimated, associated with the known zero-inflation covariate vector $Z_j^T=(1,Z_{j1},\dots,Z_{jq})$, where q is the number of the covariates Z's not including the intercept. In the terminology of generalized linear models (GLMs) $\log(\mu_i)$ and $\text{logit}(\omega_i)$ are the natural links for the negative binomial mean and Bernoulli probability of success [29].

$\log(\mu_i)=\beta_0+\beta_1x_{i1}+\cdots+\beta_px_{ip}$ and

$\text{logit}(\omega_i)=\gamma_0+\gamma_1z_{i1}+\cdots+\gamma_pz_{iq}$ ……….. **(Eq 13)**

Where X_i and Z_i are respectively vectors of covariates for the negative binomial and the logistic components, and β and γ are the corresponding vectors of regression coefficients.

2.2.4 Hurdle Regression Models:

Hurdle regression is also known as two-part model which is originally developed by Mullahy (1986) [41]. Mullahy states, "The idea underlying the hurdle formulations is that a binomial probability model governs the binary outcome of whether a count variate has a zero or a positive realization. If the realization is non- zero (positive), the "hurdle is crossed", and the conditional distribution of the positives is governed by a truncated-at-zero count data model." The attraction of Hurdle regression is that it reflects a two-stage decision-making process in most human behaviors and therefore has an appealing interpretation [4].

For instance, it is pregnant mother's decision whether to contact the doctor's office and to make the initial visit. However, after the pregnant mother's first visit, doctor plays a more important role in determining if the pregnant mother needs to make follow-up visits. Therefore, in a regression setting, the first decision might be reflected by a Logit or Probit regression, while the second one can be analyzed by a truncated Poisson or Negative binomial regression. Moreover, different explanatory variables are allowed to have different impacts at each decision process.

2.2.4.1 Hurdle Poisson (HP) Regression Model:

The most popular formulation of a Hurdle regression is called Logit-Poisson model, which is the combination of a Logit regression modeling zero vs. non-zero outcomes and a truncated Poisson regression modeling positive counts conditional on non-zero outcomes. Its probability density function is given as:

$$p(y_i/x_i)=\begin{cases}\omega_i & for\ \ y_i=0\\ \dfrac{(1-\omega_i)Exp(\mu_i)\mu_i{}^{y_i}}{(1-Exp(-\mu_i))y_i!} & for\ \ y_i>0\end{cases}$$

$Where: \omega_i = p(y_i = 0),\ \mu_i = Exp(x_i\beta),$

$log\left(\frac{\omega_i}{1-\omega_i}\right) = Z_i{}^T\gamma\ and\ log(\mu_i) = X_i{}^T\beta$ [41].. (**Eq 14**)

The log-likelihood function of a Logit-Poisson regression therefore can be expressed as the sum of log-likelihood functions of two components as below:

$LL = \sum_{i=1}^{n}[I_{y_i=0}\log(\boldsymbol{\omega_i}) + I_{y_i>0}\log(\mathbf{1} - \boldsymbol{\omega_i}) - \boldsymbol{\mu_i} + \boldsymbol{y_i}\log(\boldsymbol{\mu_i}) - \log(\mathbf{1} - \boldsymbol{Exp}(-\boldsymbol{\mu_i}) - \log(\boldsymbol{y_i}!))]$... (**Eq 15**)

Unlike Poisson and Negative binomial regressions, Hurdle regression can only be modeled through log-likelihood function.

2.2.4.2 Hurdle NB Regressions Model:

We consider a hurdle negative binomial (HNB) regression model in which the response variable $y_i(i = 1, \dots, n)$ has the distribution

$$p(Y_i) = \begin{cases} \omega_i, & when\ y_i = 0 \quad \dots (\mathbf{Eq\ 16}) \\ (1-\omega_i)\frac{\Gamma(y_i+k^{-1})}{\Gamma(y_i+1)\Gamma(k^{-1})}\frac{(1+k\mu_i)^{-k^{-1}-y_i}k^{y_i}\mu_i^{y_i}}{1-(1+k\mu_i)^{-k^{-1}}}, & y_i > 0 \end{cases}$$

Where $(k \geq 0)$ is a dispersion parameter that is assumed not to depend on covariates [41]. In addition, we suppose $0 < \mu_i < 1$ and $\omega_i = \omega_i(z_j)$ satisfy

$logit(\theta_i) = \log\left(\frac{\boldsymbol{\omega_i}}{1-\boldsymbol{\omega_i}}\right) = \sum_{i=1}^{q}\boldsymbol{Z_i\gamma},$

$log(\mu_i) = \sum_{i=1}^{p}\boldsymbol{X_i\beta}$ ………………………….. (**Eq 17**)

Where Z_i and X_i are the i^{th} row of covariate matrix Z and X as well as β and γ are the independent variables in the regression model. We now obtain the log-likelihood function for the hurdle negative binomial regression model, we have:

$LL = \sum_{i=1}^{n}\{(1-d_i)[I_{y_i=0}\log\omega_i + I_{y_i>0}\{log(1-\omega_i) + log - log(1-(1+k\mu_i)^{-k^{-1}})\}] + d_i log\sum_{j=y_i}^{\infty}pr(Y_j = j)\}$ (**Eq 18**)

In many applications, extra zeros (relative to the Poisson model) generated by the above models are insufficient to account for the full amount of zeros in the data. All single index models have to compromise between the large proportion of zeros, which tends to lower the mean, and a right-skewed distribution of counts with large non-zero values, which tends to increase it. Moreover, one often has a substantive interest in treating the zero-generating process separately from the process for strictly positive outcomes, which requires different sets of parameters.

2.5 Goodness of Fit:

2.5. 1 Likelihood and Deviance Residual:

The likelihood function can be used to assess the goodness of fit of a model, and several further measures of model performance are based on it. It is to note that this assumes mutual independence of observations. In case the observations are not mutually independent, the likelihood will be overestimated. This will have the effect of exaggerating differences in log-likelihood and so will tend to favor elaborate models unduly.

Deviance provides an alternative to likelihood. The deviance is used as a measure of discrepancy of a generalized linear model; each unit i of observation contributes an amount D_i as an increment to total deviance. For the Poisson model with observed number y_i and corresponding estimated number u_i, residual deviance is given by:

$D_i = sign(y_i - u_i)\sqrt{d_i^2}$ [21]. ………………….. (**Eq 19**)

Where d_i^2 is the squared deviance residual which can be obtained according to the distribution as follows:

Poisson regression:

$$d_i^2 = \begin{Bmatrix} 2u_i & if\ y_i = 0 \\ 2\left\{y_i \ln\left(\frac{y_i}{u_i}\right) - (y_i - u_i)\right\} & otherwise \end{Bmatrix}$$ [21]...(**Eq 20**)

NB regression:

$$d_i^2 = \begin{cases} \frac{2ln(1+\alpha u_i)}{\alpha} & if\ y_i = 0 \quad \ldots.(\textbf{Eq 21}) \\ 2y_i\, ln\left(\frac{y_i}{u_i}\right) - \frac{2}{\alpha}(1+\alpha y_i)ln\left(\frac{1+\alpha y_i}{1+\alpha u_i}\right) & else \end{cases} [21]$$

Where α is the over-dispersion parameter. The standardized residuals were obtained by multiplying the deviance residual D_i by the factor $(1-h_i)^{-\frac{1}{2}}$ where h_i is the leverage, which indicates the influence of observation i.

The total residual deviance D of the model is given by summation over all units:

$D = \sum_{i=1}^{n} D_i$. For Poisson, a properly fitted model the expected value of residual deviance should be approximately equal to the residual degrees of freedom [37].

2.5.2 Likelihood Ratio Test:

The maximum likelihood estimation method is used to assess the adequacy of any two or more than two nested models by using the likelihood ratio test. it compares the maximum likelihood under the alternative hypothesis with the null hypothesis. For instance, the null hypothesis can be the overdispersion parameter is equal to zero (i.e. the Poisson distribution can be fitted well the data) and the alternative hypothesis can be the data would be better fitted by the Negative binomial regression (i.e. the overdispersion parameter is different from zero). The likelihood ratio test is defined as:

$$R_\omega = -2 \times [l(\hat{\mu}) - l(\hat{\mu}, \hat{\omega})] \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (\textbf{Eq 22})$$

$l(\hat{\mu})\ and\ l(\hat{\mu}, \hat{\omega})$ are the maximized log-likelihood of models under the alternative and null hypothesis respectively. From the earlier computations, this likelihood ratio test can be written as:

$$R_\omega = 2 \times \{n_0\, ln\left(\frac{n_0}{n}\right) + (n-n_0)\left(\ln\left(\frac{\bar{y}}{\hat{\mu}}\right) - \hat{\mu}\right) + n\bar{y}\,(ln\,\hat{\mu} + 1 - ln\bar{y}], \ldots\ldots\ldots\ldots\ldots.. (\textbf{Eq 23})$$

Where $\bar{y}$ is the mean of the observations under H_0 and $\hat{\mu}$ is the estimated positive mean counts under H_1. This test statistic R_ω approximately follows chi-square distribution on 1 degree of freedom (d.f) under the null hypothesis.

This has a chi-square distribution. As a result this test of statistics will be compare with the tabulated chi-square with a degree of freedom, the difference between the degree of freedom of the model under null hypothesis and the alternative hypothesis respectively. This method is not appropriate for models which are not nested one on the other, in such situation; we will use another method such as the Akaike information criteria (AIC) and Bayesian information criteria (BIC) [25].

In this study a likelihood ratio was used to compare the Poisson with the negative binomial and zero-inflated Poisson with zero-inflated negative binomial as well as Hurdle Poisson with Hurdle Negative Binomial since Poisson is nested on negative binomial and zero-inflated Poisson is nested in zero-inflated negative binomial; However this will not be used to compare Poisson or negative binomial with the zero inflated Poisson and negative binomial as long as these models are not nested one on the other.

2.5.3 Information Criteria:

If there are several models to be compared in order to select the best model which fits the data instead of using the likelihood ratio test, it can be easily select by using the Akaike information criteria (AIC) and Bayesian information criteria (BIC).

2.5.3.1 Akaike Information Criteria (AIC):

AIC is the most common means of identifying the model which fits well by comparing two or more than two models. It is trying to balance the goodness of fit against the complexity of the model It is similar as of the coefficient of multiple determination (R^2); however, it penalized by the number of parameter included in the model (i.e. the complexity of the model). Unlike the R^2, the good model is the one which has the minimum AIC value. It is given by the following formula:

$$AIC = -2l + 2k. \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (\textbf{Eq 25})$$

Where l are the log likelihood of a model that will compare with the other models and k is the number of parameter in the model including the intercept [21].

2.5.3.2 Bayesian Information Criteria (BIC):

Unlike the Akaike information criteria the Bayesian information matrix (BIC) takes in to account the size of the data under considered. It is given by:

$BIC = -2l + klog(n)$. [21] ……………………….. (**Eq 26**)

Where l are the log likelihood of a model that will compare with the other models, n is the sample size of the data and k is the number of parameters in the model including the intercept.

For this study the AIC is preferred over the BIC as it is more stringent and has a stricter entry requirement than BIC for additional parameters when large datasets are used. This helps to resolve over-fitting of models where many additional parameters are added to increase the likelihood, so AIC helps to promote a parsimonious model [51].

2.5.4 Voung 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 [53]. For instance, comparisons between Zero-inflated count models with ordinary Poisson, or Zero-inflated negative binomial against ordinary negative binomial model can be done using Voung test. This test is used for model comparison. Let's define: $m_i = \left(\frac{P_1(Y_i|X_i)}{P_2(Y_i|X_i)}\right)$. Where $P_N(Y_i|X_i)$ is the predicted probability of observed count for case i from model N, then Vuong test statistic test the hypothesis of $E(m_i = 0)$ given as:

$V = \frac{\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} m_i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(m_i - \bar{m})^2}}$. …………………………….. (**Eq 28**)

The test statistic provides evidence of the superiority of model 1 over model 2. If $V > 1.96$, the first model is preferred. But if $V < 1.96$, the second model is preferred.

2.6 Software:

Almost all statistical computation was carried out using SAS version 9.2. For all regression modeling we used Proc NLMIXED, specifying the likelihood equations, and maximizing them directly using numerical methods. Maximization began from various starting points and the final gradient vectors and hessian matrices were investigated to ensure proper convergence of estimated model parameters. In addition, all hypotheses were tested at 0.05 level of significance. R statistical software version 3.0.3 was used for graphical purpose.

3. RESULTS

3.1. Descriptive Analysis:

The descriptive statistics given in table 2 below shows the number and percentage of ANC visits that the pregnant mothers in the sample have encountered in their nine months of pregnancy period. It can be seen that 580 (51.5%) of the pregnant mothers have not visited ANC service during their periods of pregnancy months, whereas 125 (11.1%) of them visited only once, 91(8.1%) of them visited twice, 77(6.8%) visited three times, 82 (7.3%) visited four times and etc.

Table 2: Number of mothers that experienced ANC visits

Number of ANC visits	Percent	Cumulative Percent
0	51.5	51.5
1	11.1	62.6
2	8.1	70.6
3	6.8	77.5
4	7.3	84.7
5	4.2	88.9
6	3.5	92.4
7	2.5	94.9
8	2.0	96.8
9	1.1	97.9
10	0.4	98.3
11	0.6	98.9
12	0.8	99.7
13	0.3	100

These were again supported by figure 1 below. Since there is large number of zero outcomes, the histograms are highly picked at the very beginning (about the zero values). However large observations (i.e. large number of ANC visits) are less frequently observed. This leads to have a positively (or right) skewed distribution.

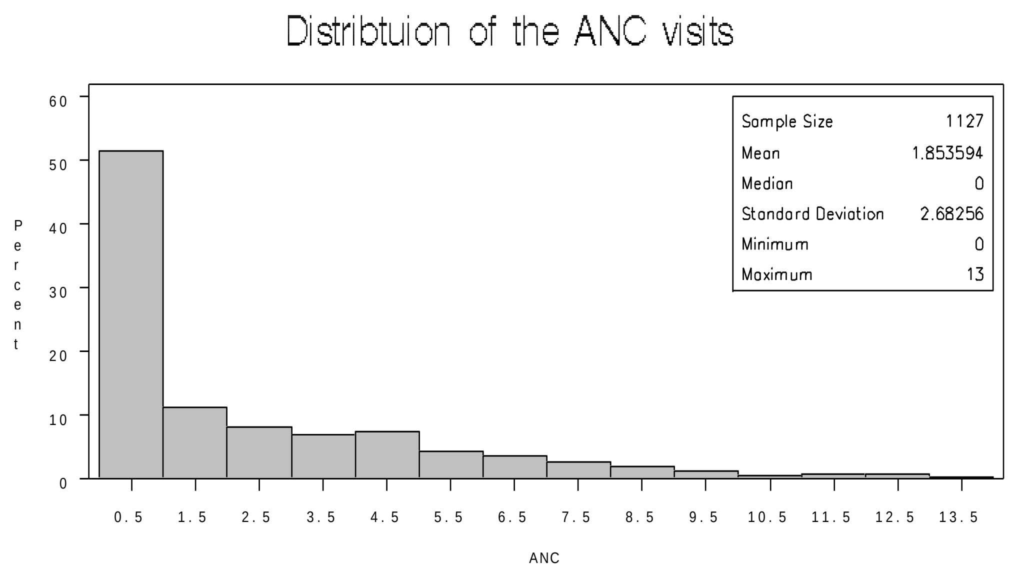

Figure 1: Histogram of ANC service visits

This was an indication that the data could be fitted better by count data models which takes into account excess zeroes.

Table 3 below presents summary statistics of the variables that are assumed to affect the number of ANC visits and its distributions for each levels of the variables.

Accordingly, less than one-third, 345 (30.6%) of the mothers can read and write while more than two-third, 782 (69.4%) of them have no education. Figure 2a & 2b below again confirms that the distribution of the number of ANC visits per Region and Education status of Mother in each group differs considerably.

Table 3: Descriptive Statistics of ANC Services Utilization among Pregnant Women in Rural Ethiopia

Variable	Category	Mn (Mx)	N (%)	Mean(St.dev)
MEDUC	No education	0(12)	782 (69.4)	0.98 (1.713)
	Can read & write	0(13)	345 (30.6)	3.84 (3.348)
REGION	BP Regions	0(13)	712 (63.2)	2.43 (3.008)
	The other region	0(11)	415 (36.8)	0.86 (1.570)
RESID	No	0(13)	476 (42.2)	1.22 (2.443)
	Yes	0(12)	651 (57.8)	2.32 (2.755)
WLOAD	No problem	0(13)	468 (41.5)	3.62 (3.189)
	Problem	0(8)	659 (58.5)	0.60 (1.138)
WEALTH	Poor	0(12)	588 (52.2)	0.61 (1.295)
	Middle	0(13)	426 (37.8)	2.90 (3.050)
	Rich	0(12)	113 (10.0)	4.40 (3.061)
HPOST	No problem	0(13)	526 (46.7)	3.28 (3.175)
	Problem	0(6)	601 (53.3)	0.60 (1.152)
AWARN	No	0(9)	597 (53.0)	0.51 (1.088)
	Yes	0(13)	530 (47.0)	3.37 (3.102)
SIGN	No	0(4)	169 (15.0)	0.62 (1.134)
	Yes	0(13)	958 (85.0)	2.07 (2.815)
PWANTD	No	0(13)	174 (15.4)	4.19 (3.516)
	Yes	0(12)	953 (84.6)	1.43 (2.255)

Mn-Minimum; mx-Maximum; BP-Better Progressed

Since there are a number of ANC visit outcomes for mothers from better progressed regions and educated mothers, the plots looks like obese after 5 ANC in both groups. However, large numbers of ANC visits are less frequently observed.

The number of participated pregnant women from regions that need special aids (Afar, Somali, Benishangul-Gumuz, and Gambella) found to be lower 415 (36.8%) than the number of mothers from better progressed regions (Tigray, Amhara , Oromiya, and SNNPR), 712 (63.2%). It was also observed that husband or partner of 476 (42.2%) pregnant mothers were not living with them, 651 (57.8%) of them were residing with their husband or partner during the time of their pregnancy periods. About 588 (52.2%) of sampled pregnant mothers were poor, 426 (37.8%) had middle income, and 113 (10.0%) were rich [Figure 5]. The number of pregnant mothers who had a problem of workload inside and/or outside home was 659 (58.5%) and those pregnant mothers who had no problem of workload inside and/or outside home was found to be 468 (41.5%). The frequency of pregnant mothers who became pregnant unexpectedly was 174 (15.4%) and majority of them, 953 (84.6%) became pregnant eagerly.

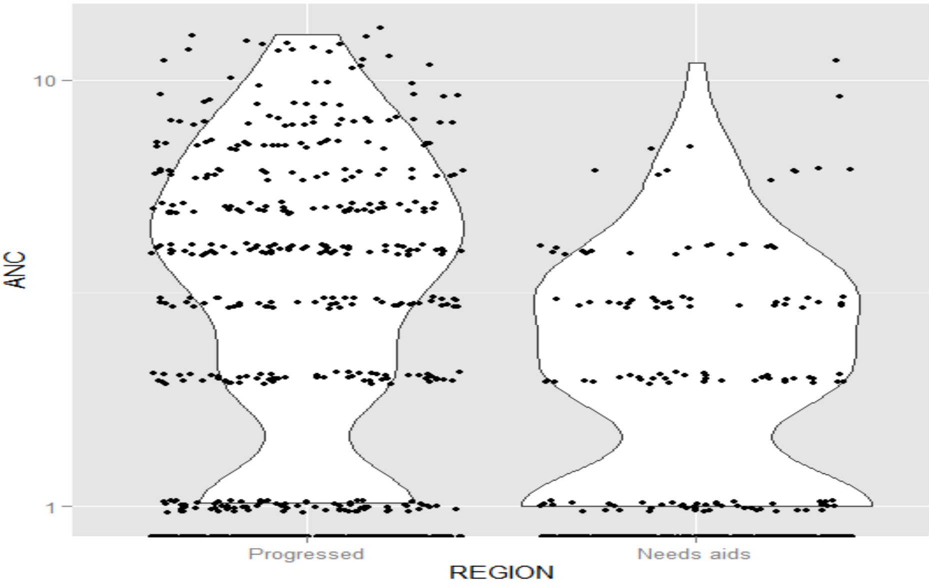

Figure 2a: Profile plot of ANC visits in Region

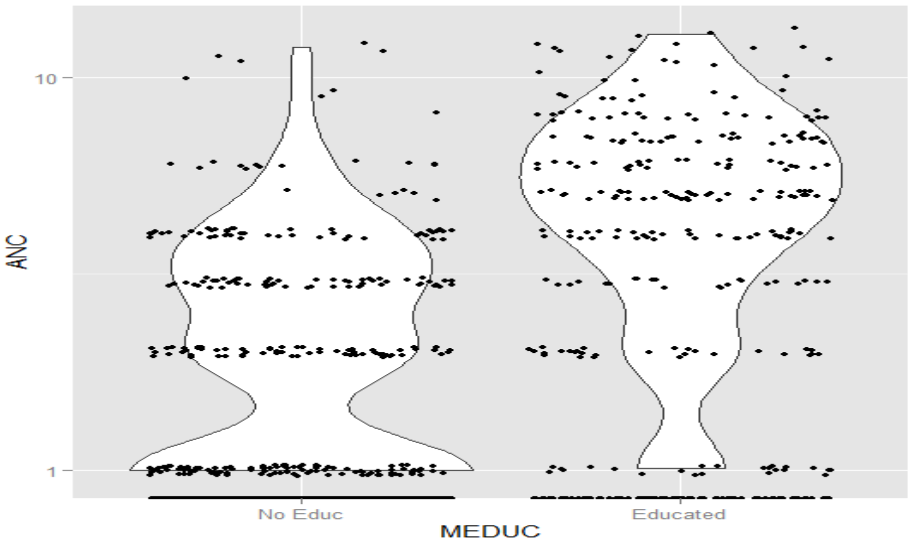

Figure 2b: Profile plot of ANC visits per Education Status

This table (Table 3) again reflects that nationally pregnant mothers use ANC visits approximately twice (1.85 visits) per their duration of pregnancy periods with standard deviation of 2.683, which is more than the mean indicating overdispersion. The number of ANC service visits during pregnancy for educated pregnant mothers is 3.84 ≈ 4visits and 0.98≈ 1 visits for that of non-educated mothers. The average number of ANC service visits for pregnant mothers from better progressed regions is two (2.43) times while the average number of ANC visits for pregnant mothers who are from regions who need special aids such as Afar, Somali, Benishangul-Gumuz and Gambella is found to be only once (0.86). The Table reveals that the happening of the signs of pregnancy complications during their pregnancy periods such as such as vaginal bleeding, vaginal gush of fluid, severe head ache, blurred vision, fever, abdominal pain had made variations among pregnant women. Hence, pregnant women who had ever seen the signs of pregnancy complications during their pregnancy periods used ANC service visits for more than twice (2.07) at average, despite the fact that the pregnant mother who had not seen the signs were visited only less than once, (0.62).

The descriptive statistics of Table 3 further illustrates that pregnant mother who had no radio or television at home and who was not visited by family planning worker last 12 months as well as not told about pregnancy complications(average ANC visits of 0.51), mother who had a problem of the availability of nearby health post and/or a problem of access to means of transportation (average ANC visits of 0.60), mothers who had a problem of workload inside and /or outside home (average ANC visits of 0.60), and poor pregnant mothers (average ANC visits of 0.61) were found to be the least ANC service users respectively. Therefore, the average number of ANC utilization ranges from pregnant mothers who had lack of awareness about the use of ANC and pregnancy complications (0.51 visits at average) to rich pregnant mothers (4.40 ≈4 visits) correspondingly.

3.2. Modeling the Number of ANC Service Visits:

3.2.1. Model Selection Summary Information:

As a starting point, the NB regression model has smallest AIC compared with the others when the whole explanatory variables are fitted under the six proposed models. Thus, it was used as the starting point for model selection. Then by penalizing a model with additional parameters, ten (10) models were refitted again under NB regression and the covariates with the largest p-value of Wald test is removed and refitted the model with the rest of the covariates sequentially. Finally, the status of the pregnant mother, either she is residing with her husband or not, (RESID) and whether the pregnancy is wanted when become pregnant (PWANTD) are the covariates excluded from the model; with Wald test p-value for the given covariates are large (P-value > 0.05).

3.2.2. Over dispersion and Poisson Regression:

In Poisson regression analyses, Table 4, deviance and Pearson Chi-square goodness of fit statistics indicating over dispersion was obtained as 1688.1931 and 1758.7784, respectively. Since the Pearson chi-square statistic divided by the degrees-of-freedom is higher than one and the observed value of 1.1268 is significantly different from one, with P-value 0.0019, then the mentioned goodness of statistics represents that there was an overdispersion in data set. Even if the Deviance and Pearson chi-square goodness of fit statistics of 1210.3476 and 1257.4983 respectively in Negative Binomial regression is dropped considerably but still an indication of significant overdispersion exists; because we would like this value divided by the degrees of freedom to be close to 1.

Table 4: Test for Overdispersion

Criteria	Models	DF	Value	Value/DF	p-value
Deviance	Poisson	1116	1688.1931	1.5127	<.0001
	NegBin	1116	1210.3476	1.0845	0.0252
Scaled	Poisson	1116	1688.1931	1.5127	<.0001
Deviance	NegBin	1116	1210.3476	1.0845	0.0252
Pearson	Poisson	1116	1758.7784	1.5760	<.0001
Chi-Square	NegBin	1116	1257.4983	1.1268	0.0019
Scaled	Poisson	1116	1758.7784	1.5760	<.0001
Pearson X2	NegBin	1116	1257.4983	1.1268	0.0019

The detailed parameter estimates and standard errors for each model are provided in Table 9, (on Appendix A). In this study we have considered different possible count data models. Likelihood ratio test (LR), (AIC), (BIC) and Vuong test were used to compare the candidate models to identify the most parsimonious model. The overdispersion parameter (k^{-1}) is significantly different from zero in NB and in both hurdle models (HP and HNB) regression models. As it can be seen from the table (Table 9), all covariates included in the standard Poisson model are significantly associated with the number of ANC visits even at 1% significance level. However in the case of the NB model only some of them are significant at 1% significance level.

3.2.3. Model Selection Criteria:

As shown in table 5, ZIP and ZINB regression models as well as HP and HNB were better fitted than Poisson and NB respectively based on their corresponding AIC. It was found out that the model with the smallest AIC and BIC was HP regression followed by ZIP regression model since their LR, χ^2 =3048.2 and χ^2 = 3049.6 both were highly significant (p-value<0.0001) supported by the information criteria's.

Table 5: Model Selection Criteria for the Regression Models

Criteria	P	NB	ZIP	ZINB	HP	HNB
-2 Log	3310.976	3230.0	3049.6	3063.9	3048.2	3059.0
AIC	3347.144	3254.0	3093.6	3109.9	3092.2	3105.0
AICC	3347.381	3254.3	3094.5	3110.9	3093.1	3106.0
BIC	3402.445	3314.4	3204.2	3225.5	3202.8	3220.6

The plots of predicted probability from each model against the observed probability of the outcome shows that the Poisson and the NB model under-estimated zero counts and the zero inflated and the hurdle models captured almost all zero values. Based on predicted probabilities, the differences in model fit between the six models were remarkable. Still the standard Poisson model and the NB model do not fit the data reasonably well; the standard Poisson predicted about 42% zeros and NB model predicted about 45% zeros compared to 51.5% observed zeros.

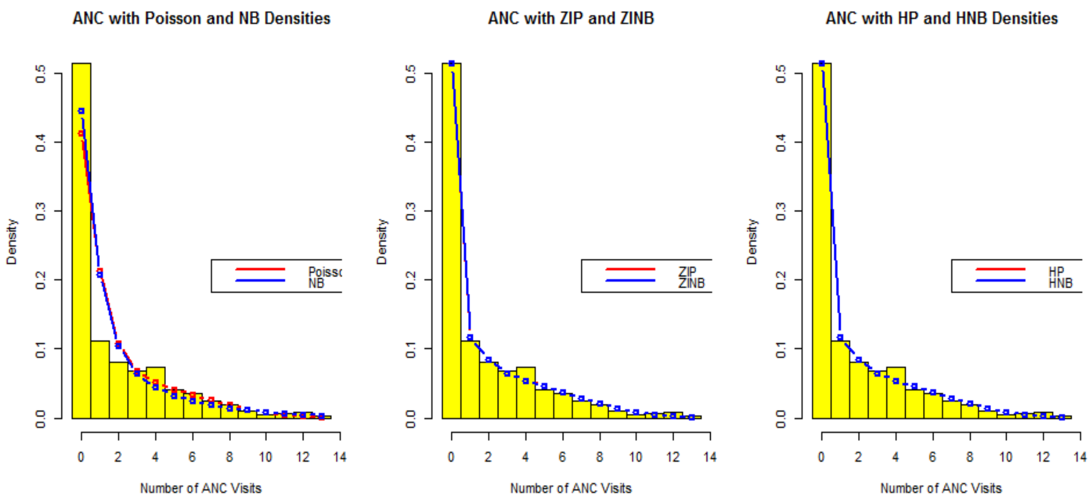

Figure 3: Comparison of the Densities of Each Model Fits

Table 6 (at appendix) clearly shows an improvement in model fitting from Poisson (LL = -0.2270) and negative binomial (LL = -0.1876) to zero-inflated Poisson (LL = -0.1284), zero-inflated negative binomial (LL= -0.1328), hurdle Poisson (LL= -0.1171) and hurdle negative Binomial (LL= -0.1345) models. The Vuong test statistic [53] found on Table 6 result reflected that all the candidate models, NB, ZIP, ZINB, HP, and HNB performed better than the standard Poisson model. zero-inflated Poisson performed better than NB (V=8.0639, P=6.661E^{-16}), better than Zero Inflated NB (V=1.9815, P=0.0307), better than Hurdle NB (V=1.9934, P=0.0464), which also holds for zero-inflated negative binomial vs. Hurdle negative binomial (V=-0.0423, P=1.0337). However, the hurdle Poisson model performed better than the ZIP model (V=1. 9704, P=0.04781).

The estimated value of k^{-1} (the overdispersion parameter) is 0.1476, 0.1574, 0.5727, 0.6179, 0.8757, and 0.8756 for NB, ZIP, ZINB, PH, and NBH respectively. Zero Inflated models are better in handling zero counts, but the AIC and LL values for ZIP and HP models were smaller compared to the others. In such situation, it would be better to use the model which takes into account the excess zeros and high variability due to non-zero outcomes. Therefore, since it has the smallest AIC (3092.2) as well as BIC (3202.8) values as presented in Table 5, HP regression model was chosen as the most parsimonious model which fits the data better than the other possible candidate models.

To better illustrate this fact, as well as to provide a more intuitive presentation of variables' influence on the expected count in the different models, Figure 4 presents the change in the value of μ as the value of the variable indicating a declaration of unconstitutionality goes from zero to one. While this variable is not, in fact, continuous, the graph does serve the useful purpose of outlining the general shape the variable's effects on the observed count. Each line was calculated holding all other variables at their mean values. We therefore turn to the hurdle and ZIP specifications, both to obtain more accurate results, and to examine the properties of the models described above.

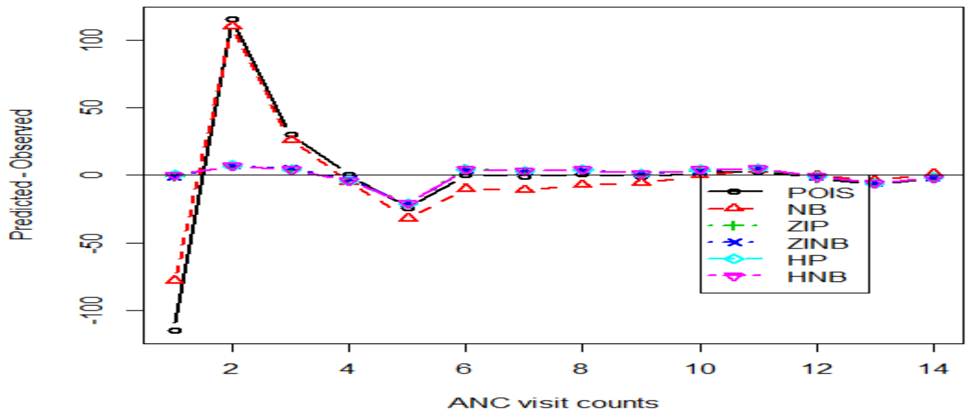

Figure 4: Observed vs Predicted Plots of the six Model Fits

Therefore, the final hurdle Poisson regression model proposed for number of ANC service utilization of pregnant mothers was in the next sub topic.

3.3. Parameter Estimates and Interpretation:

In the binary (logistic) portion of the ZIP model in Table 6 provides that all variables emerged as statistically significant predictors of number of ANC visits, since their p-values are less than 5%.

Table 6: Results for ZIP and Hurdle Poisson Model Estimates

	Zero Inflated poison		Hurdle Poisson	
	Poisson	Inflated part	Poisson	Inflated part
Variables	Estim (s.e)	Estim (s.e)	Estim (s.e)	Estim (s.e)
Intercept	0.4792 (0.3748)	0.6511 (0.1553)***	0.6993 (0.3066)*	0.6407 (0.1583)***
$MEDUC_{edu}$	-0.6590 (0.2243)**	0.3661 (0.0558)***	-0.7763 (0.1713)***	0.3596 (0.05593)***
$REGION_{oth}$	0.6808 (0.2032)***	-0.2171 (0.06965)**	0.6352 (0.1499)***	-0.2312 (0.07010)**
$WLOAD_{pro}$	-0.2154 (0.3786)	-1.0136 (0.1700)***	0.6260 (0.2506)*	-1.0168 (0.1668)***
$WEALTH_{ri}$	-0.4553 (0.1657)**	0.1435 (0.04046)***	-0.4593 (0.1308)***	0.1407 (0.04028)***
$HPOST_{prob}$	0.02906 (0.3731)	-0.7426 (0.1549)***	0.6267 (0.1978)**	-0.7579 (0.1535)***
$AWARN_{yes}$	-0.9398 (0.3046)**	0.3047 (0.1084) **	-0.9758 (0.2662)***	0.2935 (0.1085)**
$SIGN_{yes}$	-0.6941 (0.2796)*	0.3506 (0.1208) **	-0.7375 (0.2116)***	0.3844 (0.1234)**

** refers to p<0.05. ** refers to p<0.01. *** refers to p<0.001.*

When interpretating the binary portion of the model in different way from the interpretation of the count portion, the percentage changes in the factors are largely changed; and are more realistic than that of the Poisson model. Although we are still trying to estimate the relationship between each of the ANC variables and a binary outcome, here the two levels of the binary variable consist of either structural (or true) zeroes or sampling zeroes that follow the Poisson distribution. The percentages changes of the factors pregnant mother who have no education, mother from better progressed region, mother with a problem of work load inside and/or outside home, poor pregnant mother, mother with a problem of access to health post, mother who have lack of awareness about pregnancy complications as well as ANC utilization and pregnant mother who had not seen sign of pregnancy complications are 44.21%, 80.48%, 36.29%, 15.43%, 47.59%, 35.62%, 41.99%, 89.83%, and 62.26% respectively.

Consequently, the negative relationship between WLOAD since workload problem and the "no ANC visits" portion of our outcome indicates an inverse relationship between no workload problem of the women and "true" zeroes. That is, as workload problem inside outside home decreases, there is a greater likelihood of a positive number of ANC visits in the future. (OR =0.36). Similarly, the REGION and HPOST have a negative signs. A positive change in these factors induces then an increase in the number of ANC visits. The percentage change of the factor REGION is 80.48% (OR=0.8048) this means that the number of ANC service visited by the regions that need special aids is about one more likely to have zero visits than the better progressed region. The fitted model again suggests that the rate of non-zero ANC visits in educated mother was exp(0.3661)=1.44 times the rate of non zero ANC visits in non educated holding all other predictors constant. The rate of non-zero ANC visits for women who had seen signs of pregancy complications was exp(0.3506)=1.42 times the rate of non zero ANC visits in women who had no signs pregancy complications holding all other predictors constant. The presence of a statistically significant interaction term indicates that rich women with having awareness about ANC utilization have a 89.84% higher odds of having non zero ANC visits compared to poor women with lack awareness about ANC utilization in this study population.

Accordingly the percentages change of the factor HPOST is around 47.59%. That means, the number of non-zero ANC service visited by pregnant mother that have a problems of accessibility and/or availability of health post is 47.59% less non-zero ANC visits than that of women who have no problems related to health post. Whereas the percentages change of the factor MEDUC is around 44.21%. Hence, pregnant mothers who can read and write are 1.44 times more user of ANC service than mothers who have no education. For AWARN, the percentage change is around 35.62%. Thus, mothers who had lack of awareness about ANC utilization and pregnancy complications are showing less participation in ANC service to that of pregnant mothers who had awareness of ANC use. The presence of a statistically significant interaction term also indicates that pregnant mother in rural Ethiopia who had no health post related problems with not residing with her husband/ partner have 62.26% less likely ANC visits than pregnant mothers that have health post related problems and residing with her husband/ partner. If we consider the significance level of 5%, we conclude easily that there is no striking difference between zero-inflated Poisson regression fits and the hurdle Poisson (HP) model fits and they are better than the standard Poisson regression, Negative Binomial, ZINB, and NBH. But, the ZIP model is suitable only for handling zero inflation. However, the hurdle model is also suitable for modeling zero deflation. This tells us that even when a test shows significant evidence of zero inflation, the ZIP model may still not be suitable to fit the data. Since the hurdle Poisson (HP) model had the best fit than all the rest models, we interpreted the results from this model (Table 6).

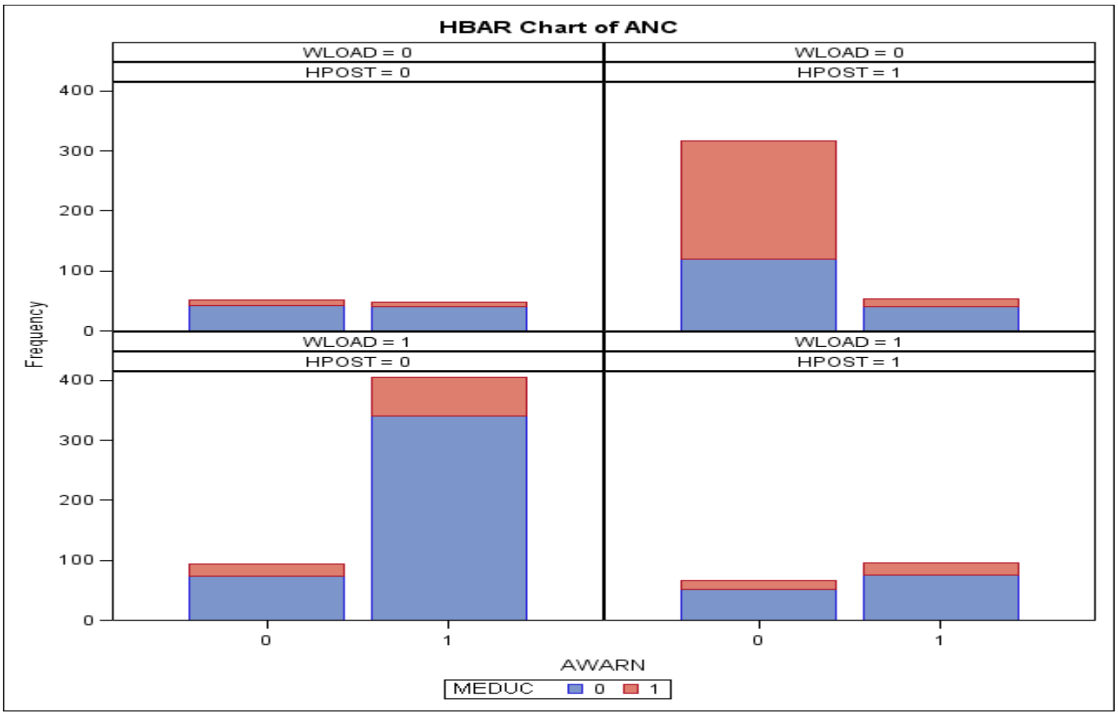

Figure 5: Bar Chart of WLOAD vs HPOST nested in AWARN

This figure shows a large numbers of women with adequate awareness about ANC usage with heavy workload and lack of nearby health post. In addition, many mothers who had no awareness about ANC usage and pregnancy complications have no workload problem inside and /or outside home but they have shortage of nearby heath post.

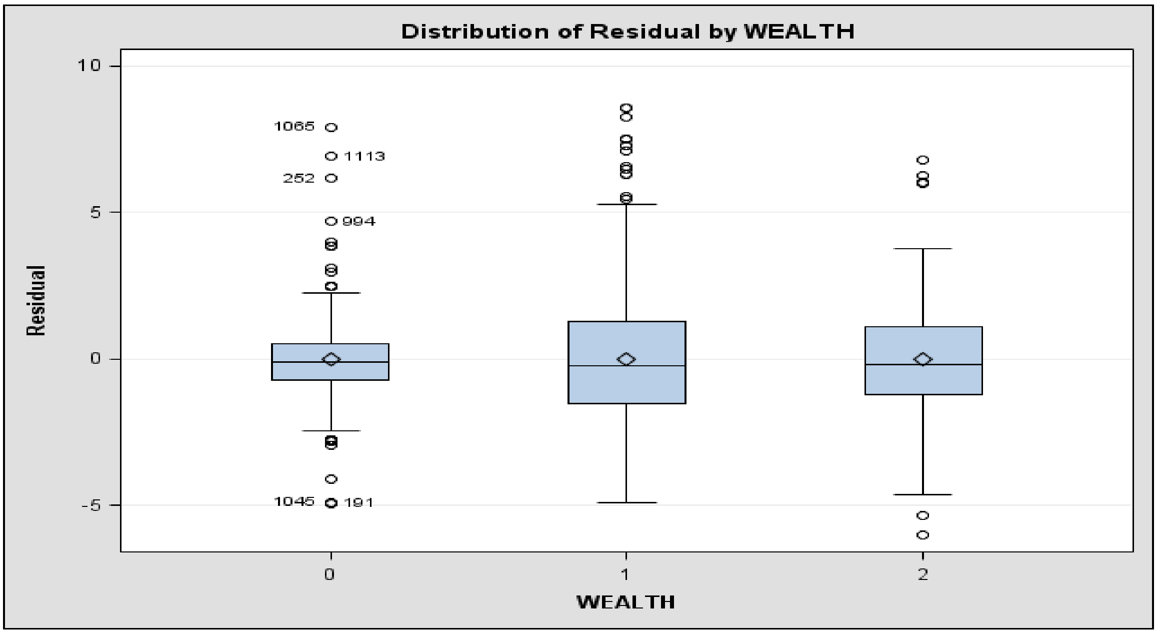

Figure 6: Both Plots of Wealth Index

In the same way, Figure 6 indicates that large numbers of women who utilize ANC have middle income and there were smaller numbers of women who make use of ANC visits having poverty. It certainly looks as if median ANC visit numbers are higher in rich women than in middle income, but the range of counts is very large in middle income earner women, so the significance of the difference is certain.

3.3.1. Hurdle Poisson Model Parameter Estimate:

The hurdle Poisson-logit model suggested that non educated pregnant mothers have a higher probability of not visiting ANC service and a higher expected number of zero visits than educated pregnant mothers. The non-zero part of HP-logit model fitting confirms this conclusion, as $\hat{\beta} = 0.3596$ has a standard error of 0.05593 (found in Table 7). The estimated odds that the number of ANC visits become zero with non educated (mothers who cannot read and write) are $\exp(0.3596\ ) = 1.43$ times the estimated odds for educated pregnant mothers. This estimate has the almost same order of magnitude as the estimate from the binary part of the ZIP model. The impact of covariates on the odds of visiting the ANC service for a less visit versus a more visits is quite different. For example, being residing with her husband/partner is not associated with the likelihood of ANC service utilization in the analysis characterized by slight number of ANC visits. However, better progressed regions such as Tigray, Amhara, Oromiya and SNNPR are statistically significantly associated with increased odds of at least one ANC visits in the analysis than not good enough regions. That means, better progressed regions are 0.79 times positive ANC visits than those regions that need special aids. This result indicates the importance of stratifying our analyses according to the severity of the growth level of rural Ethiopian regions, as the factors influencing the pregnancy complications and ANC service utilization.

The impact of access to a severe workload inside and/or outside home on ANC service utilization is an interesting finding in our analysis. Not having access to a severe workload inside and/or outside home did not influence the odds of ANC service utilization in those who demonstrated a number of ANC service visits over the study interval. For less number of ANC service visits, we estimate that having access to a primary care provider significantly reduces the likelihood (OR = 0.47) of a visit. Having signs of pregnancy complications such as vaginal bleeding, vaginal gush of fluid, severe head ache, blurred vision, fever, abdominal pain and others significantly increases the likelihood (OR = 1.47) ANC visits. That means, pregnant mothers who have cases of pregnancy complications have 1.47 times more non zero ANC visits than those who did not seen the signs of pregnancy complications. The interaction between access to a severe workload inside and/or outside home and lack of awareness about ANC utilization and pregnancy complications are statistically significant. Hence, a pregnant mother from rural Ethiopia who has no problem of workload inside and/or outside home as well as have good awareness about ANC utilization is 1.91 times more positive ANC visits than a mother with severe workload inside and/or outside home as well as lack of awareness about ANC utilizations.

3.4. Model Diagnostics:

From Figure 9 (found at Appendix), it seems that the variance stay constant as the fitted values vary, while there exist 3 outliers as labeled on Figure 9. The visual inspection plot of equal Cook's distance are shown in Figure 9 and Figure 7d to identify if any problem existed in the model. There are points having cook's distance larger as labeled. We plot the standardized deviance residuals (SDR) against the fitted rates. An informal procedure that is used to check for systematic departures from the Poisson regression is based on four regression diagnostic plots. Figure 7 contains the four regression diagnostic plots. A plot of the standardized deviance residuals (SDRs) against the fitted rate is shown in Figure 7a.

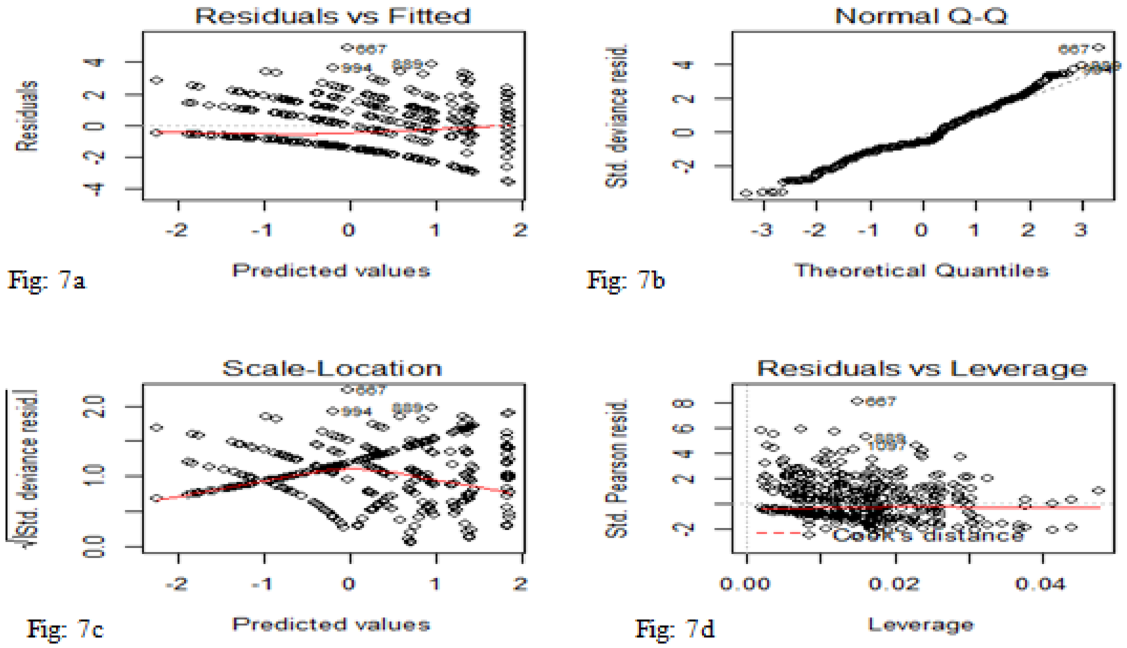

Figure 7: Residuals versus Fitted for Poisson Regression

The R function was used to calculate the solid line, and the two dashed lines correspond to the 0.005 and 0.995 quantities of the standard normal distribution, i.e. if the SDRs are approximately N(0,1) about 99% of these residuals should be between the dashed lines. The seven SDRs outside the 99% limits are identified with their observation numbers (also found at Appendix B, Figure 9). For model checking purposes, a normal Q-Q plot is used to identify extreme values which would appear in the upper right and/or lower left portion of the plot (Figure 6b). The solid line in Figure 7b corresponds to the standard normal distribution. The SDR-Leverage plot in Figure 7c identifies four points (especially observation number 667, 994, 889 and 1097.) with both $h_i > 2$ (i.e. to the right of the solid vertical line. A plot of the absolute value of the SDRs against the fitted values (see Figure 7d) gives an informal check of on the adequacy of the assumed variance function. The null pattern will not show a trend , and smoothing (shown by the solid line) is used to identify a possible pattern , in this case a positive trend. Though 3 outlier cases need more investigation, the results from diagnostics of the Poisson model in Figure 1 indicate that Poisson did not fit well. Conclusively speaking, the apparent trend in Figure 7d indicate over -dispersion, and the other three diagnostic plots do not indicate that "outliers" are a problem.

The parameter estimates of the final model before and after excluding the outlying observations were close to each other. Thus in this study the zero-inflated Poisson (ZIP) regression model was robust to the outlying observation.

The first graph of Figure 8 shows the deviance residuals plotted against fitted values. It is observed that the plot of deviance residuals against fitted values appears to show some trend of falling variation with increase in estimated value. In the second graph of Figure 7, the normal density plot and in the third a normal quantile plot of standardized deviance residuals is shown. The quantile plot appears to follow a reference line except in the upper right portion. This verifies the assumptions of normality of the residuals for most of the range of values. Some deviations are observed especially at the high end which suggests the data distribution has a long tail at that end.

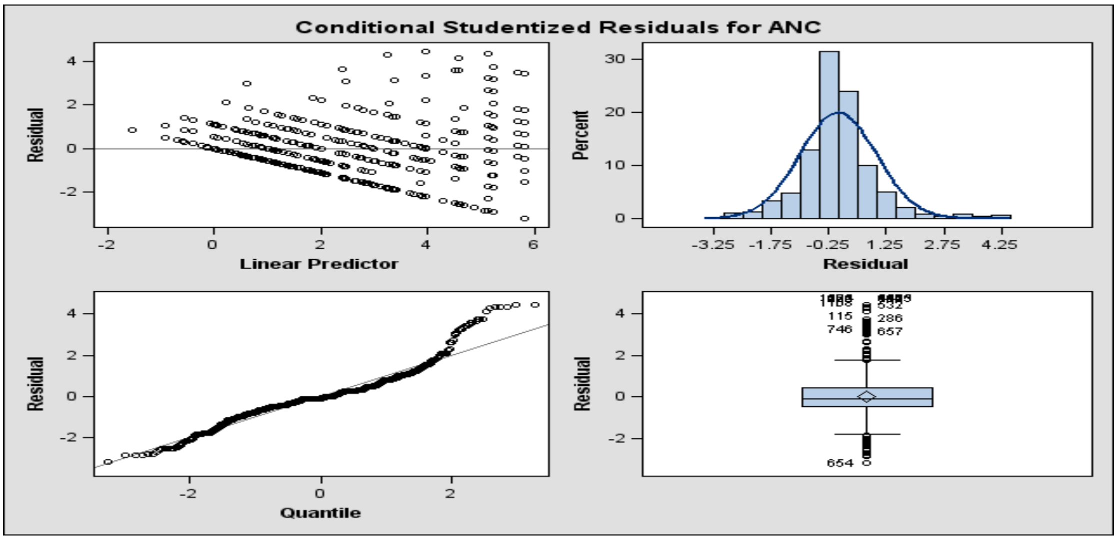

Figure 8: Conditional Studentized Residuals for ANC Visits

3.5. Discussions:

In health service studies, Antenatal care service visits could be a relevant metric to quantify efficiency of maternal care utilization. This thesis, which focused on an efficient statistical modeling for number of Antenatal care service visits, propose a GLM, zero-inflated and hurdle modeling approach to estimate parameters of demographic, socio-economic, health and environmental related factors. World Health Organization recommends a minimum of four ANC visits initiated during their pregnancy periods. In this study the ANC service utilization rate in rural Ethiopia was found to be 48.5%. Although this shows a low level of ANC service utilization, educated pregnant mothers, and mothers from better progressed regions, mothers who have no severe workloads, mothers that could get a nearby health post and have awareness about ANC use attends more than 3 times. Moreover, a significant proportion (77.5%) of the attendees had less than four visits which is less than the recommended.

The finding of this study significantly differs with that of EDHS 2005 which showed 21.6% attendance of ANC in the rural areas of Ethiopia. This could be attributed to the fact that DHS covered more remote areas where distance from health institution could be a major predictor of ANC utilization. It is also important to note the time gap between the EDHS and the current study. A study conducted in Northern Ethiopia (2004) showed that the magnitude of ANC attendance was 45%. With regard to the determinants of ANC service utilization; this study revealed that ANC service utilization is significantly influenced by mother education, region, workload, economic status, access to health post, awareness about pregnancy complications, and manifestation of pregnancy complications. Non educated mothers were less likely to utilize ANC service than educated women. This finding is consistent with the findings of previous studies conducted in Addis Ababa [52]. Moreover, in this study the use of antenatal care was found to be related to economic status; Mothers with middle and rich economic level were more likely to attend ANC than poor women. This is also in line with other studies conducted in Southern Ethiopia [59].

This finding differs with the study conducted in Metekel zone which confirms that being awared about ANC utilization were more than two times (OR=2.32) more likely to use ANC visits [18]. In our case, awared women were one times (OR=1.47) more likely to utilize ANC visit than non awared ones. This study again found that education status of secondary school and above had three times (OR=3.68) more ANC visits than non educated ones, but here 1.44 times more likely to be visited. Gurmessa's report determined that having monthly family income of 500 Ethiopian Birr and above (OR=1.53) were positively associated with antenatal care service utilization [18]. This agrees with our finding that rich and middle incomer women were more likely to attend ANC visits; the odds ratio in our case is 1.15 (slightly different with this report).

A report from Samre Saharti district in Tigray region of Ethiopia found differently from our finding that being resing with her husband or partner have significant association with ANC utilization [58]. According to this finding, this is not the factor. The reason for the difference could be that our study was conducted in a rural area, while the regional report included urban areas. Unwanted pregnancy were not the determinant to receive ANC visits in this study. This was dissimilar to other studies conducted in the Saharti Samre district, Tigray, Ethiopia [58]. This could be due to fear of stigma because a pregnancy without marriage is not accepted by the community in the study area. Therefore it appears rational to see that most of single and widowed mothers might be faced unwanted pregnancies. In addition mothers from low socio-economic status (poor mothers) are unlikely to afford the cost of transport and could have limited access to ANC utilization, and low health seeking behavior [58]. Other studies have shown comparable results with this [2, 18, 23, 58, 59]. As part of enabling factors, distance from health post were found to be predictor of antenatal care service utilization where women who live within nearby distance from the health facility were about 0.48 times more likely to visit ANC than women who live at distance far from health post(OR=0.48). this was line with the study Yem special woreda, southwestern Ethiopia [3].

Hurdle Poisson model assumes that all zero data are from one “structural” source. The positive (i.e., non-zero) data have “sampling” origin, following either truncated Poisson or truncated negative binomial distribution. For example, consider a study of ANC visit users in which a secondary outcome is a number of ANC visits during last nine months. In this case, it is safe to assume that only non ANC users will visit zero ANC visits during the last nine months and ANC users will score some positive (non-zero) number of ANC service visits during last nine months. Hence the zero observations can come from only one “structural” source, the non ANC users. If a pregnant mother is considered as ANC user, they do not have the ‘ability’ to score zero ANC visits during the last nine months and will always score a positive number of ANC visits in a hurdle model with either truncated Poisson or truncated negative binomial distributions.

4. CONCLUSIONS

In conclusion, the ANC service utilization rate in rural Ethiopia is lower than the national figures available to date. In addition, it is worth nothing that majority of the mothers who attend ANC did not receive adequate number of visits recommended by the World Health Organization. Furthermore, maternal education, workload inside and/or outside home, availability and accessibility of health post, regions, and awareness about pregnancy complications were major predictors of ANC service utilization. Therefore, efforts to bring about changes in these major predictors at individual and community level through behavioral change communication are recommended.

In this study, it was found that ZIP and hurdle Poisson regression models were better fitted the data than NB, ZINB, HNB and Poisson. This may be due to the high variability of the number of ANC visits. Hurdle Poisson regression model was better fitted the data which is characterized by excess zeros and high variability in the non-zero outcome than any other models and therefore it was selected as the best parsimonious model.

ACKNOWLEDGMENT

Primarily and foremost I would like to express my appreciation and gratitude to my Lord and Savior, Jesus Christ, for giving me life breath and the strength to complete this thesis. I owe you my life Lord! Next, I want to thank Dr Wondwosen Kassahun, who provided valuable guidance and inspiration, encouragement, sound advice, and accessible help in writing up of this paper. Finally my sincere thanks are extended to all my beloved and cooperative families and friends for their continuous encouragement. Particularly, I must make mention of my wife Obse Tesfaye and my son, Siyosan Dawit. I thank my wife for her constant support, patience and encouragement to me as my life partner. I also express my appreciation to her for all the time she availed to me to complete the paper, by taking care of our son. Thank you Siyosan! Your loving care and splendid sense of funniness is of great importance to me.

This paper is dedicated anybody doing anything to reduce maternal mortality in Ethiopia and/or in Africa and Wollega University, the basement of mine to be here.

Dawit Sekata

REFERENCES

[1] Agresti, A., & Finlay, B. (1997). Statistical methods for the social sciences. (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

[2] Alemayehu Tariku. (June, 2008). Why Pregnant Women Delay to Attend Prenatal Care? Cross Sectional Study on Timing of First Antenatal Care Booking at Public Health Institutions in Addis Ababa, Ethiopia; Journal

[3] Bahilu Tewodros, et al. (March 2009). Factors Affecting Antenatal Care Utilization in Yem Special Woreda, Southwestern Ethiopia; Journal

[4] Cameron, A. C. and P. K. Trivedi (1998). Regression Analysis of Count Data, Cambridge University Press, Cambridge.

[5] Cameron, A. C., & Trivedi, P. K. (2010). Micro econometrics using stata (revised Ed.). Stata Press Books. Consul, P. C., & F

[6] Carla A, Tessa W, Blanc A, Van P, et al. ANC in developing countries, promises, achievements and missed opportunities; an analysis of trends, levels and differentials, 1990-2001. WHO Geneva, 2003.

[7] Carson, J., and Mannering, F. (2001). "The effect of ice warning signs on ice-accident frequencies and severities." Accid. Anal. Prev., 33(1), 99-109.

[8] Casella G. and R.L. Berger (2002). Statistical Inference, (2^{nd} ed). Thomson Learning: Pacific Grove, CA.

[9] Central Statistical Authority (CSA) and ORC Macro. Ethiopia Demographic and Health Survey 2005, Addis Ababa, Ethiopia, and Calverton, Maryland, USA: CSA and ORC Macro, 2006.

[10] Chatterjee S and Hadi AS, (2006). Regression analysis by example, Fourth edition, Wiley Interscience.

[11] Cohen, A. (1963). "Estimation in mixtures of discrete distributions." Proc., Int. Symp. On Classical and Contagious Discrete Distributions, Pergamon Press, New York,351-372.

[12] CSA, ORC Macro, author. Ethiopia Demographic and Health Survey, Addis Ababa, Ethiopia and Calverton, Maryland, USA. Sep, 2006.

[13] Dietz, E. and Bohning, D. (2000). on estimation of the Poisson parameter in zero modified Poisson models. Computational Statistics and Data Analysis. 34(4): 441-459.

[14] Edward N., Bernardin S. and Eric A. (2012). Determinants of utilization of antenatal care services in developing countries Recent evidence from Ghana. Department of Economics, University of Ghana, Accra, Ghana.

[15] Eyerusalem Dagne. (2010). Role of socio-demographic factors on utilization of maternal health care services in Ethiopia. UMEA Universitet.

[16] Germu, S. and Trivedi, P. K. (1996). Excess zeros in count models for Recreational Trips. Journal of Business and Economic Statistics. 14: 469-477.

[17] Greene, William H. (1994). "Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models." New York University Department of Economics Working Paper EC-94-10.

[18] Gurmesa Tura. (July 2009). Antenatal Care Service Utilization and Associated Factors in Metekel Zone, Northwest EthiopiaJournal.

[19] Gurmu, S. and P.K. Trivedi (1996). “Excess Zeros in Count Models for Recreational Trips”, Journal of Business and Economic Statistics, 14, 469-477.

[20] Hardin, J & Hilbe, J(2001).Generalized linear models & extensions. Stata Press, Texas, USA.

[21] Hardin, J and Hilbe, J (2003). Generalized estimation equations. Chapman and Hall/CRC

[22] Hayelom Kassyou. (2008). Factors Affecting Antenatal Care Attendance in Maichew Town, Southern Tigray. Addis Ababa University, school of graduate studies.

[23] Heilbron, D. (1994). Zero-altered and other regression models for count data with added zeros. Biometrical Journal, 36, 531–547.

[24] Hinde, J. P. and Demetrio, C. G. B. (1998). Overdispersion: models and estimation. Computational Statistics and Data Analysis. 27: 151-170.

[25] Ismail, N. and Jemain, A. A. (2007). Handling Overdispersion with Negative Binomial and Generalized Poisson Regression Models. Casualty Actuarial Society Forum,103-158.

[26] Jansakul N and Hinde JP (2002). Score tests for zero-inflated Poisson models. Computational Statistics and Data Analysis40, 75–96.

[27] Kumara, S., and Chin, H. (2003). "Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros." Traffic Inj. Prev., 4(1), 53-57.

[28] Kutner, M and Neter J (2004). Applied linear regression models. McGraw-Hill Irwin.

[29] Lambert, D. (1992), Zero-inflated Poisson regression, with an application to defects in manufacturing, Technometrics, 34, 1-14.

[30] Lee, J., and Mannering, F. (2002). "Impact of roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis." Accid. Anal. Prev., 34(2), 149-161.

[31] Long, S. (1997). Regression models for categorical and limited dependent variables. Newbury Park, CA: SAGE.

[32] Lord, D., (2006). Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis & Prevention 38 (4), 751–766.

[33] Lord, D., Bonneson, J.A., (2005). Calibration of predictive models for estimating the safety of ramp design configurations. Transportation Research Record 1908, 88–95.

[34] Lord, D., Bonneson, J.A., (2007). Development of accident modification factors for rural frontage road segments in Texas. Transportation Research Record 2023, 20-27.

[35] Lord, D., and Mahlawat, M. (2009). "Examining Application of Aggregated & Disaggregated Poisson-Gamma Models Subjected to Low Sample Mean Bias." Transp. Res. Rec., 2136, 1-10

[36] Lord, D., Washington, S. P., and Ivan, J. N. (2005). "Poisson, Poisson-gamma and zero inflated regression models of motor vehicle crashes: balancing statistical fit and theory." Accid. Anal. Prev., 37(1), 35-46.

[37] McConway, KJ, Jones, MC and Taylor PC (1999). Statistical modeling using GenStat. Arnold, United Kingdom.

[38] McCullagh, P., & Nelder, J. A. (1989).Generalized linear models. London: Chapman & Hall.

[39] Md Abdullah al Mamun. (May 2014). Zero-inflated regression models for count data: an application to under-5 deaths. Ball State University, Muncie, Indiana.

[40] Ministry of Health (2006).AIDS in Ethiopia, Sixth Report, Addis- Ababa, Ethiopia

[41] Mullahy, J. (1986). Specification and testing of some modified count data models. Journal of Econometrics, 33, 341–365.

[42] Nishat Fatema.(2010). Importance of Antenatal care. MO, CMH, Chittagong, Bangladish.

[43] Ornella L. et al. (2011). Antenatal Care. Opportunities for Africa's Newborns; Journal.

[44] Paternoster R, Brame R, Bachman R, Sherman L (1997). Do fair procedures matter? The effect of procedural justice on spouse assault. Law Soc Rev 31:163–204

[45] Regassa N. (2011). Antenatal and postnatal care service utilization in southern Ethiopia: a population-based study. African Health Sciences. 11(3): 390 – 397

[46] Rider, P. R. (1961). "Estimating the parameters of mixed Poisson, binomial and Weibull distributions by the method of moments." Bulletin De l'Institut International De Statistiques.

[47] Ridout, M.S., Demétrio, C.G.B. and Hinde, J.P. (1998). Models for counts data with many zeroes. Proceedings of the XIXth International Biometric Conference, Cape Town, Invited Papers, pp. 179-192. Paper retrieved March 13, 2006 from http://www.kent.ac.uk/ims/personal/msr/zip1.html

[48] Shankar V, Milton J and Mannering F (1997). Modeling accident frequencies as zero-altered probability processes: an empirical inquiry. Accident Analysis and Prevention29, 829–37.

[49] Shankar, V.N., Ulfarsson, G.F., Pendyala, R.M., Nebergal, M.B., (2003). Modeling crashes involving pedestrians and motorized traffic. Safety Science 41(7), 627-640.

[50] STATA Press (1985). Reference manual. Longitudinal/ Panel data, Release 9, USA.

[51] STATA Press (2001). Reference manual. Generalized Linear Models and Extensions. USA.

[52] Trends in Maternal Health in Ethiopia (December 2012), Challenges in achieving the MDG for maternal mortality. Addis Ababa; Journal.

[53] Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica: Journal of the Econometric Society, 57, 307–333.

[54] Washington, S.P., Karlaftis, M., Mannering, F.L., (2003). Statistical and Econometric Methods for Transportation Data Analysis. Chapman and Hall, Boca Raton, FL.

[55] WHO, UNICEF, and UNFPA, author. Maternal Mortality in 2000: Estimates Developed by WHO, UNICEF, and UNFPA. Geneva: 2003.

[56] WHO & UNICEF. Antenatal care in developing countries: promises, achievements and missed opportunities. An analysis of trends, levels, and differentials 1990-2001. Geneva: WHO & UNICEF, 2003.

[57] World Health Organization, (2004). World Report on Antenatal Care Service Visits: Summary. World Health Organization, Geneva.

[58] Yalem Tsegay Assfaw. (2010). Determinants of Antenatal Care, Institutional Delivery and Skilled Birth Attendant Utilization in Samre Saharti District, Tigray, Ethiopia. Umeå University, Sweden

[59] Zeine Abosse, et al. (July 2010). Factors Influencing Antenatal Care Service Utilization in Hadiya Zone, Southern Ethiopia; Journal.

APPENDIX

Table 7: Parameter Estimations and S.E for the Models of PR, NB, ZIP, ZINB, HP & HNB

Parameters	Basic Count Models		Zero Inflated Models		Hurdle Models	
	Poisson	NB	ZIP	ZINB	PH	NBH
	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)	Estimation (St. Error)
Intercept	-0.2638 (0.1388)	-0.2621 (0.1654)	0.5872 (0.1608)***	0.2036 (0.1937)	0.5756 (0.1640)***	0.1781 (0.0687)
MEDUC	0.7542 (0.0867) ***	0.7782 (0.1121) ***	0.4987 (0.0999)***	0.5507 (0.1027)***	0.4937 (0.1009)***	0.5820 (0.1022)***
REGION	-0.4352 (0.0614) ***	-0.4287 (0.0769) ***	-0.2249 (0.0698)***	-0.2482 (0.0744)***	-0.2381 (0.0703)***	-0.1997 (0.0708)**
WLOAD	-0.9568 (0.1212) ***	-0.9600 (0.1391)***	-1.0231 (0.1726)***	-0.7014 (0.1744)***	-1.0142 (0.1669)	-0.8168 (0.1644)***
WEALTH	0.3844 (0.0565) ***	0.4271 (0.0732)**	0.2292 (0.0660)***	0.2612 (0.0679)***	0.2253 (0.0663)***	0.2850 (0.0666)***
HPOST	-0.7527 (0.1061)***	-0.7551 (0.1197)**	-0.7251 (0.1555)***	-0.6621 (0.1529)***	-0.7430 (0.1538)***	-0.6813 (0.1543)***
AWARN	0.5495 (0.0980)***	0.5437 (0.1212)***	0.2976 (0.1083)***	0.4899 (0.1276)***	0.2853 (0.1086)**	0.4866 (0.1034) ***
SIGN	0.6106 (0.1022)***	0.5723 (0.1209)***	0.3432 (0.1204)***	0.5318 (0.1480)***	0.3785 (0.1234)**	0.6775 (0.1055) ***
Dispersion (k^{-1})		0.2862 (0.0471)***	0.5670 (0.1085)	0.0071 (0.0154)	0.8792 (0.0264)***	0.0077 (0.0152)*

Table 8: Model Selection: Voung test, AIC, Log-Likelihood, and Inflation Probabilities

	Poi	NB	ZIP	ZINB	HP	HNB
Poi	AIC= 3359.7					
	LL=-0.2270					
	InfPr= 0.1476					
NB	V=3.3374	AIC= 3261.6				
	P=0.0008	LL=-0.1876				
	Prefers NB	InfPr= 0.1574				
ZIP	V=7.4161	V=8.0639	AIC= 3093.1			
	P=1.206E^{-13}	P=6.661E^{-16}	LL= -0.1284			
	Prefers ZIP	Prefers ZIP	InfPr= 0.5727			
ZINB	V=7.0799	V=7.9618	V=1.9815	AIC= 3107.7		
	P=1.44E^{-12}	P=1.776E^{-15}	P=0.0307	LL= -0.1328		
	Prefers ZINB	Prefers ZINB	Prefers ZIP	InfPr=0.6179		
HP	V=7.4068	V=8.1066	V=1. 9704	V=2.0304	AIC= 3091.9	
	P=1.295E^{-13}	P=4.441E^{-16}	P=0.04781	P=0.0423	LL= -0.1171	
	Prefers HP	Prefers HP	Prefers HP	Prefers HP	InfPr= 0.8757	
HNB	V=7.2661	V=7.9646	V=1.9934	V=-0.0423	V=-2.1681	AIC=3107.9
	P=3.6993E^{-13}	P=1.554E^{-15}	P=0.0464	P=1.0337	P=1.9698	LL= -0.1345
	Prefers HNB	Prefers HNB	Prefers ZIP	Prefers ZINB	Prefers HP	InfPr= 0.8756

Note: V=Vuong Test, P= P-value, LL=Log-Likelihood, InfPr= Estimated Proportion of Zeros

Table 8: Estimates of Hurdle Negative Binomial Models with Logit Link Function

	Poisson Hurdle(PH)		NegBin Hurdle (NBH)	
	Poisson part	Inflated part	Negative Binomial	Inflated part
Variables	Beta(S.E)	Beta(S.E)	Beta(S.E)	Beta(S.E)
Intercept	0.6993(0.3066)*	0.6407(0.1583) ***	-0.6992(0.3066)*	0.2541(0.2636)
MEDUC	-0.7763(0.1713)***	0.3596(0.0559) ***	0.7761(0.1713)***	0.3794 (0.0578)
REGION	0.6352(0.1499) ***	-0.2312(0.0701) ***	-0.6355(0.1499)***	-0.1847 (0.0708)
WLOAD	0.6260(0.2506)*	-1.0168 (0.1668) ***	-0.6260(0.2506)*	-0.7841 (0.1651)
WEALTH	-0.4593(0.1308) ***	0.1407 (0.0403) ***	0.4593(0.1308)***	0.1590 (0.0417)
HPOST	0.6267(0.1978) ***	-0.7579 (0.1535) ***	-0.6264(0.1978)**	-0.6976 (0.1544)
AWARN	-0.9758(0.2662) ***	0.2935 (0.1085)**	0.9759(0.2662)***	0.5355 (0.1038)
SIGN	-0.7375(0.2116) ***	0.3844 (0.1234)**	0.7375(0.2116)***	0.7474 (0.1032)

** refers to p<0.05. ** refers to p<0.01. *** refers to p<0.001.*

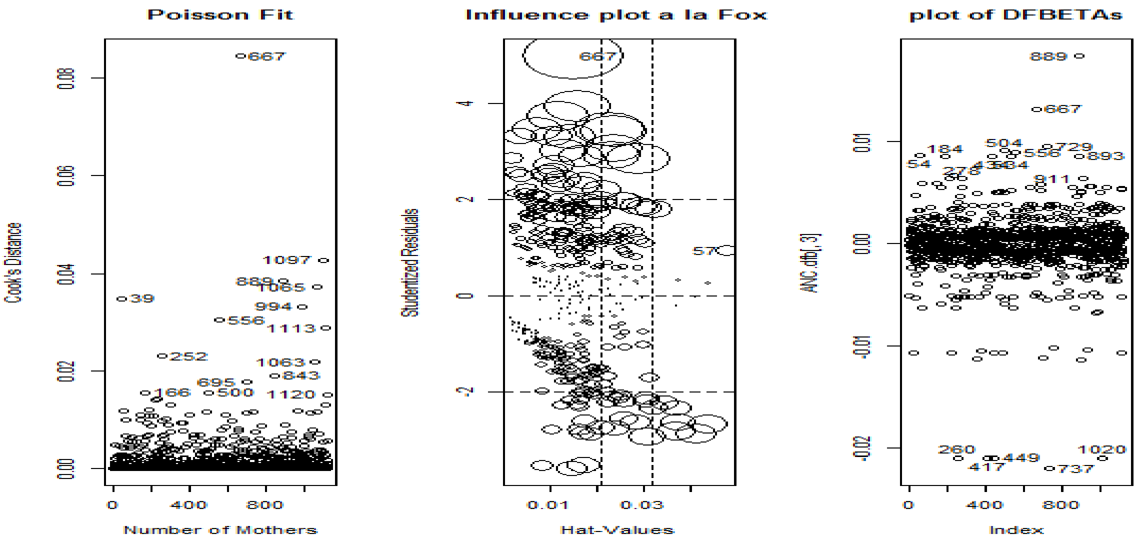

Figure 9: Cook's Distance for Poisson Fit

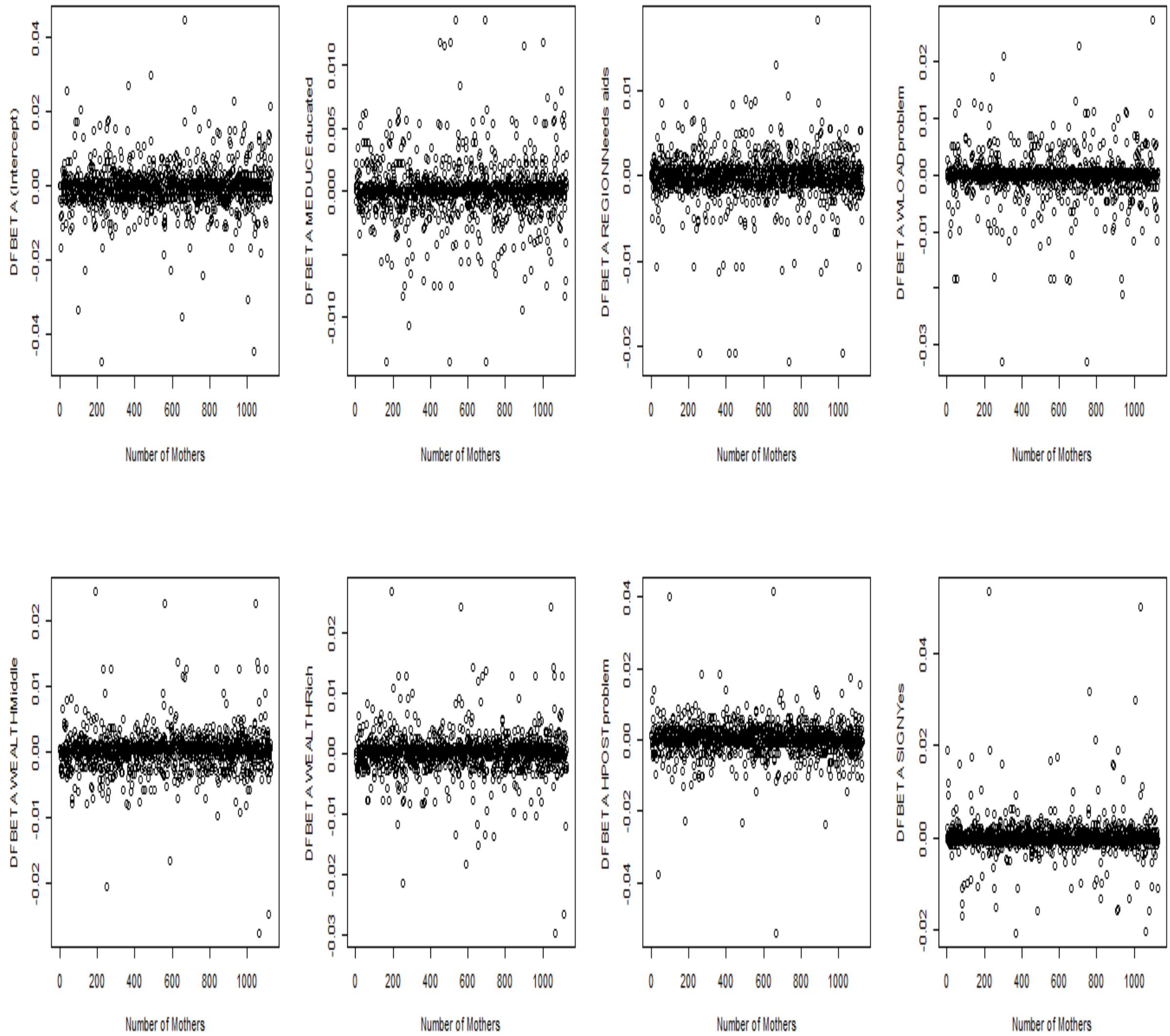

Figure 10: DFBETAs of Different Explanatory Variables