**ADDIS ABABA UNIVERSITY**

**SCHOOL OF GRADUATE STUDIES**

**COLLEGE OF NATURAL SCIENCE**

**DEPARTMENT OF STATISTICS**

# ANALYSIS OF CAUSES OF CATTLE DEATH USING DISCRETE REGRESSION MODELS IN ETHIOPIA

**By**

**Abdulaziz Shiffa**

**A Thesis submitted to the School of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics**

**June, 2015**

**Addis Ababa, Ethiopia**

ADDIS ABABA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

COLLEGE OF NATURAL SCIENCE

DEPARTMENT OF STATISTICS

# ANALYSIS OF CAUSES OF CATTLE DEATH USING DISCRETE REGRESSION MODELS IN ETHIOPIA

**By**

**Abdulaziz Shiffa: ID GSR/0567/06**

**Advisor: Dejen Tesfaw (PhD)**

**A Thesis submitted to the School of Graduate Studies of Addis Ababa University in partial fulfillment of the requirements for the Degree of Masters of Science in Statistics**

**June, 2015**

**Addis Ababa,**

**Ethiopia**

**ADDIS ABABA UNIVERSITY**

**GRADUATE STUDIES PROGRAM**

**DEPARTMENT OF STATISTICS**

This is to certify that the thesis prepared by Abdulaziz Shiffa, entitled: Analysis of Causes of Cattle Death Using Discrete Regression Models in Ethiopia and submitted in partial fulfillment of the requirements for the Degree of Master of Science complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the Examining Committee:

Advisor Dejen Tesfaw (PhD) Signature _________ Date _________

Examiner Prof. Eshetu Wencheko Signature _________Date _________

Examiner Prof. M.K Sharma Signature _________Date _________

Chair of Department or Graduate Program

# DECLARATION

I, the undersigned, declare that the thesis is my original work, has not been presented for degrees in any other University and all sources of materials used for the thesis have been duly acknowledged.

**Name**: Abdulaziz Shiffa Yimam

**Signature**: ___________________

**Place of submission:** Department of Statistics, College of Natural Sciences, Addis Ababa University

**Date of submission**: June, 2015

This thesis has been submitted for examination with my approval as a University advisor.

**Dejen Tesfaw (PhD)** ______________________

Advisor's Name signature

**Abstract**

*Analysis of Causes of Cattle Death Using Discrete Regression Models in Ethiopia*

*Abdulaziz Shiffa*

*Addis Ababa University, 2015*

*Ethiopia is believed to have the largest cattle population in Africa. The cattle sub-sector has been contributing a considerable portion to the economy of the country, food supply and foreign currency income. For the majority of smallholder farmer's cattle are livelihood and for some farmers the sub-sector is the only livelihood. The objective of this study is to identify the most important factors affecting cattle death in Ethiopia. The dataset in this study is based on the 2014 agricultural survey obtained from Central Statistical Agency (CSA), Ethiopia. Four count models: Poisson, NB, ZIP and ZINB model were applied in order to identify the best fit model for cattle death in Ethiopia. The preliminary predicted values and model selection criteria such as AIC, LRT, deviance and Vuong statistic showed that ZINB regression model is the most appropriate model to fit cattle death. The results of ZINB analysis revealed that age, household size, educational background, farming type, agro-ecology zone and land holding size were found to have statistically significant effect on cattle death. Based on the findings, both governments and stakeholders should work hard to reduce cattle death by promoting mixed farming system and exercising continuous trainings on handling cattle for small holding farmers.*

## AKNOWLEDGMENTS

First, and foremost, I thank Allah for giving me the opportunity to pursue my graduate study at Department of Statistics, Addis Ababa University.

My appreciation and full respect goes to my advisor and instructor Dejen Tesfaw (PhD), for his precious comments, suggestions and patience during the entire time of the research work.

I would like to gratefully and sincerely thank the Central Statistical Agency of Ethiopia and Addis Ababa University, department of Statistics for having given me a study opportunity in statistics.

I would like to thank beloved instructors all individuals in the Department of Statistics for their moral support and positive attitude when I needed them during the course time, and research period.

My special thanks goes to, Abiy Wogderes, his support in material and data linking Agro-ecology data, Ermyas Arega for data processing, Habtamu Eshetu, Tesema Urgessa and Meron Legesse for their support in data cleaning.

My heart-felt thanks go to my mother, Amsale G/Mariam who is supportive in my life and professional carrier. I also thank my friends for their believing in team work during the study time and research work. Lastly but not the least Professor Joseph Hilbe (President of American Astro-stat Association) for his support on stata coding and idea on count models.

**Acronyms**

| | |
|---|---|
| CSA | Central Statistical Agency |
| FAO | Food and Agriculture Organization |
| HSI | Heat Strain Index/ Heat Strain Intensity |
| IRR | Incident Rate Ratio |
| LRT | Likelihood Ratio Test |
| MoFED | Ministry of Finance and Economic Development |
| MoA | Ministry of Agriculture |
| MoE | Ministry of Energy |
| MoARD | Ministry of Agriculture and Rural Development |
| NSDS | National Strategy for the Development of Statistics |
| NBD | Negative Binomial Distribution |
| SNNP | Southern Nations, Nationalities and Peoples Region |
| ZINB | Zero-Inflated Negative Binomial |
| ZIP | Zero-Inflated Poisson |

# CONTENTS

## List of Tables

## List of Figures

# CHAPTER ONE

# INTRODUCTION

## 1.1 Background of the study

In developing countries like Ethiopia, livestock is an important and, in some situations the only, cash-income source. If a household loses cattle, the result may be loss asset for a considerable period of time. This holds true for many households in rural parts of Ethiopia. Loss of cattle can have serious consequences for a rural household's livelihood, given the pivotal role played by livestock in the farming systems of Ethiopia. Livestock death is considered to be one of the main factors contributing to poverty (World Bank, 2013).

Ethiopia is believed to have the largest cattle population in Africa (CSA, 2014). Cattles are integral components of the Ethiopian farming systems, and perform multiple functions at different levels of aggregation. At individual smallholders' level, cattle are important source of food (meat and milk), cash income, services (transport and traction) and manure (for soil fertility management and fuel). Cattle also have social and cultural values among producers, particularly among pastoralists

As it is a livelihood asset for most households in Ethiopia, this sub-sector is a source of producing income. Livestock is widely used in emergency income needs. Apart from these livestock and livestock products can be valued for supply protein which are very important in tackling and preventing malnutrition and alleviating normal growth.

The sector of agriculture is the backbone of the country, Ethiopia. It contributes about 85% of employments and 44% of the share of GDP (CSA, 2013). As it is a vast sector, agriculture is being practiced in a traditional way and is still backward compared to other sectors like service, manufacturing and industry. Almost 95% of the agricultural sector is run by the smallholder farmers. Government official statistics shows the foreign currency earned by exporting agricultural raw materials, crop products and livestock products (MoFED, 2012). Agricultural activities are studied by dividing crop items, livestock and mixed farming. In the last five years, increasing the productivity of livestock sector has not been given proper attention compared to the crop items in terms of quality and minimizing death. This shows that a lot activity should be done regarding to this area (CSA, 2013).

This cattle sub-sector has been contributing a considerable portion to the economy of the country. Cattle are the livelihood of many smallholder farmers in Ethiopia. It is evident that cattle products in the form of meat, milk supply and milk products. These provide the needed animal protein that contributes to the improvement of the nutritional status of the people. Cattle also play an important role in providing export commodities such as live animals, hides and skins (CSA, 2008). Furthermore, cattle provide farmyard manure that is commonly applied to improve soil fertility and also used as a source of energy (MoE, 2010).

The livestock sub-sector also provides wide and year-round employment opportunities for surplus family labor in rural Ethiopia (MEDaC, 2007). Cash income from cattle production is especially important for the poor and landless Ethiopian households, particularly women, as is also true in many other developing countries (Delgado et al. 1999). Income from livestock production is also used for income diversification investment activities Little et al. (2001). For the average rural farm household with limited investment alternatives, livestock are used as a store of wealth and hedge against inflation.

As national level, cattle are addressing critical bottlenecks and lay the foundation for the accelerated development of the cattle export earnings part of Ethiopia's livestock sector. Livelihoods of pastoralists rely on a path of life and parallel to the major stockholders in earning of hard currency in the economy as a whole benefits from this sector. This can increase the incomes of smallholders and pastoralists incomes, as well as in the food security sectors specially supplying live cattle and their products to the rest (CSA, 2014).

The pastoralist areas are well known in the production of cattle, and a significant number of livestock are available in the highland areas (MOA, CSA and FAO, 2008).The challenge of production in lowland area/pastoralist area is most of the time prone to drought and frequent failure of rain leads to death of cattle. Furthermore, the coping mechanism of this area to resist the drought is small. When cattle marketing is affected due to unpredictable weather conditions, disease outbreaks, or conflict, the livelihoods of many people and families connected to the different parts of the livestock chain will be affected, naturally many of the people in the demand chain earn money only when there are transactions. Absence of financial transactions means loss of livelihood for many of them. Therefore, a challenge in the cattle directly or indirectly affects the livelihood of the human that rely on it (Alemu et al., 2009).

National agricultural surveys on cattle death analysis in Ethiopia were mainly descriptive in nature and limited to the study of association between cattle death with certain cattle death related variables. Few studies have been done about risk factors of cattle death using statistical methods such as diagnostic regression and most studies are based on small-scale survey data concentrated in certain regions. This study is based on the Ethiopian 2014 agricultural sample survey dataset with reference to agricultural holder, in order to identify cattle death factors using count regressions.

## 1.2 Statement of the problem

In Ethiopia livestock represents a major national resource and forms an integral part of the agricultural production system. Ethiopia has the largest livestock population of Africa with 55.03 million heads of cattle (CSA, 2014).In spite of this, productive and reproductive performance is accompanied with very low, poor health care, high diseases incidence, poor management condition, unpredictable climatic condition causing a significant cattle death.

In Ethiopia, the aggregate annual economic losses from animal disease through direct mortality and reduced productive and productive performance were estimated over US $150 million. A number of livestock diseases in Ethiopia are still causing a negative influence on national economic activities (MoFED, 2013).

Identification of factors that are responsible for the death of cattle is an important prerequisite for avoiding excessive mortality. An epidemiological study usually involves collecting and evaluating pre-existing data to find associations with the problem. But with the existence of cattle death dataset through count regression model can also be another approach to understand factors for cattle death. Therefore, in this study the researcher tries to fill the information gap by fitting count regression models.

## 1.3 Objective of the study

**Main objective:** To identify the most important factors variable that affect cattle death based on agricultural survey 2014 dataset in Ethiopia using count regression models.

**Specific objectives**

1. To apply count regression models for analyzing determinants of cattle death in Ethiopia.
2. To identify an appropriate count regression model for analysis cattle death dataset.

## 1.4 Significance of the study

As the livelihood of rural population of Ethiopia depend on agriculture, improving the productivity and identifying problems on this sector is the crucial part of rising and scaling up the livelihood of peoples living in rural parts of the country. This study is an attempt to determine factors that have strong association with cattle death so that policy directives act on accordingly.

The end user governmental and non-governmental organizations could take intervention measures and set appropriate plans to tackle and improve the existing problems by identifying and giving priority for the very poor and vulnerable groups.

The study is expected to provide information for future research in livestock health issues. This study is likely to contribute its part by filling the information gap. Finally, the study could be used as a stepping-stone for further studies.

## 1.6 Definition of terms

Agriculture items have to be distinctly defined and identified, so that the information about the data items becomes useful. The correct way of stating related terms is a prerequisite to make standards and definitions for the collection and compilation of agricultural survey data. The purpose of using standard concepts and definitions is not only to provide quality data but also to ensure that the right items are enumerated and measured accurately to reflect the situation reality.

**Agricultural Household** is a household is considered as agricultural household when at least one member of the household is engaged in growing crops and/or raising livestock in private or in combination with others.

**Holding** is all the land and/or livestock kept, which is used wholly or partly for agricultural production and is operated as one legal entity by one person alone, or with others without regard to management, organization, size or location.

**Holder** is a person who exercises management control over the operation of the agricultural holding and makes the major decision regarding the utilization of the available resources. S/he has primary technical and economic responsibility for the holding. S/he may operate the holding directly as an owner or a manager. Under conditions of traditional agricultural holding the holder may be regarded as the person, who with or without the help of others, operates land and/or raises livestock in his/her own right, i.e. the person who decides on which, where, when, and how to grow crops or raise livestock or both and has the right to determine the utilization of the products.

**Enumeration Area (EA)** in the rural parts of the country is a locality that is, in most of the cases smaller than, and only in some cases equal to a farmers' association in geographical area and usually consists of 150-200 households.

**Death of Cattle** covers cattle that die on farm, either an unassisted death or by euthanasia (killing animal for some special reason), but not cattle that leave the farm to be slaughtered.

**Livestock Extension Package** is an integrated program supported by Ministry of Agriculture in order to increase the productivity of livestock, by promoting basic skill on livestock health, introducing new breeds to the private farmers; it could be either on dairy, fattening or improved food supply for the household.

**Agro-ecological zone (AEZ)** can be defined as a set of core applications, leading to an assessment of land suitability and potential productivity, and a further set of advanced or peripheral applications, which can be built on inventories. It can be used as core applications include maps showing, agro-ecological zones and land suitability, and quantitative estimates on potential crop areas, yields and production. Such information provides the basis for advanced applications such as land degradation assessment, livestock productivity, population support capacity assessment and land-use optimization.

**Table 1. 1:** Major agro-ecological zones

| No. | Agro-Ecological Zone | Definitions |
|---|---|---|
| 1 | Dega | Highlands (2,300 - 3,200 meters, identified by growing barley, wheat, highland oilseeds, highland pulses) |
| 2 | Lower Kola | Hot lowlands (<500 meters, in the arid east, crop production is very limited, in the humid West root crops and maize are largely grown) |
| 3 | Unsuite Wurch | Kur (highland >3, 700 meters, primarily for grazing). |
| 4 | Upper Kola | Lowlands (500-1,500 meters, identified by sorghum, fingermillet, sesame, cowpeas, groundnuts) |
| 5 | Woina Dega | Woina Dega (midlands, 1,500 - 2,300 meters, identified by wheat, teff, barley, maize, sorghum, chickpeas , haricot beans are major crops) |
| 6 | Wurch | Wurch (highlands, 3,200 - 3,700 meters, identified by barley growing is common) |

**Source:** CSA and Ministry of Agriculture, 2006

# CHAPTER TWO

# LITERATURE REVIEW

Livestock systems in developing countries are characterized by rapid change, driven by factors such as population growth, increases in the demand for livestock products as incomes rise and urbanization. Livestock currently contribute about 30 percent agricultural gross domestic product in developing countries, with a projected increase to about 40 percent by 2030 (FAO, 2010) is becoming the fastest growing sub-sector of agriculture in Ethiopia (Kefyalew and Tegegne, 2012).

Smallholder farmers predominate in developing countries and they are entirely dependent on agriculture for their livelihoods (Dixon et al., 2001). About 76 percent of the poor in developing countries live in rural areas and two-thirds of the rural people in these countries keep livestock (Owen et al., 2005). Africa hosts 305 million cattle, representing 17.9 percent of the world total cattle population (FAOSTAT, 2014). The population of small ruminants in sub-Saharan Africa is estimated to be 274 million (Samson and Frehiwot, 2010).

In Ethiopia, more than 85 percent of the human population depends on agriculture for their livelihoods (Azage et al., 2013) and usually keep livestock as pastoralists or in mixed crop livestock systems. The livestock population of Ethiopia is currently estimated at 55.03 million cattle, 27.35 million sheep, 28.16 million goats, 1.96 million horses, 0.36 million mules and 6.95 million donkeys excluding pastoralist areas (CSA, 2014) and is diverse genetically. Despite the large livestock population of Ethiopia, the economic benefits remain marginal due to prevailing diseases, poor nutrition, poor animal production systems, reproductive inefficiency, management constraints and general lack of veterinary care (Sisay, 2007).

Chaudhary et al. (2013) used logistic regression model to show that sex, poor management, and age had significant effect on cattle death. Hossain et al. (2014) employed the Poisson regression model to analyze death of dairy cattle. According to their study, age of the animal was found to be the most important risk factors of dairy cattle death followed by breed and season. The study also suggested that greater attention should be paid to the time of colostrum's feeding, proper timing and management of calves.

Shimelis et al. (2011) who conducted research from November, 2008 to December, 2008 in four selected districts of North Gondar zone, northwest Ethiopia comprises three agro-ecology zones (Dabat from highland, Gondar Zuria and Dembia from midland and Chillga from lowland), by using multiple logistic regressions. The results revealed that small ruminant helminthiasis is found to be an important problem in the study areas. They concluded that species, age and agro-ecology should be considered as potential risk factors for the occurrence of helminthiasis disease.

Bernard (2008) employed a Cox regression model with a gamma-frailty to explore death on smallholder dairy cattle farms in Tanga and Iringa regions of Tanzania. The result showed that zero-grazing was highly protective of TBDs (tick-borne disease) while agro-ecological zone (AEZ), age and district were risk factors of cattle death.

George et al. (2011) conducted a research to identify factors of livestock death during drought time in pastoral area of Kenya using binary probit model for explaining the probability of livestock death. They found that decision on migration, herd size, gender and age of household were statistically significant.

John (2009) used a multiple logistic regression model to see if mixed farming was determinant of cattle death in Kenya. The study findings identified that traditional mixed crop-livestock farming system was an important practice in the prevention of cattle death.

Godfrey (2010) conducted a study to assess the effect of water supply infrastructures on cattle death by collecting cross sectional data and using Tobit model. Grazing land area, green fodder, water and climate condition were identified factors that have effect on cattle death.

Morignat et al. (2014) used classical Poisson regression model to model the national baseline for cattle mortality between 2004 and 2005 and to predict the weekly number of expected deaths in 2003 and 2006 for the whole cattle population and by subpopulation based on age and type of production in France. They found excessive cattle mortality of 24% for the two-week heat wave of 2003, and 12% for the three-week heat wave of 2006 .

Michael (2013) conducted a research to identify factors of cattle death by using panel survey data and fitting with linear regression model. It was found that the agro-ecology zone was a determinant factor of cattle death. Using the same model and survey, Hossian et al. (2014) found that educational status of the small holder farmers was a factor associated with cattle death.

Karin (2014) conducted a study to identify factors associated with cattle death in dairy farming in Sweden using negative binomial regression model. The study showed that herd level, cattle size, calving interval, season, educational level and milk yield were associated with the death of dairy cattle.

Hossain et al. (2014) used linear regression model using sixteen years of data. The results showed that age of cattle was the most important factor and significantly associated with cattle mortality.

Cheryl et al. (2009) considered culling and mortality in beef cows from western Canada to estimate the proportion of total variation explained by herd effects for total losses from inventory, culling, and mortality. The study used a Poisson model and link function. From this study factors relating to cow nutrition accounted for 25% of the deaths, emphasizing the importance of feeding management as a determinant of cow health in western Canada.

Thumbi et al. (2013) conducted a longitudinal cohort study to assess mortality in East African shorthorn Zebu cattle under one year: 548 zebu cattle were under the study between 2007 and 2010. Each calf was followed during its year of life or until lost from the study. Using Cox-proportional hazard model, the study revealed that farmer's age, gender, education level, herd size, land size were found to be statistically significant for surviving of the calf under one year.

# CHAPTER THREE

# STATISTICAL DATA AND METHODOLOGY

## 3.1 Source of data

The dataset in this study is based on 2014 Agricultural sample survey obtained from the Central Statistical Agency (CSA), Ethiopia. This survey has the national mandate to undertake agricultural survey every year to produce official statistics on number of livestock, production, birth and death for the following domains: agricultural sample survey for Ethiopia as a whole; rural areas of Ethiopia (each as a separate domain); and 10 geographic areas namely: Tigray; Afar; Amhara; Oromyia; Somali; Benshangul-Gumuz; Southern Nations, Nationalities and Peoples (SNNP); Gambella; Harari, and Dire Dawa. The agricultural sample survey used a stratified two-stages sampling. In the 2014 Agricultural Survey a representative sample of approximately 27,424 were selected from the list of enumeration areas (EA) that is 2227 from the 2007 Population and Housing Census sampling frame.

The survey was conducted from November 12, 2013 to November 11, 2014.

## 3.2 Variables included in the study

As demonstrated in the literature review and from the agricultural survey. the following variables are considered in this study.

### 3.2.1 The Response variable

The response variable of the study is “number of dead cattle” which is a count variable and it assumes non-negative values.

### 3.2.2 Explanatory variables

The factors variable to be analyzed as determinants of cattle death are:

- Age of the agricultural holder,
- Gender of the agricultural holder,

- Educational status,
- Household size,
- Participations in extension package program,
- Land size of the household,
- Agro-ecological zones,
- Farming type.

### 3.2.3 Description and coding of the study variables

Detailed description of explanatory variables related to cattle mortality is presented as follows.

**Table 3. 1:** Description of variables

| No | Variables | Category |
|---|---|---|
| 1 | Age of the agricultural holder | Numeric (continuous) |
| 2 | Gender of the agricultural holder | 0. Female<br>1. Male |
| 3 | Educational status | 0. No education<br>1. Primary education<br>2. Secondary and above |
| 4 | Household size | Numeric(count) |
| 5 | Participations in extension package program | 0. No<br>1. Yes |
| 6 | Land size of the household | 0. Under 0.10 hectare<br>1. Hectarage 0.10-0.50<br>2. Hectarage 0.51-1.00<br>3. Hectarage 1.01-2.00<br>4. Hectarage 2.01-5.00<br>5. Hectarage 5.01-10.00<br>6. Over 10 hectares |
| 7 | Agro-ecological zone | 0. Dega<br>1. Lower Kola<br>2. Unsuit Wurch<br>3. Upper Kola<br>4. Woina Dega<br>5. Wurch |
| 8 | Farming type | 0. Crop<br>1. Livestock<br>2. Crop & livestock (mixed) |

## 3.3 Methodology

Linear models, such as linear regression and analysis of variance are used to represent the relationship between the response variable and one or more explanatory variables which can be measured on a continuous scale. One of their main assumptions is that the residual error follows a normal distribution. To meet this assumption, when a continuous response variable is skewed, a transformation of the outcome variable can produce errors that are approximately normal. There are, however, many experimental situations where the outcome variable is not continuous but categorical or discrete. In this case, a simple transformation cannot produce normally distributed errors. A common example is when the outcome variable is a count number of occurrences of an event that occur randomly over time or space. The distribution of counts is discrete and limited to non-negative integer values.

Based on the basic concepts of generalized linear models, the relationship between explanatory variables and of Poisson counts can be described by Poisson regression or Poisson log-linear models. Poisson model is formed under two principal assumptions. The first is that events occur independently over a given time or exposure period. The second assumption is that the conditional mean and variance are equal. However, in practice, the equality of the mean and variance rarely occurs; the variance may be either greater or less than the mean. If the variance is greater than the mean, it means that counts are more variable than specified by the Poisson events and this phenomenon is described as overdispersion. If the variance is less than the mean, it means that counts are less variable than specified by the Poisson events and the situation is described as underdispersion. However, in practice, underdispersion is less common than overdispersion (McCullagh and Nelder, 1989).

Overdispertion can be caused by excess number of observed zero counts, since the excess zeros will give smaller conditional mean than the true value. The count data with excess zeros is known as zero-inflated counts. Of course it is possible to have fewer zero counts than expected, but this is less common in practice (Ridout et al., 1998; Hilbe, 2011).

To summarize, when there is overdispertion or zero-inflation and we fail to take into account that, it can lead to misinterpretation of the estimate parameter (Jansakul and Hinde, 2002), since the overdispertion or zero-inflation produces:

1. smaller standard errors of the parameter estimates than the true values. Therefore, we may incorrectly choose explanatory variables for the model that are not required; and
2. a reduction of deviance associated with model selection tests. This leads to selecting more complicated models.

In the literature of statistical modeling for counts a number of models and associated estimation methods have been proposed to handle overdispersion and zero-inflated count data; for example, negative binomial and mixed Poisson regression models (Cameron and Trivedi, 1998; Dean and Lawless, 1989) and a large proportion of zero values is observed. In such case, zero inflated Poisson (ZIP) regression model is an appropriate approach for analyzing outcome variables having excess number of zero observations (Lambert, 1992; Ridout et al., 2001 Yesilova et al., 2010). In addition, overdispertion is likely in datasets having excess zero observations. In such cases, ZINB regression model is an alternative method (Ridout et al., 2001; Hilbe, 2011).

### 3.3.1 Poisson regression model

The Poisson regression model has been widely used for the analysis of count data. The Poisson distribution was first used in a regression context by letting the mean parameter  in the Poisson distribution depend on some covariates (Frome et al., 1973). In this model the Poisson parameter corresponding to the expected number of occurrences is modeled as a function of explanatory variables.

Suppose we have an independent sample of pairs of observations where denotes the value of a count event outcome variable occurring in a given time or exposure period with mean parameter and is the value of the explanatory variables for the subject. Assume The probability density function of the Poisson random variables, is characterized by

with mean and variance

One specification that is mostly used for the mean parameter is the exponential specification. This specification ensures that is non-negative and it is given as

where is a vector of explanatory variables and is a dimensional column vector of unknown parameters to be estimated. The estimation is done by using maximum likelihood method. The likelihood function is given by

Taking log on both sides and we get the log likelihood

It can be verified that the first two partial derivatives of the log-likelihood function exist and are given as follows:

Hence, equation or is nonlinear in so that they need to be solved by using an iterative algorithm. The iterative algorithms commonly used are either Newton-Raphson or Fisher scoring. Therefore, will be obtained by maximizing using a numerical iterative method (McCullagh and Nelder, 1989).

### 3.3.2 Overdispertion model

Given a distributional assumption in a regression model, we find overdispersion if the observed variance of the data is greater than the mean supposed by the model. Given a count variable, its dispersion index, is usually defined as a measure to detect departures from the Poisson distribution. The variable is overdisperesed if

There are two basic criteria commonly used to check the presence of overdispertion: the deviance, or the Pearson chi-square statistic, (Hilbe, 2011). For the Poisson regression, and are, respectively, defined in expression and below.

However, these two rules of thumb can yield misleading inferences from a direct likelihood point of view. Therefore, selecting between Poisson regression and an overdisperesed Poisson model should be performed using some appropriate modeling procedure.

#### 3.3.2.1 Negative binomial regression model

A negative binomial is one count distribution used in analyzing overdisperesed data Lawless (1987). The negative binomial regression model is given by

where the mean of the model, , given below , and is called the dispersion parameter. The mean and variance of the model in are given by;

Note that we can allow to vary with the observations but will be an overall parameter in the model that is assumed not to depend on the covariate. Comparing the first two moments with those of the Poisson distribution, we see that the mean when remain the same while difference appears in the variance. As the variance is greater than the mean when this distribution is able to account for overdispersion in the data with respect to the classical Poisson assumption.

The log-likelihood function, of the negative binomial regression model is given by:

since

then

The gradient is

### 3.3.3 Zero-inflated count data regression model

For count data with an excessive number of zero observations zero-inflated models can be used. These models were first introduced by Lambert (1992) to provide a method of accounting for excessive zero counts. We consider zero counts as excessive if the number of observed zero counts exceed the number of zero counts expected by the standard models. Two types of zeros can occur: one comes from the zero state and the other from the standard count distribution state. Thus, the resulting distribution is a mixture of a standard count model, such as the Poisson or negative binomial distribution, with one that is degenerate at zero (Lambert, 1992; Agarwal et al., 2002).

Let be a distribution function for count data, such as the Poisson and NB distribution with an unknown parameter Then a zero-inflated distribution with extra proportion of zeros is obtained from

More specifically, the zero-inflated distribution, denoted as ZI is given by Agarwal et al. (2002)

The mean and variance of the ZI distribution are given by

and

Due to its complicated form, it is not easy to make comparisons between the variances of the underlying count data distribution and the zero-inflated models for the general case. Hence, we will discuss each case separately, and we specify alternatives for the underlying count data distribution and the corresponding zero-inflated versions. From these alternatives the mixture with Poisson and the mixture with negative binomial are the most commonly used (Hilbe, 2011).

#### 3.3.3.1 Zero-inflated Poisson regression model

The zero-inflated Poisson model is a simple mixture model for count data with excess zeros introduced by Lambert (1992). The model is a combination of a Poisson distribution and a degenerate distribution at zero. Specifically, if are independent random variables having a zero-inflated Poisson distribution, the zeros are assumed to arise in two ways corresponding to distinct underlying states. The first state occurs with probability and produces only zeros, while the other state occurs with probability and leads to a standard Poisson count with mean and hence a chance of further zeros. In general, the zeros from the first state are called structural zeros and those from the Poisson distribution are called sampling zeros (Jansakul and Hinde, 2002). This two-state process gives a simple two-component mixture distribution with probability mass function

which we denote by . The mean and variance of are and

To apply the zero-inflated Poisson model in practical modeling situations, Lambert (1992) suggested the following joint models for and ,

where and are covariate matrices andare and vector of unknown parameters, respectively. Note that the vector of covariates and can be the same or different.

For a random sample of observations the log-likelihood function is given by

where is the indicator function for the specified event, equal to 1 if the event is true and 0 otherwise.

#### 3.3.3.2 Zero-inflated negative binomial regression model

Gurmu and Trivedi (1996) used the zero-inflated negative binomial (ZINB) regression to model overdisperesed data with an excess number of zeros. This regression model is given by whereis a dispersion parameter and is assumed not to depend on covariates. Here  is set to model overdispertion case. The mean and variance of the ZINB regression model are

The formulations for and  are the same as those used in the zero-inflated Poisson regression model in. This distribution reduces to zero-inflated Poisson distribution in when

since

The log-likelihood function is

Using, we have the following relationship

The gradient for this model is given by

### 3.3.4 Maximum likelihood estimation of the parameters

The method used to estimate the regression coefficients, is the maximum likelihood estimation (MLE). Statistical software finds model parameter estimates using a numerical iterative algorithm. The most common iterative algorithms are the Newton-Rapson algorithm and Fisher information matrix. The algorithm starts at an initial guess for the parameter values that maximize the likelihood function. The Fisher scoring algorithm for doing this was first proposed by R. A. Fisher. For detail properties and statistical inference, including the maximum likelihood estimation of the parameters for Poisson, negative binomial, ZIP or ZINB models, refer to Lambert (1992) and Dejen and Muniswamy (2012). In this study the parameters of the models have been estimated by maximum likelihood estimation method using statistical software STATA 13.1 which has built-in function.

### 3.3.5 Goodness of fit test

Once a model has been developed, we would like to know how effective the model is in describing the outcome variable. This is referred to as goodness of fit.

Tests for the validity for can be based on any one of the following two principles of

- Likelihood ratio test (LRT); and
- Wald test;

#### 3.3.5.1 Likelihood-ratio test (LRT)

Letdenote the value of the log-likelihood function evaluated at the restricted maximum likelihood estimates (for instance, the Poisson model), and the value of the log-likelihood function evaluated at the unrestricted maximum likelihood estimates (for instance, the negative binomial model), and let denote the number of restrictions ( in a test of the Poisson model against the negative binomial model). Then, under (if the restriction is correct):

where  is a chi-square distribution with  degree of freedom (Hilbe  ,2011). If the computed value of LRT the test statistic exceeds the critical value  the null hypothesis is rejected.

#### 3.3.5.2 Wald test

The Wald test statistic is also an alternative test which is commonly used to test the significance of individual parameter regression coefficients for each independent variable that is, = *0* versus. The Wald test statistic is given as follows:

For the  coefficient

For large sample size this statistic has an approximate chi-square distribution with one degree of freedom assuming that  holds.

## 3.4 Models comparison

If there are several models to be compared, in order to select a more appropriate model from a given number of candidate models which fits the data additional to using the log-likelihood and deviance it can be also easily selected by using the Akakie information criterion (AIC).

### 3.4.1 Akakie information criterion

Akakie information criterion (AIC) is the most common means of identifying the model which fits well by comparing two or more than two models. AIC is useful in balancing the goodness of fit against the complexity of the model. It is given as

(3.30)

where are the log likelihood of a model that will compare with the other models and  is the number of parameters in the model including the intercept (Ismail and Jemain, 2007).

### 3.4.2 The Deviance

A measure of discrepancy between observed and fitted values in the deviance and the expression is given by,

(3.31)

For large samples, the distribution of the deviance under the null hypothesis, is approximately chi-squared with degrees of freedom, where is the number of observations and the number of parameters. Thus, the deviance can be used directly to test the goodness of fit of the model.

An alternative measure of goodness of fit is Pearson's chi-squared statistic, which is defined as:

(3.32)

In large samples, the distribution of Pearson's statistic is also approximately chi-squared with degrees of freedom under the null hypothesis, . An appropriate model is the one which has a lower deviance value than those other models include.

### 3.4.3 Vuong statistic: selecting inflated model over standard model

Vuong test is used to select appropriate inflated models over standard models. Suppose and denote the probability density function of a zero-inflated model (ZIP or ZINB) and a standard model (Poisson or Negative Binomial), respectively; and and denote their corresponding cumulative distribution functions. We want to test the following hypothesis:

Two distribution functions are equivalent

Two distribution functions are different (two tail test)

In order to perform this test we start by defining

where and are predicted probabilities of the corresponding models and respectively. Let = and denotes the mean and standard deviation of the measurements . Then the Vuong test statistic is defined as

For large sample size and under the null hypothesis the test statistic follows an asymptotic standard normal distribution (Vuong, 1989). A computed value of the Vuong test statistic whose value is positive and larger than a tabulated z-value favors the zero-inflated model and vise-versa.

# CHAPTER FOUR

# STATISTICAL DATA ANALYSIS

## 4.1 Descriptive results

It can be seen from Table 4.1 that 76.36% of the households did not face any cattle death through the survey period of time while 0.61% of them faced more than seven cattle death. Figure 4.1 shows the histogram of the total cattle death of the agricultural holder. The graph indicates a high frequency of zero and is highly skewed to the right giving as evidence for the existence of overdispersion in the data.

**Table 4. 1:** Observed frequency of cattle death and percent over the last 12 months.

| **Number of cattle death** | **Observed Frequency** | **Observed Percent** |
|---|---|---|
| 0 | 20,942 | 76.36 |
| 1 | 4,241 | 15.46 |
| 2 | 1,231 | 4.49 |
| 3 | 432 | 1.58 |
| 4 | 224 | 0.82 |
| 5 | 122 | 0.44 |
| 6 | 66 | 0.24 |
| 7+ | 166 | 0.61 |
| Total | 27,424 | 100.0 |
| Mean | 0.4034058 | |
| variance | 0.938 | |
| Skewness | 3.754267 | |
| Kurtosis | 20.69516 | |

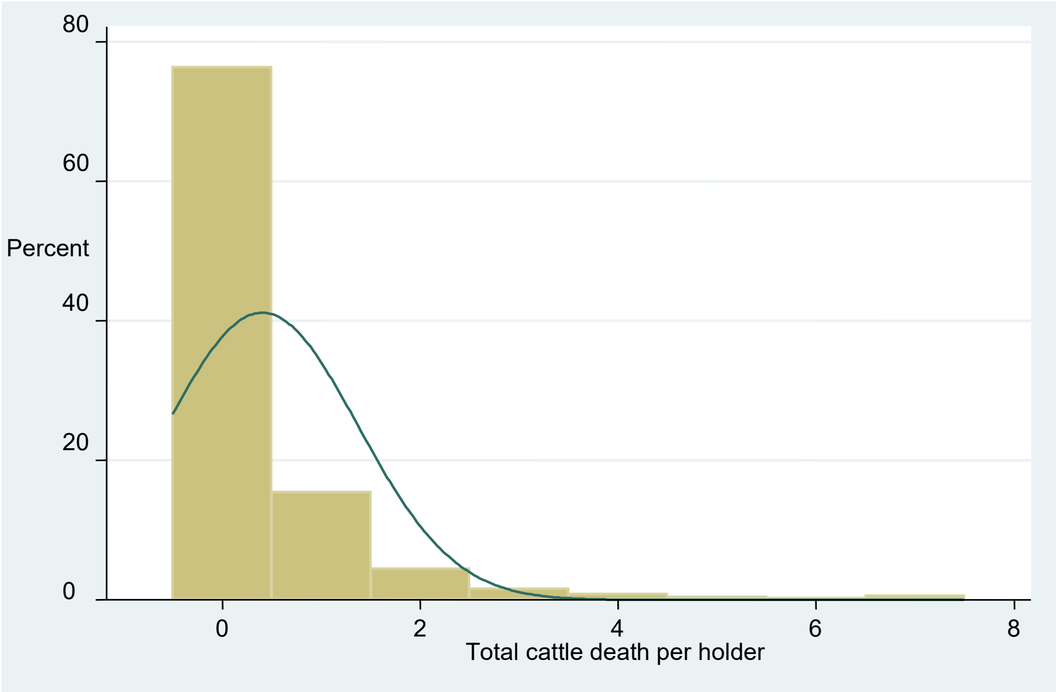


**Figure 4. 1:**Distribution of total number of cattle death per holder

The descriptive results from the agricultural survey presented in Table 4.2. The results indicate that in Gambella (1.4), Afar (1.027) and Benshangul Gumuz (0.5594) the highest mean number of cattle death occurred. Somali (0.2208), Dire Dawa (0.2727), Harari (0.301,) experienced the lowest mean number of cattle death. Amhara (0.3058), SNNPR (0.3380), Oromyia (0.3674) and Tigray (0.3895) had the intermediate mean cattle death.

The mean number of cattle death for female agricultural holders (0.4239) exceeded that of male holders (0.3994). The results also revealed that the owner of agriculture holder who without formal education (0.4271) had the highest mean number of cattle death than that holder of agriculture who had completed primary education (0.3484) and for secondary and above education (0.401). For those participating in livestock activities growing crops and both (mixed) the mean losses, respectively, were 0.7782, 2944, and 0.3923

**Table 4. 2:**Descriptive results of the number of cattle death

| Variables | Categories | Total Observations | Mean | Std. D |
|---|---|---|---|---|
| Region | Tigray | 1918 | 0.3895 | 0.90096 |
| | Afar | 777 | 1.027 | 1.74703 |
| | Amhara | 5016 | 0.3058 | 0.75397 |
| | Oromyia | 9814 | 0.3674 | 0.83226 |
| | Somali | 530 | 0.2208 | 0.6752 |
| | Benshangul G. | 581 | 0.5594 | 1.17091 |
| | SNNPR | 7334 | 0.338 | 0.86012 |
| | Gambella | 933 | 1.4009 | 2.09539 |
| | Harari | 279 | 0.3011 | 0.80625 |
| | Dire Dawa | 242 | 0.2727 | 0.52332 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Gender | Female | 4520 | 0.4239 | 1.04171 |
| | Male | 22904 | 0.3994 | 0.95363 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Education background | No Education | 18228 | 0.4271 | 1.02115 |
| | Primary | 7812 | 0.3484 | 0.82881 |
| | Sec & above | 1384 | 0.401 | 0.98008 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Farming type | Crop | 625 | 0.2944 | 0.80447 |
| | Livestock | 947 | 0.7782 | 1.61043 |
| | Mixed | 25852 | 0.3923 | 0.93776 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Extension Package | No | 4520 | 0.4239 | 1.04171 |
| | Yes | 22904 | 0.3994 | 0.95363 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Agro-ecological zone | Dega | 2536 | 0.2472 | 0.63183 |
| | Lower Kolla | 1943 | 1.1853 | 1.93819 |
| | Woina Dega | 31 | 0.0968 | 0.30054 |
| | Unsuit Wurch | 7398 | 0.438 | 0.98926 |
| | Upper kolla | 15424 | 0.3128 | 0.746 |
| | Wurch | 92 | 0.7174 | 1.49964 |
| | Total | 27424 | 0.4034 | 0.96872 |
| Land holding size | Under 0.10 hectare | 1467 | 0.5174 | 1.28064 |
| | Hectare 0.10-0.50 | 6272 | 0.4219 | 1.101 |
| | Hectare 0.51-1.00 | 6519 | 0.366 | 0.88636 |
| | Hectare1.01-2.00 | 7401 | 0.3489 | 0.81566 |
| | Hectare 2.01-5.00 | 4751 | 0.4365 | 0.92299 |
| | Hec5.01-10.00 | 582 | 0.732 | 1.2909 |
| | Over 10 | 432 | 0.4398 | 1.12804 |
| | Total | 27424 | 0.4034 | 0.96872 |

Livestock extension package is a program introduced by Ministry of Agriculture targeting the livestock sub-sector to be productive by introducing improved breeds, livestock health facility and training of the livestock holders. So, agricultural holders were asked whether or not they participated in the extension package. Those who did not participate in the extension package were found to have experienced higher mean number of cattle death compared with those who participated in the extension package.

The lower kola encountered the highest mean number of cattle deaths (1.1853), and the second highest mean number of cattle death was in Wurch (0.7174), followed by unsuited wurch (0.438). The results also revealed that the lowest mean number of cattle death occurred in Dega (0.2472) and Woina dega (0.0968).

The land holding size was measured based on agricultural holder which s/he manages or controls over the land during the study period. The results indicate that the highest mean number of cattle death was reported to lie between land holding size 5.01 and 10.00 hectares which was 0.732. Whereas, land holding size areas between 0.51 and 1.00 hectares had the lowest mean number of cattle death (0.366).

## 4.2 Parameter estimation of NB, ZIP and ZINB regression models

The analysis attempted to pick the most appropriate statistical model in order to identify significant predictor variables of cattle death. Among these, negative binomial, ZIP, and ZINB regression models are used for analysis of count data. The statistical analyses are performed using Statistical Package for Social Sciences (SPSS), version 20 and STATA, version, 13.1.

In the following table we attempted to put parameter estimation of the three regression models; as it can be seen from Table 4.3, NB, ZIP and ZINB. Due to overdispersion we need to use negative binomial model since it allows overdispersion with a parameter  , beyond 5% significant. Overdispersion is occurred due to excess zeros in the dataset and in this case we include a zero-inflation parameter.

**Table 4. 3:** Parameter estimation of NB, ZIP and ZINB

| Variables | NB IRR | NB Std. Err. | ZIP IRR | ZIP Std. Err. | ZINB IRR | ZINB Std. Err. |
|---|---|---|---|---|---|---|
| **Age** | 1.005* | 0.001 | 1.005* | 0.001 | 1.006* | 0.001 |
| **Household size** | 1.067* | 0.006 | 1.040* | 0.004 | 1.059* | 0.010 |
| **Education background** | | | | | | |
| No education(reference) | | | | | | |
| Primary education | 0.909* | 0.030 | 0.899* | 0.025 | 0.914* | 0.030 |
| Secondary and above | 0.958 | 0.061 | 0.928 | 0.047 | 0.951 | 0.060 |
| **Farm Type** | | | | | | |
| Livestock (reference) | | | | | | |
| Crop | 0.694* | 0.086 | 0.695* | 0.070 | 0.469* | 0.091 |
| Mixed (both) | 0.767* | 0.062 | 0.772* | 0.041 | 0.501* | 0.054 |
| **Land holding size** | | | | | | |
| Hec 5.01-10.00(reference) | | | | | | |
| Under 0.10 hectare | 0.459* | 0.049 | 0.614* | 0.046 | 0.459* | 0.050 |
| Hectarage 0.10-0.50 | 0.493* | 0.043 | 0.608* | 0.037 | 0.485* | 0.041 |
| Hectarage 0.51-1.00 | 0.496* | 0.042 | 0.569* | 0.034 | 0.490* | 0.041 |
| Hectarage 1.01-2.00 | 0.504* | 0.043 | 0.582* | 0.035 | 0.501* | 0.042 |
| Hectarage 2.01-5.00 | 0.628* | 0.054 | 0.701* | 0.042 | 0.624* | 0.053 |
| Over 10 hectarage | 0.573* | 0.076 | 0.756* | 0.078 | 0.573* | 0.075 |
| **Agro-ecological zone** | | | | | | |
| Lower kolla (reference) | | | | | | |
| Dega | 0.213* | 0.014 | 0.244* | 0.019 | 0.205* | 0.014 |
| Woina dega | 0.085* | 0.055 | 0.0379* | 0.022 | 0.082* | 0.052 |
| Unsuit wurch | 0.381* | 0.019 | 0.454* | 0.016 | 0.362* | 0.018 |
| Upper kolla | 0.271* | 0.013 | 0.316* | 0.012 | 0.260* | 0.013 |
| Wurch | 0.708 | 0.145 | 0.939 | 0.139 | 0.681 | 0.138 |
| **Constant** | 1.571* | 0.211 | 3.496* | 0.313 | 2.637* | 0.446 |
| **lnalpha** | 0.804* | 0.028 | | | 0.718* | 0.077 |
| **Alpha** | 2.234* | 0.062 | | | 2.049* | 0.159 |
| **Inflate** | | | | | | |
| Household size | | | -0.036* | 0.008 | -0.162** | 0.070 |
| **Farm type** | | | | | | |
| Livestock | | | | | | |
| Crop | | | | | -2.311 | 2.396 |
| Mixed (both) | | | | | -3.239 | 2.271 |
| **Agro-ecological zone** | | | | | | |
| Dega | | | 0.177 | 0.123 | | |
| Unsuit wurch | | | 0.241* | 0.063 | | |
| Upper kolla | | | 0.190* | 0.062 | | |
| Wurch | | | 0.430 | 0.255 | | |
| **Constant** | | | 0.430* | 0.069 | 0.217 | 0.402 |

* Significance at 1% level, **Significance at 5% level,

For each model IRR (incident rate ratio) and the standard error of the estimates (Std. Err) are given in Table 4.3. ZINB model has the largest standard error compared to others. It should be noted that unlike most statistical models, in count models, the smaller standard error of a parameter may not give a precise estimate for count data. We would like to mention that the overdispersion parameters were statistically significant and their value in the NB case was 2.23 and that of ZINB case was 2.05.

## 4.3 Model Comparisons

The following table displays the calculated values of the Akakie information criterion (AIC), Vuong statistic, deviance, and LRT.

**Table 4. 4 :**Model comparison of cattle death

| Criteria | Count Regression Models | | |
|---|---|---|---|
| | **NB** | **ZIP** | **ZINB** |
| **LRT** | 1,401 | 1,584 | 1,349 |
| **AIC** | 43,429 | 44,363 | 43,387 |
| **Vuong** | - | 22.85 | 3.52 |
| **Deviance** | 43,392 | 44,313 | 43,341 |
| **Log likelihood** | -21,695 | -22,156 | -21,670 |

Vuong statistic was used to select non-nested models; Poisson versus ZIP and NB versus ZINB. The result revealed that both ZIP and ZINB models better fitted than did Poisson and NB models, respectively. Furthermore, ZINB had the least AIC, indicating that the ZINB was the best model to fit cattle death data. Moreover, the above conclusion was further supported by the deviance and the log-likelihood statistics. On the basis of all the above criteria the ZINB model was found to be the most appropriate to fit the cattle death data.

## 4.4 Predicted value

Once a model has been selected and fitted it necessary important to visualize how the model adequately explains the dependent variable. One way to do this is to look the observed and predicted probabilities produced by each count model. Table 4.5 indicates that ZINB and NB give similar results for almost all number cattle death categories in predicting the observed values.

**Table 4. 5:** Observed vs predicted values from the three models

| | **Predicted probabilities** | | | |
|---|---|---|---|---|
| **Number of cattle death** | **Observed** | **NB** | **ZIP** | **ZINB** |
| 0 | 0.764 | 0.765 | 0.763 | 0.764 |
| 1 | 0.155 | 0.149 | 0.137 | 0.149 |
| 2 | 0.045 | 0.049 | 0.063 | 0.049 |
| 3 | 0.016 | 0.019 | 0.023 | 0.019 |
| 4 | 0.008 | 0.009 | 0.008 | 0.008 |
| 5 | 0.004 | 0.004 | 0.003 | 0.004 |
| 6 | 0.002 | 0.002 | 0.001 | 0.002 |
| 7 | 0.006 | 0.001 | 0.001 | 0.001 |

Table 4.6 revealed that the maximum difference encountered in prediction of the three models. The maximum difference occurred at the observation at value 'one'. NB and ZINB models seem to have the smallest difference.

**Table 4. 6 :** Maximum difference of the three count model at "one"

| Model | Maximum Difference | At value | Mean \|Dif\| |
|---|---|---|---|
| NB | 0.006 | 1 | 0.002 |
| ZIP | -0.018 | 2 | 0.005 |

| | | | |
|---|---|---|---|
| ZINB | 0.005 | 1 | 0.002 |

## 4.5 Plots of differences between observed and predicted value

Figure 4.2 given below shows the fitted count models. There is a very close overlap of NB regression model and ZINB models. This tells us that both the predicted and the observed value of the cattle death count do not differ much.

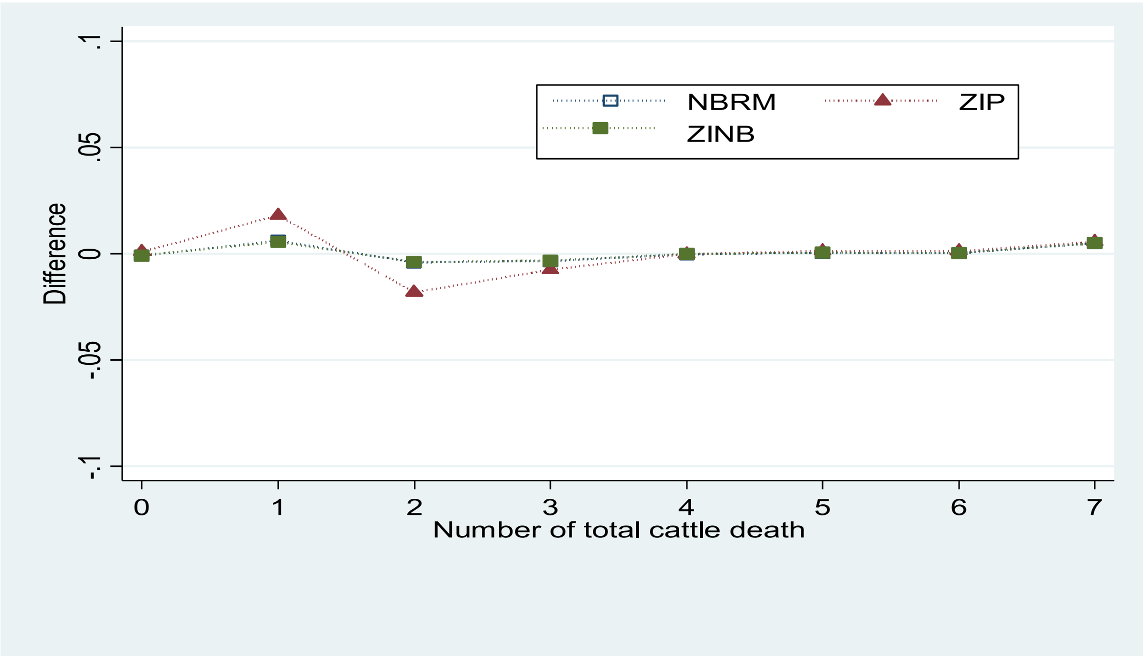


**Figure 4. 2:** Difference b/n observed and predicted probabilities for number of cattle death

## 4.6 Discussion on the ZINB regression model

The issue that we had been dealing with the count model was to identify determinant factors/variables of the cattle death data. Interpreting the results obtained using the appropriate model is crucial to choose the factors/variable so that these could be used in the cattle subsector research.

**Table 4. 7 :**Model coefficients of ZINB model and percentage change in expected count

| | Coefficient | Wald | P-value | Percentage change |
|---|---|---|---|---|
| Age | 0.006 | 5.904 | 0.000 | 0.600 |
| Household size | 0.058 | 6.120 | 0.000 | 5.900 |
| **Educational background** | | | | |
| No education (reference) | | | | |
| Primary education | -0.090 | -2.748 | 0.006 | -8.600 |
| Secondary and above | -0.050 | -0.791 | 0.429 | -4.900 |
| **Farm type** | | | | |
| Livestock (reference) | | | | |
| Crop | -0.757 | -3.882 | 0.000 | -53.100 |
| Mixed (both) | -0.692 | -6.429 | 0.000 | -49.900 |
| **Land holding size** | | | | |
| Hectarage 5.01-10.00 (reference) | | | | |
| under 0.10 hectare | -0.777 | -7.178 | 0.000 | -54.000 |
| Hectarage 0.10-0.50 | -0.723 | -8.508 | 0.000 | -51.500 |
| Hectarage 0.51-1.00 | -0.713 | -8.471 | 0.000 | -51.000 |
| Hectarage 1.01-2.00 | -0.691 | -8.302 | 0.000 | -49.900 |
| Hectarage 2.01-5.00 | -0.472 | -5.595 | 0.000 | -37.600 |
| Over 10 hectare | -0.556 | -4.256 | 0.000 | -42.600 |
| **Agro-ecology zone** | | | | |
| Lower kola (reference) | | | | |
| Dega | -1.587 | -23.702 | 0.000 | -79.600 |
| Woina dega | -2.504 | -3.920 | 0.000 | -91.800 |
| Unsuit wurch | -1.015 | -20.302 | 0.000 | -63.800 |
| Upper kolla | -1.346 | -27.881 | 0.000 | -74.000 |
| Wurch | -0.385 | -1.901 | 0.057 | -31.900 |
| **Constant** | 0.970 | 5.741 | 0.000 | . |
| **Alpha** | | | | |
| lnalpha | 0.718 | . | . | . |
| Alpha | 2.049 | . | . | . |

| **Binary Equation: factor change in odds of always zero** | | | | |
|---|---|---|---|---|
| **Household size** | -0.1624 | -2.329 | 0.02 | -15 |
| **Farm Type** | | | | |
| Livestock (reference) | | | | |
| Crop | -2.3114 | -0.965 | 0.335 | -90.1 |
| Both(Mixed) | -3.2392 | -1.427 | 0.154 | -96.1 |

In the discussion of the count data results we use incident ratio rate (IRR) to interpret the coefficients directly interpreted (Hilbe, 2008).

In Ethiopia, agricultural activities are divided into three farming activities: crop growing, livestock and mixed farming. Table 4.7 shows that farming type is statistically significant in reducing cattle death. Those who participated in crop cultivation and mixed farming activities were able to reduce cattle death by 53% and 50%, respectively, compared to the livestock farming activities.

Education is a key factor for the change in attitude of agricultural holder with regard to agricultural activities. Farmers are exposed to policies and strategies through trainings, manuals and researches on to improve day to day agricultural activities. The result showed that educational level had statistically significant effect on the number of cattle death. Those who had completed primary education could decrease cattle death by 8.6% when compared to those with no formal education.

A five year increase in age of agricultural holder is associated with a 3% increase of cattle death. Agricultural holding level, grazing land and crop land (permanent crop and temporary crops) are very crucial to integrate livestock farming system and other farming activities. Land size groups under 0.10 hectare, 0.10-0.50 hectare, 0.51-1.00 hectare, 1.01-2.00 hectare and 2.01-5.00 hectares are associated with reduction of cattle death by 54%, 51.5%, 51%, 49% and 37%, respectively, compared to land holding groups 5.01-10.00 hectares.

Agro-ecology zone is a determining factor associated with death of cattle because it is directly related to the pastureland availability, water availability and distribution of cattle diseases which varies from one agro-ecological zone to another agro-ecological zone. The results showed that the number of cattle death decreased by 79.6%, 91.8%, 63.8%, and 64.5% for the agro-ecology zone Dega, Woina dega, unsuited wurch and upper kola, respectively, compared to lower kola.

As shown in Table 4.7, the “always zero group” can be interpreted as follows. For a unit increase of household size, the odds of no occurrence of cattle death (being always zero group) decreased by 15%.

# CHAPTER FIVE

# CONCLUSION AND RECOMMENDATIONS

## 5.1 Conclusion

This study identified some determinants of the cattle death in Ethiopia. The data were taken from CSA and MoA. In this study, we considered three count regression models to identify determinant factors of cattle death. Since the cattle death data is overdisperesed. Poisson regression model is not appropriate. Instead, we fitted count regression models-NB, ZIP and ZINB that accommodate overdispersion. These models provided better fit the data. Comparing these models based on the deviance, log-likelihood and Vuong statistic, the result revealed that ZINB model was found to be the most appropriate.

The findings of this study showed that agricultural holder’s age, farming type, educational background, agro-ecological zone and land size have statistically significant effect on cattle death. Mixed farming contributed to decreasing the number of cattle death compared with livestock farming activities; primary education is associated with reducing cattle death compared to no formal education; small size land holding categories contributed to decrease in the number of cattle death as compared to 5.01-10.00 hectares; agro-ecological zone categories: unsuited wurch, woina dega, and dega are found to be associated with the number of cattle death compared to lower kola.

## 5.2 Recommendations

Based on the research findings, we make the following recommendations:

- Stakeholders should work to reduce cattle death by preparing continuous trainings and capacity buildings related to cattle for smallholder farmers.
- Promote mixed farming system.
- Priority should be given for development of livestock infrastructure in lowland areas to reduce cattle death.
- Variables such as seasonality, breed type, death history of the cattle, vaccination and age of cattle can be included in other similar studies in the future.

# References


Agarwal D.K., Gelfand, A.E. and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. Environmental and Ecological Statistics, 9: 341–355.

Agresti A. (2007). Introduction to categorical analysis. Wiley Series in Probability and Statistics.

Alemu Bekele, Tebikew, J., Chamberlin W., and Benson T. (2009). Impacts of the productive safety net program in Ethiopia on livestock and tree holdings of rural households. Ethiopia Environmental Economics Policy Forum for Ethiopia (EEPFE) Ethiopian development research institute (EDRI/AAU) Addis Ababa, Ethiopia.

Azage T., Yenesew A., Solomon M. and Firew T. (2013). Assessment of sheep production system in Burie district North-Western Ethiopia. Global journal of agriculture research, 1(2):29-47.

Bernard Joachim (2008). Epidemiology of morbidity and mortality on small holder dairy farms in Eastern and South Africa. Institute of veterinary, Massey University, Palmerstone, North New Zealand

Cameron A.C. and P.K. Trivedi (1998).Regression Analysis of Count Data, New York: Cambridge University Press.

Central Statistical Agency and Ministry of Agriculture (2006). Atlas of Agro-ecology of Ethiopia. Addis Ababa, Ethiopia

Central Statistics Agency with MoARD and FAO (2008). Agricultural Statistics Methodology in Ethiopia. Addis Ababa, Ethiopia.

Central statistical Agency (2013). Annual Abstract. Addis Ababa, Ethiopia.

Central Statistical Agency (2008). National strategy for the development of statistics. Addis Ababa, Ethiopia.

Central Statistical Agency (2014). Livestock and Livestock characteristics statistical Report. Addis Ababa, Ethiopia.

Chaudhary, A. Jain, S. K., Keshri, R., Goswami, A., and Sarkar, A., (2013). Identification of drought-vulnerable areas and factors for cattle death. Pakistan Veterinary Journal, 40: 41-43.

Cheryl L. Waldner, Richard I. Kennedy, Leigh Rosengren and Edward G. Clark (2009). A field study of culling and mortality in beef cows from Western Canada. Canadian Veterinary Journal, 50:491-492.

Dean C. and Lawless J.F. (1989). Test for detecting overdispertion in Poisson regression model. Journal of American Statistical Association. 84(4):282-284.

Dixon J., Gulliver and A., Gibbon D. (2001). Farming systems and poverty, FAO/World bank, Rome/Washington DC.

Dejen T. and B. Muniswamy (2012). Power of test for negative binomial regression coefficients in count data. International Journal of Mathematical Archive, 3(8):3136-3142.

Delgado, C., Staal, S., and Nicholson, C. (1999). Livestock to 2020 – The Next Food Revolution. Food, Agriculture and the Environment Discussion Paper 28. IFPRI, Washington D.C.

Food and Agriculture Organization (FAO) (2010). Breeding strategies for sustainable management of animal genetic resources. Animal Production and Health Guidelines, No.3.Food and Agriculture Organization of the United Nations, Rome, Italy.

Food and Agriculture Organization (FAO) (2010). Food Insecurity in the World, Addressing Food Insecurity in Protracted crises, FAO (2010). Rome, Italy.

Food and Agriculture Organization (FAO). (2014). Livestock statistics in 2013 FAOSTAT. Agriculture Organization of the United Nations, Rome, Italy further information www.faostat.org

Frome, E. L., Kurtner, M. H. and Beauchamp, J. J. (1973). Regression analysis of Poisson distribution data, Journal of American Statistics association, 68:935-940.

Godfrey, S. (2010). Sustainability of water supply infrastructures: A good practice from Guro district in central Mozambique. s.l.: UNICEF, 2010.

Gurmu, S. and Trivedi, P. K. (1996). Excess zeros in count models for recreational trips. Journal of Business and Economic Statistics,14: 469–477.

Hilbe J. M. (2008).Brief overview on interpreting count model risk ratio: addition to negative binomial regression model. Cambridge University Press

Hilbe J. M. (2011). Negative binomial regression 2nd edition. Published in the United States of America by Cambridge University Press, New York.

George k., Bernard Njehia and Abdi Y. Guliye (2011).Effects of cattle rustling and household characteristics on migration decisions and herd size amongst pastoralists in Baringo District, Kenya. Research, policy and practice 2011, (1): 18-35.

Hossain M.M., Islam A.H. kemal A.K. Rahman and Cho H. S. (2014). Small scale Dairy farming practice in a selective area of Bangladish. Department of Livestock service, Pakistan Journal of Nutrition, 4:215-216.

Ismail, Noriszura and Jemain, Abdulaziz (2007). Construction of insurance scoring system using regression models. Journal of Modern Applied Statistical Methods, 7(2):25-35, article 25.

Jansakul N. and Hinde J. P. (2002). Score tests for extra-zero models, Journal of Computational statistics and Data Analysis, 55: 15-16.

John M. G. (2009). Estimating the seroprevalence of T.parvain infection and determination of associated risk factors in the cattle population of Mebeere district , Kenya, Department of Public Health, Pharmacology and Toxicology, Faculty of Veterinary medicine, University of Nairobi

Karin A. (2014). On farm cow mortality in Swedish dairy herds. Faculty of Veterinary Medicine and Animal Science. Department of Clinical Science, Uppsala, Sweden.

Kefyalew A. and Tegegne F. (2012). The effect of climate change on ruminant livestock population dynamics in Ethiopia. Bahir Dar University, Bahir Dar, Ethiopia and Mizan-Tepi University, Mizan Ethiopia, Livestock Research for Rural Development, 24-10.

Lambert D. (1992). Zero-inflated Poisson regression with an application to random defects in manufacturing Technometrics,34 (1):14-28.

Lawless J. F. (1987). Negative binomial and mixed Poisson regression. The Canadian journal of statistics,15(3):209-225.

Little P.D., Bailey D.V. Barrett C.B. and chabari F. (2001). Cross border livestock trade and food security in the horn of Africa: Accessed online at February 15, 2015.
http://www.ies.wisc.edu/ltc/live/bashorn0109a.pdf

McCullagh P. and Nelder J. A. (1989). Generalized Linear Models. Second edition. London: Chapman and Hall.

Michael R. M. (2013). Small holder market participation and welfare effects: evidence from the Kenya dairy sector. Cornell University, Cornell.

Ministry of Energy (2010). Annual energy review. Addis Ababa, Ethiopia, 10:23-24.

Ministry of Economic Development and Cooperation (MEDaC) (2007). Survey of the Ethiopian Economy, Review of post reform developments (2002/03-2005/06).

Ministry of Finance and Economic Development (2013). Yearly economic report, 20: 60-61.

Morignat E., Maugey X., Vinard J.L., Hendrikx P. Ducrot C., and Calavas D.,(2014). Assessment of the Impact of the 2003 and 2006 Heat Waves on Cattle Mortality in France. PLoS ONE 9(3): e93176.doi: 10.1371/journal. pone.0093176.

Owen E., Smith T., Steele M.A., Anderson S., Duncan A.J., Herrero M., Leaver J.D., Richards J.I. and Ku-Vera J.K.(2005).Responding to the Livestock Revolution: The Role of Globalization and Implications for Poverty Alleviation Nottingham University Press, Nottingham.

Ridout M., Demetrio C. G. B. and Hinde J. (1998). Models for count data with many zeros. Invited paper presented at the nineteenth international biometric conference, Cape Town, South Africa, 179-190.

Samson L. and Frehiwot M. (2010).Prevalece of Small ruminant Trypanosomosis and Tsetse fly challages in upper Didessa valley, Ethiopia Global Veterinaria. 5(4): 215-219 IDOSI publications.

Shimelis, M., A. Mekonnen and F. Abebe (2011). Epidemiology of gastrointestinal helminthiasis of small ruminants in selected sites of North Gondar zone, Northwest Ethiopia, Journal of Ethiopian Vetenary science, 15(2): 57-68.

Thumbi Samuel M., Mark BMdec Bronsvoort, Henry Kiara, Jane Poole, Mary Ndila, Ilana Conradie Amy Jennings, Ian G Handel, JAW Coetzer, Johan Steyl, Olivier Hanotte and Mark EJ Woolh (2013). Mortality in East African shorten Zebu cattle under one year. BMC Veterinary Research, 9:175-195.

Sisay, M.M. (2007). Helminth Parasites of Sheep and Goats in Eastern Ethiopia. Epidemiology and Anthelmintic Resistance and its Management. Faculty of Veterinary Medicine and Animal Science. Department of Biomedical Sciences and Veterinary Public Health Division of Parasitologyand Virology Uppsala, Sweden Doctoral thesis Swedish University of Agricultural Sciences,11-21.

Vuong Q.H. (1989). Likelihood ratio tests for model Selection and non-nested hypotheses. Econometrica, 57:307–333.

World Bank (2013). Ethiopian economic review and progress report. and accessed from www.worldbank.org home page October 12,2014.

Yeşilova, A., Kaydan, B. and Kaya, Y. (2010). Modeling insect-egg data with excess zeros using zero-inflated regression models. Hacettepe Journal of Mathematics and Statistics, 39(2):273-282.