DEBRE BERHAN UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCES

DEPARTMENT OF STATISTICS (MSC.BIOSTATISTICS)

FACTORS ASSOCIATED WITH ILLEGAL MIGRANTS FROM RAYA KOBO, ETHIOPIA TO SAUDI ARABIA

M.SC. THESIS

BY

MESERET YIRGA

DEBRE BERHAN, ETHIOPIA

FEBRUARY, 2023

FACTORS ASSOCIATED WITH ILLEGAL MIGRANTS FROM RAYA KOBO, ETHIOPIA TO SAUDI ARABIA

BY

MESERET YIRGA

PRINCIPAL ADVISOR: AYELE ADMASSU (Ass. PROF)

CO-ADVISOR: SINTAYEHU WORKU (MSc.)

A THESIS SUBMITTED TO THE DEPARTMENT OF STATISTICS IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS

DEBRE BERHAN, ETHIOPIA

FEBRUARY, 2023

DECLARATION

I hereby declare that this Master of Science thesis is my original work and has not been presented for a degree in any other university, and all sources of material used for this thesis have been appropriately admitted.

Name: **Meseret Yirga**

Signature: ____________

Date: February, 2023

Approved by:

Ayele Adimassu (Ass.Prof)	----------------------	2/11/2023
Principal Advisor	Signature	Date
Sintayehu Worku (MSc.)	----------------------	02/11/2023
Co-Advisor	Signature	Date

THESIS APPROVAL SHEET

We undersigned members of the Board of Examiners of the final open defense by Meseret Yirga Dubale have read and evaluated his thesis entitled **"FACTORS ASSOCIATED WITH ILLEGAL MIGRANTS FROM RAYA KOBO, ETHIOPIA TO SAUDI ARABIA"** and examined the candidate. This is, therefore, to certify that the thesis has been accepted in partial fulfillment of the requirement of the Degree of master of sciences in statistics with specialization of biostatistics.

Approved by:

Ayele Adimassu (Ass.Prof)	----------------------	------------------
Principal Advisor	Signature	Date
Sintayehu Worku (MSc.)	-----------------------	---------------------
Co-Advisor	Signature	Date
----------------------	-----------------------	----------------------
External examiner	Signature	Date
-----------------------	-------------------------	-----------------------
Internal examiner	Signature	Date
-----------------------	-------------------------------	---------------------
Chairperson	Signature	Date

THESIS APPROVAL SHEET

We, the undersigned member of the broad examiners of the final open defense by Meseret Yirga Dubale have read and evaluated her thesis entitled **"FACTORS ASSOCIATED WITH ILLEGAL MIGRANTS FROM RAYA KOBO, ETHIOPIA TO SAUDI ARABIA"** and examined the candidate. This is therefore to certify the thesis has been accepted in partial fulfilment of the requirement for the degree of master of sciences in statistics with specialization of biostatistics.

Approved by:

______________________	________________	_________
Name of Advisor	Signature	Date
_________________	________________	_________
Name of External examiner	Signature	Date
_________________	________________	_________
Name of Internal examiner	Signature	Date
_________________	________________	_________
Associate Dean, College Res/CS& Postgraduate	Signature	Date
_________________	________________	_________
Dean collage of Graduate Studies	Signature	Date

Final approval and acceptance of the thesis is contingent up on the submission of the thesis of the final copy of the thesis to the college of graduate studies (CGS) through the department of graduate committee (DGC) of the candidate's

ACKNOWLEDGMENT

Above all my thanks is to Almighty GOD, for giving me good health, strength, and well-being during all my works and in all my life. Next, it gives me a great pleasure to come to this final point to express my deepest gratitude to all individuals and institutions who contributed in the realizing of this thesis paper.

My special gratitude goes to Ayele A. (Ass. prof), my main advisor for his supervision, invaluable advice, professional guidance and my thanks goes to my Co-advisor Mr. Sintayehu W. (MSc) for his invaluable advice, professional guidance, useful comments and suggestions. And also I would like to acknowledge Aksum University which gave me this chance to learn my M.Sc. degree in Biostatistics and all members of Debre Berhan University, instructors and all individuals in the Department of Statistics of for their support and other facilities.

Contents

LIST OF FIGURE

LIST OF TABLE

ABBREVIATIONS

AIC	Akaike information criterion
BIC	Bayesian information criterion
CSA	Central Statistical Agency of Ethiopia
HH	Household
HHH	Household head
IOM	International Organization of Migration
IRR	Incidence Rate Ratio
LRT	Likelihood-ratio test
MMC	Mixed Migration Centre
NB	Negative binomial
RDH	Regional Data Hub
ZI	Zero-Inflated
ZIP	Zero-Inflated Poisson
ZINB	Zero-Inflated Negative Binomial
HP	Hurdle Poisson
HNB	Hurdle Negative binomial

ABSTRACT

The youth migration affects the socioeconomic development of the people in developing countries like Ethiopia. The goal of this study was to explore the determinants that affect the number of migrants per household in raya kobo woreda, north wollo zone. To achieve the objective, the investigator used primary data and that primary information was collected from the households by using questionnaires. A cross-sectional study was carried out in August 2022 to October 2022 and a two stage cluster sampling was also applied to get the eligible respondents. Among the total of 49 kebeles, 7 kebeles (clusters) were selected at the first stage and for the second stage, a total 577 respondent household heads were selected from these selected kebeles (clusters).Descriptive statistical method was employed to analyze quantitative data by using STATA and SPSS whereas, to explore the determinant factors of migration Count Regression model was applied by using R and STATA software. In this study the hurdle negative binomial (HNB) model was found to be the most appropriate and fitted model for analyzing the number of migrants. Finally Sex of the household head, age of the household head, education level, religion of the household head and residence place of the household were the significant factors of migrants per household in the Raya kobo Woreda.

Key words:*Labor power migration, Out-migrants, pull factors, Push factors, Return migrants*

1 INTRODUCTION

1.1 Background of the Study

Migration is the term used to describe a person's or a group's movement inside a country or across international borders. International migration is one of the key aspects of globalization that characterizes every part of the globe. It encompasses movement of refugees, displaced people, economic migrants, and people traveling for other reasons, such as family reunion, regardless of its length, content, and causes (Yousaf, 2018).

According to Schewel & Bahir (2019), More than before, there has been a steady increase in the number of migrants internationally, who presently make up 271,642,105 (3.5% of the world's population) (Araya & Meles, 2021). And also Africa is a continent where cross-border illegal migration has become a major phenomenon (McAuliffe & Koser, 2017). According to international migration reports, there were several factors responsible for the increase in illegal migration in Africa. Among them, the issue of conflict, political discourse, bad environmental conditions, absence of human security, and unavailability of job opportunities were the main ones (Carruth & Smith, 2022).

One of the most important migration routes in the world runs from the Horn of Africa to Saudi Arabia is the Eastern Migration Route. It is among the busiest maritime migration routes; reports on 2019movements indicate an average of11,500 people boarding every month , with 63% of the migratory movement tracked in the Horn of Africa region (Fernandez, 2017).

Irregular migration from Ethiopia towards Djibouti and Yemen and onwards to Saudi Arabia is now larger than migration towards Europe (Tinti, 2017). According to IOM, over 150,000 migrants are estimated to have entered Yemen in 2018 and a majority of these migrants are undocumented (Mihai, 2018).The main goal of a migrant to the Middle East is to enhance their standard of living. While traveling via Djibouti and Yemen, two transit nations, migrants run the risk of being exploited, robbed, and sexually abused by human traffickers (Ebrahim, 2016).

At least 400,000 Ethiopians migrated illegally to the Arabian Peninsula between 2017 and early 2020 (Marchand, 2015). According to a report by (MMC),18,320 refugees and migrants entered Yemen in April of 2019; 92% of them were Ethiopians. And over 2,000 migrants mainly from Ethiopia were apprehended by Yemeni security forces during the same month and imprisoned in a soccer stadium in Aden under inhumane conditions. According to the same MMC report, these comprised 1,789 adult men, 389 boys, and 28 girls under the age of 18 as cited in (Corona et al., 2021).

Despite the high risks associated with irregular migration, Ethiopians still see Saudi Arabia as an escape from dire living conditions because employment opportunities are available there. Since November 2020, an average of 6,500 Ethiopians per month–have attempted to migrate to Saudi

Arabia(Carruth & Smith, 2022).

Many of the young Ethiopian migrants traveling the Eastern Route to the Middle East were uninformed of the dangers of the journey, according the IOM's New Study carried on Ethiopian Migrants to the Gulf. These dangers include the increased potential for hunger, dehydration, and gastrointestinal and water-borne illnesses while traveling, as well as possibility of being abused (Latham et al., 2019).

Still, there is disagreement among numerous studies regarding the socioeconomic effects. Some have stated that socio-economic remittances including the change in ideas, technologies, and traditions have helped a country's welfare development. However, others hold that illegal migration has many challenges to countries of origin, transit, and destination; mainly, migrants, as they are victims of exploitation, oppressions, abuse and crime (Araya & Meles, 2021).

As illegal migration affects each and every one of us in a variety of way, finding the socioeconomic aspects of illegal migration is seen to be crucial. Since illegal migration is also a very common practice where the researcher is from, this fact serves as motivation for the researcher to conduct a study on this topic. The investigation was carried out in Raya Kobo Woreda of North Wollo, which is a major source of illegal migrants to the Middle East mainly to Saudi Arabia.

1.2 Statement of the problem

As much as the researchers' assessment, various Studies have been conducted on illegal migration at the international and national level. For example, (Araya & Meles, 2021; and Momo et al., 2019) revealed that rapid population growth, child marriage, and religious similarity were considered as push factors of illegal migration in Ethiopia). However, these reports paid little consideration for socio¬economic related pull factors such as, socioeconomic differences between Ethiopia and Saudi Arabia, personal motivation to increase additional assets, and the existence of family network with country of destination.

Moreover , Yilma (2013) has studied on school girls in Raya Kobo, Amhara Regional State and (Mesfin, 2011) studied on women returnees in Raya Alamata. Here, both researchers neglected to include men illegal migrants. However, based on the new researcher's personal experience indicated that men also experienced illegal migration in the study area. In such a case, the researcher seeks to include both sexes to fill a gender-based knowledge gap that existed in earlier studies.

One of the main factors influencing migration from Ethiopia is a lack of employment opportunities. Natural hazards and armed conflict are two factors that contribute to the lack of employment opportunities. In June 2021, most migrants reported economic reasons as their primary motivation for migration, followed by conflict and family reasons (Alrababah et al., 2021).This further aggravates the movement of people in the search for jobs, education and a better life. Conflict and hunger in the northern regions of Ethiopia are also likely to increase migration patterns (Devaraj et al., 2021).

Although exact number of Ethiopian migrants is not as such easily known, every year close to 120,000 people flee to the gulf states of whom; 60-70% are undocumented illegal migrants (Zeleke & Minaye, 2015).In 2019, of the arrivals into Yemen, around 92% were Ethiopians. The majority of Ethiopian migrants were from the Amhara, Oromia, and Tigray regions as cited (Schewel & Bahir, 2019).

The eastern route through Bosaso is longer, more dangerous and strenuous for migrants. In June 2021,993 migrants, the majority of whom (60%) were Ethiopian, arrived in Yemen from Somalia. This number is a 400% increase compared to May 2021 (King & Kuschminder, 2022). Based on the two consecutive research projects conducted by a RDH to identify communities of high emigration in four regions of Ethiopia: Raya Kobo in Amhara ,Deder and Setema in Oromia, Erer in Harari and Misha in Southern Nations, Nationalities, and People's Region with migrants most prevalent in Raya Kobo (60%) (Naomi et al., 2021).But this RHD project doesn't tell about what factors motivate migrants to leave their residence. Thus, this study inspired to see the determinant factors about this issue in raya kobo.

Furthermore, studies have been investigated risk factors of migration through binary logistic regression (Lambore & Taye, 2017; Teshome et al., 2022). However, binary logistic regression undercounts the total number of migrants per household since multiple migrants are collapsed into a single unit to fulfill the requirements of binary logistic regression. Besides, binary logistic regression may not provide sufficient information for studying the pattern of multiple migrants per house. That means it only predicts whether or not a household had a migrant rather than the number of migrants per household.

Despite the fact that, the number of migrants is a count data with a positively-skewed distribution may not fit well in the logistic regression model. Additionally, to analyze the number of migrants per household by logistic regression model may lead incorrect conclusions, inefficient and the least square estimates of dependent variable suffer from these problems and are biased and inconsistent as well. Therefore, Count regression models are more appropriate than the previous models for studying the number of migrants per household.

1.3 ObjectiveoftheStudy

1.3.1 General Objective

The general objective of this study is to investigate determinant factors associated with migrants to Saudi Arabia, in case of Raya kobo.

1.3.2 Specific Objective of the Study

- To see how the demographic characteristics of the household head affects the migration with the households;
- To identify the motivating pressures that make the society to migrate (pull and push factors);

1.4 Significance of the Study

The finding of this study has importance in producing information to the Ethiopian government about the main drivers of labor power of Migration from Raya kobo Woreda to the Saudi Arabia. Here also the study provides information that would be helpful for decision makers in formulating policies as well as in developing solutions to the issues of illegal migration that predominates in the study area. Finally, it can serve as a spring board for further studies on the problem raised.

1.5 Scope of the Study

This study is limited to looking for some household factors that contribute for illegal migration to Saudi Arabia. The study also tried to investigate both Push factors, that force or initiate youths to migrate from their place of origin to another area, as well as pull factors which are benefits or opportunities that attract people to move from their place of origin into the destination country. Moreover, the study was also restricted to a few chosen kebeles in Raya kobo woreda.

1.6 Ethical Considerations

To form the empirical investigation, the researcher had taken major ethical considerations into account in the whole process of the research undertaking. The analyst respected the rights, needs, and values of participants. Members too had been informed that the information they give for researcher would be used only for the purpose of this study. Moreover, the purpose of the study was also referred in the introductory section of the questionnaire, and the researcher assured participants‟ confidentiality. And their participation was primarily based on their willingness.

2 REVIEWOF LITRTURES

2.1 Definition and Concept of Migration

Migration is an age-old concept of human civilization in different parts of the globe(Rossi, 2018). According to (Buffone et al., 2017), people may migrate for better education and sustainable paid work. He also adds that besides humans other animals like birds also migrate from one place to another in search of food and water but, Animal migration is guided or initiated by seasonal changes and other natural occurrences.

It causes numerous impacts upon the social and economic lives of the migrants as well as their families (Ratha et al., 2011).The phenomenon of migration is characterized by high coordination of human histories, cultures, and civilization. This implies that migration by its nature is not literally a simple phenomenon, rather, it is a strategic, historical, and tradition of a given society (Araya & Meles, 2021).

The concept of illegal is the pre-existing term that shifts attitudes regarding the moral related worthiness of illegal migrants. While in the past, immigration laws were not widely regarded as criminal, those who enter and remain without authorization are increasingly understood as illegal on current basis due to conditions of political and public acceptance in a wide range of measures, on the other hand, illegal is one of the most minimizing one's dignity and given to border crossing migration (Dauvergne, 2018).

Illegal mobility has different challenges to countries of origin, transit, and destination; particularly migrants are vulnerable to discrimination, exploitation, and abuse. Moreover, illegal migrants are also faced to a serious violation of the human rights of its victims (McAuliffe & Khadria, 2019).

2.2 Causes of Migration

Internal and transnational migration is taking place when economic inequalities happened between countries (Massey, 1988). The economic related causes such as poverty, economic crisis, the burden of foreign debt recorded as the main causes that lead to illegal migration. Moreover, people need to move in order to give response to economic and social differences in a progressively integrated and inadequate world (Mulugeta & Makonnen, 1970).

According to Massey (1988) report, Economic motives are the root causes of illegal migration. Illegal migration is caused by the absence of paid work, especially for a job; accordingly, unemployment by men and women is high in many places of sending countries. This indicates that illegal mobility happens due to economically elicited determinants that are assumed to alleviating poverty by migrants' minds.

In Ethiopia, economic-based motives to the Gulf regions report that women domestic workers to the Middle East and estimated that 170-180,000 women deported each year among them, 60-70% are estimated to be illegal migrants and the opening of job opportunities in the receiving country (Araya & Meles, 2021). This is evident in Saudi Arabia and other Gulf countries, where there is a strong structural demand for the kinds of low-skilled labor that Ethiopian men and women undertake reputedly (IOM, 2020).

Push and pull factors are pressures that contributing to illegal migration in countries of origin and destination. For example, lack of employment and low living standards leads to pushing out from the country of origin. The pull factor on the other part happens when the need for economic-related gains in the host. Job opportunities in abroad countries would be important examples (Prieto et al., 2015).

Social causes are roots of migration determinants that taking place at a wider level of the social system. For example, high familial attachments in destination countries and social networks with other migrants in the host contributed to illegal migration. Families and communities related factors are more attached to symbols is termed as the culture of migration. Social determinants can change migration knowledge and information (Immersion, 2017).

Peace Research Institute reports that motivation for illegal migration is caused by sociopolitical factors that induce mobility. For example, oppression, and conflict are basic driving forces of migration. Moreover, socio-political determinants of migration were also elicited through indigenous conflict that initiated violence and individual disorder. Additionally, oppressive political regimes and violation of human rights also cause mobility (Araya & Meles, 2021).

Authoritarian lead regimes are political determinants that are considered as push factors of migration. Mainly, Authoritarian lead implemented countries contribute significantly to the expansion of considerable illegal migration due to political reasons (Immersion, 2017).

2.3 Migration in Ethiopia

Migration has negative consequences. The loss of human resource power is one of the negative consequences. Professionals mainly teachers have been leaving their jobs and migrating to the Arabian Countries. Unexpectedly, students and even young children dream of moving out of the country, just as children commonly dream to be a doctor, an engineer and the like (Admasu, 2018).

Lambore & Taye (2017) tried to investigate socio-demographic and economic characteristics of a household on international migration from southern Ethiopia. It is found that the likelihood of a household of having international migrant increases with head's age and family size. Household characteristics: age, educational level and occupation of head, and family size are determinants of international migration.

The variable occupation of household heads in categories of farmer is not large enough to provide strong evidence to be significant. Whereas according to (Bassie et al., 2022), Occupation of the household head has significant effect on migration. Household heads occupied in agriculture have relatively high number of migrants than household's heads rely on trade. On the other hand the probability to be migrated for people in trade was higher than probability to be migrated for people engaged in agriculture (Teshome et al., 2022).

According to Admasu (2018),Sex of household head, age of household head, land size, educational level, family size, place of residence and occupation were found to be important determinants of migration. His finding also explained that the highest number of migrants per household occurred in the male headed household. Education of the household head significantly affects number of migrants in household. Household with educated household head were fewer number of migrants than uneducated household head.

Yilma (2013) aimed to assess hazards and associated factors among returned migrants in Bati-Woreda, Amhara National Regional State and used cross-sectional survey with a sample size of 390 returnees was made in five kebeles of the woreda using a structured questionnaire. Multiple logistic regression analysis was used to assess the relative importance of associated factors. Socio-economic variables with respect to age, education and unemployment were predictors of the migration phenomenon.

The study of Haile (2015) indicated that lack of job opportunities, family or peer pressure, poverty, unemployment, brokers, and lack of commitment of the local government officials to create jobs are the key push factors of migration. On the other hand, job opportunities, better income, social networks and smugglers at destination country are identified as pull factors of migration.

According to Araya & Meles (2021), presence of family network the need for better quality of life, poverty, pressure from (parents, family, & peers) and better job opportunity in the host country were considered to be fundamental instigators of illegal migration. Moreover, (Hadis, 2017) reported that the religious proximity were the leading factor of illegal migration.

Taking into account the reviewed literature, in this study the most important expected factors that influence migration of people were examined using count regression models. Based on reviewed literatures stated in this section, the explanatory variables which are expected to affect numbers of migrants per household are categorical and continuous.

3 METHODOLOGY

3.1 Description of the Study Area

This study was conducted on the determinant factors associated with the number of illegal migrants from Raya Kobo to the Saudi Arabia. Raya Kobo Woreda is one of the Woreda among the eleven Woredas in the administrative zone of North Wollo, Amhara National Regional State. At present, Raya Kobo Woreda contains a total of 49 kebeles and it is 570 Km far from Addis Ababa (Yilma, 2013).

Based on the Population Projection of Ethiopia at Woreda Level for (2014 – 2017), the Raya Kobo Woreda has a total population of 260,170,with a total of 58,420 households (CSA, 2013). Two consecutive research projects conducted to identify communities of high emigration in four regions of Ethiopia: Raya Kobo from Amhara region ranked as first (Immersion, 2017) .According to IOM , Raya Kobo Woreda is also one of the hotspot regions for illegal migration to Gulf States like Kuwait, Saudi Arabia, and Dubai as cited in (Araya & Meles, 2021).

3.2 Study Population

The term “study population” refers to a target population that includes a list of people in a population from which the investigator can gather the necessary data(Buffone et al., 2017).Therefore for this study, the target population were all hosehold heads in Raya Kobo Woreda, While the study population were household heads from some selected kebeles in the study area.

3.3 Sources of Data

The researcher used primary sources to collect the data through house to house visit. And this study also used secondary data sources from the Woreda officers as well as kebele officers. The investigator had taken the list and the estimated numbers of the households for each kebeles from these officers. Books, websites, and other relevant publications served as secondary data sources for the researcher.

3.4 Methods of Data Collection

A household based cross sectional study design of data collection was held from Aug 06-19/2022 and again from Oct 2-18/2022. Primary data had been collected through structured questionnaire to selected households from the selected kebeles.

3.5 Sampling Technique

The sampling techniques in this study were two-stage cluster sampling techniques. Cluster sampling is a type of sampling method in which we split a population into clusters, then randomly select some of the clusters and include all members from those clusters in the sample. An extension of this is known as two-stage cluster sampling, and Selections of eligible participants will be as follows:

First, divide a population into clusters and then randomly select some of the clusters (choose n clusters). Then, within each chosen cluster, randomly select some of the members to be included in the survey (choose m units from each cluster).Thus, Simple random sampling is used to select units at each stage (Peter & Wright, 2010). The reason that why this study used a two stage cluster sampling is when the sizes of the clusters are large, making it difficult or expensive to observe all the units inside them.

Therefore, in this study the total population of the woreda was split into 49 clusters (kebeles), and then by using simple random sampling some clusters (kebeles) selected. Out of these forty nine kebeles, seven kebeles were randomly selected using simple random sampling and an estimate of the number of households in each Kebele was obtained from the Kebele office. Again a simple random sampling was applied to include the eligible household heads in the sample from the selected kebeles.

3.6 Sample Size Determination

Prior to every survey, a formal sample size should be estimated. In this study, the formula proposed by (Cochran, 1977) was used to determine the sample size required for the study.

$$n = \frac{p(1-p)(z\alpha/2)^2}{e^2} \tag{1}$$

Where $n =$ the size of the sample

$z_{\alpha/2}$= A desired level of the confidence of, 95% was assumed (zα/2 = 1.96)

$P = 0.6$, the proportion of out-migrants for both sexes in raya kobo

$e^2 =$ Margin of error, here, $e^2 = 0.04^2$

$n_h = \frac{0.6(1-0.6)(1.96)^2}{0.04^2} = 577$ belongs to the minimum sample size.

Even if the formula is used for a binary response variable,here researcher used it for a count response with a modification of the parametre 'p' (Admasu, 2018).

Table 3.1: The Sample Size Allocation for the Selected Kebeles

number of hhs for each kebeles	sample size of hhs for each kebels
900	108
750	90
560	68
600	72
840	101
500	60
650	78
sum = 4800	sum =577

Since the size of each kebele varies, a two-stage cluster sampling using probability proportional to the size for the primary unit was used to distribute the given sample size to these selected kebeles (Peter & Wright, 2010). Using the formula:

$$n_h k_i = n_h \,/\, N_h * N_h k_i ...$$

Where,

N_h = total number of household for these selected kebeles

$N_h k_i$ = total number of households in the i^{th} kebele

$n_h k_i$ = the sample size of households for the i^{th} kebele

3.7 Variables of the Study

Based on lessons learned from the available similar studies, the dependent and independent variables were chosen.

3.7.1 The Response Variable

The response variable of this study is a count variable, the number of migrants from each household.

3.7.2 Explanatory Variables/Factors

Some of the common predictors that are expected to influence the numbers of Illegal migrants per households in the study area are given below.

Table 3.2: Description of The Predictor Variables

Variable	Description
1. Age of household head(year)	continous
2. Sex of household head	Female(0) Male(1)
3. Marital Status of household head	Single(0) widowed(2)
	Married(1) divorced(3)
4. The highest level of education attained	Illiterate(0) Primary School (1)
by household head	High School(2)Higher Education(3)
5. household size	Number of household member(discrete)
6. Main economic activity or job of	Merchant(0) Agriculture(1)
household head	Government Employee(2)Other(3)
7. Religion of household head	Orthodox(0) Protestant(1)
	Muslim(2)
8. previous experience of migration for	Migrated(0) not migrated(1)
the household head	
9 .Residence place of household	Urban(0)Rural(1)
10. Home ownership of household	Own(0) By rent(1) Other(2)
11. push factors	Poverty(0) unemployment(1)
	Peer pressure(2) Family pressure(3)
12. Pull factors	High Income In Destination (0)
	Social networks(1) job opportunities(2)
	smugglers(3)

3.8 Methods of Data Analysis (statistical models)

3.8.1 Introduction to Count Regression Model

Count data regression models frequently used in statistics to model response variable which has a form of count. When the response or dependent variable (number of migrants per house in this study) is a count (which can take on non-negative integer values, it is appropriate o use non-linear models based on non-normal distribution to describe the relationship between the dependent variable and a set of predictor variables.

The three appropriate models for count data are divided into the following categories: count model for equal dispersion; Poisson Regression model; count model common for over dispersion includes; Negative Binomial Regression model; Zero-Inflated Count Models for the existence of excess zeroes; Zero-inflated Poisson model and Zero-inflated Negative Binomial model. Previously, linear regression has been used to analyze count variables as continuous variables. However, count data with a positively-skewed distribution may not fit well in the linear regression model (Dauvergne, 2018).

3.8.2 Poisson regression model

Poisson regression model is commonly used to investigate the relationship between the count outcome variable and covariates. Poisson Regression Model provides a standard framework for the analysis of count data. However, its restrictive assumptions often make it inadequate in real-life applications. There are two strong assumptions for Poisson model to be checked: one is the events that occur independently over time or exposure period, while the other is the conditional mean and variance is equal. In practice, counts have greater variance than the mean described as over-dispersion. This indicates Poisson regression is not adequate.

There are two common causes that can lead to over-dispersion: additional variation to the mean or heterogeneity, a negative binomial model is often used and other cause counts with excess zeros or zero-inflated counts, since the excess zeros will give smaller mean than the true value, this can be modeled by using ZIP or ZINB models (Cameron & Trivedi, 1998). Let Y_irepresent counts of eventsoccurring in a given time or exposure periods or area with rateμ_i,Y_i are Poisson random variables with probability mass function (pmf) given below:

$p(y, \mu) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}$, where y_i=0,1,2,3,and $_i$ =1,2,3…where, Y denotes the number of migrants per household, is the rate parameter which is non-negative and it is given as: estimated and p is number of predictors.

The estimation is undertaken by using maximum likelihood method.$E(y_i) = \mu_i = exp(X^T\beta)$ Where $X^T = (1, X_1i^T, X_2i^T....)$ And $\beta = p + 1$ dimensional column vector of unknown parameters to be estimated and p is number of predictors. The estimation is undertaken by using maximum likelihood method. The first two moments of the Poisson random variable Y are $E(Y) = \mu$ and $V(Y) = \mu$.If both are equal that means conditional mean is equal to conditional variance this shows the well-known equ-dispersion (equal mean and variance) property of the Poisson distribution.

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

$$\ell(y, \beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \tag{2}$$

The log-likelihood function of Poisson is $l = Log(L(\beta)) = \sum_{t=1}^{n}[y_i ln(\lambda_i) - (\lambda_i) - ln(y_i!)]$

The likelihood equations for estimating the parameter is obtained by taking the partialderivations of the log-likelihood function and solve them equal to zero. Thus, we obtain the first derivatives of ℓ with respect to the parameters β as follows:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{i=1}^{n} (y_i - \lambda_i)\, x_{ij} \mid \tag{3}$$

1. Limitations of Poisson regression model

Poisson regression model is suitable for modeling count data when successive events occur independently and at the same rate. However, in reality, the features of data often violate these assumptions. Usually, the variance of count data exceeds its mean, leading to over-dispersion. Given that the rate parameter is influenced by both a deterministic function and a random (unobserved) component, this may be the result of unobserved heterogeneity.

Excess zeros, which occur when observed zeros exceed those predicted by the assumed distribution, may also contribute to the over-dispersion. Furthermore, excessive dispersion will lead to deflated parameter estimate standard errors and consequently inflated t-statistics. Thus, it is always necessary to conduct a test of over dispersion after the development of Poisson regression If $E(yi) < var(yi)$ then we speak about over-dispersion, and when $E(yi) > var(yi)$ we say we have under-dispersion. Next, we employed two tests of over dispersion where the Null Hypothesis (H_0) is: mean and variance of the response variable are equal against the Alternative Hypothesis (H_1): Variance exceeds the mean. There are two basic criteria commonly used to check the presence of over dispersion:

1. Deviance,$D(y_i, \hat{\lambda}_i)$,is given by

$$D\left(\left(y_i, \hat{\lambda}_i\right) = 2 * \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{y_i}{\hat{\lambda}_i}\right) - \left(y_i - \hat{\lambda}_i\right)\right\}\right. \tag{4}$$

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

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

$$\mathrm{x}^2 = \sum_{\mathrm{i}=1}^{\mathrm{n}} \left(\frac{\left(\mathrm{y_i} - \hat{\lambda}_\mathrm{i}\right)^2}{\hat{\lambda}_\mathrm{i}} \right) \tag{5}$$

If the model fits the data, both deviance and Pearson Chi-square statistic divided by the degrees of freedom are approximately equal to one. Values greater than one indicate the variance is an over dispersion, while values smaller than one indicate an under-dispersion. This Over-dispersion may be a result of high occurrence of zero counts and subject heterogeneity. This study used these two statistics to determine the more appropriate model between the Poisson Regression and Negative Binomial Regression.

In the general, Poisson Regression Model, we think of λ_i as the expecteddesired number of migrants from i^{th} household and the total number migrants from i^{th} household isN_i. This means, parameter will depend on the population size and the total number of migrants from individual household. Thus, the distribution of Y_i can be written as:

$Y_i \sim poisson(N_i, \mu_i)$

Where N_i are the total migrant of i^{th} household and μ_i is the logarithm of the migrant is introduced in the regression model as an offset variable. By including:

$$Log(\mu_i) = log(\mu_i) + X^T\beta \tag{6}$$

The link between the expectation of the dependent variable and the linear predicator is a logarithmic function and the linear predictor contains a known part or offset. This allows for estimation of maximum likelihood, standard errors and the likelihood ratio goodness of fit chisquare statistics (Agresti, 2007). The model suggests that both set of the parameters are dependent on the covariates.

Since $Log(N_i)$ is a constant, any variation in the coefficients of the independent variables will show up affecting the dependent variable and not the number of household.The procedure therefore allows us to obtain the maximum likelihood regression coefficients that can be easily interpreted in terms of differentials in the dependent variables.

3.8.3 Negative Binomial Regression Model

The NB Regression Model is used when count data are over dispersed (i.e when the variance exceeds the mean). Over dispersion, caused by heterogeneity or an excess number of zeros (or both) to some degree is inherent to most Poisson data. By introducing a random component into the conditional mean, the Negative Binomial Regression Model addresses the issue of over-dispersion. However, it equally models both zero and nonzero counts, which might result in a poor fit for data with excessive number of zeros. Therefore, it is always necessary to check the proportion of zero counts before developing a Negative Binomial Regression Model.

The NB regression model is:

A random variable $y_i, i = 1, 2, ...$, is called a negative binomial distributed count with parameter λ and α the probability density function $(\alpha > 0)$

$$P\left(Y_i = y_i, \lambda_i, \alpha\right) = \frac{\mathrm{r}\left(y_i + \left(\frac{1}{\alpha}\right)\right)}{y_i\,\mathrm{lr}\left(\frac{1}{\alpha}\right)}\left(1 + \alpha\lambda_i\right)^{-1/\alpha}\left(1 + 1/\alpha\lambda_i\right)^{-y_i} \tag{7}$$

Then

$E(y_i) = \lambda_i = exp^(X_i)^T\beta$ And $var(y_i) = \mu_i(1+\alpha\mu_i)$, Where, shows the level of over-dispersion and (.) is the gamma function. If =0 NB Regression Model will reduce to Poisson Regression Model. Often data will show over-dispersion (Variance > mean) or under-dispersion (Variance< mean). With over-dispersed data we may well use the Negative Binomial Regression model. This Model adds unobserved heterogeneity by specifying $E\left(y_i\right) = \lambda_i = \exp^{X_i^T}\beta$ Where ${X_i}^T$ is 1xp row vector of covariate (including an intercepts), p is the number of covariate in the model and px1 column vector of unknown regression parameters.Thelikelihood function of the NB model based on a sample of n independent observations is given by:

$$L\left(Y_i = y_i, \lambda_i, \alpha\right) = \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \left(\frac{1}{\alpha}\right)\right)}{y_i!\Gamma\left(\frac{1}{\alpha}\right)} \left(1+\alpha\lambda_i\right)^{-1/\alpha} \left(1+1/\alpha\lambda_i\right)^{-y_i} \tag{8}$$

3.9 Zero-inflated Regression Models

In some cases, excess zeros exist in count data and considered as a result of over dispersion. In such a case, the NB model cannot be used to handle the over-dispersion which is due to the high amount of zeros. To do this, zero-inflation (ZI) can be alternatively used. Real-life count data are frequently characterized by over-dispersion and excess zero (Gurmu & Trivedi, 1996). Zero inflated count models provide powerful way to model this type of situation. Such models assume that the data are a mixture of two separate data generation processes: one generates only zeros, and the other is either a Poisson or negative binomial data-generating process. ZIP) and ZINB regression are frequently used to model zero-inflated count data.

3.9.1 Zero-inflated Poisson Regression Model

Suppose the mean of the underlying Poisson distribution is and the probability of an observation being drawn from the constant distribution that always generates zeros is ω_i. The parameter ω_i is often called the zero inflation probability (Agresti, 2007). In ZIP regression, the responses $Yi = (Y1, Y2, ..., Yn)$ are independent distributed. One assumption of this model is that with probability p the only possible observation is 0, and with probability $(1-p)$, a Poisson (λ) random variable is observed in Y. In order to explain the occurrence of extra zeros in the variable Yi , The ZIP regression model is given by (Dauvergne, 2018),

$$P(yi) = \begin{cases} \omega i(1+\omega i)e^{-\lambda i}, & yi = 0 \\ (1-\omega i)\frac{e^{\lambda i}\lambda i^{yi}}{yi!}, & yi = 1, 2\ldots \end{cases} \quad 0 \leq \omega i \leq 1\ldots \tag{9}$$

Where,$\sim ZIP(\lambda i, \omega i)$

The mean and variance of Zero-inflated (ZIP) distribution is given as

$(Yi) = (1-\omega i)$ and $Var(Yi) = (1-\omega i)\lambda i(1+\omega i\lambda i)$

The excess zeros are a form of over-dispersion and fitting a zero inflated Poisson model can account for the excess zeros, but there are also other sources of over-dispersion that must be considered. If there are sources of over-dispersion that cannot be attributed to the excess zeros, failure to account for them constitutes a model misspecification, which results in biased standard errors. In ZIP Models, the underlying Poisson distribution for the first subpopulation is assumed to have a variance that is equal to the distribution‘s mean. If this is an invalid assumption, the data exhibit over-dispersion (or under-dispersion). (Agresti, 2007) A useful diagnostic tool that can aid in detecting over dispersion is the Pearson chi-square statistic defined as

$$x^2 = \sum_{i=0}^{n} \frac{(y_i - \mu_i)^2}{V(\mu_i)} \tag{10}$$

Comparing the computed Pearson chi-square statistic to an appropriate chi-squared distribution with n-p df constitutes a test of over-dispersion. If over dispersion is detected, the ZINB Model often provides an adequate alternative.

The parameter λ_i and ω_i can be obtained by using the link functions,

w$Log(\lambda_i) = {x_i}^T \beta$ and $(\frac{\omega_i}{(1-\omega_i)}) = {z_i}^T \gamma, i = 1, 2, ...$

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

$$\ell(\lambda, y) = \sum_{i=1}^{n} \left\{ \begin{array}{c} \ln\left(\left[\omega_i + (1-\omega_i) e^{-\lambda_i}\right] I_{(y_i=0)} \right. \\ + \left[\ln(1-\omega_i) - \lambda_i + y_i \ln(\lambda_i) - \ln y_i!\right]_{(y_i>0)} \end{array} \right\} \tag{11}$$

Where, I (.) is the indicator function for the specified event, i.e. equal to 1 if the event is true and 0 otherwise. To obtain the parameter estimates of ZIP regression models,(β) and γ,the Newton-Raphson method can be used. The first derivatives with respect to β and γ are

$$\begin{aligned} \frac{\partial \ell}{\partial \beta_j} &= \frac{\partial \ell \partial \lambda_i}{\partial \lambda_i \partial \beta_j} = \sum_{i=1}^{n} \left\{ I_{(y_i=0)} \left[\frac{-(1-\omega_i)\lambda_i e^{-\lambda_i}}{\omega_i + (1-\omega_i) e^{-\lambda_i}} \right] + (y_i > 0)(y_i - \lambda_i) \right\}, \mathrm{j} = 0, 1, 2 \ldots, \mathrm{p} \\ \frac{\partial \ell}{\partial \gamma r} &= \frac{\partial \ell \partial \omega_i}{\partial \omega \partial \gamma r_i} = \left\{ (y_i = 0) \left[\frac{-(1-\omega_i)\lambda_i e^{-\lambda_i}}{\omega_i + (1-\omega_i) e^{-\lambda_i}} \right] - (y_i > 0) \left[\frac{1}{(1-\omega_i)} \right] \right\} z_{ir}, \mathrm{r} = 0, 1, 2 \ldots, \mathrm{q} \end{aligned} \tag{12}$$

3.9.2 Zero-inflated negative binomial regression model

Zero-inflated negative binomial (ZINB) regressions have been used by researchers for handling both zero-inflation and over-dispersion in count data. This model provides a way of modeling the excess number of zeros (with respect to a Poisson distribution or negative binomial distribution) in addition to allow for count data that are skewed and over dispersed (Dauvergne, 2018). The ZINB distribution is a mixture distribution, similar to ZIP distribution, where the probability ‘p’ for excess zeros and with probability (1- p) the rest of the counts followed negative binomial distribution. Note that the negative binomial distribution is a mixture of Poisson distributions, which allows the Poisson, mean λ to be distributed as Gamma, and in this wayover dispersion is modeled (Johnson, 2012).

The ZINB regression model is given by

$$P(yi \mid \omega i, \alpha, \lambda) = \begin{cases} \omega i + (1-\omega i)(1+\alpha\lambda i)^{-\frac{1}{\alpha}}, & yi = 0 \\ (1-\omega i)\frac{\mathrm{r}(yi+(\frac{1}{\alpha}))}{yi!r(\frac{1}{\alpha})}(1+\alpha\lambda i)^{-\frac{1}{\alpha}}\left(1+\frac{1}{\alpha\lambda i}\right)^{-yi}, yi > 0 \end{cases} \tag{13}$$

Where, λi is the mean of the underlying negative binomial distribution, $\alpha > 0$ is the over dispersion parameter and is assumed not to depend on covariates and $0 \leq \omega i \leq 1$. Also the parameters λi and ωi depend on vectors of covariates xi and zi, respectively. The formulations for λi and ωi are the same as those used in the zero-inflated Poisson regression model. In this case, the mean and variance of the Y_i are

$E(Yi) = (1–\omega i\lambda i)$ and $Var(Yi) = (1–\omega i\lambda i(1 + \omega i\lambda i + \alpha\lambda i$

ZINB approaches ZIP and NB as $\alpha = 0$ and $\omega i = 0$, respectively. If both α and $\omega i = 0$, then ZINB reduces to Poisson. The parameter λi is modeled as a function of a linear predictor, that is,

$$\lambda_i = exp(xi^T\beta) \tag{14}$$

Where, β is the $(p+1) \times 1$ vector of unknown parameters associated with the known covariate vector $xi^T = (1, xi1, ..., xip)$,p is the number of covariates not including the intercept. The parameter ωi, which is often referred as the zero-inflation factor is the probability of zero counts from the binary process. For common choice and simplicity, ωi is characterized in terms of a logistic regression model by writing as

$logit(\omega i) = log(\omega i/(1-\omega i)) = Z_i^T\gamma$ γWhere, γ is the $(q + 1) \times 1$ vector of zero-inflated coefficients to be estimated, associated with the known zero-inflation covariate vector $Z_i^T = (1, Z_i1, ..., z_iq)$, where $'q'$ is the number of the covariates without including the intercept. ZINB is also used to analyze exploratory data. When all the covariates are included in the log link model, as in the case of ZIP, the estimate of the inflated parameter was found to be zero.

The log likelihood function $\ell = \ell(\alpha, \lambda i, \omega i; y)$, for the ZINB model is given below

$$\begin{aligned}\ell &= \ell(\alpha, \lambda\mathrm{i}, \omega\mathrm{i}; \mathrm{y}) \\ &= \sum_{I=1}^{n} \left\{ I_{(y_i=0)} \log \left(\omega i + (1-\omega i) + (1+\alpha\lambda_i \right)^{-1/\lambda} \\ &+ I_{(y_i>0)} \log(1-\omega i) \frac{\Gamma\left(y_i + \frac{1}{\alpha}\right)}{y_i! \Gamma\left(\frac{1}{\alpha}\right)} \left((1+\alpha\lambda_i) - \frac{1}{\alpha}\left(1 + \frac{1}{\alpha\lambda i}\right)^{-yi} \right] \right\}\end{aligned} \tag{15}$$

since $\frac{\gamma(yi+\frac{1}{\alpha})}{yi\,!(\frac{1}{2})} = \prod_{k=1}^{yi}\left(y_i + \frac{1}{\alpha} - k\right) = \alpha^{-yi}\prod_{k=1}^{yi}(\alpha y_i - \alpha k + k + 1)$

Furthermore, it can be written as

$$\begin{aligned}\ell = \sum_{I=1}^{n} \Big\{ I_{(y_i=0)} \log \Big(\omega i + (1-\omega i) + (1+\alpha\lambda_i \Big)^{-1/\lambda} + I_{(yi>0)} \log(1-\omega i) - \log(\mathrm{yi!}) + \\ \sum_{k=1}^{n} (\alpha y_i - \alpha k + k + 1) - \left(y_i + \frac{1}{\alpha} \right) \log\left(1 + \alpha\mu_i\right) + \log yi! + y_i \log \mu_i \Big\}\end{aligned} \tag{16}$$

3.10 Hurdle regression models

A hurdle model is a class of statistical models where a random variable is modeled using two parts, the first which is the probability of attaining value 0, and the second part models the probability of the non-zero values. It was introduced by (Cameron & Trivedi, 1998), where the non-zero values of x was modeled using a normal model, and a probit model was used to model the zeros. The probit part of the model was said to model the Presence of “hurdles” that must be overcome for the values of x to attain non-zero values, hence the designation hurdle model.

The Hurdle model may be defined as a two part model where the first part is binary outcome model, and the second part is a truncated count model. As per (Cameron & Trivedi, 1998) such as a partition permits the interpretation that positive observation arise from crossing the zero hurdles or the zero thresholds. The zero value has special appeal because in many situations it portions the population in to subpopulations in a meaningful way. So, a data set is split in to zero and non-zero (positive) values to fit two different models with associated covariates in regression.

Zero-inflated models and hurdle models provide a way of modeling the excessive proportion of zero values and allow for over dispersion. Especially when there is a large number of zeros, these techniques are much better able to provide a good fit than Poisson or negative binomial models (Dauvergne, 2018). Suppose that (0) is the probability value when the value for response variable is zero and that (k), k = 1, 2 … is a probability function when the response variable is a positive integer. Therefore, the probability function of the hurdle-at-zero model is given by:

$$\mathrm{P}\left(Y_i = \mathrm{k}\right) = \begin{cases} g_1(0) \text{ for k} = 0 \\ \left(1 - g_1(0)\right) g_2(\mathrm{k}) \text{ for } k = 0, 1 \end{cases} \tag{17}$$

(Bedrick & Hossain, 2013) discussed the hurdle-at-zero model and he assumes that both parts of the hurdle model are based on probability functions for non-negative integers such as f_1 and f_2 . In terms of the general model above,

Let $g_1(0) = f_1(0)$ and $g_2(k) = (\frac{f_2(k)}{((1-f_2(0))})$. In the case of g_2, normalization is required because f_2 is defined over the non-negative integers $(k = 0, 1, 2...)$ whereas the support of g_2 must be over the positive integers $(k = 1, 2...)$. This means that we need to truncate the probability function f_2. However, this is a theoretical concept, i.e., truncation on f_2 does not mean that there is truncation of the population here. All we need to do is to work with a distribution with positive support, and the second part of hurdle model can use a displaced distribution or any distribution with positive support as well. Under the (Dauvergne, 2018) assumptions, the probability distribution of the hurdle at zero models is given by

$f(y = 0) = f_1(0)$

$f(y = k) = (f_1(0))/((1 - f_2(0)))f_2(k) = \theta f_2(k), k = 1, 2...$

Where f_2 is referred to as percent-process. The numerator of θ presents the probability of crossing the hurdle and the denominator gives normalization that accounts for the (purely technical) truncation of f_2. It follows that if $f_1 = f_2$ or, equivalently, $\theta = 1$ then the hurdle model collapses to the parent model. The expected value of the hurdle model is given by

$$E(Y) = \sum_{k=1}^{\infty} k f_2 \tag{18}$$

The difference between this expected this expected value and the expected value of the parent model is the factorθ. In addition, the variance value of the hurdle model is given by $Var(Y) = \theta \sum_{k=1}^{\infty} k^2 f_2(k) - \left[\theta \sum_{k=1}^{\infty} k f_2(k)\right]^2$

If θ exceeds 1, it means that the probability of crossing the hurdle is greater than the sum of the probabilities of positive outcome in the percent model. Therefore, increasing the expected value of the hurdle model is related to the expected value of the percent model. Alternatively, if θ is less than 1 (that is the usual case in an application with excess zero), it means that the probability of notcrossing the hurdle is greater than the probability of a zero in the parent model. Therefore, this model gives a new explanation of excess zeros as being a characteristic of the mean function rather than a characteristic of the mean function rather than a characteristic of the variance function.

3.10.1 Hurdle Poisson regression model

The hurdle model is flexible and can handle both under- and over dispersion problems. A generalized hurdle model is introduced by (Gurmu, 1998)for the analysis of over dispersed or under dispersed count data. (Dauvergne, 2018) has discussed the comparison between hurdle and zero-inflated models as two part-models. A Poisson model typically is assumed for count data. In many cases because of many zeros in the response variable, themean is not equal to the variance value of the dependent variable.

Therefore, the Poisson model is no longer suitable for suitable for this kind of data. Thus, we suggest using a hurdle Poisson regression model to overcome the problem of over dispersion.We start with the binomial process, which determines whether the dependent variable takes on the value zero or a positive value. The probability mass function is

$$\mathrm{P(Y=y)} = \begin{cases} \pi, \quad y=0 \\ 1-\pi \quad y=1,2,3,... \end{cases} \tag{19}$$

The zero-truncated Poisson process has probability mass function

$$\mathrm{P(Y=y)} = \begin{cases} \pi, \quad y=0 \\ (1-\pi)\frac{\lambda^y}{(e^{\lambda-1})y!}, \quad y=1,2,3,... \end{cases} \tag{20}$$

And the log likelihood for the i^{th} observation, assuming the observations are independent and identically distributed, is

$$\mathrm{Ln\,L}\left(\pi_i,\lambda_i,y_i\right) == \begin{cases} \ln(\pi), \quad y=0 \\ \ln\left\{(1-\pi)\frac{\lambda^y}{(e^{\lambda}-1)y!}\right\}, \quad y=1,2,3,... \end{cases} \tag{21}$$

3.10.2 Hurdle negative binomial regression model

In many case because of many zeros in the response variable, the mean is not equal to the variance value of the dependent variable. In this case we suggest using a hurdle negative binomial regression model to overcome the problem of over dispersion. Let $Y_i(i=1,2,...n)$ be anon negative integer-valued random variable and suppose $Y_i=0$ is observed with a frequency significantly higher than can be the usual model. We consider a hurdle negative binomial regression model in which the response variable $Y_i(i=1,2,...n)$ has the distribution

$$\mathrm{P}\left(Y_i=\mathrm{k}\right) = \begin{cases} g_1(0) \text{ for k} = 0 \\ \left(1-g_1(0)\right)g_2(\mathrm{k}) \text{ for } k=0,1 \end{cases} \tag{22}$$

3.11 Estimation Techniques

Thecount regression models are a nonlinear regression models.The maximum likelihood estimate of a parameter is the parameter value for which the probability of the observed data takes its greatest value.However, Count data models are not standard GLM type exponential family distribution and standard GLM fitting methods are not applied. To obtain the parameter estimates of these regression models the Newton-Raphson method can be used.The likelihood equations for estimating the parameter of these models is obtained by taking the partialderivations of the log-likelihood functions for each model and solve them equal to zero.

3.12 Methods of variable selection

A variable selection method is a way of selecting a particular set of independent variables for use in a regression model. It is intended to select the best subset of predictors. This selection might be an attempt to find a best model, or it might be an attempt to limit the number of independent variables when there are too many potential independent variables.

The commonly used variable selection methods are forward selection, which is used to provide an initial screening of the candidate variables when a large group of variables exists. It adds the most significant variable and Stop adding variables when none of the remaining variables are significant. The other selection method is backward selection; model starts with all candidate variables in the model and at each step, the variable that is the last significant is removed. This process continues until no non-significant variables remain. This method is less popular because it begins with a model in which all candidate variables have been included.

There is also a combination of these two selection methods called Stepwise selection method. It was very popular at one time, Stepwise regression is a modification of the forward selection so that after each step in which a variable was added, and all candidate variables in the model are checked to see if their significance has been reduced below the specified tolerance level. If non-significant variable is found, it is removed from the model. Stepwise regression requires two significance levels: one for adding variables and one for removing variables. The cutoff probability for adding variables should be less than the cutoff probability for removing variables so that the procedure does not get in to infinite loop.

This study used Stepwise variable selection method to identify the predictors in the model. This was done on Poisson regression model as it is the bench mark for other count regression models.

3.13 Model Fitting Test (Goodness of Fit Tests)

There are different count regression models to be compared in order to select the appropriate fitted model, which fits the data well.This was done using likelihood-ratio test (LRT),Akakie information criteria (AIC) and Bayesian information criteria (BIC).A model with the smallest

AIC,BIC and the largest LRT value fits the given data well. (AIC) is the most common means of identifying the model which fits well by comparing two or more than two models. The formula is given as:

$$AIC = -2\ell + 2k$$

, Where ℓ is 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. Bayesian information criteria (BIC), takes in to account the size of the data under consideration.

$BIC = -2\ell + klog(n)$, where ℓ, is 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. The comparison will start from the model without anyindependent variable with the model with adding the independent variable one by one through the full model. The model which has the minimum value of AIC and BIC is the most appropriate fitted model to the dataset.

3.14 Tests for the Comparison of the Models

3.14.1 Tests for comparison of nested models

Likelihood Ratio test

The likelihood-ratio test is used to assess the adequacy of two or more than two nested models. It compares the maximized log-likelihood value of the full model and reduced model. For instance, the null hypothesis can be stated as the over dispersion parameter is equal to zero (i.e. the Poisson model can be fitted well the data) versus the alternative hypothesis can be stated as the over dispersion parameter is different from zero (i.e. the data would be better fitted by the negative binomial regression).The likelihood-ratio test is given by:

Likelihood-ratio test: $G2 = -2(\ell nul - \ell k) \sim \chi^2_{p-1}$

$\ell null$ is the log-likelihood of the null model and ℓk is the log-likelihood of the full model comprising k predictors,is number of parameter and $\chi^2(p-1)$ is a chi-square distribution with p -1 degree of freedom. If the test statistic exceeds the critical value, the null hypothesis is rejected. That means the overall model is significant. In this study, to compare different modern Count regression models,Count regression models, we use dsignificance of dispersion parameter and likelihood ratio (LRT) testas criterions. The statistic of likelihood ratiotest for α is given by the following equation

$$LRT\alpha = -2(LL1–LL2)$$

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

3.14.2 Test for comparison of Non-nested models

Vuong Test

The Vuong test is anon-nested test that is based on a comparison of the predicted probabilities of two models that do not nest (Schneider et al., 2018) that means In order to compare the adequacy of zero-inflated models to the standard count models, vuong test statistics are required. For example, comparisons between zero-inflated count models with ordinary Poisson, or zero-inflated negative binomial against ordinary negative binomial model can be done using Vuong test. This test is used for model comparison. For testing the relevance of using zero-inflated models versus Poisson and NB regression models, the Vuong statistic is used. Let's define:

$$Mi = log(\frac{(p1Yi/Xi)}{(p2Yi/Xi)})$$

Where, $P1(Yi/Xi)$ and $P2(Yi/Xi)$ are probability mass functions of zero-inflated and Poisson or NB models, respectively. In general,(Yi/Xi) is the predicted probabilities of observed count for case from i model N, then the Vuong test statistic is simply the average log-likelihood ratio suitably normalized. The test statistic is

$$V = \sqrt{n}\frac{(\sum_{i=1}^{n} mi)/n)}{((\sum_{i=1}^{n}(mi - m)^2)}$$

Where, mean of mi,sm standard deviation and samle size. The hypotheses of the Vuong test are: $Ho : [mi] = 0$ vs $H1 : [mi] \neq 0$

The null hypothesis of the test is that the two models are equivalent. Vuong showed that asymptotically, has a standard normal distribution (Schneider et al., 2018). If $V > Z\alpha/2$, the first model is preferred. If $V < -Z\alpha/2$, the second model is preferred. If $|V| < Z\alpha/2$, none of the models are preferred

3.15 Statistical software packages

In this study, Statistical Package for Social Science (SPSS), Excel, South Texas Art Therapy Association (STATA), and R were used to conduct the statistical analyses.

4 RESULTS AND DISCUSSION

This chapter discusses the results of the study showing how selected factors affect number of migrants to Saudi Arabia among households in some selected kebeles of raya kobo Woreda. The Poisson, negative binomial, the zero inflated and hurdle models were used to analyze the data.

4.1 Descriptive Statistics

To obtain an overall picture of the distribution of the number of migrant in the house hold to Saudi Arabia, the following descriptive analyses were performed.

Table 4.1: Summary statistics of predictor variables related to number of migrants per household

	At least one migrant per household			
variables	catagory	mean	observation	sdv
Sex	Male	1.17	483	1.549
	Female	1.30	94	1.487
Marital status	Single	0.35	31	0.839
	Married	0.92	389	1.443
	Widowed	1.89	82	1.440
	Divorced	2.20	75	1.644
Education level	Illiterate	1.64	212	1.645
	Primary	1.34	198	1.601
	Secondary	0.57	110	1.079
	Higher	0.19	57	0.515
Economic activity	Agriculture	1.54	350	1.628
	Merchant	0.92	127	1.355
	Government	0.23	62	0.777
	Other	0.50	38	1.084
Religion	Orthodox	1.26	472	1.571
	protestant	0.88	76	1.433
	Muslim	0.83	29	1.104
Residence	Urban	1.03	307	1.483
	Rural	1.38	270	1.582
Home ownership	Own	1.30	398	1.579
	By rent	1.13	124	1.556
	Other	0.55	55	0.939
Migration status	Not migrated	1.47	461	1.590
	Migrated	0.09	116	0.449
Push factors	Poverty	1.21	108	1.529
	Unemployment	1.04	158	1.427
	Family pressure	1.69	35	1.745
	Peer pressure	1.01	134	1.492
	Large family size	1.39	142	1.624
Pull factors	High income in Saudi Arabia	1.30	82	1.638
	Social networks	1.14	185	1.519
	Job opportunity	1.42	192	1.610
	Smugglers	0.82	118	1.305

Table 4.1 showed that the highest mean number of migrants was occurred in households headed by females (1.3) than households headed by male(1.17) . The table also showed that the average number of migrants within the divorced household heads(2.2) were higher than others. Moreover, the mean number of migrants for illiterate household heads (1.64) is higher than primary and above educated household heads.

As shown in the above table, the highest mean number of migrants was occurred in household heads engaged on agriculture (1.54) whereas the lowest was in government employee household heads (1.75).And also households in which the ownership of the house is own, as well as by rent were exposed for more number of migrants than other means of the ownership. Table 4.1 also showed that the mean number of migrants per household from rural area (1.38) had higher than households from urban area (1.03).

On the other hand the mean number migrants were higher in households whose household heads did not migrated (1.47) than households heads experianced migration. The number of migrants per household was also affected by the push and pull factors. Family pressure dominated the categories of push factors (1.69), whereas job opportunity in the destination (1.42) contributed the most migrants on average per household to the pull factor categories.

Table 4.2: Frequency Distribution for the Number of Migrants in Raya Kobo

Num.of migrants per hh	Freq.	Percent
0	294	50.95
1	94	16.29
2	71	12.31
3	49	8.49
4	40	6.93
5	29	5.03
Total	577	100.00

The result of table 4.2 showed that among the 577 households, 294 (50.95 %) never have any migrants while 283(49.05%) of the households experienced at least one migrants.

Table 4.3: Summary statistics of dependent variable

Variable	Obs	Mean	Variance
Total migrants	577	1.192	2.367

Table 4.3 shows the descriptive statistics of the dependent variable. Here, the mean number of migrants per household is 1.192, whereas the variance of total migrants of household is 2.367.

Since The variance of the outcome variable (total number of migrant) for entire data, 2.367 is greater than the mean 1.192, and the ratio $\frac{2.367}{1.192} = 1.986 > 1$, the result suggesting over dispersion in the data set.

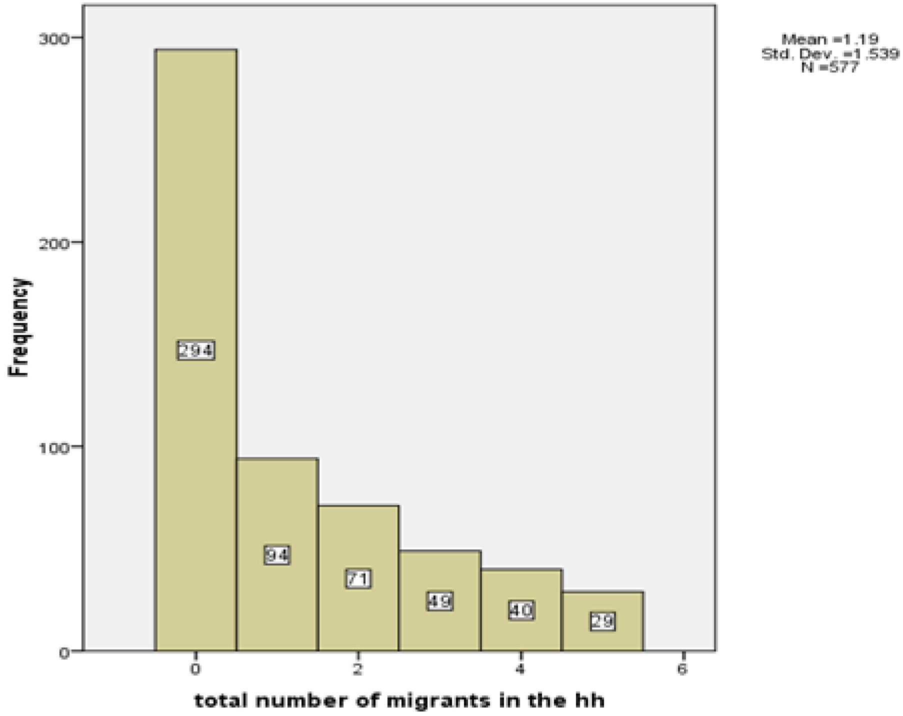

Figure 4.1: Histogram for the Number of Migrants in Raya Kobo

As Figure 4.1, the outcome variable contains many counts of zero, and the histograms are heavily biased toward zero values at the beginning (zero values). However large observations are less frequently observed. This leads to have a positively (or right) skewed distribution. This is an indication that the data could be fitted better by count data models which takes into account excess zeros models.

4.2 Count Regression Model Results

4.2.1 Variable selection

This study used Stepwise variable selection method before fitting the Poisson regression model as it is the bench mark for other count regression models. Stepwise selection method addresses where variables were added or removed with respect to the p-value in the process. Based on stepwise variable selection procedure; sex of household head, age of household head, marital status of household head, education level of household head, economic activity of the household head, religion of the household head, place of residence , home ownership, Migration status of the household head and pull factors were included in the model. Further analyses were made only on those variables which are found significant factors for the number of migrants per household.

4.2.2 Formal Test of Over-Dispersion after fitting Poisson regression model

Table 4.4: Test of over-dispersion

Statistics	value	df	value/df	p-value
Deviance test statistics	699.53	556	1.26	0
Pearson Chi-square statistic	821.27	556	1.48	0

The result of the above table shows the values of Deviance and Pearson Chi-square statistic were greater than one with their p-value less than 0.05; then the null hypothesis that there is equi-dispersion is rejected, and conclude that there is significant over-dispersion in the dataset. Therefore the Negative Binomial regression model is preferred over the Poisson model.

4.3 Model Fitting Test and Model Comparison for nested models

4.3.1 Information Criteria's and LRT Values

In data analysis, selecting the right models for a given study is a crucial issue. Several criteria can be used to compare and select among considered models. In this study, model selection criteria: the Log likelihood, Akaike information criterion (AIC) and Bayesian information criterion (BIC) were used to identify the most appropriate fitted model. The Model with the smallest AIC, BIC and the largest Likelihood values is the most appropriate model for the data.

Table 4.5: Model Selection Criteria for the Count Regression Models

	Models					
criteria	Poisson	NB	ZIP	ZINB	HP	HNB
AIC	1632.186	2423.909	1591.875	1423.865	1471.177	1410.335
BIC	1723.701	2519.782	1774.904	1611.252	1654.206	1597.722
Likelihood	-795.090	-719.090	-753.900	-668.900	-693.600	-662.200

As shown in the above table, the HNB has the smallest AIC, smallest BIC and maximum likelihood among these six models. So the HNB is the most suitable and preferred for describing the number of migrants per household.

4.3.2 Vuong Test for Non-Nested Models

The Vuong test is a non-nested test that compares the probabilities predicted by two independent models. That means vuong test statistics were used to provide the appropriateness of zero-inflated and hurdle models against the standard count models. (Schneider et al., 2018).

Table 4.6: Vuong non-nested tests results

Model comparison	Vuong Test Statistic	P-value	Preferable Modes
Poisson vs. ZIP	-3.6238	1e-04	ZIP Preferable
NB vs. ZINB	-5.8961	0e+00	ZINB Preferable
HP vs. ZIP	3.0694	1e-03	HP Preferable
HNB vs. ZINB	3.0661	2e-03	HNB Preferable

Among the preferable models in the vuong test the HNB model was supported by the likelihood, AIC and BIC to fit the data, thus the HNB model was selected.

Table 4.7: Observed and predicted probabilities of number of migrants per household

		Predicted probabilities					
no.mig	Obs. pr	Poisson	NB	ZIP	ZINB	HP	HNB
0	0.5095	0.6243	0.4360	0.6370	0.5132	0.5095	0.5095
1	0.1629	0.1724	0.2345	0.1565	0.1558	0.3112	0.2838
2	0.1231	0.0715	0.1429	0.0703	0.1357	0.0788	0.0842
3	0.0849	0.0362	0.0835	0.0369	0.0948	0.0315	0.0380
4	0.0693	0.0213	0.0468	0.0222	0.0547	0.0169	0.0210
5	0.0503	0.0145	0.0257	0.0152	0.0270	0.0112	0.0132

Since the predicted probability for HNB model is closed to the observed probability, the HNB model is a better choice than the other count models. Therefore, we concluded that the HNB model is more appropriate than the other count model to fit the number of migrants per household.

4.4 Parameter Estimation of HNB Model

After fitting the HNB Model,the estimated coefficients can be interpreted as : for every one unit increase of the significant predictors, the log number of response variable is expected to increase or decrease by approximately the corresponding coefficient.

On the other hand, in count data we use the incidence rate ratios $(IRR = exp^((coef.)))$ to interpret the coefficient directly (Johnson, 2012). This is important to explain the change in percentage (IRR-1) of significant predictors.

As table 4.8, the variables Sex of the household head, age of the household head, education level, religion of the household head and residence place of the household were the significant factors of migrants per household.

Table 4.8: The Estimated Zero Truncated (Count) Part of HNB Model

						CI of IRR	
Variables	Estimate	Std. Error	z value	Pr(>|z|)	IRR	Lower	Upper
Intercept	-3.8259	0.6634	-5.767	0.0000	0.0218	0.0059	0.0800
sex							
female	1.1606	0.2285	5.080	0.0000	3.1917	2.0396	4.9945
Age	-0.0221	0.0075	-2.944	0.0032	0.9781	0.9638	0.9926
Marital status							
Married	-0.8895	0.5718	-1.556	0.1198	0.4109	0.1339	1.2602
widowed	-0.7643	0.5909	-1.293	0.1959	0.4657	0.1463	1.4827
Divorced	-0.3893	0.5858	-0.665	0.5064	0.6775	0.2149	2.1360
Education							
Primary (1-8)	-0.4364	0.1745	-2.501	0.0124	0.6463	0.4591	0.9100
Sendary (9-12)	-0.6390	0.1696	-3.767	0.0002	0.5278	0.3306	0.7982
Higher	-0.8420	0.4037	-2.086	0.0370	0.4308	0.0318	0.6322
Economic activity							
Merchant	-0.4419	0.2516	-1.756	0.0790	0.6428	0.3926	1.0525
gov Employee	0.1822	0.8097	0.225	0.8220	1.1998	0.2454	5.8653
other	-0.1676	0.4449	-0.377	0.7065	0.8457	0.3536	2.0228
Religion							
Protestant	0.4675	0.2669	1.751	0.0799	1.5960	0.9458	2.6930
Muslim	1.0235	0.3938	2.599	0.0094	2.7829	1.2861	6.0215
Residence							
Rural	0.4871	0.1605	3.035	0.0024	1.6276	1.1883	2.2292
Home ownership							
By rent	-0.0492	0.2182	-0.226	0.8216	0.9520	0.6208	1.4600
Other	-0.8732	0.4713	-1.853	0.0639	0.4176	0.1658	1.0518
migration experience							
Migrated	-0.9189	0.7504	-1.225	0.2208	0.3990	0.0917	1.7365
Pull factors							
Social networks	-0.0205	0.2322	-0.088	0.9298	0.9797	0.6215	1.5444
Job opportunity	-0.3095	0.2222	-1.393	0.1636	0.7338	0.4747	1.1342
Smugglers	0.1346	0.2750	0.489	0.6246	1.1441	0.6674	1.9612

4.4.1 Interpretation of the HNB Model Results for Zero Truncated Part

Number of migrants in the household does differ by sex of household head. The coefficient for female is positive and statistically significant. Female headed households are expected to have about 219 % more number of migrants than male headed household, holding all other variables in the model constant.

However ,the estimated coefficient for age of household head is negative and statistically significant. This implying that, forone unit increment of house hold's age, the log odds number of migrants per households is decreased by 0.0221, holding all other variables in the model constant.

Additionally, in the fitted model, the coefficients for allcategories of education are negative and statistically significant. The results reveal that the household in which primary, secondary and higher educated household heads; are expected to have about 35 percent, 47 percent and about 57 percent respectively fewer expected number of migrants than uneducated household head, holding other variables constant in the model.

Here also the estimated coefficient for the religion category of Muslim household head is positive and had a significant impact on the number of migrants in the household. That is, households with Muslim household head are expected to have 178% more number of migrants as compared to the expected number of migrants with Orthodox households heads, while setting all variables in the model as constant.

Furthermore,the coefficient for rural is positive and statistically significant. This implying that the expected number of migrant in rural households are expected to have 62.76% more number of migrants than the urban households, holding other variables in the model are constant.

4.4.2 Interpretation of the HNB Model Results for Zero Inflated Part

For Zero-Inflation part of the model,the estimated coefficients describe that change in the odds of being in the always 0 group compared to the not always 0 group. As table 4.9, Age of the household head , Marital status of the household head, Education level of the household head ,Job of the household head,Migration experience of the household head and Residence of the household were significantly associated with the probability of being in the always zero group . And the interpretation is as follows:

When the age ahousehold increased by one unit, the odds of being in zero group (zero migrants) is increased by 1.033 ,while setting all variables in the model as constant. The variable marital status also had a positive significant effect on probability of being in the always zero group.Therefore ,the odds of no occurrence of migrants for households headed by widowed,divorced were increased by 9.29 , 8.52 respectively than households headed by single, holding other variable constant in the model .

Table 4.9: The Estimated HNB Model of Zero Inflated Part

Variables	Estimate	Std. Error	z value	Pr(>|z|)	IRR	Lower	Upper
Intercept	-1.6973	0.8471	-2.004	0.0451	0.1832	0.0348	0.9636
sex							
female	0.3792	0.3524	1.076	0.2820	1.4610	0.7323	2.9149
Age	0.0331	0.0130	2.550	0.0108	1.0336	1.0077	1.0602
Marital status							
Married	0.4093	0.5747	0.712	0.4764	1.5058	0.4881	4.6447
widowed	2.2290	0.6728	3.313	0.0009	9.2907	2.4853	34.7311
Divorced	2.1426	0.6629	3.232	0.0012	8.5218	2.3240	31.2478
Education							
Primary (1-8)	-0.2459	0.2804	-0.877	0.3805	0.7820	0.4513	1.3548
Sendary (9-12)	-0.8420	0.4037	-2.086	0.0370	0.4308	0.1953	0.9505
Higher	0.6034	0.7702	0.783	0.4334	1.8283	0.4041	8.2728
Economic activity							
Merchant	-0.9253	0.3572	-2.591	0.0096	0.3964	0.1968	0.7983
gov Employee	-3.3283	0.8111	-4.104	0.0000	0.0359	0.0073	0.1757
other	-1.1065	0.5151	-2.148	0.0317	0.3307	0.1205	0.9077
Religion							
Protestant	-0.1736	0.3459	-0.502	0.6158	0.8407	0.4267	1.6561
Muslim	0.0743	0.5056	0.147	0.8831	1.0772	0.3999	2.9015
Residence							
Rural	1.3273	0.2908	4.564	0.0000	3.7707	2.1325	6.6674
Home ownership							
By rent	-0.3154	0.3024	-1.043	0.2969	0.7295	0.4033	1.3196
Other	-0.2055	0.4270	-0.481	0.6303	0.8142	0.3526	1.8802
migration experience							
Migrated	-3.2832	0.4893	-6.710	0.0000	0.0375	0.0144	0.0979
Pull factors							
Social networks	-0.0801	0.3498	-0.229	0.8188	0.9230	0.4650	1.8321
Job opportunity	0.4067	0.3508	1.159	0.2463	1.5018	0.7552	2.9867
Smugglers	-0.1972	0.3897	-0.506	0.6129	0.8211	0.3826	1.7622

However,the main economic activity of the household head had a negative impact these households being in zero group.the result showed that the odds of no occurrence of migrants for household headed by a merchant ,government employee,other were decreased by 0.3963, 0.0358, 0.3307 respectively as compaired to households engaged in agriculture, holding other variable constant in the model.

Furthermore , the odds of being in zero group of migrants in rural households were increased by 3.77 as compared with urban households ,holding other variable constant in the model.Where as the migration experience of the household had a negative impact on the odds of no occurrence of migrants.This means household heads who had an experience of migration were decreased the odds of no occurrence of migrants by 0.0375 rather than household heads who had not experienced,setting other variables in the model as constant.

4.5 Discussion of the Results

Using a count regression model, this study tried to point out the variables that influence the number of migrants per household in Raya Kobo, North Wollo Zone. Thus, the study found that Sex of the household head, age of the household head, education levelof the household head, religion of the household head and residence place of the household were the significant factors of migrants per household in the Woreda.

Based on the study result, sex of household head was a significant factor for number of migrants per household. The highest number of migrants per household occurred in the female headed household than male headed households. This might be due to that more of the Woreda's community is depending on the agricultural activity the household members headed by female are more exposed for the migration and youths in these household wants to migrate in order to change the life style of the household and they want to fill their household's gab as good as the households headed by male. However the findings in (Admasu, 2018; Lambore & Taye, 2017) explained that households headed by male had more number of migrants than households headed by female.

The study also revealed that, education level of the household head is the important variable for the number of migrants in the household. Households with the educated household head are expected to have fewer expected number of migrants than households headed by un educated person. As Admasu (2018) education improves the systematical thinking and decision making skill, the educated heads are better in discussing issues with their youths. Therefore these households have a chance to decrease migrants by talking about the risks and consequences of this viral issue. This study is in lined with (Admasu, 2018; Lambore & Taye, 2017) .

According to the findings of this study, age of the household head was also a significant predictor for the number of migrants per household. As the age of the household head increases, the log odds of the number of the migrant in household will decreased. This might be when the age of household increased, the responsibility of the house hold becomes to the remaining members of the household. But the findings of(Lambore & Taye, 2017) explore that as the age of the household head increases the number of migrants for that household also increased.

The result of this study also showed that, the household's religion had a significant effect on the number of migrants per the household. That is, households with Muslim household head are expected to have more number of migrants than Orthodox household heads. The reason may be migrants expect less risk and think as they will be survived easily if their religion is the same with their destination. Therefore the religious similarity between the migrant and the destination country has its own contribution in increasing the number of migrants. The result is in lined with (Hadis, 2017).

The study also suggested that, residence of the household was found to statistically significant impact on the migrants. This study revealed that the expected number of migrant that rural households are expected to have more expected number of migrants than the urban households. This finding is consistent with other related study (Teshome et al., 2022).On the other hand, this study also revealed that ownership of home, migration experience of the household head, main economic activity of the household head, push and pull factors were not the significant predictors at 5% level of significance.

However, according to Admasu (2018) the variable home ownership had a significant effect on the number of migrants per household. That is, the coefficients for the category of home ownership by rent is positive and statistically significant whereas home ownership by other means is negatively affect the number of migrants.

As shown in different studies, Occupation of the household head had also significant effect on migration. Household heads occupied in agriculture have relatively high number of migrants than household's heads rely on trade (Bassie et al., 2022). Another research said that the probability to be migrated for people in trade was higher than probability to be migrated for people engaged in agriculture (Teshome et al., 2022).

Additionally, previous migration experience of household head was also an important predictor for the number of migrants per household. The coefficient for migrated household head is positive and statistically significant implying that the migrated household heads had more expected number of migrants than non migrated household heads as documented in (Admasu, 2018).

Furthermore, the major cause of Illegal Migration was push factors. Such as; Poverty, Unemployment, family pressure, and peer pressure (Admasu, 2018; Haile, 2015; and Prieto et al., 2015). Whereas, job opportunities, better income, social networks and smugglers at destination country were identified as pull factors of migration (Araya & Meles, 2021 ; Haile, 2015; and Prieto et al., 2015).

5 CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The objective of this study was to identify the determinant factors of the number of migrants per household using the count regression model. The source of the data was mainly primary data, collected through house to house visit. And the investigation was done via questionnaire from 577 randomly selected households in raya kobo Woreda, north wollo Amhara region. The descriptive results showed that 51% of households had no migrants in their family and 49 % households had at least one migrant. The descriptive results also suggested that there is high variability in the non-zero values. The variance of the number of migrants was larger than its mean, suggesting the possibility of over-dispersion.

In addition, the values of deviance and pearson chi-square were found to be significantly different from one which is an indication of an overdispertion in the data set. Using different comparison techniques like LRT, AIC,BIC for nested model and Vuong test for non-nested model,the HNB model was found to be the most appropriate model to predict the number of migrants pere household . Therefore, Hurdle negative binomial regression model is better fitted the data which is characterized by excess zeros and high variability in the non-zero outcome than any other count regression models.

Finally,this study captured predictor variables that had significant effects on the number of migrants. For selected HNB model, the truncated negative binomial part, Sex of the household head, age of the household head, education level, religion of the household head and residence place of the household were statistically significant factors influencing the number of migrants that households would like to have.While for the zero inflated part of the model,Age of the household head , Marital status of the household head, Education level of the household head ,Job of the household head,Migration experience of the household head and Residence of the household had a significanteffect on the probability of being in the always zero group.

5.2 Recommendations

After analyzing the main causes that has high contribution to migration of labor power migration from the Raya kobo woreda, north wollo zone to the Saudi Arabia, the researcher proposes the following suggestions that could be implemented by policy makers, government officials, and leaders or preachers of the religions as well as the community at large in the study area. On the basis of the findings obtained in the study, investigator recommends the following issues:

- The governmental bodies such as the minister of women and social affairs consider the identified major factors while designing policy that will impact migration decisions the most.

- The Raya Kobo Woreda's women and social affairs office have to conduct public discussions and awareness creation programs especially with the religious leaders of the Muslim communities as well as with female household heads.
- The researcher also strongly suggested that additional, in-depth research be done on the factors that lead people to migrate abroad,impacts on their families, and the difficulties that migrants face during their journey as well as their stay in the destination.

REFERENCES

Admasu, S. (2018). *Determinants of migration to south africa using count regression model: The case of soro and misha districts in snnp.*

Agresti, A. (2007). *An introduction to categorical data analysis*. John Wiley & Sons, Inc. https://doi.org/10.1002/0470114754

Alrababah, A., Dillon, A., Williamson, S., Hainmueller, J., Hangartner, D., & Weinstein, J. (2021). Attitudes toward migrants in a highly impacted economy: Evidence from the syrian refugee crisis in jordan. *Comparative Political Studies*, *54*(1), 33–76. https://doi.org/10.1177/0010414020919910

Araya, & Meles, J. (2021). *Socio-economic aspects of illegal migration from raya kobo district of north wollo to the middle east. Haramaya university.*

Bassie, H., Sirany, T., & Alemu, B. (2022). Rural-urban labor migration, remittances, and its effect on migrant-sending farm households: Northwest ethiopia. *Advances in Agriculture*, *2022*, 1–8. https://doi.org/10.1155/2022/4035981

Bedrick, E. J., & Hossain, A. (2013). Conditional tests for homogeneity of zero-inflated Poisson and Poisson-hurdle distributions. *Computational Statistics & Data Analysis*, *61*, 99–106. https://doi.org/10.1016/j.csda.2012.11.009

Buffone, A., Anderson, N. R., & Hammer, D. A. (2017). Migration against the direction of flow is LFA-1 dependent in human hematopoietic stem and progenitor cells. *Journal of Cell Science*, jcs.205575. https://doi.org/10.1242/jcs.205575

Cameron, A. C., & Trivedi, P. K. (1998). *Regression analysis of count data*. Cambridge University Press. https://doi.org/10.1017/CBO9780511814365

Carruth, L., & Smith, L. (2022). Building one's own house: Power and escape for Ethiopian women through international migration. *The Journal of Modern African Studies*, *60*(1), 85–109. https://doi.org/10.1017/S0022278X21000434

Cochran, W. (1977). *Sampling techniques. 3d ed.*

Corona, S., Bhabha, J., Wickramage, K., Wood, L. C. N., Erragne, L., Ortega García, O., Burgess, R., Digidiki, V., Aldridge, R. W., & Devakumar, D. (2021). International migration of unaccompanied minors: Trends, health risks, and legal protection. *The Lancet Child & Adolescent Health*, *5*(12), 882–895. https://doi.org/10.1016/S2352-4642(21)00194-2

CSA. (2013). *Population projection of ethiopia for all regions at wereda level from 2014 – 2017*.

Dauvergne, P. (2018). Why is the global governance of plastic failing the oceans? *Global Environmental Change*, *51*. https://doi.org/10.1016/j.gloenvcha.2018.05.002

Devaraj, S., Malkapuram, R., & Singaravel, B. (2021). Performance analysis of micro textured cutting insert design parameters on machining of Al-MMC in turning process. *International Journal of Lightweight Materials and Manufacture*, *4*(2), 210–217. https://doi.org/10.1016/j.ijlmm.2020.11.003

Ebrahim, S. M. (2016). *Migration of ethiopian youth to the middle east and its impacts on the migrants' sending community : Raya- mehoni town*.

Fernandez, B. (2017). *Irregular migration from ethiopia to the gulf states*. Gulf Research Centre Cambridge. http://hdl.handle.net/11343/191705

Gurmu, S., & Trivedi, P. (1996). Excess zeros in count models for recreational trips. *Journal of Business &Amp; Economic Statistics*, *14*(4), 469–477. https://EconPapers.repec.org/RePEc:bes:jnlbes:v:14:y:1996:i:4:p:469-77

Hadis, C. (2017). *Socio-cultural aspects of illegal migration in raya azebo woreda, tigray national regional state* [Thesis, Addis Ababa University]. http://etd.aau.edu.et/handle/123456789/929

Haile, H. (2015). *Socio economic impacts of migration of ethiopians to the south africa and its implications for ethio rsa relations: The case of kembata tembaro and hadiya zones* [Thesis, Addis Ababa University]. http://etd.aau.edu.et/handle/123456789/4442

Immersion. (2017). *International conference on 3D immersion (ic3d)*.

IOM (Ed.). (2020). *The future of migration: Building capacities for change*. IOM.

Johnson, T. R. (2012). J. M. Hilbe(2011) negative binomial regression, second edition. : New york: Cambridge university press. 576 pages. Us$85. 00. Isbn: 0521198151. *Psychometrika*, *77*(3), 611–612. https://doi.org/10.1007/s11336-012-9263-7

King, R., & Kuschminder, K. (Eds.). (2022). Introduction: Definitions, typologies and theories of return migration. In *Handbook of Return Migration*. Edward Elgar Publishing. https://doi.org/10.4337/9781839100055.00008

Lambore, T., & Taye, A. (2017). Logistic mixed modelling of determinants of international migration from the southern ethiopia: Small area estimation approach. *American Journal of Theoretical and Applied Statistics*.

Latham, J., David, F., Bryant, K., & Larsen, J. (2019). *Migrants and their vulnerability to human trafficking, modern slavery and forced labour*. http://hdl.handle.net/1854/LU-8636089

Marchand, J. R., Inez Roosen. (2015). *Irregular migration from and in the east and horn of africa,Maastricht grduate school of governance*.

Massey, D. S. (1988). Economic development and international migration in comparative perspective. *Population and Development Review*, *14*(3), 383. https://doi.org/10.2307/1972195

McAuliffe, M., James, & Khadria, B. (2019). *World migration report 2020*. International Organization for Migration.

McAuliffe, M., & Koser, K. (Eds.). (2017). *A long way to go: Irregular migration patterns, processes, drivers and decision-making* (1st ed.). ANU Press. https://doi.org/10.22459/LWG.12.2017

Mesfin, D. (2011). *Challenges and prospects of female labour migration to the arab middle east: A case study of women returnees in the town of girana, north wollo, ethiopia*.

Mihai, I. (2018). *Turkey between migrations: Social and political aspects of the phenomenon* [SSRN Scholarly Paper]. https://doi.org/10.2139/ssrn.3143145

Momo, M. S. M., Cabus, S. J., De Witte, K., & Groot, W. (2019). Systematic review of the literature on the causes of early school leaving in Africa and Asia. *Review of Education*, *7*(3), 496–522. https://doi.org/10.1002/rev3.3134

Mulugeta, M., & Makonnen, T. (1970). Illicit cross-border migration in ethiopia: Causes, patterns and policy responses. *Ethiopian Journal of the Social Sciences and Humanities*, *13*(2), 53–80. https://doi.org/10.4314/ejossah.v13i2.3

Naomi, L., Nistri, G., Pennetti, C., & Rodotà. (2021). Return migration dynamics in five ethiopian communities of high emigration. *International Organization for Migration*.

Peter, M., & Wright, J. (2010). *Handbook of survey research. 2nd ed.*

Prieto, R., Victoria, L.-G., & Antonio. (2015). *Push and pull factors of latin american migration*. *5*. https://doi.org/10.1007/978-3-319-12361-5

Ratha, D., Mohapatra, S., & Scheja, E. (2011). *Impact of migration on economic and social development : A review of evidence and emerging issues*. World Bank. https://doi.org/10.1596/1813-9450-5558

Rossi, B. (2018). Migration history and historiography. In *Oxford Research Encyclopedia of African History*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190277734.013.325

Schewel, K., & Bahir, A. L. (2019). *Migration and social transformation in ethiopia*.

Schneider, L., Chalmers, R. P., Debelak, R., & Merkle, E. C. (2018). *Model selection of nested and non-nested item response models using vuong tests*. https://doi.org/10.48550/ARXIV.1810.04734

Teshome, Haymanot, & Bizuayehu. (2022). Rural-urban labor migration, remittances, and its effect on migrant-sending farm households: Northwest ethiopia. *Department of Economics, Debre Markos University, Debre Markos, Ethiopia*.

Yilma, G. (2013). *Migration of school girls from raya kobo wereda to middle east: Antecedents and consequences*.

Yousaf, F. N. (2018). Forced migration, human trafficking, and human security. *Current Sociology*, *66*(2), 209–225. https://doi.org/10.1177/0011392117736309

Zeleke, W., & Minaye, A. (2015). Mental health and somatic distress among ethiopian migrant returnees from the middle east. *International Journal of Mental Health & Psychiatry*, *01*(02). https://doi.org/10.4172/2471-4372.1000106

APPENDICES

Table A Estimates and standard error for Zip model

	Count part of zip			
Variables	Estimate	Std.Error	z-value	Pr.z
Intercept	-3.2577	0.3740	-8.7116	0.0000
sex				
female	0.9139	0.1264	7.2305	0.0000
Age	-0.0248	0.0042	-5.9646	0.0000
Marital status				
Married	-1.3021	0.3191	-4.0803	0.0000
widowed	-1.0239	0.3305	-3.0977	0.0020
Divorced	-0.8119	0.3273	-2.4807	0.0131
Education				
Primary (1-8)	-0.5154	0.0962	-5.3580	0.0000
Sendary (9-12)	-0.3655	0.1607	-2.2743	0.0229
Higher	-0.1599	0.5031	-0.3178	0.7506
Economic activity				
Merchant	-0.5266	0.1308	-4.0263	0.0001
gov Employee	0.2276	0.4451	0.5114	0.6091
other	-0.1012	0.2488	-0.4068	0.6841
Religion				
Protestant	0.3455	0.1395	2.4767	0.0133
Muslim	1.0426	0.2144	4.8620	0.0000
Residence				
Rural	0.3384	0.0899	3.7623	0.0002
Home ownership				
By rent	-0.0854	0.1154	-0.7399	0.4594
Other	-0.5053	0.2172	-2.3262	0.0200
migration experience				
Migrated	-1.1917	0.3297	-3.6142	0.0003
Pull factors				
Social networks	-0.0323	0.1275	-0.2535	0.7999
Job opportunity	-0.2583	0.1211	-2.1322	0.0330
Smugglers	0.1692	0.1525	1.1095	0.2672

	zero inflated part of zip			
Variables	zero. Estimate	zero.Std.Error	zero.z.value	zero.Pr.z.
Intercept	19.1640	7.4487	2.5728	0.0101
sex				
female	4.6347	1.9884	2.3309	0.0198
Age	-0.3827	0.1426	-2.6836	0.0073
Marital status				
Married	-3.5734	2.7325	-1.3078	0.1909
widowed	-10.0555	4.5131	-2.2281	0.0259
Divorced	-7.6648	3.9909	-1.9206	0.0548
Education				
Primary (1-8)	-6.6454	2.7605	-2.4073	0.0161
Sendary (9-12)	0.8860	2.0926	0.4234	0.6720
Higher	-5.2112	2.9769	-1.7505	0.0800
Economic activity				
Merchant	-5.5236	2.6989	-2.0466	0.0407
gov Employee	16.6557	5.9849	2.7829	0.0054
other	-1.2048	3.0787	-0.3913	0.6955
Religion				
Protestant	-11.2356	4.8467	-2.3182	0.0204
Muslim	6.0467	3.1737	1.9052	0.0568
Residence				
Rural	-5.2411	2.6455	-1.9812	0.0476
Home ownership				
By rent	0.7119	1.8515	0.3845	0.7006
Other	-30.4502	1343.3915	-0.0227	0.9819
migration experience				
Migrated	9.7928	3.8532	2.5415	0.0110
Pull factors				
Social networks	-8.0266	3.3444	-2.4001	0.0164
Job opportunity	-7.0004	2.6240	-2.6678	0.0076
Smugglers	1.4420	1.4924	0.9662	0.3339

Table B. Estimates and standard error for ZINB model

	Count part of zinb			
Variables	Estimate	Std.Error	z-value	Pr.z
Intercept	-0.0284	0.4105	-0.0692	0.9448
sex				
female	0.0693	0.1227	0.5652	0.5719
Age	0.0142	0.0044	3.1999	0.0014
Marital status				
Married	0.1925	0.3610	0.5332	0.5939
widowed	0.1013	0.3638	0.2784	0.7807
Divorced	0.4539	0.3635	1.2487	0.2118
Education				
Primary (1-8)	0.1152	0.1005	1.1464	0.2516
Sendary (9-12)	0.0963	0.1891	0.5091	0.6102
Higher	-1.6325	0.5343	-3.0552	0.0022
Economic activity				
Merchant	-0.1609	0.1455	-1.1058	0.2688
gov Employee	0.5402	0.4588	1.1775	0.2390
other	-0.3930	0.2890	-1.3595	0.1740
Religion				
Protestant	0.0438	0.1524	0.2874	0.7738
Muslim	-0.1842	0.2844	-0.6476	0.5173
Residence				
Rural	0.2101	0.0911	2.3075	0.0210
Home ownership				
By rent	-0.0125	0.1227	-0.1018	0.9189
Other	-0.9514	0.2058	-4.6231	0.0000
migration experience				
Migrated	-1.3050	0.3919	-3.3301	0.0009
Pull factors				
Social networks	-0.2330	0.1325	-1.7588	0.0786
Job opportunity	-0.1343	0.1246	-1.0783	0.2809
Smugglers	-0.2646	0.1561	-1.6950	0.0901
Log(theta)	16.4931	0.0000	0.0000	0.0000

	zero inflated part of zinb			
Variables	Estimate	Std.Error	z-value	Pr.z
Intercept	1.8013	1.2466	1.4450	0.1485
sex				
female	-0.7177	0.7152	-1.0035	0.3156
Age	-0.0412	0.0203	-2.0315	0.0422
Marital status				
Married	-0.5778	0.8110	-0.7125	0.4761
widowed	-24.0312	29381.2300	-0.0008	0.9993
Divorced	-2.8725	1.0977	-2.6168	0.0089
Education				
Primary (1-8)	0.6637	0.4596	1.4441	0.1487
Sendary (9-12)	2.0454	0.7271	2.8129	0.0049
Higher	-3.4189	1.7351	-1.9704	0.0488
Economic activity				
Merchant	1.0061	0.5445	1.8479	0.0646
gov Employee	5.8593	1.6175	3.6225	0.0003
other	0.1422	0.8976	0.1584	0.8741
Religion				
Protestant	0.1242	0.4901	0.2535	0.7999
Muslim	0.0391	1.0639	0.0368	0.9707
Residence				
Rural	-1.7821	0.5435	-3.2787	0.0010
Home ownership				
By rent	0.5634	0.4396	1.2818	0.1999
Other	-20.5407	3758.8730	-0.0055	0.9956
migration experience				
Migrated	3.0609	0.8449	3.6226	0.0003
Pull factors				
Social networks	-0.4488	0.5356	-0.8379	0.4021
Job opportunity	-0.6198	0.4988	-1.2424	0.2141
Smugglers	0.1933	0.5751	0.3361	0.7368

Table C. Hurdle poisson Regression model

	Count part of hp				Zero inflated part of hp			
Variables	c. Estimate	std. erro	Z-value	pr. Z	Estimate	Std.Error	z-value	Pr.z
Intercept	-3.7426	0.4827	-7.7543	0.0000	-1.6973	0.8471	-2.0037	0.0451
sex								
female	0.9978	0.1698	5.8754	0.0000	0.3792	0.3524	1.0759	0.2820
Age	-0.0203	0.0050	-4.0713	0.0000	0.0331	0.0130	2.5497	0.0108
Marital status								
Married	-1.1125	0.4253	-2.6159	0.0089	0.4093	0.5747	0.7122	0.4764
widowed	-1.0928	0.4421	-2.4719	0.0134	2.2290	0.6728	3.3132	0.0009
Divorced	-0.7056	0.4357	-1.6194	0.1054	2.1426	0.6629	3.2320	0.0012
Education								
Primary (1-8)	-0.4442	0.1192	-3.7273	0.0002	-0.2459	0.2804	-0.8771	0.3805
Sendary (9-12)	-0.4169	0.2093	-1.9925	0.0463	-0.8420	0.4037	-2.0856	0.0370
Higher	-0.9187	0.7938	-1.1574	0.2471	0.6034	0.7702	0.7834	0.4334
Economic activity								
Merchant	-0.6390	0.1696	-3.7670	0.0002	-0.9253	0.3572	-2.5906	0.0096
gov Employee	0.4177	0.4796	0.8710	0.3837	-3.3283	0.8110	-4.1037	0.0000
other	-0.0012	0.3212	-0.0037	0.9971	-1.1065	0.5151	-2.1479	0.0317
Religion								
Protestant	0.4921	0.1711	2.8761	0.0040	-0.1736	0.3459	-0.5017	0.6159
Muslim	1.1246	0.2892	3.8890	0.0001	0.0743	0.5056	0.1470	0.8831
Residence								
Rural	0.3104	0.1119	2.7729	0.0056	1.3273	0.2908	4.5640	0.0000
Home ownership								
By rent	-0.1068	0.1454	-0.7347	0.4625	-0.3154	0.3024	-1.0430	0.2969
Other	-0.5506	0.3704	-1.4865	0.1371	-0.2055	0.4270	-0.4814	0.6303
migration experience								
Migrated	-0.7591	0.4736	-1.6028	0.1090	-3.2832	0.4893	-6.7103	0.0000
Pull factors								
Social networks	0.0094	0.1561	0.0605	0.9518	-0.0801	0.3498	-0.2291	0.8188
Job opportunity	-0.3098	0.1452	-2.1329	0.0329	0.4067	0.3508	1.1595	0.2463
Smugglers	-0.0058	0.1946	-0.0299	0.9761	-0.1971	0.3897	-0.5059	0.6129

Debre Berhan University,

College of Natural and Computational Science

Department of Statistics (Graduate program (R))

QUESTIONER:

Dear respondents: Thank you very much for your time! The purpose of this questionnaire is to gather information to conduct research for the partial fulfillment of the requirements for the degree of Master of Bio Statistics. This questionnaire is designed for the purpose of gathering information concerning the determinants of migration in raya kobo woreda to the Republic of Saudi Arabia with these 5 years. The information that you will have provide is intended to serve for identification of determinant factors of youth migration, and I confirm you that all data given by you will be treated confidentially. Therefore, you are kindly requested to provide accurate information as much as possible. Instruction: Tick in the box or write the answer as may be necessary and Do not write your name

I would like to Thank you again very much for your time!

1. Sex of household head
   □ Male
   □ Female
2. Age of household head______
3. Marital Status ofhousehold head □Single □ Married □ widowed □ divorced
4. What is highest level of education for HHH? □ Illiterate □ Primary (1-8) □ Secondary (9-12) □ Higher
5. Main economic activity of the household head □ Agriculture □ Merchant □ Government Employee□ Other
6. House hold member size_______
7. What is the household heads religion? □ Orthodox □ Protestant □ Muslim □ Other
8. Current residence place of the household□ Rural □ urban
9. Households home ownership □ Own □ By rent □ Other
10. Does the household head experienced migration to Saudi Arabia?
    □ Not migrated
    □ Migrated
11. How many total illegal migrants (return migrants and out migrants) are present in your household except the hhh with the previous 5 years? ———

12. What is the main “push” factors which motivate people to migrate to the Saudi Arabia in your Environment? Poverty Unemployment Family pressure Peer pressure
13. What is main the “pull” factors that motivate people, including you to migrate to theSaudi Arabia in your Environment? High income in Saudi Arabia Social networks Job opportunity s mugglers